FEATURE EXTRACTION FOR CANCER PREDICTION BY TENSOR DECOMPOSITION OF 1D PROTEIN EXPRESSION LEVELS

Ivica Kopriva^{1*}, Ante Jukić¹ and Andrzej Cichocki^{2,3} ¹Division of Laser and Atomic Research and Development, Ruđer Bošković Institute Bijenička cesta 54, P.O. Box 180, 10002 Zagreb, Croatia ikopriva@irb.hr, ajukic@irb.hr ²Laboratory for Advanced Brain Signal Processing Brain Science Institute, RIKEN 2-1Hirosawa, Wako-shi, Saitama, 351-0198, Japan ³Warsaw University of Technology and Systems Research Institute, PAN, Poland.

cia@brain.riken.jp

ABSTRACT

Tensor decomposition approach to feature extraction from one-dimensional data samples is presented. Onedimensional data samples are transformed into matrices of appropriate dimensions that are further concatenated into a third order tensor. This tensor is factorized according to the Tucker-2 model by means of the higher-orderorthogonal iteration (HOOI) algorithm. Derived method is validated on publicly available and well known datasets comprised of low-resolution mass spectra of cancerous and non-cancerous samples related to ovarian and prostate cancers. The method respectively achieved, in 200 cross-validations, independent two-fold average sensitivity of 96.8% (sd 2.9%) and 99.6% (sd 1.2%) and average specificity of 95.4% (sd 3.5%) and 98.7% (sd 2.9%). Due to the widespread significance of mass spectrometry for monitoring protein expression levels and cancer prediction it is conjectured that presented feature extraction scheme can be of practical importance.

KEY WORDS

Cancer prediction, mass spectrometry, feature extraction, tensor decomposition, pattern recognition.

1. INTRODUCTION

Feature extraction and selection are essential problems in the analysis of datasets with a large number of variables. Typical areas in which given problems arise include text mining in internet documents, combinatorial chemistry, proteomics, genomics and computational biology [1]. Extraction of suitable features is considered to have a major effect on classification performance, and matter more than the classifier used [2]. Problem of data overfitting arises when a large number of features and a relatively small number of samples are available for learning from data, since the classifier tunes to the idiosyncrasies of the training set. In this case perfect results can be obtained on the training data, but the trained classifier will generally not perform well on the test data, i.e. classifier exhibits poor generalization for unseen data samples. Decreasing number of features remedies the problem of data overfitting, while simultaneously reducing memory requirements and time needed for data analysis. Even training techniques that utilize regularization to prevent overfitting, such as support vector machines (SVM), benefit from dimensionality reduction [3]. Methods for dimensionality reduction by feature extraction are commonly divided into two approaches: feature selection and feature transformation [4]. Methods for feature selection do not alter original data, but merely select subset of the original large-scale dataset. Depending on the criteria used for selection, they can be divided into three different categories: filter, wrapper and embedded methods [1]. Transformation methods construct reduced feature set as a linear or nonlinear transformation of the original data. In both approaches, classifier is trained on the reduced feature set. In this paper we propose general method for feature extraction, based on the linear transformation constructed by tensor decomposition.

Multimodal or multi-way datasets with high dimensionality are often used in modern applications. Due to the multi-way structure, tensors are natural way for representing such data. Hence, tensor decompositions are appropriate tools for analysis of such higher-dimensional datasets [5], [6]. Methods for tensor decomposition are used in various areas, especially for supervised and unsupervised dimensionality reduction, classification and multi-way clustering [5], [6], but also in the analysis of multi-spectral [7], [8], and hyper-spectral images [9]. Various tensor decompositions for model reduction and feature extraction are used for classification of different types of multi-way data, such as images of objects, handwritten digits and multi-channel EEG data [10], [11], as well as music genre classification [12].

Original contribution of this paper is application of tensor representation for analysis of one-dimensional (one-way) data, more specific for protein expression patterns of cancerous and non-cancerous samples acquired by the low resolution surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) mass spectrometry. For this type of data each sample is represented as a single vector (one-way data sample), containing large number (typically 5000 to 20000) mass to charge m/z ratios representing features or variables. Each sample corresponds to a single patient, so the number of available samples is typically rather small, being several orders of magnitude smaller than the number of variables. Thus, a method for dimensionality reduction is needed to obtain efficient classification on small number of discriminative features. Use of tensor decomposition for feature extraction requires one-way sample (vector) to be transformed onto multi-way structure, e.g. matrix or higher-order tensor. Standard approach in analysis of multispectral image using matrix factorization is to transform a two-way image to a vector. This procedure is commonly known as vectorization [13]. Here we use reverse procedure known as matricization to transform vector to matrix. Method for feature extraction from one-way data is demonstrated on prediction of prostate and ovarian cancers from mass spectra of biological samples. Low resolution SELDI-TOF mass spectra available at [14] are used for this purpose. This type of data has previously been used, with results reported in literature, so our method can be compared with others. Several papers address the problem of sample classification by the analysis of low-resolution SELDI-TOF data, including [15], [16], [17] for prostate cancer, and [17]-[21] for ovarian cancer prediction.

2. METHODS

This section presents basics of the tensor notation. Brief overview of the Tucker model is also given followed by algorithms for tensor decomposition, and transformation of one-way samples to the format appropriate for tensorbased analysis.

2.1 Basics of tensor algebra, Tucker model and HOOI algorithm

Tensor or multi-way array is a generalization of the concepts of vector and matrix. For example, vector is a one-way tensor and matrix is a two-way tensor. Throughout this paper a vector with *I* elements will be denoted as $\mathbf{a} \in \mathbb{R}^{I}$ and matrix with I_{1} rows and I_{2} columns as $\mathbf{A} = \begin{bmatrix} \mathbf{a}_{1}, \mathbf{a}_{2}, \dots, \mathbf{a}_{I_{2}} \end{bmatrix} \in \mathbb{R}^{I_{1} \times I_{2}}$. *N*-way tensor will be denoted as $\mathbf{X} \in \mathbb{R}^{I_{1} \times I_{2}}$. *N*-way tensor will be denoted as $\mathbf{X} \in \mathbb{R}^{I_{1} \times I_{2} \times \dots \times I_{N}}$ with elements $\{x_{i_{1}i_{2}\cdots i_{N}}\}_{i_{1},i_{2},\dots,i_{N}=1}^{I_{1},I_{2},\dots,I_{N}}$. Each index in tensor is called way or mode and number of levels on certain mode is called

dimension in that mode. This is the standard notation adopted in multi-way analysis [22]. Subtensor can also be defined as a subset of the original tensor, obtained by fixing certain indices. For example, for third-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ frontal, lateral and horizontal slices are defined as $\mathbf{X}_{i_1} = \underline{\mathbf{X}}_{::i_1}$, $\underline{\mathbf{X}}_{:i_2:}$ and $\underline{\mathbf{X}}_{i_1::}$ (respectively), where each slice is a second order tensor, i.e. matrix. The product of tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times .. \times I_N}$ and a matrix $\mathbf{A} \in \mathbb{R}^{I_n \times I_n}$ along the mode-*n* is denoted as:

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_{n} \mathbf{A} \in \mathbb{R}^{I_{1} \times \ldots \times I_{n-1} \times J_{n} \times I_{n+1} \ldots \times I_{N}}$$

which on component level is defined as:

$$y_{i_{1}i_{2}\cdots i_{n-1}j_{n}i_{n+1}\cdots i_{N}} = \sum_{i_{n}=1}^{I_{n}} x_{i_{1}i_{2}\cdots i_{N}} \cdot a_{j_{n}i_{n}}$$

Multiplication of tensor and a matrix can be successively applied in several distinct modes $(m \neq n)$ and it can be shown to be commutative:

$$(\underline{\mathbf{X}} \times_n \mathbf{A}) \times_m \mathbf{B} = (\underline{\mathbf{X}} \times_m \mathbf{B}) \times_n \mathbf{A} = \underline{\mathbf{X}} \times_n \mathbf{A} \times_m \mathbf{B}$$

where matrices A and B have appropriate dimensions according to dimension of modes m and n in tensor. Repeated mode-n product can be expressed as:

$$(\underline{\mathbf{X}} \times_{n} \mathbf{A}) \times_{n} \mathbf{B} = \underline{\mathbf{X}} \times_{n} (\mathbf{B} \cdot \mathbf{A})$$

where matrices **A** and **B** have appropriate dimensions. Multiplication of tensor $\underline{\mathbf{X}}$ in all modes (*n*=1,2,...,*N*) by a set of matrices $\mathbf{A}^{(n)}$ is denoted as:

$$\underline{\mathbf{X}} \times \{\mathbf{A}\} = \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_N \mathbf{A}^{(N)}$$

and can be expressed in matrix form as:

$$\begin{bmatrix} \mathbf{X} \times \{\mathbf{A}\} \end{bmatrix}_{(n)} =$$
$$= \mathbf{A}^{(n)} \mathbf{X}_{(n)} \begin{bmatrix} \mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \end{bmatrix}^{\mathrm{T}}$$

where \otimes denotes Kronecker product and $[.]_{(n)}$ is matricization of tensor in mode *n* [5].

Basic model for decomposition of *N*-way tensor is Tucker model, that in general form can be expressed as [5]:

$$\begin{split} \mathbf{\underline{X}} &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N} \left(\mathbf{a}_{j_1}^{(1)} \circ \mathbf{a}_{j_2}^{(2)} \circ \cdots \circ \mathbf{a}_{j_N}^{(N)} \right) + \mathbf{\underline{E}} \\ &= \mathbf{\underline{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} + \mathbf{\underline{E}} \\ &= \mathbf{\underline{G}} \times \left\{ \mathbf{A} \right\} + \mathbf{\underline{E}} = \mathbf{\underline{\hat{X}}} + \mathbf{\underline{E}} \end{split}$$
(1)

where \circ denotes outer product of vectors. In (1) $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ denotes data tensor, $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_N}$ is the core tensor of model with reduced dimension in each mode, and $\mathbf{A}^{(n)} = \left[\mathbf{a}_{1}^{(n)}, \mathbf{a}_{2}^{(n)}, \dots, \mathbf{a}_{J_{n}}^{(n)}\right] \in \mathbb{R}^{I_{n} \times J_{n}}$, (n=1,2,...,N), are factor matrices for each mode. Tensor $\hat{\mathbf{X}}$ is approximation of the original data tensor \mathbf{X} , while \mathbf{E} denotes the approximation error. For three-way tensor \mathbf{X} , Tucker-3 model is represented by three factors $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}, \mathbf{A}^{(3)}$, and the core tensor \mathbf{G} as:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} .$$
 (2)

This model can be reduced to Tucker-2 model by incorporating one of the factors into the core tensor, for example:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{F}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)}$$
(3)

where new core tensor is $\mathbf{F} = \mathbf{G} \times_3 \mathbf{A}^{(3)}$. In this scenario dimensionality reduction is not performed in mode 3. Analogous reduction of Tucker-*N* to Tucker-(*N*-1) model can be done for higher order tensors by including one of the factors into the core tensor. Throughout this paper we will assume that dimensions of the core tensor are all equal, i.e. $J_1=J_2=\ldots=J_n=J$, where $J \leq \min\{I_1,\ldots,I_N\}$, and only third-order tensors and their corresponding Tucker-2 model will be used for experiments.

Decomposition of an N-way tensor according to the Tucker model (1) is in general not unique. Therefore, it is necessary to impose additional constraints on the factor matrices and the core tensor of the model. Typically, tensor decomposition algorithms are obtained by minimizing cost function equal to some of divergences between original tensor \mathbf{X} and its approximation by model tensor $\hat{\mathbf{X}}$. By adding regularization term to the cost function constraints on the core tensor and the factor matrices can be enforced. Sparseness, nonnegativity or orthogonality constraints narrow down solution space and that often results in decomposition that is virtually unique [5]. Type of algorithm that can be used for decomposition depends on the properties of the given data tensor \mathbf{X} , for example, it is meaningful to decompose data tensor according to Tucker model using nonnegativity constraints only if the elements of the data tensor are nonnegative. Here, decomposition according to the Tucker model is performed using the higher order orthogonal iteration (HOOI) algorithm [23]. The HOOI algorithm minimizes Euclidean distance between the data tensor **X** and its Tucker model **X**:

$$D\left[\mathbf{\underline{X}}\|\mathbf{\underline{\hat{X}}}\right] = \left\|\mathbf{\underline{X}} - \mathbf{\underline{\hat{X}}}\right\|_{H}^{2}$$

where orthogonality constraints are imposed on the factor matrices $\mathbf{A}^{(n)}$ and all-orthogonality and ordering constraints on the corresponding core tensor in the model (1). Restriction to all-orthogonal core tensor means that

all subtensors $\underline{\mathbf{G}}_{i,=k}$ and $\underline{\mathbf{G}}_{i,=l}$ must be orthogonal for all n, k, and l ($k \neq l$), while ordering implies $\left\|\underline{\mathbf{G}}_{i_{n}=1}\right\|_{F} \geq \left\|\underline{\mathbf{G}}_{i_{n}=2}\right\|_{F} \geq \ldots \geq \left\|\underline{\mathbf{G}}_{i_{n}=J_{n}}\right\|_{F} \quad \text{for all } n=1,2,\ldots,N.$ Experimental results in section 3 were obtained using the implementation of the HOOI algorithm provided through tucker function made available as a part of N-way toolbox for MATLAB [24], [25]. Above mentioned implementation allows to fix a single mode during decomposition, thus enabling to decompose N-way tensor according to Tucker-(N-1) model. As an example, in our case three-way tensor can be decomposed according to Tucker-2 model.

After decomposition of the third-order data tensor, approximation of core tensor is calculated from the data $\underline{\mathbf{X}}$. Orthogonality of the factors obtained by HOOI algorithm enables calculation of the core tensor as:

$$\hat{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)\mathrm{T}} \times_2 \mathbf{A}^{(2)\mathrm{T}} \times_3 \mathbf{A}^{(3)\mathrm{T}}$$
(4)

in case of the Tucker-3 model (2), or as:

$$\hat{\mathbf{\underline{F}}} = \mathbf{\underline{X}} \times_1 \mathbf{A}^{(1)\mathrm{T}} \times_2 \mathbf{A}^{(2)\mathrm{T}}$$
(5)

in case of the Tucker-2 model (3) that is used herein for decomposition of mass spectra of cancerous and non-cancerous samples.

2.2 Feature extraction

Herein we present basic concept for feature extraction from two-way samples as in [10]. Let $\mathbf{X}^{(k)} \in \mathbb{R}^{I_1 \times I_2}$, k=1,...,K, be K given data matrices (two-way samples) with dimensions I_1 and I_2 , with associated class labels $\{c_k\}_{k=1}^{K}$. Without loss of generality, we can assume binary classification problem, e.g. class labels c_k are equal to 1 or -1. In order to construct reduced feature set for each sample and to reduce dimensionality joint matrix factorization is performed [10]:

$$\mathbf{X}^{(k)} \approx \mathbf{A}^{(1)} \mathbf{F}^{(k)} \mathbf{A}^{(2)T}$$

for all *k*. Matrices $\mathbf{A}^{(1)} \in \mathbb{R}^{I_1 \times J}$ and $\mathbf{A}^{(2)} \in \mathbb{R}^{I_2 \times J}$, $J \leq I_n$, are two common factors used for obtaining reduced feature set by linear transformation of the given data matrices $\mathbf{X}^{(k)}$. Matrix $\mathbf{F}^{(k)} \in \mathbb{R}^{J \times J}$ represents a feature matrix, containing features extracted from the *k*-th training sample. It is obvious that the given $I_1 \cdot I_2$ features are reduced to J^2 extracted features, with typically $J^2 << I_1 \cdot I_2$. Note that, because of simplicity, we assumed that matrix $\mathbf{F}^{(k)}$ is square, although that is not necessary condition. Joint matrix factorization is equivalent to tensor decomposition of three-way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times K}$ according to the Tucker-2 model (3) as:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{F}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \,.$$

Data matrices $\mathbf{X}^{(k)}$ are combined into three-way tensor \mathbf{X} so that each sample $\mathbf{X}^{(k)}$ corresponds to k-th frontal slice of **X**. The core tensor $\mathbf{F} \in \mathbb{R}^{J \times J \times K}$ is composed of *K* feature matrices, where k-th frontal slice matches $\mathbf{F}^{(k)}$. After HOOI-based decomposition of tensor X to its Tucker-2 model, we obtain orthogonal factor matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$. Approximation of the core tensor can be calculated directly from the data tensor as in (5). Extracted features $\left\{ \hat{\mathbf{f}}^{(k)} \in \mathbb{R}^{J^2} \right\}_{k=1}^{K}$ for each of the K training data samples $\mathbf{X}^{(k)}$ are obtained by transforming frontal slices $\hat{\mathbf{F}}_{k} = \hat{\mathbf{F}}^{(k)}$ to vector representation. This $S = \left\{ \left(\hat{\mathbf{f}}^{(1)}, c_1 \right), \dots, \left(\hat{\mathbf{f}}^{(K)}, c_K \right) \right\}$ set vields containing extracted features paired with labels for K training samples. The set S is used for training of the selected classifier.

The factor matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ obtained by decomposition are used to extract features from unseen data samples that are used to validate classification performance. Lets us denote two-way test samples as $\mathbf{X}_{test}^{(m)} \in \mathbb{R}^{I_1 \times I_2}$, m=1,..,M. By concatenation we form tensor $\underline{\mathbf{X}}_{test} \in \mathbb{R}^{I_1 \times I_2 \times M}$ containing each of the *M* test samples as single frontal slice. Using orthogonal factors obtained by decomposition of the training data, core tensor for the test data is calculated as in (5):

$$\hat{\mathbf{\underline{F}}}_{test} = \mathbf{\underline{X}}_{test} \times_{1} \mathbf{A}^{(1)T} \times_{2} \mathbf{A}^{(2)T}.$$

Vectorization of the frontal slices yields a set of extracted features used to validate previously trained classifier.

If available data samples are one-way, i.e. vectors, a transformation is necessary before described feature extraction procedure can be applied. We propose to use matricization to form two-way data samples. Let $\mathbf{x}^{(k)} \in \mathbb{R}^{l_1 l_2}$ be vector for *k*-th sample. Then sample matrix $\mathbf{X}^{(k)} \in \mathbb{R}^{l_1 \times l_2}$ can be formed by rearranging the elements of vector $\mathbf{x}^{(k)}$, as shown on an example in Figure 1.



Figure 1. Example of transformation of vector to matrix

Mapping pattern presented in Figure 1 is not specific for the feature extraction scheme presented in the paper i.e. other arrangement of the indices in matricized version of the data sample are possible as well.

3. EXPERIMENTS

Previously described method for feature extraction from one-way data samples is demonstrated on prediction of prostate and ovarian cancer from mass spectrometry data [14]. Features obtained by proposed procedure are used to train and test several classifiers, whereas performance is estimated in a statistically rigorous manner, using ten- and two-fold cross-validation (CV). All computations were performed in MATLAB environment (version 7.11), on desktop computer with 2.4GHz clocked quad-core processor and 4GB of RAM.

3.1 Description of data

Low-dimensional mass spectroscopy data samples, available at [14], are obtained by analysis of proteins in serum samples of different patients. Appropriate classes are assigned to the samples in dataset, describing each sample as a control or disease. In this way each set of data samples is divided into two populations: one corresponding to healthy individuals (controls) and other to patients with biopsy confirmed cancer. Prostate cancer dataset consists of 69 disease samples and 63 control samples. Another 100 disease data samples and 100 controls are available for ovarian cancer dataset (labeled Ovarian 4-3-02 Study set). In both datasets each sample is represented by a vector containing $T_0=15154$ elements, representing intensity level for all m/z ratios in range available by used mass spectrometer. Before transforming each one-way sample from vector to matrix, last 25 elements (corresponding to the highest m/z ratios) were truncated. Number of removed elements is negligible in relation to the total number of elements in a sample, and thus does not affect final performance. In this way, each sample is represented as a square matrix with dimensions $I_1=I_2=123$. It is also important to emphasize that the original samples were preprocessed by hand and the baseline was subtracted, creating the negative intensities seen for some values [14]. This is the reason why nonnegative tensor decompositions are not appropriate for analysis of this data, and, thus, only HOOI algorithm was used herein.

3.2 Results

Classification performance is evaluated by estimating sensitivity and specificity through ten-fold and two-fold cross-validation with 200 random partitions. Several classifiers were used: *k*-nearest neighbor (kNN), linear support vector machine (linSVM), and nonlinear support vector machine with polynomial kernel (polySVM) and

Gaussian radial basis function kernel (rbfSVM). Classification and cross-validation have been carried out using MATLAB functions *svmtrain*, *svmclassify*, *knnclassify*, and *classperf*.

Parameters of the kNN and nonlinear SVM classifiers were optimized by cross-validation (number of neighbors from 2 to 12; width of Gaussian kernel from 0.8 to 20, with the step size of 0.2; order of polynomial kernel from 2 to 7). Feature extraction was performed using HOOIbased tensor decomposition according to the Tucker-2 model. The core tensors with dimension in modes 1 and 2 equal to $J = \{6, 10, 16, 20, 25\}$ were extracted, yielding J^2 = {36, 100, 256, 400, 625} extracted features per data sample. The results for prediction of prostate cancer in terms of mean values of sensitivity and specificity and their standard deviations are shown in Table 1, while Table 2 presents prostate cancer prediction results reported in the literature. Results for detection of ovarian cancer obtained herein are shown in Table 3, while results reported in literature are presented in Table 4. Performance of prostate and ovarian cancer prediction obtained on publicly available datasets shows that presented method is comparable with the ones reported in the literature. Additionally, for some of the results from literature, shown in Tables 2 and 4, there is no information about used CV method, or CV is not used at all, while our result is obtained under rigorous CV procedure. Effect of the selected CV method on the final result is not negligible, which can be seen from Tables 1 and 3, where estimation of sensitivity and specificity based on ten-fold CV yields too optimistic impression about performance. Results of cancer prediction can also be over- or underestimated if only one random partition was used. Because of that, we believe that results reported herein for the two-fold CV are realistic.

4. CONCLUSION

Extraction of features is of vital importance for accurate class prediction in many application areas. While tensor decompositions are useful tools in analysis of multi-way large-scale datasets, they operate exclusively on higherorder data. Here we propose a method for tensor-based feature extraction from one-way datasets such as mass spectra of biological samples. Proposed feature extraction scheme is validated on prediction of prostate and ovarian cancers from publicly available and known datasets. Obtained results imply that described approach is of practical importance for prediction of carcinoma through analysis of mass spectra of biological samples. Presented method is not limited for use solely on proteomic expression profiles. It is conjectured that it can also be useful in analysis of other types of data such as gene expression levels acquired by DNA microarrays.

ACKNOWLEDGEMENT

The work of I. Kopriva and A. Jukić has been supported through grant 098-0982903-2558 funded by the Ministry of Science, Education and Sports, Republic of Croatia.

	Prediction of prostate cancer					
	J^2	kNN	linSVM	polySVM	rbfSVM	
	36	88.8±11.3 /	95.1±8.8 /	97.5±6.0 /	96.1±7.2 /	
		88.3±11.7	94.7±8.2	89.9±11.8	91.3±10.9	
		k=4		d=3	σ=3.2	
	100	94.6±8.3 /	99.3±3.5 /	99.6±2.7 /	98.3±5.1 /	
		88.7±11.6	96.9±7.2	93.1±9.3	93.1±9.9	
		k=3		d=3	σ=6.8	
р	256	96.4±6.7 /	99.9±1.0 /	99.7±2.1 /	99.8±1.7 /	
-fo		92.2±10.5	97.7±5.6	93.7±9.9	93.0±10.0	
10		k=4		d=3	σ=13	
		95.4±8.1 /	100±0 /	99.6±2.5 /	99.8±2.3 /	
	400	93.7±9.2	100±0	94.0±10.1	93.2±9.7	
		k=3		d=2	σ=15.6	
	625	96.0±7.2 /	100±0 /	100±0 /	100±0 /	
		94.6±8.9	100 ± 0	94.4±9.6	93.8±9.4	
		k=3		d=2	σ=19.8	
	36	85.4±7.1 /	91.9±5.6 /	94.4±4.7 /	93.2±5.1 /	
		85.1±5.6	91.4±4.8	89.0±5.9	89.7±5.1	
		k=4		d=3	σ=3.2	
	100	89.3±5.4 /	98.2±2.8 /	98.3±2.7 /	94.6±5.2 /	
		86.7±5.3	95.6±3.6	89.2±5.5	90.3±4.7	
		k=2		d=3	σ=6.6	
р	256	90.2±5.4 /	98.8±2.2 /	97.0±3.3 /	96.2±3.9 /	
fol-		89.7±5.1	96.5±3.4	90.0±5.4.	91.0±4.9	
7		k=4		d=2	σ=12.6	
	400	92.6±4.6 /	99.3±1.6 /	98.2±2.7 /	96.4±4.3 /	
		90.1±4.2	97.8±3.3	90.7±5.9	91.5±4.8	
		k=2		d=2	σ=13.2	
	625	93.0±4.9 /	99.6±1.2 /	98.9±2.1 /	98.0±3.2 /	
		91.8±4.6	98.7±2.9	91.4±5.1	93.2±4.5	
		k=2		d=2	σ=19	

Table 1 Prediction of prostate cance

Sensitivity and specificity in % (mean values \pm standard deviation) estimated by two-fold and ten-fold CV. Optimal values of the classifier parameters are given in the table, where 'k' denotes number of neighbors of the kNN classifier, σ denotes standard deviation of the Gaussian kernel of the radial basis function (RBF) SVM classifier and 'd' denotes degree of the polynomial of the polynomial SVM classifier.

Table 2
Prostate cancer prediction results reported in the literature

Petricoin et al.[15]	sensitivity: 94.7%; specificity: 75.9%; 253 benign and 69 cancers. CV details not	
	reported	
	sensitivity: 97.1%; specificity: 96.8%; 253	
Xu et al. [16]	benign and 69 cancers. CV details not	
	reported	
Vana et al [17]	Average error rate of 28.97 on four class	
r ang <i>et al</i> . [17]	problem with 3-fold CV.	

[I ²	kNN	linSVM	nolvSVM	rbfSVM
	J	72 4 12 6 /	99 1 1 0 0 /	POIVS VIVI	28 4 10 2 /
	36	(2.4 ± 15.0)	00.1±9.97	80.1±10.//	00.4±9.37
		03.9±15.1	84.0±10.4	85.8±12.1	81.1±11.7
		K=4		d=3	σ=4.2
	100	87.0±10.4 /	92.8±8.4 /	93.1±8.5 /	93.9±7.7 /
		65.0±14.1	89.4±9.2	88.6±10.2	87.9±10.2
		k=6		d=3	σ=6
p		87.4±10.8 /	95.3±6.4 /	94.5±7.0 /	95.2±6.6 /
of-(256	69.8±14.2	95.4±7.4	90.0±11.1	91.1±9.3
10		k=4		d=3	σ=12.2
		90.2±10.2 /	96.8±5.5 /	94.8±7.4 /	95.9±6.4 /
	400	71.0±14.7	95.0±6.4	91.5±9.3	92.5±8.8
		k=6		d=2	σ=14.4
	625	93.8±7.5 /	97.9±4.9 /	96.5±5.9 /	96.3±6.1 /
		71.0±14.2	97.9±4.8	95.2±6.6	94.5±7.5
		k=8		d=2	σ=19.4
	36	66.7±8.4 /	84.3±6.1 /	83.4±6.5 /	84.6±5.6 /
		64.7±7.2	81.4±5.9	79.8±7.2	77.3±6.7
		k=12		d=3	σ=4.4
	100	77.0±7.0 /	91.1±4.6 /	90.7±4.5 /	92.1±4.2 /
		64.6±6.7	87.7±5.0	82.7±6.4	84.5±5.4
		k=4		d=3	σ=8.2
p	256	82.7±6.2 /	93.6±3.7 /	93.6±3.5 /	93.9±3.0/
2-fold		67.7±6.5	91.9±4.1	85.1±5.8	87.7±4.8
		k=4		d=3	σ=14.2
	400	85.4±5.7 /	95.5±3.2 /	93.4±3.6 /	95.1±2.7 /
		67.2±7.3	93.4±3.6	85.4±5.5	90.1±4.4
		k=4		d=2	σ=17.4
	625	87.8±5.3 /	96.8±2.9 /	94.2±3.4 /	95.0±2.7 /
		68.7±6.2	95.4±3.5	87.1±5.7	91.9±4.1
		k=4		d=2	σ=18.6

Table 3 Prediction of ovarian cancer

Sensitivity and specificity in % (mean values \pm standard deviation) estimated by two-fold and ten-fold CV. Optimal values of the classifier parameters are given in the table, where 'k' denotes number of neighbors of the kNN classifier, σ denotes standard deviation of the Gaussian kernel of the radial basis function (RBF) SVM classifier and 'd' denotes degree of the polynomial of the polynomial SVM classifier.

Table 4

Ovarian cancer prediction results reported in the literature				
	sensitivity: 100%; specificity: 95%			
Petricoin et al. [18]	(one partition only: 50/50 training; 66/50			
	test).			
Assemble of al [10]	accuracy averaged over ten 10-fold partitions:			
Assaren <i>et at</i> . [19]	98-99% (SD: 0.3-0.8)			
L:	sensitivity: 98%; specificity: 95%			
Li <i>et al</i> . [20]	(2-fold CV with 100 partitions)			
Qiu et al. [21] 97.63% accuracy (other details not spec				

Yang *et al.* [17] average error rate of 4.1 % with 3-fold CV.

REFERENCES

[1] I. Guyon, and A. Elisseeff, An Introduction to Variable Selection and Feature Selection, *Journal of Machine Learning Research*, *3*, 2003, 1157-1182.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (New York, NY: Springer, 2009).

[3] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene Selection for Cancer Classification Using Support Vector Machine, *Machine Learning*, 46 (1-3), 2002, 389-422.

[4] K.-L. Tang, T.-H. Li, W.-W. Xiong, and K. Chen, Ovarian cancer classification based on dimensionality reduction for SELDI-TOF data, *BMC Bioinformatics*, *11*, 2010.

[5] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorization* (Chichester: John Wiley & Sons, 2009)

[6] T. Kolda, and B.Bader, Tensor decompositions and applications, *SIAM Review*, *51* (3), 2009. 455-500.

[7] I. Kopriva, and A. Cichocki, Blind Multi-spectral Image Decomposition by 3D Nonnegative Tensor Factorization, *Optics Leters*, *34* (14), 2009, 2210-2212.

[8] I.Kopriva, Tensor Factorization for model-free spacevariant blind deconvolution of the single- and multi-frame multi-spectral Image, *Optics Express*, *18* (17), 2010, 17819-17833.

[9] S. Bourennane, C. Fossati, and A. Cailly, Improvement of Classification for Hyperspectral Images Based on Tensor Modeling, *IEEE Geoscience and Remote Sensing Letters*, 7 (4), 2010, 2407-2417.

[10] A. H. Phan, and A. Cichocki, Tensor Decomposition for Feature Extraction and Classification of High Dimensional Datasets, *IEICE Nonlinear Theory and Its Applications*, 1 (1), 2010, 37-68.

[11] A. H. Phan, and A. Cichocki, Extended HALS Algorithm for Nonnegative Tucker Decomposition and Its Applications for Multiway Analysis and Classification, *Neurocomputing*, In Press, doi:10.1016/j.neucom.2010.06.031.

[12] E. Benetos, and C. Kotropoulos, Non-Negative Tensor Factorization Applied to Music Genre Classification, *IEEE Transactions on Audio, Speech, and Language Processing*, 18 (8), 2010, 1955-1967.

[13] I. Kopriva, A. Cichocki, Blind decomposition of lowdimensional multi-spectral image by sparse component analysis, *Journal of Chemometrics*, 23 (11), 2009, 590-597.

[14] Center for Cancer Research, National Cancer Institute. Available:

http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.as p

[15] E.F. Petricoin *et al.*, Serum proteomic patterns for detection of prostate cancer, *Journal of National Cancer Institute*, 94, 2002, 1576-1578.

[16] Q. Xu *et al.*, Mass spectrometry-based proteomic pattern analysis for prostate cancer detection using neural networks with statistical significance test-based feature selection, *Proceedings of the 2009 IEEE Conference Science and Technology for Humanity (TIC-STH)*, 2009, 837-842.

[17] P. Yang *et al.*, A clustering based hybrid system for biomarker selection and sample classification of mass spectrometry data, *Neurocomputing*, *73*, 2010, 2317-2331.

[18] E. F. Petricoin *et al.*, Use of proteomic patterns in serum to identify ovarian cancer, *The Lancet*, *359*, 2002, 572-577.

[19] A. Assareh *et al.*, Fuzzy rule based classifier fusion for protein mass spectra based ovarian cancer diagnosis, *Proceedings of the 2009 IEEE Symposium Computational Intelligence in Bioinformatics and Computational Biology* (*CIBCB'09*), 2009, 193-199. [20] L. Li *et al.*, Application of the GA/KNN method to SELDI proteomics data, *Bioinformatics*, 20 (10), 2004, 1638-1640.

[21]P. Qiu *et al.*, Ensemble dependence model for classification and prediction of cancer and normal gene expression data, *Bioinformatics*, *21* (14), 2005, 3114-3121.

[22] H. A. L. Kiers, Towards a standardized notation and terminology in multiway analysis, *Journal of Chemometrics*, 14 (3), 2000, 105-122.

[23] L. De Lathauwer, B. De Moor, and J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal on Matrix Analysis and Applications*, *21* (4), 2000, 1253-1278.

[24] C. A. Andersson and R. Bro, The N-way Toolbox for MATLAB, *Chemometrics & Intelligent Laboratory Systems*, *52* (1), 2000, 1-4.

[25] http://www.models.life.ku.dk/source/nwaytoolbox/