# Melanoma Recognition Using Representative and Discriminative Kernel Classifiers

Tatiana Tommasi[1], Elisabetta La Torre[1], and Barbara Caputo[2]

[1] University of Rome La Sapienza, P.le A. Moro 5, 00185, Rome, Italy
{tatiana.tommasi, elisabetta.latorre}@uniroma1.it
[2] NADA/CVAP, KTH, SE-100 44, Stockholm, Sweden
caputo@nada.kth.se

**Abstract.** Malignant melanoma is the most deadly form of skin lesion. Early diagnosis is of critical importance to patient survival. Existent visual recognition algorithms for skin lesions classification focus mostly on segmentation and feature extraction. In this paper instead we put the emphasis on the learning process by using two kernel-based classifiers. We chose a discriminative approach using support vector machines, and a probabilistic approach using spin glass-Markov random fields. We benchmarked these algorithms against the (to our knowledge) state-of-the-art method on melanoma recognition, exploring how performance changes by using color or textural features, and how it is affected by the quality of the segmentation mask. We show with extensive experiments that the support vector machine approach outperforms the existing method and, on two classes out of three, it achieves performances comparable to those obtained by expert clinicians.

## 1 Introduction

Malignant melanoma is a spreading disease in the western world. Its incidence has been increasing over the past decades; currently 132,000 melanoma skin cancer occurs globally each year. One in every three cancers diagnosed is a skin cancer and, according to Skin Cancer Foundation Statistics, one in every five Americans will develop this kind of tumor in their lifetime [15]. Management of melanoma is a complex issue requiring a multidisciplinary approach. The most effective method of protection against the development of skin cancer is minimization of ultraviolet exposure from sunlight. Since advanced melanoma is still practically incurable, early detection and treatment are critical steps towards a reduction in mortality. Surgical excision remains the mainstay of treatment [9]. In northern Europe a deceleration in the incidence and mortality trends occurred recently in persons aged under 70, whereas in southern Europe both incidence and mortality rates are still increasing [4]. The most plausible explanations for the deceleration in these trends in northern Europe are earlier detection, more frequent excision of pigmented lesions and a growing public awareness of the dangers of excessive sunbathing [4].

Epiluminescence Microscopy (ELM or dermoscopy) is the most used diagnostic technique used by clinicians to reveal malignant melanoma. It is non-invasive and allows for a detailed surface analysis of a suspicious skin lesion by using hand-held device emitting incident light from a light source penetrating the epidermal skin layer. Physicians visually inspect dermoscopic images for abnormal morphologic and chromatic features that indicate malignancy. They commonly use the ABCD (Asymmetry, Border, Color, Dimension and Dermoscopic structures) method as guideline. Due to the subjective nature of examination the accuracy of diagnosis is highly dependent upon physician's expertise.

There is a growing awareness that one of the weakest links in the biomedical interpretation process is the perception of details and the recognition of their meaning by the dermatologists. An automatic system for melanoma recognition would constitute a valuable support for physicians in every day clinical practice. Such a system should reproduce the perceptual and cognitive strategy followed by doctors, and should allow the dermatologist to trace each step of the process which led to a given diagnosis, so to leave space for exploring multiple interpretations. Recently numerous research on this topic have been proposed (for a more comprehensive discussion of the most significant literature we refer the reader to section 2); a key factor for the development and evaluation of these systems is the availability of a statistically significant database. One of the largest databases of melanoma images available to the research community was contributed by H. Ganster et al. [5]. That paper presented a database of 5363 images, accompanied by: (a) a segmentation algorithm for isolating the potential melanoma from the surrounding skin, determined by several basic segmentation algorithms combined together with a fusion strategy [5]; (b) a set of features containing shape and radiometric features as well as local and global parameters, calculated to describe the malignancy of a lesion, from which significant features are selected by application of statistical feature subset selection methods [5]; (c) a nearest neighbor classification algorithm [5]. In that work the authors concentrated particularly on the segmentation technique and the features selection process, obtaining results that, to the best of our knowledge, represent the state-of-the-art on this topic. Here we focus instead on the classification algorithm, proposing to use kernel methods for classification of skin lesion images. Specifically, we selected a discriminative method and a probabilistic one. As discriminative method we chose Support Vector Machines (SVM, [12]), a state-of-the-art large margin classifier, where the optimal separating surface is defined by a linear combination of scalar products between the view to be classified and some support vectors [11][12]. By introducing a Mercer kernel, a non-linear SVM can be constructed replacing the scalar products in the linear SVM via the kernel function. SVMs have demonstrated remarkable performance on object recognition and categorization [13] and biomedical imaging [14]. As probabilistic method we chose Spin Glass-Markov Random Fields (SG-MRF, [2]), a fully connected MRF which integrates results of statistical mechanics with Gibbs probability distributions via non linear kernel mapping [2]. Experiments have shown the robustness and categorization capabilities of this algorithm for object recognition [2] and its applicability for

biomedical applications [3]. We conducted an experimental evaluation of these two techniques on the Ganster's database[3], which allows for a straightforward benchmarking of our algorithms against theirs. We tested out two methods on two different types of features, Color Histograms (CH) and Multidimansional receptive Fields Histograms (MFH, [10]). These features reproduce two of the criteria followed by dermatologists for diagnosis, respectively "C" for color variegation and "D" for differential local structures. Several series of experiments were performed for selecting optimal feature descriptors. We also evaluated the influence of the segmentation method by running two series of experiments: the first using the segmentation masks obtained by Ganster, the second using an hand-made rectangular mask which roughly contains the whole lesion while minimizing the amount of surrounding skin in the image. In order to have a fair comparison, we replicated the experimental setup used in [5] for a benchmark evaluation. Our results show that SVM obtains remarkably better performances than SG-MRF and Ganster's method with both feature types and regardless of the segmentation method. More important, on two classes out of three, SVM achieves recognition results comparable to those obtained by skilled clinicians.

The rest of the paper is organized as follows: section 2 reviews the state of the art in computer-assisted melanoma recognition. Then we briefly review the theory behind SG-MRF (section 3) and SVMs (section 4). Section 5 describes the experimental setup and reports on our findings. The paper concludes with a summary discussion and some possible directions for future research.

## 2 Related Work

Recently there has been an increasing interest in developing algorithms for melanoma classification. Grana et al. [6] provided mathematical descriptors for the border of pigmented skin lesion images and assessed their efficacy for distinction among different lesion groups. They introduced new descriptors such as lesion slope and lesion slope regularity and define them mathematically, then they employed a new algorithm based on the Catmull Rom spline method and the computation of the gray-level gradient of points extracted by interpolation of normal direction on spline points [6]. The efficacy of these descriptors was tested on a data set of 510 pigmented skin lesions, composed by 85 melanomas and 425 nevi, by employing statistical methods for discrimination between the two populations [6]. Grzymala-Busse et al. [7] used discretization based on cluster analysis, LEM2 algorithm for rule induction, and standard LERS classification scheme to check whether the ABCD formula is optimal [7]. The data consisted in total of 276 cases of benign nevus, blue nevus, suspicious nevus, and malignant melanoma [7]. Lefevre et al. [8] proposed a theory used in different fields such as data fusion, regression or classification: the Dempster-Shafer's theory, or evidence theory [8]. They applied the classification process on a training set of

[3] We gratefully thank H. Ganster and A. Pinz for making the database and their segmentation masks available to us.

81 lesions: 61 benign lesions (nevi) and 20 malignant lesions (melanoma) and a test set of 209 lesions: 191 nevi and 18 melanoma [8].

Ganster et al. [5] presented a system where as initial step the binary mask of the skin lesion was determined by several basic segmentation algorithms combined together with a fusion strategy [5]. The algorithms used to segment the lesion are: global thresholding, dynamic thresholding, and a 3-D color clustering concept [5]. A set of features was then calculated to describe the malignancy of a lesion: global features (size and shape descriptors), color features and local features [5]. Significant features were then selected from this set by application of statistical feature subset selection methods [5]. The classification experiments were performed with a 24-NN classifier based on the derived features [5]. A notable characteristic of this work is the large dimension of the database. They had at their disposal overall 5363 skin lesion images, categorized into three classes. The three classes are: clearly benign lesions, dysplastic lesions and malignant lesions [5]. The training set for the classifier was a set of 270 lesions (90 images for each class). The test set was the entire database of 5363 lesions in three categories [5]. They obtained a mean recognition rate of 61%. To the best of our knowledge, this is the largest existing database on skin lesions, and these results constitute the state of the art in the field. This is the database on which we ran our experiments, and the results with which we compare our performance.

## 3 Spin Glass - Markov Random Fields

Consider a visual class $\Omega_j$ and a set of $k$ observations $\{\boldsymbol{x}^1 \dots \boldsymbol{x}^k\}, \boldsymbol{x} \in \Re^m$, that we consider random samples from the underlying, unknown, probability distribution $P(\boldsymbol{x})$ defined on $\Re^m$. Consider also $\mathcal{K}$ different visual classes $\Omega_j, j = \{1, \dots \mathcal{K}\}$ (here, $\mathcal{K}$ will be 3, corresponding to the visual labels "benign", "dysplastic" and "malignant"). Given an observation $\hat{\boldsymbol{x}}$, our goal is to classify $\hat{\boldsymbol{x}}$ as a sample from $\Omega_{j^*}$, one of the $\Omega_j$ visual classes. Using a Maximum A Posteriori (MAP) criterion we have

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\Omega_\kappa|\boldsymbol{x}) = \underset{j}{\operatorname{argmax}}\{P(\boldsymbol{x}|\Omega_j)P(\Omega_j)\}$$

using Bayes rule, where $P(\boldsymbol{x}|\Omega_j)$ are the Likelihood Functions (LFs) and $P(\Omega_j)$ are the prior probabilities of the classes. Assuming that $P(\Omega_j)$ are constant, the Bayes classifier simplifies to

$$j^* = \underset{j}{\operatorname{argmax}}\, P(\boldsymbol{x}|\Omega_j)\ . \tag{1}$$

Spin Glass-Markov Random Fields (SG-MRFs) [2] are a new class of MRFs which connect SG-like energy functions (mainly the Hopfield one [1]) with Gibbs distributions via a non linear kernel mapping. The resulting model overcomes many difficulties related to the design of fully connected MRFs, and enables to use the power of kernels in a probabilistic framework. The SG-MRF probability

distribution is given by

$$P_{SG-MRF}(\boldsymbol{x}|\Omega_j) = \frac{1}{Z}\exp\left[-E_{SG-MRF}(\boldsymbol{x}|\Omega_j)\right] \ , \tag{2}$$

$$Z = \sum_{\{\boldsymbol{x}\}} \exp\left[-E_{SG-MRF}(\boldsymbol{x}|\Omega_j)\right] \ , \tag{3}$$

with

$$E_{SG-MRF} = -\sum_{\mu=1}^{p_j}\left[K(\boldsymbol{x},\tilde{\boldsymbol{x}}^{(\mu)})\right]^2 \ , \tag{4}$$

where the function $K(\boldsymbol{x},\tilde{\boldsymbol{x}}^{\mu})$ is a Generalized Gaussian kernel [11]:

$$K(\boldsymbol{x},\boldsymbol{y}) = \exp\{-\rho d_{a,b}(\boldsymbol{x},\boldsymbol{y})\} \ , \qquad d_{a,b}(\boldsymbol{x},\boldsymbol{y}) = \sum_i |x_i^a - y_i^a|^b$$

and $\{\tilde{\boldsymbol{x}}^{\mu}\}_{\mu=1}^{p_j}, j \in [1,\mathcal{K}]$ are a set of vectors selected (according to a chosen ansatz, [2]) from the training data that we call *prototypes*. The number of prototypes per class must be finite, and they must satisfy the condition:

$$K(\tilde{\boldsymbol{x}}^i,\tilde{\boldsymbol{x}}^k) = 0 \ , \tag{5}$$

for all $i,k = 1,\ldots p_j, i \neq k$ and $j = 0,\ldots\mathcal{K}$ (the interested reader can find a detailed discussion regarding the derivation and properties of SG-MRF in [2]). Thus, the Bayes classifier (1) will become

$$j^* = \underset{j}{\operatorname{argmin}}\, E_{SG-MRF}(\boldsymbol{x}|\Omega_j) \ . \tag{6}$$

## 4 Support Vector Machines

Support Vector Machines are state-of-the-art large margin classifiers which have gained popularity within visual pattern recognition. Here we provide a brief review of the theory behind this type of algorithm. For a more detailed treatment, we refer to [12].

Suppose we are in the two class case. Consider the feature vector $\boldsymbol{x} \in \Re^N$ and its class label $y \in \{-1,+1\}$. Let $(\boldsymbol{x}_1,y_1)$, $(\boldsymbol{x}_2,y_2)$, $\ldots$,$(\boldsymbol{x}_m,y_m)$ denote a given set of $m$ training examples. If we assume that the two classes are linearly separable, there exists a linear function

$$f(\boldsymbol{x}) = \boldsymbol{w}\cdot\boldsymbol{x} + b \tag{7}$$

such that for each training example $\boldsymbol{x}_i$, it yields $f(\boldsymbol{x}_i)\geq 0$ for $y_i = +1$ and $f(\boldsymbol{x}_i\mathrm{x})\leq 0$ for $y_i = -1$. The optimal separating hyperplane is the one which has maximum distance to the closest points in the training set. Mathematically this

hyperplane can be found by solving a constrained minimization problem using Lagrange multipliers $\alpha_i$ $(i = 1, \ldots, m)$. It results in a classification function

$$f(\boldsymbol{x}) = \mathrm{sgn}\,(\sum_{i=1}^{i=m} \alpha_i y_i \boldsymbol{w} \cdot \boldsymbol{x} + b)\ , \tag{8}$$

where $\alpha_i$ and $b$ are found by using an SVC learning algorithm [12]. It turns out that a small number of the $\alpha_i s$ are different from zero; their corresponding data $\boldsymbol{x}_i$ are called support vectors [12].

SVM can be extended to non-linear problems by using a non-linear operator $\Phi(\cdot)$ to map the input feature vectors $\boldsymbol{x}_i$ from the original $\Re^N$ into a higher dimensional feature space $\mathcal{H}$ by $\boldsymbol{x} \rightarrow \Phi(\boldsymbol{x}) \in \mathcal{H}$. Here the mapped data points of the two classes become linearly separable. Assuming there exists a kernel function $K$ associated with the inner product of the desired nonlinear mapping such that $K(\boldsymbol{x},\boldsymbol{y}) = \Phi(\boldsymbol{x}) \cdot \Phi\ (\boldsymbol{y})$, then a non linear SVM can be obtained by replacing $\boldsymbol{x} \cdot \boldsymbol{y}$ by the kernel $K(\boldsymbol{x},\boldsymbol{y})$ in the decision function, obtaining then

$$f(\boldsymbol{x}) = \mathrm{sgn}\,(\sum_{i=1}^{i=m} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b)\ . \tag{9}$$

This corresponds to constructing an optimal separating hyperplane in the feature space. In this paper we consider four kernel types:

| | |
|---|---|
| Polynomial kernel ("poly") | $K(\boldsymbol{x}, \boldsymbol{y}) = (\gamma * \boldsymbol{x} \cdot y)^d$ |
| Generalized Gaussian kernel ("gengauss") | $K(\boldsymbol{x}, \boldsymbol{y}) = \exp\{-\gamma * \lvert\boldsymbol{x}^a - \boldsymbol{y}^a\rvert^b\}$ |
| Gaussian kernel ("gauss") | $K(\boldsymbol{x}, \boldsymbol{y}) = \exp\{-\gamma * \lvert\boldsymbol{x} - \boldsymbol{y}\rvert^2\}$ |
| Chi-squared kernel ("chi") | $K(\boldsymbol{x}, \boldsymbol{y}) = \exp\{-\gamma * \chi^2(\boldsymbol{x}, \boldsymbol{y})\}$ . |

## 5 Experiments

In this section we present experiments that show the effectiveness of kernel methods for melanoma recognition. To this purpose, in a preliminary step, we ran a first series of experiments for feature selection. Then we used the selected features for an extensive set of classification experiments. In the rest of the section we describe the database used (section 5.1), the experimental setup (section 5.2) and our experimental findings (section 5.3).

### 5.1 Database

We performed our experiments on the database created by the Department of Dermatology of the Vienna General Hospital [5]. The whole database consists of 5380 skin lesion images, divided into three classes: 4277 of these lesions are classified as clearly benign lesions (Class 1), 1002 are classified as dysplastic lesions (Class 2) and 101 lesions are classified as malignant melanomas (Class 3).[4]

[4] These numbers are not perfectly coincident with those reported in [5], where the database is said to be of 5363 images, but this difference should not affect the comparison between the two algorithms.

The lesions of the classes 2 and 3 were all surgically excised and the ground truth was generated by means of histological diagnosis [5]. In order to have statistically significant results, we ran experiments with five different partitions, then we calculated the mean and the standard deviation of the obtained recognition rates. This procedure has been adopted for all the experiments reported here.

### 5.2 Experimental Setup

The three key components for an automated melanoma recognition algorithm are: segmentation/preprocessing, features extraction and classification. We describe below the general approach followed in this paper for each of these steps:

*Segmentation/preprocessing:* Following the approach proposed in [5], we didn't implement any preprocessing step such as color normalization or hair removal. As for the segmentation procedure, we used two different methods. The first consists in simply cutting all the images with the help of a common image editor software, selecting for each image the smallest rectangle containing the lesion and keeping out as much skin as possible. We call the resulting images "hand-segmented". The second method is the one developed by Ganster et al. [5]. It consists of a binary mask determined by several segmentation algorithms combined together with a fusion strategy. We call the resulting images "mask--segmented". An example of the images obtained by these two segmentation techniques is in Fig.1. Running experiments on these two types of images allows us to explore how the classification performance is affected by the quality of the segmentation process.

*Feature Extraction:* In the ABCD rule, the color variegation and the dermoscopic structures in the skin lesion are two of the discriminant characteristics for clinical melanoma recognition, thus we decided to use CH and MFH as features able to retain chromatic and textural information respectively. The color histogram was computed by discretizing the colors within the image and counting the number of pixels for each color. We ran several experiments for selecting the best features, namely using hue, rg, RG, RB and GB color histograms. The resolution of the bin axes was varied for each representation consisting of 8, 16, 32, 64 (for bidimensional histograms we chose the resolution of each axis with the same bin value). We found that the GB representation obtained the best results for all the bin values, thus we used it in all the following experiments.
The main idea of MFH is to calculate multidimensional histograms of the response of a vector of receptive fields. A MFH is determined once we chose the local property measurements (i.e., the receptive field functions), which determine the dimensions of the histogram, and the resolution of each axis. We converted originally RGB images to gray-scale and then we used two different kinds of MFH representation: the first consisted in Gaussian derivatives along $x$ and $y$ directions and with $\sigma = 1.0$ ($D_x D_y$); the second consisted in Laplacian Gaussian operator with $\sigma_1 = 1.0, 1.5, 3.0$, and $\sigma_2 = 2.0, 3.0, 6.0$ respectively ($Lp2\sigma$). The bin axes' resolution was varied for each representation consisting of 8, 16, 32, 64 for Gaussian-filter MFH and 16, 32 for Laplacian-filter MFH.

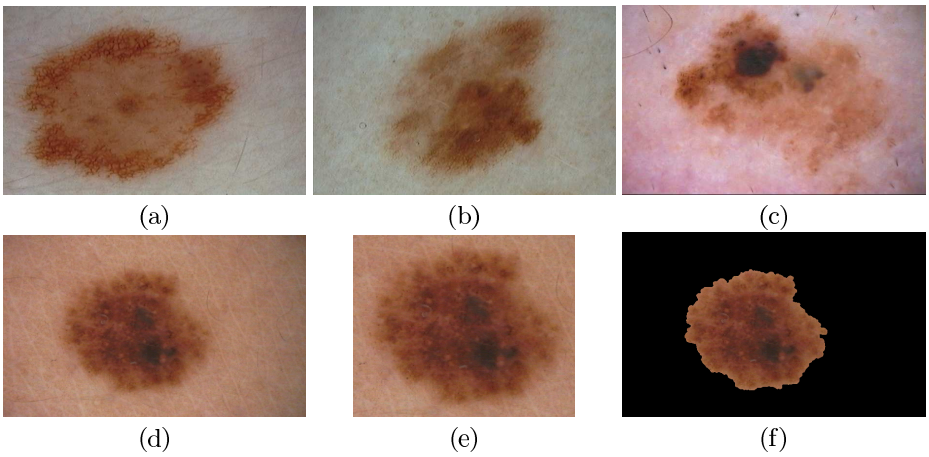


**Fig. 1.** Examples of skin lesion's images used: (a) image of a benign lesion, (b) image of a dysplastic lesion,(c) image of a malignant lesion. (d) shows an example of an entire image, (e) the same image hand-segmented, (f) the same image mask-segmented

*Classification:* We used SG-MRF and SVM algorithms (see section 3 and 4 respectively). For SG-MRF we learned the kernel parameters during the training stage using a leave-one-out strategy. For SVM we used the four kernel types described in section 4. The kernel parameters were chosen via cross validation.

All the experiments were performed respecting the procedure reported by Ganster et al. [5]. The training set consisted of 270 images (90 for each class); the test set consisted of the whole database [5]. Note that training and test set are not disjoint; once again we underline that this follows the procedure proposed in [5] which allows for benchmarking.

### 5.3 Experimental Results

A first set of experiments was ran using CH with GB. A second set of experiments was ran with MFH with $D_x D_y$ and $Lp2\sigma$ representations as features. The obtained recognition rates for hand-segmented and mask-segmented images using SG-MRF and SVM, with both features types are reported in Table 1. Results for each class are averaged on five partitions. We also report the average of the recognition rate obtained class by class ("Mean Class"), and the overall recognition rate ("Overall"). For sake of clarity we report the results obtained in [5] too; note that these results were obtained on a single run.

A first comment is that SVM obtains the best result with respect to Ganster's method and SG-MRF, for both feature types and for both segmentation strategies. The best result, in terms of overall recognition rate, is of 82.5%, obtained

using the generalized Gaussian kernel, MFH features and mask-segmented images; comparable results are obtained with color features, selected kernels and on hand-segmented images. The best result obtained by using SG-MRF is of 49.5%, obtained using mask-segmented images and MFH features; finally, the best performance obtained by using the Ganster's method is of 58%. These results clearly show the effectiveness of SVMs for melanoma recognition. A second comment is that SVM performance varies considerably depending on the kernel type used. For instance, using color features and hand-segmented images, the overall recognition rate goes from a minimum of 59.0% for the Gaussian kernel to a maximum of 76.0% for chi-squared kernel. A similar behavior is observed by using mask-segmented images, and on textural features. It is also interesting to note that with both segmentation techniques and feature types, for the over-

**Table 1.** Recognition results obtained by Ganster et al [5], with SG-MRF and SVM methods using different kernels, for hand-segmented and mask-segmented images and using CH and MFH as features. We report the recognition rates for the three classes, the mean and the overall recognition rates. Results obtained with SG-MRF and SVM are mean values from five different runs with their standard deviations. Class 1, Class 2, Class 3 correspond to the benign, dysplastic and malignant lesions respectively

| | | | Class 1 | Class 2 | Class 3 | Mean Class | Overall |
|---|---|---|---|---|---|---|---|
| **Ganster et al. [5] (%)** | | | 59 | 53 | 73 | 61 | 58 |
| **CH features** | | | | | | | |
| **hand** | **SG-MRF (%)** | | 43.2 ± 4.5 | 41.2 ± 2.1 | 95.1 ± 1.6 | 59.8 ± 17.6 | 46.1 ± 5.6 |
| | **SVM (%)** | **poly** | 91.7 ± 4.9 | 9.8 ± 7.3 | 5.5 ± 0.5 | 35.7 ± 28.0 | 74.9 ± 2.8 |
| | | **gauss** | 65.7 ± 17.1 | 31.6 ± 16.0 | 49.5 ± 26.0 | 48.9 ± 9.8 | 59.0 ± 10.3 |
| | | **gengauss** | 89.8 ± 20.4 | 15.6 ± 13.6 | 82.6 ± 14.6 | 62.7 ± 23.6 | 75.9 ± 14.0 |
| | | **chi** | 90.0 ± 20.2 | 15.0 ± 12.3 | 89.1 ± 0.0 | 64.7 ± 24.9 | **76.0 ± 13.7** |
| **mask** | **SG-MRF (%)** | | 48.6 ± 4.2 | 38.8 ± 3.4 | 94.1 ± 3.4 | 60.5 ± 17.0 | 47.7 ± 2.9 |
| | **SVM (%)** | **poly** | 80.1 ± 13.0 | 15.7 ± 13.7 | 29.5 ± 20.4 | 41.8 ± 19.6 | 67.1 ± 7.8 |
| | | **gauss** | 71.9 ± 11.1 | 24.8 ± 12.7 | 45.0 ± 28.5 | 47.2 ± 13.6 | 62.6 ± 6.2 |
| | | **gengauss** | 96.2 ± 4.0 | 11.0 ± 1.8 | 89.5 ± 0.9 | 65.6 ± 27.4 | **80.2 ± 2.8** |
| | | **chi** | 68.6 ± 17.7 | 22.4 ± 7.5 | 62.6 ± 19.7 | 51.2 ± 14.5 | 59.9 ± 12.9 |
| **MFH features** | | | | | | | |
| **hand** | **SG-MRF (%)** | | 39.2 ± 4.1 | 42.2 ± 3.1 | 94.5 ± 2.9 | 58.6 ± 18.0 | 40.8 ± 2.8 |
| | **SVM (%)** | **poly** | 85.3 ± 18.3 | 9.7 ± 8.5 | 19.8 ± 22.9 | 38.3 ± 23.7 | 66.9 ± 13.1 |
| | | **gauss** | 55.7 ± 13.9 | 31.6 ± 17.1 | 54.1 ± 19.4 | 47.1 ± 7.8 | 51.1 ± 8.6 |
| | | **gengauss** | 96.7 ± 2.8 | 11.7 ± 3.0 | 89.7 ± 0.9 | 66.0 ± 27.2 | **80.7 ± 1.7** |
| | | **chi** | 80.8 ± 2.3 | 23.1 ± 4.0 | 93.1 ± 1.4 | 65.7 ± 21.6 | 70.3 ± 1.5 |
| **mask** | **SG-MRF (%)** | | 49.3 ± 5.1 | 45.4 ± 4.0 | 94.5 ± 1.8 | 63.1 ± 15.7 | 49.5 ± 3.9 |
| | **SVM (%)** | **poly** | 80.5 ± 4.2 | 28.5 ± 14.9 | 22.0 ± 19.1 | 43.7 ± 18.5 | 69.7 ± 3.8 |
| | | **gauss** | 80.9 ± 3.6 | 27.2 ± 13.5 | 25.3 ± 23.0 | 44.5 ± 18.2 | 69.8 ± 3.7 |
| | | **gengauss** | 99.4 ± 0.1 | 9.6 ± 0.4 | 89.3 ± 0.4 | 66.1 ± 28.4 | **82.5 ± 0.1** |
| | | **chi** | 96.7 ± 0.4 | 13.0 ± 1.6 | 90.5 ± 0.5 | 66.7 ± 26.9 | 81.0 ± 0.2 |

**Table 2.** Confusion matrices for different classification methods. Top left, Ganster's method [5]; top right, clinical diagnosis performed from expert dermatologists of the Department of Dermatology at the Vienna General Hospital [5] . Middle left, SVM results with the "chi" kernel and GB CH feature for hand-segmented images; middle right, SVM results with the "gengauss" kernel and GB CH feature for mask-segmented images. Bottom left, SVM results with the "gengauss" kernel and MFH feature for hand-segmented images; bottom right, SVM results with the "gengauss" kernel and MFH feature for mask-segmented images. The number of images reported are mean value of the number obtained from five different partitions. Class 1, class 2 and class 3 identify the three classes corresponding to benign, dysplastic and malignant lesions respectively

| **Ganster et al** [5] | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 2500 | 1347 | 410 |
| **class 2** | 324 | 531 | 155 |
| **class 3** | 14 | 12 | 70 |

| **Clinicians** | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 4161 | 94 | 9 |
| **class 2** | 42 | 960 | 8 |
| **class 3** | 6 | 19 | 78 |

| **CH hand** | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 3850.6 | 259.4 | 167.0 |
| **class 2** | 798.2 | 150.4 | 53.4 |
| **class 3** | 9.8 | 1.2 | 90.0 |

| **CH mask** | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 4112.6 | 112.6 | 50.8 |
| **class 2** | 874.8 | 110.0 | 17.2 |
| **class 3** | 10.4 | 0.2 | 90.4 |

| **MFH hand** | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 4184.8 | 45.5 | 45.8 |
| **class 2** | 861.6 | 116.8 | 23.6 |
| **class 3** | 9.8 | 0.6 | 90.6 |

| **MFH mask** | | | |
|---|---|---|---|
| | **Assigned** | | |
| **True** | **class 1** | **class 2** | **class 3** |
| **class 1** | 4251.8 | 4.2 | 20.0 |
| **class 2** | 901.0 | 95.8 | 5.2 |
| **class 3** | 10.4 | 0.4 | 90.2 |

all recognition rate, the kernels which obtains the worst performances tend to have the highest standard deviations, while the kernel with the best performance has the smallest one. This illustrates the importance of doing kernel selection in the training phase; the low standard deviation of the SVM's best results also shows the stability of our findings. By comparing the hand-segmented overall best result with the mask-segmented one, we can see an improvement in recognition rate and stability passing from the first to the second, for both feature types. This is an experimental proof of the importance of using a sophisticated segmentation method. A final remark should be made on the poor performance of SG-MRF. This might be due to the dimension of the training set for each class; it could be possible that the probabilistic method needs a higher statistic in order to estimate properly the energy function.

Table 2 reports the confusion matrices for the best results obtained by each possible combination of (segmentation mask, feature type) and SVMs, plus the confusion matrix obtain by Ganster and that relative to clinicians' performance on the database [5].[5] For both segmentatation techniques and feature types, we see that SVM outperforms Ganster's method for class 1 and class 3 and it is comparable with the dermatologists' performances. It is very interesting to note that, in contrast, SVM performs poorly on class 2, which corresponds to dysplastic lesions. This might be explained considering that here we are using only one feature type for each set of experiments, while Ganster used a selection of different features and dermatologists used the ABCD rule. It is thus possible that just color/textural information is not discriminant enough in order to recognize correctly dysplastic lesions, while both feature types seem to be effective for separating benign and malignant lesions.

## 6 Conclusions

In this paper we presented a learning approach to melanoma recognition. To this purpose, we proposed two kernel-based classification algorithms: a probabilistic one, spin glass-Markov random fields, and a discriminative one, support vector machines. Both methods have proved successful on visual recognition problems like object recognition. The two classifiers were tested on a database of more than 5000 images, using two feature types and two segmentation methods. Our results show that SVM obtains an improvement in recognition rate of more than 20% compared to what reported in [5], which to our knowledge constitutes the state of the art in the field. Moreover, on two classes out of three, SVM achieves recognition results comparable to those obtained by skilled clinicians.

This work can be extended in many ways: first, we plan to repeat the experiments presented here on different partitions of the Ganster's database (disjoint training and test set, several partitions, varying number of images in training and test set), so to assess better our method's performance and the database at the same time. Second, we plan to conduct similar experiments using shape descriptors, and finally to experiment with cue integration schemes, in order to test the effectiveness of different types of information and eventually to reproduce the ABCD method followed by the dermatologists in every day clinical practice.

## References

1. D. J. Amit: Modeling Brain Function. Cambridge University Press, Cambridge, USA, 1989.
2. B. Caputo: A new kernel method for object recognition: spin glass Markov random fields. PhD thesis, Stockholm, November 2004. Available at http://www.nada.kth.se/∼caputo

[5] For more details on the number of images used in the these last two confusion matrices we refer the reader to [5].

3. B. Caputo, E. La Torre, S. Bouattour, G.E. Gigante: A New Kernel Method for Microcalcification Detection: Spin Glass- Markov Random Fields. Proc. of MIE02, Budapest, August 2002.
4. E. De Vries, F. I. Bray, J. W. W. Coebergh and D. M. Parkin: Changing Epidemiology of Malignant Cutaneous Melanoma in Europe 1953-1997: Rising Trends in Incidence and Mortality but Recent Stabilizations in Western Europe and Decreases in Scandinavia. Int. J. Cancer 107: 119-126; 2003 Wiley-Liss, Inc..
5. H. Ganster, A. Pinz, R. Rhrer, E. Wildling, M. Binder and H. Kittler: Automated Melanoma Recognition. IEEE Trans on MI, 20, 3, march 2001.
6. C. Grana, G. Pellacani, R. Cucchiara, and S. Seidenari: A New Algorithm for Border Description of Polarized Light Surface Microscopic Images of Pigmented Skin Lesions. IEEE Trans on MI, 22, 8, August 2003.
7. J. P. Grzymala-Busse, J. W. Grzymala-Busse and Z. S. Hippe: Melanoma Prediction Using Data Mining System LERS. Proc COMPSAC 2001, pp 615-620.
8. E. Lefevre, O. Colot , P. Vannoorenberghe, D. de Brucq: Knowledge modeling methods in the framework of Evidence Theory An experimental comparison for melanoma detection. Proc of Int Conf on Systems, Man, and Cybernetics, 4, pp 2806-2811.
9. D. S. Rigel and J. A. Carucci: Malignant Melanoma: Prevention, Early Detection, and Treatment in the 21st Century. CA Cancer J Clin 2000; 50:215-236.
10. B. Schiele, J. L. Crowley: Recognition without correspondence using Multidimensional Receptive Field Hisograms. IJCV, 36(1), 2000, pp 31-52.
11. B. Scholkopf, A. J. Smola: Learning with kernels. 2001, the MIT Press.
12. V. Vapnik: Statistical learning theory. Wiley and Son, 1998.
13. C. Wallraven, B. Caputo, A. Graf: Recognition with Local features: the kernel recipe. Proc. ICCV03.
14. L. Wei, Y Yang, R. M. Nishikawa and Y Jiang: A Study on Several Machine-Learning Methods for Classification od Malignant and Benign Clustered Microcalcifications. IEEE Trans. On MI, 24, 3, march 2005.
15. Informations available at the World Healt Organization website: http://www.who.int