

**Tensor decomposition-based feature extraction for noninvasive diagnosis of
melanoma from the clinical color image**

Ante Jukić^{a*}, Ivica Kopriva^a, Andrzej Cichocki^{b,c}

^a Division of Laser and Atomic Research and Development, Ruđer Bošković Institute,
Bijenička cesta 54, P.O. Box 180, 10002 Zagreb, Croatia (e-mail: ajukic@irb.hr;
ikopriva@irb.hr)

^b Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN2-1
Hiroshawa, Wako-shi, Saitama, 351-0198, Japan (e-mail: cia@brain.riken.jp)

^c Warsaw University of Technology and Systems Research Institute, PAN, Poland

* Corresponding author: e-mail: ajukic@irb.hr; phone: +385-1-4571-241; fax: +385-1-
4680-104

Abstract

We propose a feature extraction method for noninvasive diagnosis of melanoma based on tensor decomposition of the clinical color image of skin lesion. Extracted features are elements of the core tensor in the corresponding Tucker3 model, and represent spatial-spectral profile of the lesion. In contrast to majority of methods that exploit either texture or spectral diversity of the tumor only, this method simultaneously captures spatial and spectral characteristics. The proposed procedure is demonstrated on a problem of noninvasive diagnosis of melanoma from cost-effective auto-fluorescence color images of skin lesions, with overall sensitivity 82.1% and specificity 86.9%.

Key Words: Feature extraction, noninvasive diagnosis, melanoma, multispectral imaging, multidimensional signal processing, tensor decomposition, Tucker3 model.

1. Introduction

Noninvasive diagnosis of tumor is a procedure of identification and discrimination among various types of tumors by exploiting data that is not collected invasively, e.g. by biopsy. The aim of the noninvasive diagnosis is to detect malignant tumors with high accuracy and to simultaneously reduce the number of biopsies of the innocuous benign tumors. In general, methods for automated noninvasive diagnosis rely on sophisticated techniques for processing information present in the collected data. This data can be acquired by various imaging modalities, such as multispectral/hyperspectral imaging (MSI/HSI) [1], [2]. Diagnosis is usually obtained by classifying a set of features extracted from the image of the tumor. Consequently, extraction of tumor-specific features is of central importance for accurate diagnosis. That is in line with the reasoning in the machine learning community that feature extraction matters more than the classifier used [3], [4]. Aim of this paper is to present a novel method for feature extraction suitable for analysis of the MSI data, and to demonstrate it on automated noninvasive diagnosis of cutaneous melanoma from color red-green-blue (RGB) images.

Malignant melanoma is presently among the leading cancers among the white-skinned population, with rapidly increasing incidence and mortality rates over the last decades [5], [6], [7]. While advanced form of the cutaneous melanoma is still practically incurable, early diagnosis can significantly increase probability of survival. In fact, very high degree of curability can be achieved if the surgical excision is performed early enough [8]. This increased occurrence along with the high lethality of the advanced melanoma implies a demand for a simple and accurate screening test as an alternative to biopsy.

In spite of best efforts of researchers, the accuracy of the noninvasive diagnosis of the melanocytic lesions is still far from ideal and differentiating a malignant melanoma from benign melanocytic lesions without histological examination represents a challenge [9]. One of the most widely used methods for preliminary diagnosis based on visual inspection is the so called ABCDE rule. It is a semi-quantitative diagnosis scheme based on the asymmetry (A), border sharpness (B), color variation (C), number of differential structures (D) present in the lesion, and evolution (E) of the lesion. Unfortunately, it has a limited sensitivity in melanoma diagnosis [7], and overall accuracy depends on the level of expertise of dermatologists. In the case of a well-trained dermatologist it typically yields an accuracy of 75% [10]. Since the naked eye inspection achieves a low accuracy of the diagnosis, epiluminiscence light microscopy (ELM) or dermoscopy was suggested to improve performance by evaluating morphological features of lesions [11]. It was reported that the accuracy of the diagnosis obtained by dermoscopy in case of expert dermatologists is around 75-84% [5], [12]. Despite the formal training of dermatologists and availability of comprehensive atlases, interpretation of the features acquired by dermoscopy is often subjective and not reproducible, especially when performed by inexperienced clinician. Still, standard approach for classification of skin lesions in clinical practice is visual inspection by dermoscopy followed with biopsy and tissue analysis if needed [7].

To alleviate the need for highly-trained dermatologists and reduce the time required for diagnosis, procedures for computer-based automatic diagnosis are needed. The possibility to identify melanoma from visual information motivated development of methods based on image analysis. In this regard, in [9] a comparative performance analysis of twelve studies is presented, with accuracy exceeding the one attained by

visual inspection. Use of a computer-aided diagnostic system with high level of sensitivity and specificity (typically above 90%) could provide a second-opinion to the dermatologist and also reduce the number of unnecessary biopsies. Systems for automatic discrimination of skin lesions based on dermoscopic imaging are described in [9], [13], [14], [15], [16].

Opposed to inspection of morphological features, another line of research in noninvasive diagnosis of melanoma has been focused on the MSI and HSI systems, where discrimination among tumors is based on their spectral profile. In [1], a multispectral system with 10 channels in range 430-950 nm has been designed and used for automatic discrimination between benign nevus and malignant melanoma by using auto-fluorescence of the tumor. A hyperspectral system with 21 channels between 440 and 660 nm has been used in [2] for automatic discrimination among benign and malignant skin tumors on laboratory mice based on fluorescence induced by fluorophore.

While automatic systems based on dermoscopic and high spectral resolution MSI/HSI systems show great potential, it was demonstrated in [9], as well as recently in [17] and [18], that automatic diagnosis of melanoma is possible using color RGB auto-fluorescent images. Presented diagnostic methods utilizing RGB images achieved sensitivity in range 80-94% and specificity in range 46-95%. Results of the twelve studies reported in [9] as well as results from [17] and [18] are summarized in Table 1. Practical importance of these results is in demonstration that an accurate automatic diagnosis of melanoma is possible using affordable RGB auto-fluorescence imaging.

To this end, this paper proposes a method for feature extraction from multi-way data using tensor decomposition-based data analysis [19], [20]. The proposed method is

applied on low-dimensional MSI data (e.g. RGB color images). For this purpose experimental multispectral medical image is represented by a Tucker3 tensor model, and features are obtained through decomposition of a tensor. Dimensionality analysis yields that extracted features simultaneously contain spatial and spectral information about the object in the image. To account for possibly nonlinear nature of acquired MSI data, before decomposition images are nonlinearly mapped into a reproducible kernel Hilbert space (RKHS). The feature extraction scheme is demonstrated on noninvasive diagnosis of melanoma from clinical auto-fluorescent color (RGB) images, and sensitivity and specificity are estimated in a statistically rigorous manner performing two-fold stratified cross-validation (CV).

Table 1: Comparative performance analysis of fifteen studies related to digital image analysis based automated diagnoses of melanoma. Full references to first twelve studies are given in [9].

Source	Sample size	Melan. [%]	Method of analysis ¹	Sens.	Spec.
Green et al, 1991	70	7	Single set	80	91
Cascinelli et al, 1992	88	49	Single set	83	60
Claridge et al, 1992	88	49	Single set	91	69
Schindewolf et al, 1993	353	61	ten-fold CV	94	88
Green et al, 1994	164	11	Single set	89	89
Ercal et al, 1994	214	56	Single set	80	86
Schindewolf et al, 1994	404	59	ten-fold CV	90	88

¹ The method of analysis refers to a procedure used to validate statistical significance of the method. Single set means that the crossvalidation procedure was not used to assess the quality of classification, and that the reported performance was based on a result of prediction on a single test set. However, we performed a two-fold CV to obtain better estimate of the performance of our method.

Bono et al, 1996	43	42	Single set	83	72
Cristofolini et al, 1997	176	20	Single set	78	46
Seidenari et al, 1998	90	34	Single set	93	95
Smith et al, 2000	60	47	Single set	86	94
Farina et al, 2000	237	28	Single set	80	46
Tabatabaie et al, 2008, [17]	160	50	Single set	82.5	92.5
Wadhawan et al, 2011, [18]	1300	30	ten-fold CV	80.7	85.6
Tabatabaie et al, evaluated herein	180	50	two-fold CV	79.9	79.1
Proposed method	180	50	two-fold CV	82.1	86.9

2. Preliminaries

This section contains a brief review of notions and basics of multi-way analysis, as well as definition of nonlinear mappings related to kernels. Also, interpretation of MSI as a tensor is presented, with focus on low-dimensional MSI such as color RGB image. We will denote tensors by underlined bold letters (e.g. $\underline{\mathbf{X}}$), matrices by bold capital letters (e.g. \mathbf{X}), vectors by bold lowercase letters (e.g. \mathbf{x}) and scalars as lowercase italic letters (e.g. x).

2.1. Tensor Algebra and Tucker3 Model

Tensors are generalization of matrices and vectors. A tensor can be represented as a multi-way array with arbitrary number of indices, e.g. an N -mode tensor has N indices. For clarity, and motivated by application in RGB image analysis, we will mostly focus on 3-way tensors. In this paper image in RGB color space is represented as a 3-way

tensor $\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$, that consists of elements $\left\{ x_{i_1 i_2 i_3} \in \mathbb{R}_{0+} \right\}_{i_1, i_2, i_3=1}^{I_1, I_2, I_3}$, where \mathbb{R}_{0+} denotes the

set of nonnegative real numbers. Each index in tensor is called a way or mode, and number of levels on a mode represents dimension of that mode, e.g. dimension of mode-1 is I_1 . This is the standard notation used in multi-way analysis [21]. RGB image $\underline{\mathbf{X}}$ is a set of $I_3 = 3$ spectral images, corresponding to red, green and blue color channels [22]. Each spectral image consists of $I_1 \times I_2$ pixels, where each element $x_{i_1 i_2 i_3}$ represents brightness intensity. Consequently, two modes of tensor $\underline{\mathbf{X}}$ are used for the spatial information, i.e. rows and columns of the color image, and third mode is for the spectral band. Sub-tensors can be defined by fixing some of the indices for a general N -way tensor. In case of 3-way tensor we can have slices and fibers. Tensor slice is a matrix (two dimensional array) obtained by fixing a single index in a 3-way tensor. For example, frontal slice of 3-way tensor $\underline{\mathbf{X}}$ is obtained by fixing index i_3 and denoted as $\mathbf{X}_{::i_3}$ or \mathbf{X}_{i_3} . An n th mode vector of a tensor $\underline{\mathbf{X}}$ is called n th mode fiber, and it is an I_n dimensional vector obtained by fixing all indices except i_n . For example, a mode-3 fiber of $\underline{\mathbf{X}}$ at position (i_1, i_2) is a vector denoted as $\mathbf{x}_{i_1 i_2}$ or $\mathbf{x}_{i_1 i_2}$. [19]. For illustration, fibers and slices for 3-way tensor are displayed on Fig. 1.

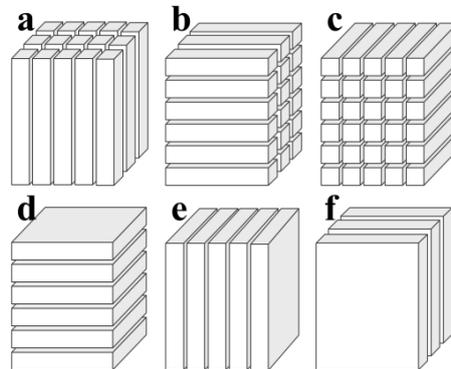


Fig. 1: Fibers in (a) mode-1; (b) mode-2; (c) mode-3; and (d) horizontal; (e) lateral; and (f) frontal slices.

Tensor is often transformed to a matrix and vice-versa. Procedure of rearranging the elements of tensor into a matrix is known as unfolding or matricization. The mode- n unfolding of tensor $\underline{\mathbf{X}}$ produces matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times \prod_{k, k \neq n} I_k}$ which consists of mode- n fibers stacked in the matrix as columns. There are various possibilities for ordering of mode- n fibers into columns of $\mathbf{X}_{(n)}$, but particular ordering is not important, as long it is consistent through all computations [23], [19]. Useful notion is the n -rank of a tensor, defined as the dimension of the space spanned with columns of $\mathbf{X}_{(n)}$. If N -way tensor has n -ranks equal to J_1, J_2, \dots, J_N we say that it is a rank- (J_1, J_2, \dots, J_N) tensor. Mode- n product of a tensor $\underline{\mathbf{X}}$ and matrix \mathbf{M} is defined when number of columns of matrix is equal to the dimension of the tensor in mode n . It results in a new tensor $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_n \mathbf{M}$, expressed in unfolded form as $\mathbf{Y}_{(n)} = \mathbf{M} \mathbf{X}_{(n)}$ [19]. For example, mode-2 product of a 3-way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and a matrix $\mathbf{M} \in \mathbb{R}^{D \times I_2}$ is a 3-way tensor $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_n \mathbf{M} \in \mathbb{R}^{I_1 \times D \times I_3}$, calculated element-wise as

$$y_{i_1, d, i_3} = \sum_{i_2=1}^{I_2} x_{i_1, i_2, i_3} \cdot m_{d, i_2} .$$

Tensor decompositions are valuable tools for analysis of multi-way data, and are used in image and signal analysis, neuroscience, chemometrics and psychometrics [19], [23]. Basic model for decomposition of general N -way tensor is Tucker model. The Tucker decomposition for 3-way tensors is often referred to as Tucker3 decomposition, and is illustrated in Fig. 2. The sample tensor $\underline{\mathbf{X}}$ is modeled by a core tensor of reduced

dimension $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$, and three factors $\{\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}\}_{n=1}^3$. The Tucker3 model of tensor $\underline{\mathbf{X}}$ on level of a single element can be expressed as:

$$x_{i_1 i_2 i_3} \approx \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} g_{j_1 j_2 j_3} \cdot a_{i_1 j_1}^{(1)} \cdot a_{i_2 j_2}^{(2)} \cdot a_{i_3 j_3}^{(3)} \quad (1)$$

where $a_{i_n j_n}^{(n)} \in \mathbb{R}$ is element of the matrix $\mathbf{A}^{(n)}$ on the position (i_n, j_n) , and J_1, J_2 , and J_3 are dimensions of the core tensor. The model (1) can be expressed in more compact form as

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

In practice, it is common that the core tensor $\underline{\mathbf{G}}$ is smaller than the original tensor $\underline{\mathbf{X}}$, i.e. $J_n < I_n$, and if not stated otherwise we will assume this is fulfilled in remainder of the paper. Then tensor $\underline{\mathbf{X}}$ is modeled by a low-rank tensor on the right side of (2), and dimensions of the core tensor $\underline{\mathbf{G}}$ determine the mode- n ranks for the low-rank model. By representing the original tensor $\underline{\mathbf{X}}$ with a rank- (J_1, J_2, \dots, J_n) tensor consisting of the core and factors we effectively perform model reduction. This is the reason why low-rank approximation is of practical interest. Decomposition (2) of $\underline{\mathbf{X}}$ can be seen as a set of directional bases $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}$ in modes 1, 2 and 3, and a set of weights contained in $\underline{\mathbf{G}}$. The elements in the core can then effectively be considered as features describing a sample contained in tensor $\underline{\mathbf{X}}$. For example, when $\underline{\mathbf{X}}$ represents an RGB image the elements of the core connect spatial bases (in modes 1 and 2) and spectral basis in mode 3.

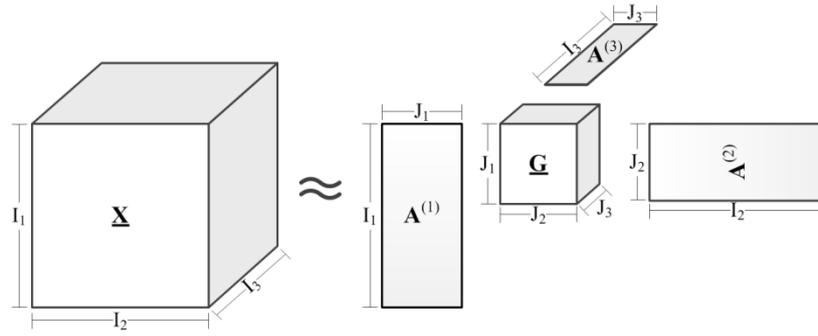


Fig. 2: Tucker3 decomposition of a 3-way tensor

The Tucker3 model is flexible in modeling complex interactions within the data because the core tensor allows interplay of factors from all modes [24]. However, the Tucker decomposition is in general not unique. Proof of uniqueness exists in the special case when the core tensor is diagonal (i.e. $g_{j_1 j_2 j_3} \neq 0$ only for $j_1 = j_2 = j_3$), resulting in the canonic polyadic decomposition (CPD), known also as parallel factor analysis PARAFAC model, that is unique (up to scale and permutation) under mild conditions [25]. Constraints such as nonnegativity, sparsity and orthogonality are commonly imposed on the factor matrices and the core tensor in order to attain unique decomposition of the data tensor $\underline{\mathbf{X}}$ into Tucker model (2). Particular constraints are used to narrow down the solution space and yield a virtually unique decomposition.

Exemplary orthogonality constrained decomposition is the higher order singular value decomposition (HOSVD), a multilinear generalization of the matrix singular value decomposition² (SVD) [26]. The HOSVD implements decomposition of the data tensor to its Tucker model (2) with $J_n = I_n$ while imposing orthogonality for factor matrices $\mathbf{A}^{(n)}$, i.e. $\mathbf{A}^{(n)T} \mathbf{A}^{(n)} = \mathbf{I}$. The factor matrix in n -th mode is estimated by performing

² Note that the HOSVD does not give a low-rank model of the data tensor. It is an exact decomposition in a set of orthonormal bases.

SVD of the n -th mode unfolding of the data tensor, i.e. $\mathbf{A}^{(n)}$ is equal to the left singular vectors of the unfolded tensor $\mathbf{X}_{(n)}$. Then the core tensor is calculated as

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

As noted earlier, we are mainly interested in low rank approximation of the original tensor. In this regard, algorithms usually seek for a rank- (J_1, J_2, J_3) tensor that is optimal approximation of the data tensor $\underline{\mathbf{X}}$ in the least-squares sense with $J_n < I_n$. This can be performed by minimizing the Frobenius norm between the data tensor $\underline{\mathbf{X}}$ and its low-rank Tucker3 model:

$$D \left[\underline{\mathbf{X}} \left\| \left(\underline{\mathbf{G}}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)} \right) \right\| \right] = \left\| \underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \right\|_F^2 \quad (4)$$

while at the same time imposing orthogonality constraints on factors $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}$.

Actually, it is enough to find factor matrices that maximize the function $g(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}) = \left\| \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \times_3 \mathbf{A}^{(3)T} \right\|_F^2$, and the core tensor is given with (3)

[27]. Several algorithms address this problem, such as the higher order orthogonal iteration (HOOI) [27], and algorithms based on the Newton method on Grassmannians (NG) [28], [29]. Both HOOI and qNG seek for the best multilinear rank- (J_1, J_2, J_3) approximation with orthogonality constraints on factors, and the HOOI has proved to be a "workhorse" algorithm in various applications of tensor decompositions [20]. However, note that minimization of (4) is a non-convex optimization problem with multiple local minima [27], [29]. Additionally, algorithms based on alternating least squares (such as HOOI) can even have problems in finding stationary points. Still, in most situations orthogonality constrained best low-rank approximation produces a virtually unique decomposition [30].

Another way to obtain a low-rank approximation of the data tensor $\underline{\mathbf{X}}$ is to use truncated HOSVD (trHOSVD) decomposition. This decomposition can be easily computed in two steps: (i) For $n=1, 2, 3$ unfold tensor to a matrix $\mathbf{X}_{(n)}$, and calculate the standard SVD, i.e. $\underline{\mathbf{X}}_{(n)} = \mathbf{U}^{(n)}\mathbf{S}^{(n)}\mathbf{V}^{(n)T}$. The orthogonal factors $\mathbf{A}^{(n)}$ are J_n leading left singular vectors of $\mathbf{X}_{(n)}$, i.e. first J_n columns of $\mathbf{U}^{(n)}$; (ii) Compute the core tensor using (3). For a general N -way tensor, this decomposition is calculated using N matrix SVD's and can be very fast, even for large-scale problems. The truncated HOSVD (as opposed to the matrix SVD) does not give the optimal rank- (J_1, J_2, J_3) decomposition, but it is a good approximation with error bound given in [26]. It has proved to be very useful in various applications since it is computationally light and does not have problems with local minima. For comparison, we performed several experiments using HOOI and qNG, but did not obtain better overall performance in terms of discrimination of malignant vs. benign skin lesions. Therefore, the experimental section of this paper we used trHOSVD for decomposition of tensors.

2.2. Dimensions of the core tensor

One of the essential problems when describing data $\underline{\mathbf{X}}$ using Tucker model is the size of the core tensor. Since in our approach elements of the core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$ will represent features, the number of features per sample directly depends on dimensions of the core tensor, i.e. rank of the Tucker3 model. Estimation of the rank of a tensor is essential but still very difficult problem in multi-way analysis. This problem is also known as intrinsic dimensionality problem or model order selection, and several methods have been developed in recent years with various applications within signal

processing and data analysis [31]. Estimation of rank in each of the modes can be based on the gap in a sequence of parameters, e.g. eigenvalues of the covariance or correlation matrix. Methods in this group are GAP [32], [33], RAE (ratio of adjacent eigenvalues) [34], SORTE (second order statistic of the eigenvalues) [33], [32], RAESORTE [34], and EIF (empirical indicator function) [35], [36]. Former methods estimate mode- n rank from the unfolded tensor $\mathbf{X}_{(n)}$ by searching for dominant eigenvalues of the Gram matrix $\mathbf{X}_{(n)}\mathbf{X}_{(n)}^T$ according to some criterion, e.g. gap between adjacent eigenvalues or ratio of adjacent eigenvalues. Also, information theoretic avenue can be used, resulting in model order selection methods based on Akaike's information criterion (AIC) [37], Kullback-Leibler information criterion (KIC) [38], and minimum description length (MDL) [39]. When applied to tensors, these methods also analyze corresponding unfolded tensor $\mathbf{X}_{(n)}$ resulting in estimation of intrinsic dimension in mode- n . Other approaches can be derived using Bayesian estimation, an example being automatic relevance determination (ARD) method for multi-way models [40]. The ARD method directly estimates dimensions of the core for the Tucker model.

All of the mentioned methods have been used in diverse applications, such as detection of number of clusters [33], dimensionality reduction using principal component analysis (PCA) [41], and the choice of number of sources in linear model [42], as well as in feature extraction [20]. However, due to different theoretical assumptions as well as noise distributions in the real-world data, they often yield significantly different estimations when applied to the same problem [34].

Another related problem arises from the fact that clinical images of skin lesions used in our experiments do not have equal spatial dimensions, i.e. dimensions in modes 1 and 2 are not the same for all images of benign nevus and melanoma. This causes large

variation among estimated dimensions of the core tensor for each sample. Therefore, results of described approaches for mode- n rank estimation will be presented and commented in more detail in the experimental section of the paper.

2.3. *Nonlinear mapping*

Representation of the MSI by the multilinear model (2) can be questioned on the basis of a number of arguments, see [43], [44] for multi- and hyperspectral remote sensing. Taking into account possible nonlinear relationships in the data is expected to improve final classification accuracy. This assumption is supported by the Cover's theorem [45]. In a nutshell, this theorem states that a set of samples that are not linearly separable in the original (low-dimensional) space, are more likely to be linearly separable after being nonlinearly mapped into some higher-dimensional space. Consequently, we expect that linear model holds with high probability for the data nonlinearly mapped in a higher-dimensional space.

In this regard, an implicit nonlinear mapping of original image into RKHS is performed, based on kernel techniques. Description of kernel-based nonlinear mapping will be given herein, but for detailed theoretical treatment interested reader is referred to [46].

Let $k : S \times S \rightarrow \mathbb{R}$ be a real, positive definite kernel, and $S \subseteq \mathbb{R}^l$ a nonempty set. Then map $\Phi : S \rightarrow \mathbb{R}^S$ can be defined as $\mathbf{s} \mapsto \Phi(\mathbf{s}) := k(\cdot, \mathbf{s})$, where $\mathbb{R}^S := \{f : S \rightarrow \mathbb{R}\}$ is a set of all functions from S to \mathbb{R} . In this way each input sample $\mathbf{s} \in S$ is mapped into a function $\Phi(\mathbf{s})$ defined on input space, which is possibly an infinite dimensional object. However, it is possible to obtain an approximation of the map Φ by evaluating it only on a set of points, since the mapped data is embedded in a subspace of RKHS [46], [47].

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_D\} \subset S$ be a set of points in the input space. Then we define mapping

$\Phi_D : S \rightarrow \mathbb{R}^D$ as

$$\mathbf{s} \mapsto \Phi_D := [k(\mathbf{v}_1, \mathbf{s}), \dots, k(\mathbf{v}_D, \mathbf{s})]^T$$

and call it the empirical kernel map with respect to a set of points $\{\mathbf{v}_1, \dots, \mathbf{v}_D\}$ [46]. In this way nonlinearly mapped data is projected onto a D dimensional subspace of RKHS, while at the same time all calculations are performed on the original data from the input set $S \subseteq \mathbb{R}^I$. It is obvious that $D \gg I$ must hold in order to exploit benefits stipulated by the Cover's theorem. Complex problem of selection of points $\{\mathbf{v}_1, \dots, \mathbf{v}_D\}$ is known as basis selection. Various basis selection methods were previously proposed, such as kernel PCA [48] and feature vector selection [49]. In our experiments we have used *k-means* clustering to estimate basis vectors as centroids of D clusters, as in [47].

In this paper we perform nonlinear mapping of the sample tensor $\underline{\mathbf{X}}$ along the mode 3, i.e. each mode-3 fiber (corresponding to a single pixel composed of intensities at wavelengths corresponding to red, green and blue colors), will be nonlinearly mapped to a new D dimensional vector, $\mathbf{x}_{i_1 i_2} \mapsto \Phi_D(\mathbf{x}_{i_1 i_2}) = [k(\mathbf{v}_1, \mathbf{x}_{i_1 i_2}), \dots, k(\mathbf{v}_D, \mathbf{x}_{i_1 i_2})]^T$. Using this procedure we can define nonlinear map for tensor objects

$$\begin{aligned} \Phi_D : \mathbb{R}^{I_1 \times I_2 \times I_3} &\rightarrow \mathbb{R}^{I_1 \times I_2 \times D} \\ \underline{\mathbf{X}} &\mapsto \Phi_D(\underline{\mathbf{X}}) \end{aligned}$$

that maps a 3-way tensor $\underline{\mathbf{X}}$ to a new 3-way tensor $\Phi_D(\underline{\mathbf{X}})$ by replacing each I_3 -dimensional mode-3 fiber $\mathbf{x}_{i_1 i_2}$ with its nonlinearly mapped image, a D dimensional

mode-3 fiber $\Phi_D(\mathbf{x}_{i_2})$ ³. As mentioned, basis points $\{\mathbf{v}_1, \dots, \mathbf{v}_D\}$ are estimated by *k-means* clustering of mode-3 fibers $\{\mathbf{x}_{i_2}\}$ into D clusters. Now, a low-rank tensor model can be used to represent nonlinearly mapped data as:

$$\Phi_D(\underline{\mathbf{X}}) \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)}. \quad (4)$$

Since $D \gg I_3$, the core tensor in the low-rank model (4) can have greater dimension J_3 than the core tensor in the low-rank model (2). In experimental section we have used most common kernel known as the Gaussian kernel or Gaussian radial basis function $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \sigma^2)$. Additionally, other kernels such as Laplacian $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\| / \sigma^2)$ and polynomial $k(\mathbf{x}, \mathbf{y}) = (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^d$ were used, but did not yield better results than the Gaussian kernel.

3. Proposed feature extraction method

The proposed feature extraction scheme is summarized in Algorithm 1. Let us assume that our dataset consists of K samples (e.g. RGB images of skin lesions), each represented by a 3-way tensor $\underline{\mathbf{X}}^{(k)}$, $k=1, \dots, K$. Also, let a class label $c^{(k)}$ be assigned to each sample, e.g. $c^{(k)} = 1$ if k -th sample belongs to class 1 (benign nevus), and $c^{(k)} = 2$ if it is from class 2 (melanoma). The proposed feature extraction procedure consists of nonlinear mapping of each sample $\underline{\mathbf{X}}^{(k)}$, followed by tensor decomposition using trHOSVD algorithm. Core tensor is calculated directly from the nonlinearly mapped

³ This is actually a slight abuse of notation, since map Φ_D was originally defined for vectors, not tensors.

The notation $\Phi_D(\underline{\mathbf{X}})$ can be interpreted in the following sense: we take mode-3 unfolded tensor $\mathbf{X}_{(3)}$, perform nonlinear map of each column, and then rearrange columns (tensorise) back to a 3-way tensor.

data sample by projecting the mapped tensor $\Phi_D(\underline{\mathbf{X}}^{(k)})$ onto factors $\mathbf{A}_k^{(1)}$, $\mathbf{A}_k^{(2)}$ and $\mathbf{A}_k^{(3)}$ obtained by the trHOSVD. Finally, features are obtained by rearranging elements of the core $\underline{\mathbf{G}}^{(k)}$ into a vector $\mathbf{g}^{(k)}$ ⁴. Note that feature extraction is performed on each sample $\underline{\mathbf{X}}^{(k)}$ separately, but with the same values of parameters.

From the feature extraction point of view it is important that the core tensor $\underline{\mathbf{G}}^{(k)}$ connects two spatial modes with spectral mode of the RGB image. In this way, the core tensor represents a spatial-spectral profile of the lesion present in the image $\underline{\mathbf{X}}^{(k)}$ while each element $g_{j_1 j_2 j_3}^{(k)}$ of the core tensor can be viewed as a spatial-spectral characteristic of the lesion. Therefore it is conjectured that the core tensor viewed as a set of features can be used for robust and accurate noninvasive diagnosis of melanoma.

Proposed feature extraction method differs from the one recently published in [20], where simultaneous tensor decomposition is performed for all samples in the training set to find projection filters that are applied on unknown data sample. Consequently, it is required that all images in the data set have the same dimensions (both in spatial and spectral modes). However, this requirement is not easily met in practice, and method proposed here allows greater flexibility when acquiring clinical images for analysis. Also, separate decomposition of each sample enables us to easily update training set when new labeled sample is available, since it is decomposed independently of other samples previously present in the training set.

⁴ The vectorization procedure was performed by stacking mode-1 fibers of the core tensor. However, ordering of the elements of the core tensor is not important, as long as the same procedure is used for each sample.

Algorithm 1: Proposed method for feature extraction

Input: sample (autofluorescent RGB image) represented by a 3-way tensor

$$\underline{\mathbf{X}}^{(k)} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$$

Parameters: kernel map (σ, D) , core size (J_1, J_2, J_3)

Output: features $\hat{\mathbf{g}}^{(k)} \in \mathbb{R}^{J_1 \cdot J_2 \cdot J_3}$

1) Nonlinear mapping

$$\underline{\mathbf{X}}^{(k)} \mapsto \Phi_D(\underline{\mathbf{X}}^{(k)})$$

2) Tensor decomposition into Tucker3 model (by trHOSVD)

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

3) Feature extraction

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

4) Vectorization of the core tensor

Classification system built around the proposed feature extraction method is displayed in Fig. 3. Feature extraction is performed for each sample in the data set with the same values of parameters σ , D , and J_n , $n=1,2,3$. In this way all of the K core tensors have the same dimensions and for each sample we obtain $L = J_1 \cdot J_2 \cdot J_3$ features used for classification. Validation step consists of partitioning the K samples $(\mathbf{g}^{(k)}, c^{(k)})$ into two sets: training set with K_{TR} samples $(\mathbf{g}_{TR}^{(k)}, c_{TR}^{(k)})$ and test set with K_{TE} samples $(\mathbf{g}_{TE}^{(k)}, c_{TE}^{(k)})$. In experimental section we used linear support vector machine (linSVM) classifier, and nonlinear SVM with the Gaussian kernel (rbfSVM) and polynomial kernel (polySVM), although other classifiers can be used as well.

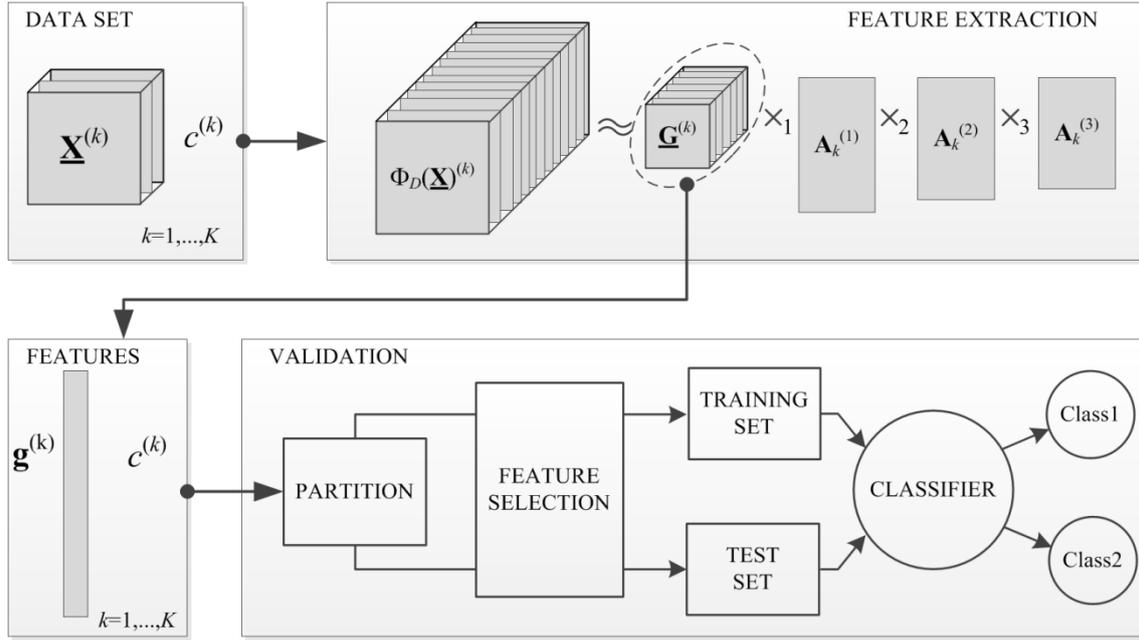


Fig. 3: Flow chart diagram for classification system based on the proposed method

In order to obtain a reliable estimate of classification performance on an independent data set, we performed a two-fold cross-validation (CV). In general, the cross-validation procedure uses a part of the data set to train the classifier and remaining part to test it [4]. In K -fold CV the data set is divided into K equally sized parts so that each contains the same proportion of each type of class labels. Then in the k -th step classifier is trained on $K-1$ parts of the data set and tested on k -th part of the data set. We repeat this for $k=1, \dots, K$ and then combine the K obtained estimates to give a final estimation of the classification performance. For example, in case of a two-fold CV, samples are divided into two equal sized data sets, each consisting of 50% of samples from class 1, and 50% of samples from class 2. In order to obtain a more reliable estimate of prediction performance, CV is repeated several times and the validation results are averaged.

Even though the number of features L is significantly smaller than the number of elements in each of the tensors $\Phi_D(\underline{\mathbf{X}}^{(k)})$, it can still be large. Thus it is reasonable to

perform a feature selection (FS) step to identify significant features obtained by the proposed method. We ranked features based on the Fisher score [20], albeit other information criteria can also be used for ranking [50]. For each feature Fisher score is calculated using only samples in the training set as:

$$\varphi(i) = \frac{K_{TR}^1 (\bar{g}_{TR,i}^1 - \bar{\bar{g}}_{TR,i})^2 + K_{TR}^2 (\bar{g}_{TR,i}^2 - \bar{\bar{g}}_{TR,i})^2}{\sum_{k=1}^{K_{TR}} (g_{TR,i}^{(k)} - \bar{g}_{TR,i}^{(k)})}$$

where K_{TR}^c denotes the number of training samples belonging to the c -th class ($c=1,2$), $g_{TR,i}^{(k)}$ is i -th feature of k -th training sample, $c_{TR}^{(k)}$ label of the k -th sample, $\bar{g}_{TR,i}^c$ is the mean of the i -th feature calculated over training samples from class c , and $\bar{\bar{g}}_{TR,i}$ is the mean of the i -th feature calculated over all training samples. Larger value of Fisher score means that the feature is more discriminative.

4. Experimental results

Proposed method for feature extraction is demonstrated on noninvasive diagnosis of melanoma. Autofluorescent RGB images of benign nevus and superficial melanoma were made available by courtesy of [51], [52], [53], [54]. All calculations in the experiments presented below have been carried out in MATLAB environment on a desktop computer with 2.4GHz clock speed and 4GB of RAM. Before feature extraction, each RGB image has been cropped around the tumor region to prevent extracting features from the non-tumor related areas of the image. A more sophisticated approach would be to use segmentation to extract the part of the image corresponding to lesion, as in [17], [18], but it would introduce additional computational cost and compromise reproducibility of the method. We performed experiments on cropped

images to show that the proposed method is robust when a small amount of image is occupied by healthy skin. Also, no hair removal was performed, since the images did not contain a large amount of hair. The data set for the experiments consisted of 180 RGB clinical images with 90 images of benign nevus and 90 images of superficial melanoma. Few examples of analyzed images are displayed in Fig. 4. Since the data was collected from various archives we had no control over parameters of the acquisition. Images in the data set differed in size and aspect ratio, depending on the size of the lesion and parameters of the acquisition procedure, where the smallest image had dimensions 38x36 pixels to largest with 912x671 pixels. Thus, mode-1 and mode-2 dimensions were in the range $I_{1,\min} = 38$ to $I_{1,\max} = 912$ and $I_{2,\min} = 36$ to $I_{2,\max} = 671$.

To demonstrate performance of the proposed method for automated diagnosis of melanoma we compared it against approach based on independent component analysis (ICA) [17]. This approach uses separate spectral and spatial features of skin lesion present in the image. Spectral features for each sample are calculated as mean and variance of red, green and blue channel from RGB image of the lesion. In order to obtain spatial characteristics of the lesion, original RGB image is converted to grayscale. Then spatial features are extracted from the grayscale image using ICA-learned filters. Firstly, a set of patches (16x16 pixels) was selected from the images in the training set. The preprocessing step was performed, including mean removal for each patch, followed with data whitening and dimensionality reduction using PCA. After preprocessing, as described in [17], ICA was used to obtain spatial filters as rows of the unmixing matrix obtained from a set of patches, for nevus and melanoma independently. Specifically, 100 ICA filters were obtained for melanoma and 100 filters for benign nevus, resulting in a filter bank with 200 filters. Finally, spatial features were

calculated as the energy of the response of each grayscale image to a filter. The original paper [17] used FastICA algorithm [55] with *tanh* nonlinearity to estimate spatial filters. However, it is important to address the problem of stability of the FastICA algorithm. Since the algorithm uses a random initial point for the optimization of the cost function, it is possible to end up in different local minima due to a different starting point. To alleviate this problem we used *Icasso* software package [56] that runs FastICA algorithm M times with different initial points, and the unmixing matrix (i.e. filters) is obtained by clustering the estimated unmixing matrices from each run⁵. In our experiments for ICA-based method we used $M=30$.

As reported in [9], the ICA-based feature extraction scheme performed better than twelve studies compared here in Table 1. Thus, we implemented and tested the described procedure on the same data set as our method. To make comparison fair, ICA approach has been applied on the same images as the proposed tensor-based feature extraction method, i.e. no hair removal preprocessing has been done. The reason is that presence of hair was very limited and we believe that good feature extraction method should exhibit robustness without requiring special type of preprocessing. Also, no image segmentation has been performed to separate lesion from the healthy skin. The reason is that the images of the melanoma and nevus were already cropped around the lesion area and, thus, area occupied by the healthy skin was small.

⁵ Although we used *Icasso*, the obtained results did not differ significantly compared with application of FastICA without stability analysis.

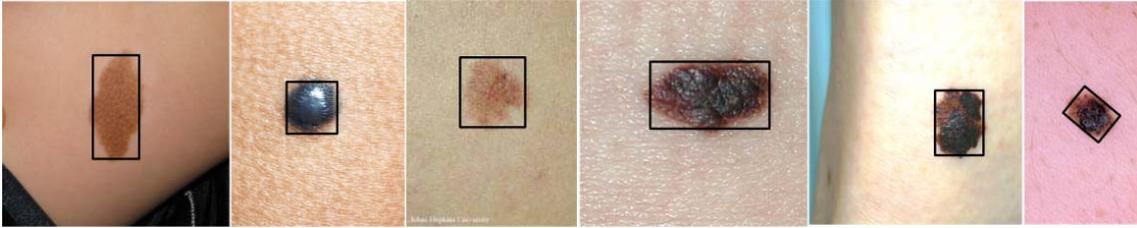


Fig. 4: Examples of clinical RGB images: three benign nevi (left) and three malignant melanomas (right). Border of the region cropped for analysis is also displayed. (Color online)

The classification performance was measured in terms of sensitivity and specificity, which are standard metrics for binary classification. Higher sensitivity and specificity mean better classification performance. For our experiments accuracy of classification can be calculated as the arithmetic mean of the two.

For a start, we tested all methods for automatic rank selection (i.e. dimensions of the core tensor) described in section 2.3. Estimations of rank in mode-1 and mode-2 for images in the data set using MDL, ARD and GAP method are displayed in Fig. 5. As expected, estimation results highly depend on the dimension of the clinical image. This is not acceptable since the proposed method requires that the core tensor has the same dimensions for all samples. Additionally, different methods produced very different results for the same image. As can be seen in Fig. 5, GAP method estimation is very close to the dimension of the image, while ARD method gives very small estimation. Results for other methods fall in between these two extreme cases (only MDL is displayed in the image).

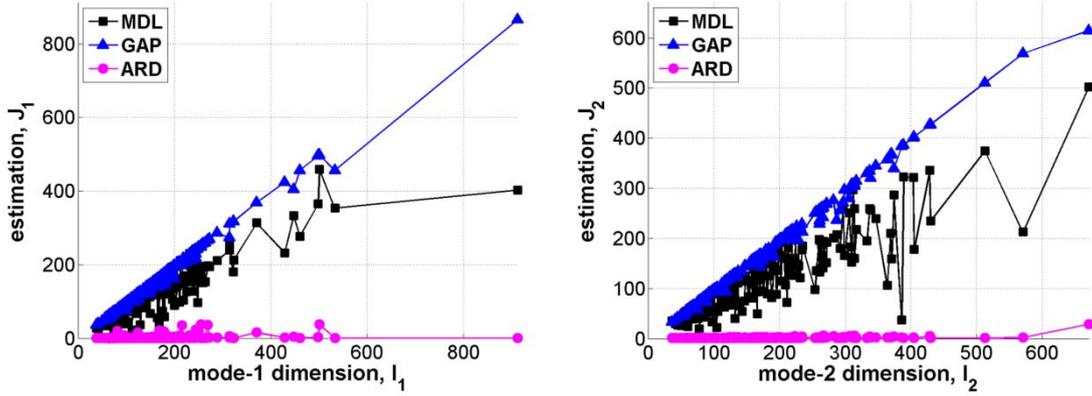


Fig. 5: Estimation of mode-1 and mode-2 rank for images in the data set using MDL (square), GAP (triangle) and ARD (circle): estimated rank \hat{J}_1 vs. mode-1 dimension I_1 (left); and estimated rank \hat{J}_2 vs. mode-2 dimension I_2 (right). (Color online)

Due to this we could not find optimal dimensions of the core tensor either by a single method or as consensus of several methods. Thus, we employed wrapper-like procedure for determination of the optimal dimensions: performance of the classification was estimated through cross-validation for different dimensions of the core tensor, and dimensions that produced the best result were selected as optimal⁶. To be more precise, feature extraction was performed with dimensions of the core tensor from a grid defined by $1 \leq J_1 \leq I_{1,\min}$, $1 \leq J_2 \leq I_{2,\min}$ and $1 \leq J_3 \leq D$, and classification performance was estimated using CV with 30 random partitions. Dimensions that produced the smallest CV error were selected for further analysis. Experiments also showed that feature extraction method is not sensitive to choice of parameter D , as long as it is large enough (in comparison with $I_3 = 3$) so we fixed it to $D=40$, while kernel parameter was

⁶ To measure performance we used rbfSVM as classifier, with parameter σ_C tuned through CV.

optimized to $\sigma=0.8$ through CV. These values for D and σ were used in further experiments.

The best classification performance was achieved for dimensions of the core tensor $10 \times 12 \times 12$ and $11 \times 12 \times 10$. We selected these dimensions as optimal, although there were several other combinations with just slightly inferior performance. The proposed method was thoroughly evaluated for core tensor with selected optimal dimensions through two-fold CV with 100 random partitions. The obtained results can be found in Table 2 in terms of sensitivity and specificity (mean \pm standard deviation). Overall, equal accuracy of 84.5% using the whole core tensor was obtained with dimensions $10 \times 12 \times 12$ and $11 \times 12 \times 10$.

Since the number of features for selected dimensions of the core was still quite large (above thousand), a feature selection procedure was utilized. Elements of the feature vector were ranked in a descending order based on their Fisher scores, and only significant features were selected for classification. Results of this analysis for core sizes $10 \times 12 \times 12$ and $11 \times 12 \times 10$ are displayed in Fig. 6. It can be seen that in the beginning adding more features improved performance significantly, but after a certain number accuracy was not drastically increased. Thus, selection step can be used to reduce the number of features without affecting the overall performance of the method. For examples in Fig. 6 we obtained almost the same performance with only 40% of the most significant features.

The ICA-based approach yielded sensitivity of 79.9% and specificity of 79.1%, exhibiting a significant decrease in performance compared to the original paper [17]. The reasons for that could possibly be a larger and more diverse dataset used here, and the rigorous two-fold CV used in the experiments. Thus, results obtained by tensor

factorization approach proposed here outperform the ones produced by the ICA-based approach. Performance of the proposed approach is also comparable with the one reported in [57], that, however, has been obtained using melanoma specific marker expressions extracted from biopsy tissues, where (depending on the type of gene marker) area under the curve varied between 0.94 and 0.5.

Table 2: Performance of the proposed method and ICA-based method estimated by two-fold CV with 100 random partitions (in %, mean±standard deviation). Parameters d (degree of the polynomial) and σ_c were selected through CV.

Method	linSVM		polySVM		rbfSVM	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
Core: 10x12x12	20.0±6.2	96.2±3.1	8.3±4.1	99.4±1.1	81.8±5.2	87.2±4.1
Core: 11x12x10	18.4±6.7	95.0±3.2	8.2±4.1	99.8±0.6	82.1±5.2	86.9±4.2
ICA	79.9±5.6	79.1±6.8	57.2±13.2	58.9±17.4	75.7±7.4	72.4±6.5

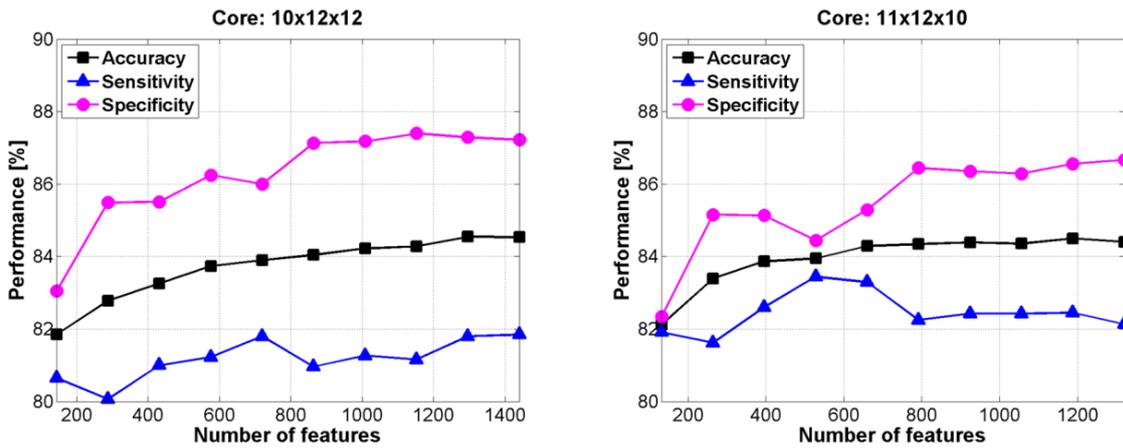


Fig. 6: Classification performance vs. number of features. Feature selection based on Fisher score was applied on the core tensor with dimensions: 10x12x12 (left) and 11x12x10 (right). Rbf SVM classifier was used, with width of the kernel tuned through CV. (Color online)

5. Conclusions

Tensor representation of a multi-spectral image enables extraction of features that simultaneously capture spatial and spectral characteristics of object present in the image. Multimodal features are obtained by decomposing the image according to the Tucker3 model, and the proposed approach uses computationally light and scalable trHOSVD algorithm. As demonstrated, the proposed method for noninvasive diagnosis of melanoma from auto-fluorescent RGB images acquired by cost-effective RGB cameras compares favorably with the state of the art results reported in the literature. Automated system based on a cost-effective, and nowadays ubiquitous, RGB camera could possibly be used for preliminary screening in distant areas and areas without appropriate medical care, e.g. by implementing software with diagnostic algorithm on some of the widespread smartphone platforms or as application-specific low-cost device.

Conflict of interest statement

None.

Acknowledgments

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

References

- [1] M. Elbaum et al., "Automatic differentiation of melanoma from melanocytic nevi

- with multispectral digital dermoscopy: a feasibility study," *J. Am. Acad. Dermatol.*, vol. 44, no. 2, pp. 207-218, 2001.
- [2] S. G. Kong and L-J Park, "Hyperspectral Image Analysis for Skin Tumor Detection," in *Augmented Vision Perception in Infrared: Algorithms and Applied Systems*, R. I. Hammoud, Ed. London: Springer-Verlag, 2009, ch. 7.
- [3] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for Cancer Classification Using Support Vector Machine," *Machine Learning*, vol. 46, no. 1-3, pp. 389-422, 2002.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*: Springer, 2009, ch. 18.
- [5] W. Stolz et al., *Color atlas of dermoscopy*, 2nd ed. Berlin: Blackwell publishing, 2002.
- [6] A. Green et al., "Computer image analysis of pigmented skin lesions," *Melanoma Res.*, vol. 1, no. 4, pp. 231-236, 1991.
- [7] E. M. Wurm and H. P. Soyer, "Scanning for melanoma," *Aust. Prescr.*, vol. 33, pp. 150-155, 2010.
- [8] A.W. Kopf et al., "Thickness of malignant melanoma: global analysis of related factors," *J. Dermatol. Surg. Oncol.*, vol. 13, no. 4, 1987.
- [9] B. Rosado et al., "Accuracy of Computer Diagnosis of Melanoma - A Quantitative Meta-analysis," *Arch. Dermatol.*, vol. 139, pp. 361-367, 2003.
- [10] T. Schindewolf, W. Stolz, R. Albert, W. Abmayr, and H. Harms, "Classification of melanocytic lesions with color and texture analysis using digital image processing," *Anal. Quant. Cytol. Histol.*, vol. 15, pp. 1-11, 1993.

- [11] H. P. Soyer, J. Smolle, H. Kerl, and H. Steetnre, "Early diagnosis of malignant melanoma by surface microscopy," *Lancet*, vol. 330, no. 8562, p. 803, 1987.
- [12] G. Argenziano and H.P. Soyer, "Dermoscopy of pigmented lesions - a valuable tool for early diagnosis of melanoma," *Lancet Oncol.*, vol. 2, no. 7, pp. 443-451, 2001.
- [13] H. Ganster et al., "Automated Melanoma Recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 233-239, 2001.
- [14] K. Hoffmann et al., "Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy.," *Br. J. Dermatol.*, vol. 149, no. 4, pp. 801-809, 2003.
- [15] S. W. Menzies et al., "The Performance of SolarScan - An Automated Dermoscopy Image Analysis Instrument for the Diagnosis of Primary Melanoma," *Arch. Dermatol.*, vol. 141, no. 11, pp. 1388-1396, 2005.
- [16] H. Iyatomi et al., "An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Comput. Med. Imag. and Graph.*, vol. 32, pp. 566-579, 2008.
- [17] K. Tabatabaie and A. Esteki, "Independent Component Analysis as an Effective Tool for Automated Diagnosis of Melanoma," in *Proceedings of the IEEE Biomedical Engineering Conference (CIBEC'08)*, Cairo, 2008, pp. 1-4.
- [18] T. Wadhawan, N. Situ, K. Lancaster, Y. Xiaojing, and G. Zouridakis, "SkinScan: A portable library," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Chicago, IL, 2011, pp. 133-136.

- [19] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorization*. Chichester: John Wiley & Sons, 2009.
- [20] A. H. Phan and A. Cichocki, "Tensor Decomposition for Feature Extraction and Classification of High Dimensional Datasets," *IEICE Nonlinear Theory and Its Applications*, vol. 1, pp. 37-68, 2010.
- [21] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 105-122, 2000.
- [22] I. Kopriva and A. Cichocki, "Blind Multi-spectral Image Decomposition by 3D Nonnegative Tensor Factorization," *Opt. Lett.*, vol. 34, no. 14, pp. 2210-2212, 2009.
- [23] T. G. Kolda and B. W. Bader., "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455-500, 2009.
- [24] E. Acar and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 1, pp. 6-20, 2009.
- [25] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95-138, 1977."
- [26] L. De Lathauwer, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. and Appl.*, vol. 21, no. 4, pp. 1253-1278, 2000.
- [27] L. De Lathauwer, "On the best rank-1 and rank-(R_1, R_2, \dots, R_N) approximation of higher-order tensors," *SIAM J. Matrix Anal. and Appl.*, vol. 21, no. 4, pp. 1324-1342, 2000.
- [28] Berkant Savas and Lek-Heng Lim, "Quasi-Newton Methods on Grassmannians and Multilinear Approximations of Tensors," *SIAM J. Sci. Comput.*, vol. 32, no. 6, pp.

3352-3393, 2010.

- [29] L. Elden and B. Savas, "A Newton-Grassmann method for computing the Best Multi-Linear Rank-(r_1, r_2, r_3) Approximation of a Tensor," *SIAM J. Matrix Anal. Appl.*, vol. 31, pp. 248-271, 2009.
- [30] D. Luo, C. Ding, and H. Huang, "Are tensor decomposition solutions unique? On the Global convergence HOSVD and parafac algorithms," in *Proceedings of the 15th Pacific-Asia conference on Advances in knowledge discovery and data mining*, 2011, pp. 148-159.
- [31] P. Stoica and Y. Selen, "Model-order selection: a review of information criteria rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36-47, 2004.
- [32] Z. He, A. Cichocki, and S. Xie, "Efficient method for Tucker3 model selection," *Electronics Letters*, vol. 45, no. 15, pp. 805-806, 2009.
- [33] Z. He, A. Cichocki, Shengli Xie, and Kyuwan Choi, "Detecting the Number of Clusters in n-Way Probabilistic Clustering," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2006-2021, 2010.
- [34] F. Cong, A. K. Nandi, H. Zhaoshui, A. Cichocki, and T. Ristaniemi, "Fast and Effective Model Order Selection Method to Determine the Number of Sources in a Linear Transformation Model," in *Proceedings of the 2012 European Signal Processing Conference (EUSIPCO-2012)*, Bucharest, Romania, 2012 (Accepted).
- [35] E. R. Malinowski, "Theory of error in factor analysis," *Anal. Chem.*, vol. 49, pp. 606-612, 1977.
- [36] E. R. Malinowski, "Determination of the number of factors and experimental error in data matrix," *Anal. Chem.*, vol. 49, pp. 612-617, 1977.

- [37] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716-723, 1974.
- [38] J. E. Cavanaugh, "A large-sample model selection criterion based on Kullback's symmetric divergence," *Stat. Probab. Lett.*, vol. 42, pp. 333-344, 1999.
- [39] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.
- [40] M. Morup and L. K. Hansen, "Automatic relevance determination for multi-way models," *Journal of Chemometrics*, vol. 23, pp. 352-363, 2008.
- [41] T. P. Minka, "Automatic choice of dimensionality for PCA," in *Neural Information Processing Systems 13*, 1999.
- [42] F. Cong, Z. He, J. Hämäläinen, A. Cichocki, and T. Ristaniemi, "Determining the Number of Sources in High-density EEG Recordings of Event-related Potentials by Model Order Selection," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP) 2011*, Beijing, China, 2011.
- [43] J. M. P. Nascimento and J. M. Bioucas Dias, "Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?," *IEEE Trans. Geosc. and Remote Sensing*, vol. 43, pp. 175-187, 2005.
- [44] T. Han and D. G. Goodenough, "Investigation of Nonlinearity in Hyperspectral Imagery Using Surrogate Data Methods," *IEEE Trans Geosc. and Remote Sens.*, vol. 46, pp. 2840-2847, 2008.
- [45] T. M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Electronic Computers*, vol. EC-14, no. 3, pp. 326-334, 1965.

- [46] B. Scholkopf and A. J Smola, *Learning with kernels*. Cambridge, MA: MIT Press, 2002.
- [47] S. Hameling, A. Ziehe, and M. Kawanabe, "Kernel-Based Nonlinear Blind Source Separation," *Neural Computation*, vol. 15, pp. 1089-1124, 2003.
- [48] B. Schölkopf, A. J. Smola, and R. K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.
- [49] G. Baudat and F. Anouar, "Kernel-based methods and function approximation," in *Proceedings of International Joint Conference on Neural Networks*, Washington, DC, 2001, pp. 1244-1249.
- [50] A. K. Jain, R. P. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, 2000.
- [51] T. L. Diepgen, G. Yihune, and et al. Dermatology Online Atlas. [Online]. <http://www.dermis.net>
- [52] DermAtlas. [Online]. <http://dermatlas.org>
- [53] An Atlas of Clinical Dermatology. [Online]. <http://dandem-pdv.is.kkh.dk/atlas>
- [54] DermNetNZ. [Online]. <http://dermnet.org.nz/lesions>
- [55] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483-1492, 1997.
- [56] Johan Himberg, Aapo Hyväinen, and Fabrizio Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *NeuroImage*, vol. 22, pp. 1214-1222, 2004.
- [57] D. T. Alexandrescu et al., "Melanoma-Specific Marker Expression in Skin Biopsy Tissues as a Tool to Facilitate Melanoma Diagnoses," *J. Invest. Dermatology*, vol.

130, no. 7, pp. 1887-1992, 2010.

- [58] S. Mori, *Introduction to Diffusion Tensor Imaging*. Elsevier Science, 2007.
- [59] T. Nakai et al., "Application of independent component analysis to magnetic resonance imaging for enhancing contrast of gray and white matter," *Neuroimage*, vol. 21, no. 1, pp. 251-260, 2004.
- [60] L. Ruskó, G. Bekes, and M. Fridrich, "Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images," *Med. Image Anal.*, vol. 13, no. 6, pp. 871-882, 2009.
- [61] G. Tao, A. Singh, and L. Bidaut, "Liver Segmentation from Registered Multiphase CT Data Sets with EM Clustering and GVF Level Set," in *Proc. of the SPIE 7623*, 2010, p. 76230V.

Summary

Noninvasive diagnosis of tumor is a procedure of discrimination among various types of tumor by exploiting noninvasively collected data, e.g. data acquired by various imaging modalities. The aim of the computer-based automatic noninvasive diagnosis is to utilize information processing algorithms to analyze collected data in order to correctly classify the tumor. Usually, automatic diagnosis is obtained by classifying a set of features extracted from the image of the tumor. In case of automatic diagnosis of cutaneous melanoma research has been done dominantly along two lines: inspection of morphological features of the suspicious lesion; and inspection of spectral characteristics of the lesion. Morphological features are usually analyzed from images obtained by skin microscopy, so called dermoscopy. On the other hand, spectral profile of the lesion is usually acquired by a multispectral or hyperspectral imaging system. However, both approaches use special imaging systems that are available only in specialized clinics, and often must be operated by a trained individual.

In contrary, recent results reported in the literature demonstrate that automatic diagnosis of melanoma is possible using red-green-blue (RGB) color autofluorescent clinical images. These findings are of great practical importance, since they demonstrate that an accurate diagnosis of melanoma is possible using affordable RGB color imaging. In this way an automated system based on cost-effective and nowadays ubiquitous RGB camera could possibly be used for preliminary screening in areas without trained dermatologist and/or specialized equipment.

To this end, we propose a feature extraction method for noninvasive diagnosis of melanoma based on tensor decomposition of the clinical color image of skin lesion. The image of the suspicious skin lesion is represented by a multi-way array, i.e. by a tensor.

The image tensor is modeled by Tucker3 model, a standard model in the multi-way analysis. Extracted features are elements of the core tensor in the corresponding Tucker3 model, and represent spatial-spectral profile of the object in the image. In contrast to majority of methods that exploit either texture or spectral diversity of the tumor only, this method simultaneously captures spatial and spectral characteristics. The proposed procedure is demonstrated on a problem of noninvasive diagnosis of melanoma from cost-effective auto-fluorescence color images of skin lesions, with overall sensitivity 82.1% and specificity 86.9%, and compare favorably with the state of the art results reported in the literature. It is also important to note that the proposed feature extraction procedure is based on a computationally efficient and scalable multilinear generalization of singular value decomposition. Thus, diagnostic system based on this feature extraction method could be easily implemented on some of the widespread smartphone platforms, or as an application-specific low-cost device.