The Ninth International Symposium on Operations Research and Its Applications (ISORA'10)
Chengdu-Jiuzhaigou, China, August 19–23, 2010

Decision Support System for Skin Cancer Diagnosis

Karol Przystalski Leszek Nowak
Maciej Ogorzałek Grzegorz Surówka

Department of Information Technologies, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland

Abstract Computer aided medical diagnosis is mostly based on very advanced analysis of huge amounts of measured data. Application of computational intelligence methods helps medical doctors to process data faster, and give better, more reliable diagnoses. In dermatology computer analysis is getting more important every year. Differentiation of malignant melanoma images requires very fast image processing and feature extraction/classification algorithms. Applying artificial intelligence algorithms to explore and search large database of dermatoscopic images allows doctors to semantically filter out images with specified characteristics. This paper describes a decision-support system which is based on semantic analysis of melanoma images and further classification of characteristic objects commonly found in pigmented skin lesions. For classification Artificial Neural Networks and Support Vector Machines are used and their performance is evaluated and compared.

Keywords Semantic analysis; melanoma; artificial neural networks; radial basis function; image analysis; image processing; support vector machines

1 Introduction

Skin cancer is a fast developing disease of modern society, reaching 20% [8] increase of diagnosed cases every year. Dermatoscopy is primary and commonly used method of diagnostics for nearly thirty years. This method is non-invasive and requires great deal of experience to make correct diagnosis. As described in [1, 4] only experts arrive at 90% sensitivity and 59% specificity in skin lesion diagnosis while for less trained doctors these figures show significant drop till around 62% - 63% for general Practitioners – see the Table 1 below.

Specificity and sensitivity rates are calculated according to equations:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

This paper proposes decision support system based on semantic analysis method for dermatoscopic images of malignant melanoma [5] cases. Further the system uses automatic classification using two approaches, namely the radial basis function kernels in artificial neural networks and four kernels (linear, polynomial, radial and sigmoid based) for support vector machines. In the analysis process all images are segmented into semantic objects containing various textures, shapes and colors.

	sensitivity	specificity
Experts	90%	59%
Dermatologists	81%	60%
Trainees	85%	36%
General practitioners	62%	63%

Table 1: Sensitivity and specificity

Calculation based on this method shows accuracy in object classification ranging from 88,58% to 97,44%, depending on train/test ratios. Application of gathered knowledge based on conducted analysis allows creation of semantic search tools that can be used for automatic classification of dermatoscopic images. The paper is organized as follows. In section 2 Malignant Melanoma images classification problem is described. Section 3 reviews the idea of proposed method for semantic objects classification. In section 4, the experimental data shows the effectiveness of proposed methods. As a conclusion of this paper, section 5 and 6 contains some future research direction.

2 Classification problem

Skin lesions are difficult to classify because of their short color ranges, instead of real-world images [2]. Malignant melanoma [5] is a kind of skin cancer that has some characteristic color groups like: black, blue-grey, red, light brown, dark brown and skin color. These colors appears on images and can depend on cancer progress stage, lesion depth and blood vessels. This features are good shown on SIAscope images [9]. These colors are also used in melanoma diagnosis scales like ABCD. Melanoma colors are shown in two color spaces on figure 1.

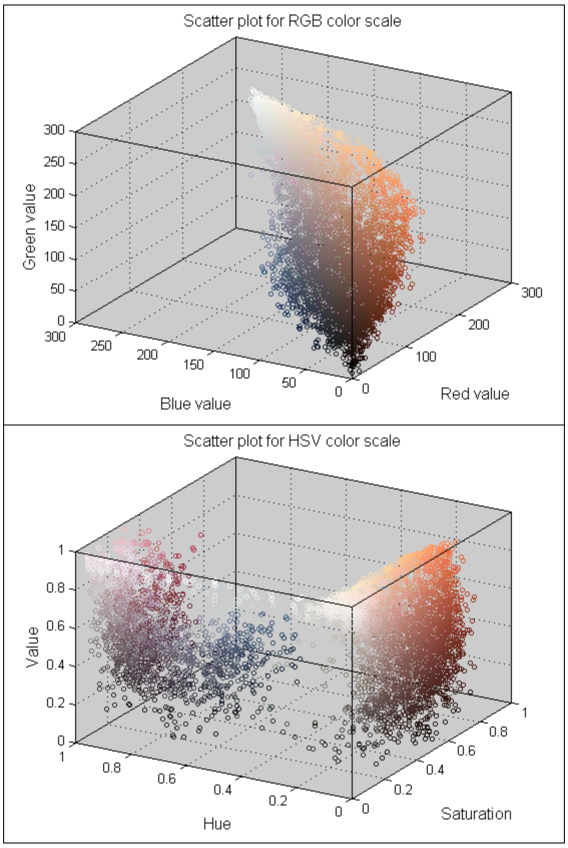

Fig. 1. Objects scatter plots for RGB and HSV color spaces

For this problem classification algorithms like Support Vector Machine [10] and Neural Networks [3] are used. SVM is used with four kernels: linear, polynomial, radial and sigmoid. For neural networks radial basis function is used. All calculations are done using Mathworks Matlab.

3 Segmentation and feature extraction

The images used in the study are diagnosed malignant melanoma cases. Every diagnosed case was confirmed by a histopathology examination. The images used are part of Jagiellonian University Collegium Medicum skin lesions database.

Proposed approach is divided in 5 parts : image segmentation, objects extraction, feature extraction, training and classification. Standard package JSEG [6] is used for image segmentation. Proposed algorithm gives satisfactory results as shown in figure 3a. Segmented image is subject to further operations that lead to border extraction. In this way the binary border mask is generated. Using such mask, objects can be extracted by running simple region growing algorithm. Such an approach is very slow. For 5198 objects this part takes about 90 minutes in MatLab. When every object is separated from the image, feature extraction can be done as a next step. Each object is represented in four color spaces: RGB, HSV, NTSC (YIQ) and YCbCr (YUV). Use of different image decompositions enables better separation experimental points in the feature space.

Each color space is represented by only a few variables: red, green and blue in RGB or hue, saturation and value for HSV. There are 12 colorimetric parameters and a geometric one namely the object size represented by simple pixel count. Together there are 13 features in the proposed features vector as specified in table 2.

Features vector	Feature	Description
RGB	R G B	Red Green Blue
HSV	H S V	Hue Saturation Value
YUV (YCbCr)	Y1 CB CR	Grayscale Blue-difference Red-difference
YIQ(NTSC)	Y2 I Q	Luminance Chrominance Chrominance
	SE	Object size

Table 2: Features Vector

In the diagnosis of malignant melanoma six characteristic color groups: black, red, light brown, dark brown, blue-grey and skin color are typically being considered. After JSEG segmentation all 5189 objects are designated to one of the groups – 7.28% objects belong to the “black” group, 1.85% belong to the “red” group, 14.42% are in the “white” group, 45.1% belong to the “light brown” group, 18.29% to the “dark brown” and 13.07% to the “bluish” group respectively.

Processing method is demonstrated in short on figure 3. Figure 3a. presents segmented image by JSEG with threshold set to 20. Next on figure 3b. feature extraction is shown. Every object features are separately calculated. Color spaces features are just color averages [7]. Figure 3b. is a classification result of support vector machine algorithm.

4 Classification Results

Classification test was performed on four different train to test data ratios (table 3). Every test was repeated 100 times, and presented results are average values of those tests for every ratio. Each test was also performed on randomly selected objects group.

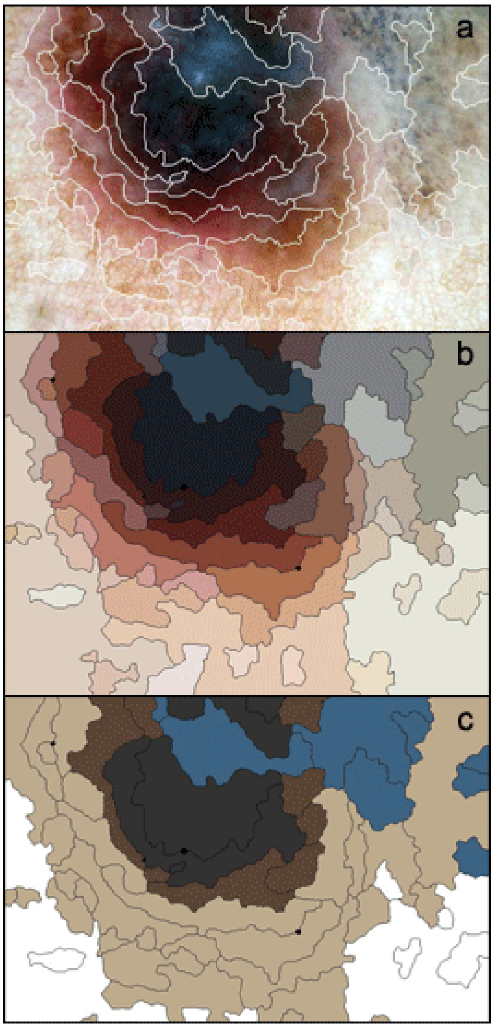

Fig. 3. a) segmented image, b) processed regions,c) classified regions

Train/Test [%]	Trained	Tested
34 / 66	1553	3635
50 / 50	2591	2595
70 / 30	3626	1563
80 / 20	4148	1041

Table 3: Trained / Tested Objects Count

Table 3 contains information about the number of objects belonging to training and testing datasets. Those datasets were prepared with different ratios to see how size of training data can influence classification results, and to determine if there can occur an over-train effect (see 80/20 ratio).

Best classification results were reached with 70/30 train to test ratio with SVM classifiers, and from further on, only this results will be discussed.

	SVM				NN
ratio	linear	polynomial	radial	sigmoid	RBF
34/66	**96,75%**	94,04%	96,45%	95,71%	92,78%
50/50	**97,12%**	94,72%	96,95%	95,90%	93,01%
70/30	**97,44%**	95,16%	97,29%	95,97%	93,08%
80/20	**95,53%**	90,54%	94,10%	92,08%	88,58%

Table 4: Classification Success Rates

Table 5 presents results of classification rates for six object groups: skin regions, red regions, black regions, light and dark brown regions, and grey-blue regions. Best results (average 98,30 %) are obtained for objects corresponding to dark-brown regions of the images which are easiest to classify, and can be mistaken by black, or light brown. Second best result is reached with standard skin region (average 97.53% for SVM classifiers). This is because of characteristic high color values and size of skin region. Most images contain only few regions that are classified as skin, and those are usually quite large corresponding to other object on that image.

Objects classified as black regions of image represent parts of the skin lesion which contain very high concentration of melanin (skin pigmentation) in multiple skin layers. Most malignant melanoma images contain few black regions. As black are classified regions with low intensity value. Due to low saturation those regions can be sometimes misclassified as blue-gray regions reaching only average of 93.89% (for SVM classifier) success rate.

Blood vessels visible on dermatoscopic images appear in pink-red color. Results in classification success rates for regions containing blood vessels reach an average of 94.33% for SVM classifiers. Number of classified objects (figure 2) shows that only few images contained blood vessels, and for testing purposes only 96 out of 5189 was classified as red. This could suggest that misclassified regions could contained brown objects that disrupted classification process.

Dark brown color represents high melanin concentration in top layers of skin, and it is found in most pigmentation skin lesions images. Those regions are easily recognized and classified, although they can be misclassified as light brown regions because it is difficult to determine when region is light brown or when it is already dark.

Second best classification rate is achieved for light brown regions with count is the biggest corresponding to other found on images. These regions represent parts of the images where it is possible to detect pigmentation network. For the same reasons as mentioned earlier some misclassification has occurred between light brown and dark brown regions.

	linear	polynomial	radial	sigmoid	NN RBF
black	93,80%	92,84%	**94,66%**	94,28%	68,62%
skin	97,49%	97,36%	**98,01%**	97,29%	93,25%
red	**95,87%**	91,43%	95,05%	94,98%	92,72%
light	**97,83%**	94,79%	97,58%	95,32%	95,55%
dark	**98,99%**	98,15%	98,80%	97,95%	97,64%
blue	**80,07%**	42,71%	77,32%	56,50%	53,75%

Table 5: Classification Success Rate For 70/30 Ratio

Last classified regions of image are those covered by blue-grey veil. Characteristic of these regions is that areas covered by them could easily belong to other region types, because blue-grey veil can appear over any other region class.

5 Conclusion

As shown in table 5 the proposed approaches allow for building efficient diagnosis support systems. The success rate in object classification of 93,08% for 70/30 train to test ratio for RBF classifier is very high. Even better success rate has been obtained for linear SVM - 97,44% for 70/30 train to test ratio. Classified objects are further analysed in terms of texture content and smaller object count. For example object classified as light brown or dark brown have high chance to be covered by pigmentation network. Again due to high pigmentation, black objects have little to none chance for containing blood vessels, or regression regions.

On the summary, the Support Vector Machines with linear kernel proved to perform best in classification task and provide very promising results in classifying regions in the malignant melanoma images. The results obtained using the proposed decision support approach are good enough to be helpful for medical doctors for semantic image filtering and can be helpful in future research allowing fast access to images that meet given criteria.

6 Future work

Further exploration of gathered data, will include extended description of analyzed classes by additional parameters that will determine standard deviation from typical object of every class. This research will be interesting and should give more precise results in terms of object classification. This solution will be used for building algorithms designed for melanocytic nevi texture recognition, texture and fractal dimension analysis.

Acknowledgments

This research has been supported by the Research Grant No. N518 419038

References

[1] S. W. Menzies, L. Bischof, H. Talbot et al., "The Performance of SolarScan. An Automated Dermoscopy Image Analysis Instrument for the Diagnosis of Primary Melanoma.", *Archive of Dermatology*, pp. 1388-1396, 2005.

[2] Chuan Yu-Chang, Hung-Jen Wang, Chi-Fang Li, "Semantic analysis of real-world images using support vector machine", *Expert Systems with Applications*, pp. 10560-10569, 2009.

[3] S. B. Park., J. W. Lee and S. K. Kim, "Content-based image classification using neural network", pp. 287-300, *Pattern Recognition*, 2004.

[4] M. Burroni, R. Corona, G. Dell'Eva, F. Sera, R. Bono, P. Puddu, R. Perotti, F. Nobile, L. Andreassi, P. Rubegni, "Melanoma Computer-Aided Diagnosis: Reliability and Feasibility Study", *Clinical cancer research*, pp. 1881-1886, 2004.

[5] F. de Braud, D. Khayat, B. B. R. Kroon, R. Valdagni, P. Bruzzi, N. Cascinelli, "Malignant melanoma", *Critical Reviews in Oncology/Hematology*, pp. 35-63, 2003.

[6] Y. Deng, and S. B. Manjunath, "Unsupervised segmentation of color-texture regions in images and video", *IEEE Transactions on Pattern Analysis and Machine Inteligence*, pp. 800–810, 2001

[7] Y. Liu, D. Zhang, G. Lu, and Y. Ma, "Region-based image retrieval with high-level semantic color names", *IEEE Proceedings of the Multimedia Modeling Conference*, pp. 180–187, 2005.

[8] J. Han, G. A Colditz and D. J. Hunter, "Risk factors for skin cancers: a nested case–control study within the Nurses' Health Study", *International Journal of Epidemiology*, pp. 1514-1521, 2006.

[9] M. Moncrieff, S. Cotton, P. Hall, R. Schiffner, U. Lepski, E. Claridge, "SIAscopy assists in the diagnosis of melanoma by utilizing computer vision techniques to visualise the internal structure of the skin", pp. 53-56, 2001.

[10] C. Wu Y., S. Lee Y. and C. Yang J., "Robust and efficient multiclass SVM models for phrase pattern recognition", *Pattern Recognition*, pp. 2874–2889, 2008