

# AN ENSEMBLE TEMPLATE MATCHING AND CONTENT-BASED IMAGE RETRIEVAL SCHEME TOWARDS EARLY STAGE DETECTION OF MELANOMA

SPIROS KOSTOPOULOS<sup>✉,1</sup>, DIMITRIS GLOTSOS<sup>1</sup>, PANTELIS ASVESTAS<sup>1</sup>, CHRISTOS KONSTANDINOU<sup>3</sup>, GEORGE XENOGIANNOPOULOS<sup>1</sup>, KONSTANTINOS SIDIROPOULOS<sup>2</sup>, EIRINI-KONSTANTINA NIKOLATOU<sup>1</sup>, KONSTANTINOS PERAKIS<sup>4</sup>, SPYROS MANTZOURATOS<sup>4</sup>, THEOPHILOS SAKKIS<sup>5</sup>, GEORGE SAKELLAROPOULOS<sup>3</sup>, GEORGE NIKIFORIDIS<sup>3</sup> AND DIONISIS CAVOURAS<sup>1</sup>

<sup>1</sup>Medical Image and Signal Processing Laboratory, Department of Biomedical Engineering, Technological Educational Institute of Athens, Greece; <sup>2</sup>European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; <sup>3</sup>Department of Medical Physics, University of Patras, 26504, Rio, Patras, Greece; <sup>4</sup>UBITECH Research Department, UBITECH Ltd., Athens, Greece; <sup>5</sup>Dermatology Center, Aegion, Greece  
e-mail: skostopoulos@teiath.gr

(Received November 27, 2015; revised March 29, 2016; revised June 15, 2016; accepted June 22, 2016)

## ABSTRACT

Malignant melanoma represents the most dangerous type of skin cancer. In this study we present an ensemble classification scheme, employing the mutual information, the cross-correlation and the clustering based on proximity of image features methods, for early stage assessment of melanomas on plain photography images. The proposed scheme performs two main operations. First, it retrieves the most similar, to the unknown case, image samples from an available image database with verified benign moles and malignant melanoma cases. Second, it provides an automated estimation regarding the nature of the unknown image sample based on the majority of the most similar images retrieved from the available database. Clinical material comprised 75 melanoma and 75 benign plain photography images collected from publicly available dermatological atlases. Results showed that the ensemble scheme outperformed all other methods tested in terms of accuracy with  $94.9 \pm 1.5\%$ , following an external cross-validation evaluation methodology. The proposed scheme may benefit patients by providing a second opinion consultation during the self-skin examination process and the physician by providing a second opinion estimation regarding the nature of suspicious moles that may assist towards decision making especially for ambiguous cases, safeguarding, in this way from potential diagnostic misinterpretations.

Keywords: content-based image retrieval, decision support system, melanoma diagnosis, self-skin examination, template matching

## INTRODUCTION

Malignant melanoma represents the most dangerous type of skin cancer with annual incidences of 48.000 new cases worldwide according to the World Health Organization (Lucas, *et al.*, 2006). Increased ultraviolet (UV) radiation has proved to be the most important risk factor of the disease (Rastrelli, *et al.*, 2014). A relative large number of inherited and non-inherited gene mutations have been implicated in the pathogenesis of melanoma. But besides UV radiation, the aetiology of the disease is largely unknown making it difficult to establish preventing strategies and effective therapies. Melanomas have good prognosis when

they are detected at early stages, since available treatments, such as surgical excision, will mostly retain affected patients disease-free for more than 5-years (Veronesi, *et al.*, 1991, Ringborg, *et al.*, 1996, Cohn-Cedermark, *et al.*, 2000, Balch, *et al.*, 2001). One of the most popular technologies that have proven to be effective in discriminating melanomas from normal moles and other skin lesions (>90% detection accuracy (Schein, *et al.*, 2009)), comprise digital dermoscopy, which allows expert physicians to visually observe suspected lesions using polarized or non-polarized light (Tenenhaus, *et al.*, 2010). On the other hand, routine eye examination has proven to be significantly less effective with detection rates approximately 65%

(Schein, *et al.*, 2009). Thus, dermoscopy may be considered as the basic instrumentation that is utilized for melanoma detection in daily practice. However, dermoscopy presents certain limitations. The quality and accuracy of diagnostic conclusions greatly depend on the experience of the observing physician. Considering that early stage melanomas present very subtle visual changes as compared to benign moles, the identification of malignancy evidence (Abbasi, *et al.*, 2004) is not straightforward. Thus, the risk of exonerating suspicious moles is accountable, endangering inappropriate patient management with debatable effects in patient prognosis (Lorentzen, *et al.*, 2001, Pfahlberg, *et al.*, 2008, Veierod, *et al.*, 2009).

Although dermoscopy may contribute towards the early detection of melanomas, it has been shown that many patients refer to the physician only when the malignancy has progressed and the visual signs are obvious, since they do not have the sensitivity of visually discriminating the disease at its early phases, when the visual signs are more subtle (Carli, *et al.*, 2002). At later stages, the detection of melanomas with dermoscopy becomes more straightforward, however, the risk for a poor prognosis increases, since late phase melanomas tend to metastasize aggressively (Rastrelli, *et al.*, 2014). Thus, it is of paramount importance to alert patients towards a visit to the physician as soon as possible.

One promising strategy towards this direction is the self-skin examination (Carli, *et al.*, 2002). Self-skin examination has been shown to improve long term survival of patients with melanoma, lowering the risk of death after 10 years of initial diagnosis by 25% (Leachman, *et al.*, 2016, Paddock, *et al.*, 2016). The patient assesses visually new and/or existing moles and refers to the physician when a suspicious pigmented mole is detected. However, self-assessment of one's skin moles may be difficult, rendering the self-skin examination an inadequate strategy for wide-spread melanoma screening. The significant value of self-skin examination has driven research towards the development of new technologies that may offer patients and physicians means for more effective, frequent and distant inspection of suspicious moles. Computer-based automated tools have been previously proposed, which can be used as a second opinion tool for self-skin examination and advice patients regarding the urgency for a physician visit. Moreover, such systems have been used to address another important liability in the early stage detection of melanoma, which is the risk of diagnostic misinterpretations (Stringa, 1988, Field, 1994, Grant-Kels, *et al.*, 1999, Ming, 2000, Zagrouba, *et al.*, 2004, Zhang, *et al.*, 2010, Abikhair, *et al.*, 2014).

Handheld devices, such as smartphones and tablets, are becoming increasingly popular. More than 1.75 billion such users have been predicted for 2015, making these devices ideal candidates for accessing moles through smartphone applications that may be used to facilitate self-skin examination and distance monitoring of patients by the expert physicians. A number of applications are nowadays commercially available for melanoma detection on the basis of smartphone-camera generated plain photography images (Robson, *et al.*, 2012, Stoecker, *et al.*, 2013, Wolf, *et al.*, 2013, Vañó-Galván, *et al.*, 2015).

A recent comprehensive review lists 39 such applications (Kassianos, *et al.*, 2015). However, most of these applications tend to conceal their algorithmic architecture due to reasons such as patenting. Moreover, scientific analysis is usually either lacking or limited, making experts sceptical regarding the effectiveness of these technologies as self-skin examination facilitators for patients or second opinion consultants for experts.

In this study, we present a decision support system for melanoma detection, which attempts to guide patients to meaningful alerts regarding the urgency for a physician visit and safeguard physicians' decisions from diagnostic misinterpretations by means of second opinion consultations. In comparison to previous studies, the proposed system differs in the following: a/ the decision support system technology relies on the combination of three different template matching and content-based image retrieval algorithms, namely the mutual information, the cross-correlation and the clustering based on image features proximity approach, which are merged in a majority vote ensemble scheme. In this way, it is possible to investigate the image content properties from different and complementary perspectives and combine information involving the image's entropy, the image's cross-correlation and the specific morphological, textural, and color characteristics of each investigated mole. To the best of our knowledge such an ensemble scheme is for the first time investigated. b/ The proposed system has been tested on plain photography images collected from different dermatological atlases. In this way, it was possible to investigate the effectiveness of the ensemble scheme in the identification of melanoma in images that have been generated under different conditions and equipment (*i.e.*, different cameras, analyses, angles, lighting etc.). c/ The proposed system has been comprehensively evaluated using an external cross-validation process in order to approximate the performance of the system to unknown data.

## MATERIAL AND METHODS

### CASE MATERIAL

The dataset consisted of 75 melanoma and 75 benign moles plain photography images, each corresponding to a different case, collected from publicly available resources/databases, such as six (6) from the Loyola University Dermatology Medical Education Website<sup>1</sup>, thirteen (13) from the Danderm Atlas of Clinical Dermatology<sup>2</sup>, three (3) from the Hellenic dermatological atlas<sup>3</sup>, three (3) from the atlasdermatologico.com.br website<sup>4</sup> and fifty (50) melanomas and seventy (75) from the DERMOFIT database<sup>5</sup> (Ballerini, *et al.*, 2013).

### IMAGE PREPARATION & PREPROCESSING

Each image was preprocessed using the DullRazor algorithm (Lee, *et al.*, 1997) that was utilized in order to eliminate hair pixels overlapping the mole region. The algorithm operates in three main stages. At the first stage the location of the pixels belonging to hair regions is identified using morphological filtering. At the second stage the pixel values of the hair regions is re-calculated by means of interpolation with nearest regions. Finally, at the third stage a smoothing filtering operation is applied to level the intensity around interpolated regions.

Following the DullRazor algorithm, images were filtered using the mean shift algorithm (Fukunaga, 1975), which is very effective in flattening the image's texture. In this way, it was possible to obtain a preliminary separation of the mole region from the surround-

ing background and prepare the image for the subsequent step of image segmentation. The illumination of the image was, then, corrected using a polynomial fitting algorithm, whose terms were estimated using a least square approach (Gonzalez, *et al.*, 2002). Finally, the image was thresholded using the minimum cross entropy thresholding method (Li, *et al.*, 1993) in order to separate pixels of the mole region from pixels of the surrounding background regions. An example of the image pre-processing and segmentation stage is illustrated at Fig. 1.

### ENSEMBLE TEMPLATE MATCHING AND CONTENT-BASED IMAGE RETRIEVAL SCHEME

The main task of the proposed ensemble template matching and content-based image retrieval scheme is twofold: a/ to retrieve the  $n$  most similar, to the unknown case, image samples from the available verified image database and b/ to provide an automated consultation regarding the nature of the unknown sample (benign case or malignant melanoma).

The ensemble scheme was designed based on three well documented methods, the mutual information (MI) (Mazurowski, *et al.*, 2011), the cross-correlation (COR) (Asgarizadeh, *et al.*, 2012), and the content based image retrieval based on clustering of image features (FC) (Yan, *et al.*, 2011) in order to investigate image content from three different and complementary perspectives, involving the image's entropy, the image's cross-correlation and the specific morphological, textural, and color characteristics of each investigated mole.

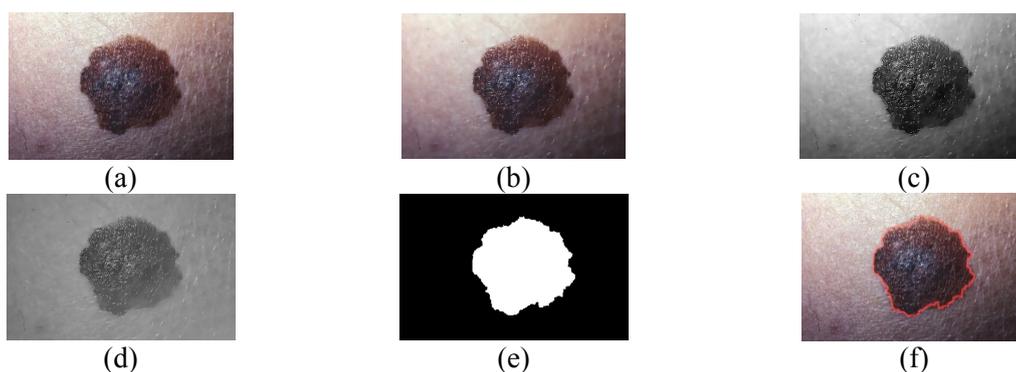


Fig. 1. Mole segmentation process (a) original RGB image, (b) filtered image, (c) gray scale transformation, (d) illumination correction, (e) thresholded image, (g) superimposition of the mole's border on the original image.

<sup>1</sup><http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/content1.htm>

<sup>2</sup><http://www.danderm-pdv.is.kkh.dk/atlas/index.html?PHPSESSID=b3fad1be23c2edb54e85b29dc7c6ba2e>

<sup>3</sup><http://www.hellenicdermatlas.com/en/>

<sup>4</sup><http://www.atlasdermatologico.com.br/>

<sup>5</sup><http://homepages.inf.ed.ac.uk/rbf/DERMOFIT/>

The determination of the most similar images using the mutual information and the cross-correlation criteria relied on testing each inputted image against all other images in the available database. The determination of the most similar images using the content based image retrieval method relied on testing each feature subset extracted from the segmented mole of the inputted image against all other feature subsets that are computed from the segmented mole of the images in the available database. An example of the content-based image retrieval process is illustrated at Fig. 2.

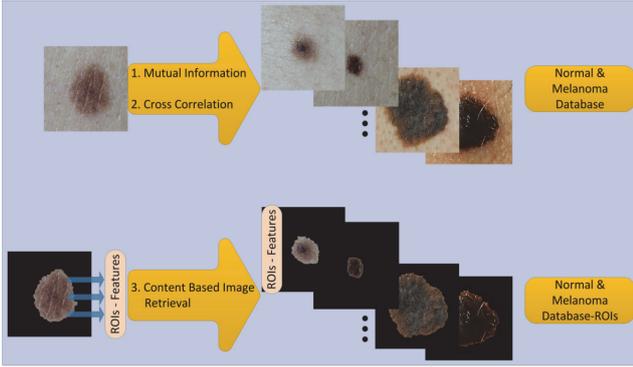


Fig. 2. *Image retrieval methods utilized in this study.*

#### A. Mutual information (MI)

Mutual Information is a method originating from information theory. It has been employed in numerous content-based image retrieval and template matching applications. Mutual information is related to the joint entropy of two images. Mutual information between an unknown image  $I_u$ , and an image from an available database  $I_j$ , may be calculated as (Russakoff, *et al.*, 2004):

$$MI(I_u, I_j) = H(I_u) + H(I_j) - H(I_u, I_j) \quad (1)$$

with

$$(I_u, I_j) = - \sum_{I_u, I_j} p_{I_u, I_j}(I_u, I_j) \log p_{I_u, I_j}(I_u, I_j)$$

and

$$H(X) = - \sum_x p_x(x) \log p_x(x)$$

where  $j = 1:N$ , and  $N$  is the number of available images in the database. The input to the mutual information algorithm comprised single-mole images (*see* Fig. 2). Mutual information was then calculated for all possible pairs that included the unknown image sample and one of the available database's image samples. The most similar images were considered those having the largest mutual information with the unknown image sample.

#### B. Cross-correlation (COR)

The cross-correlation between two images  $I_u$ , and  $I_j$ , is a measure of similarity and may be defined as (Gaidhane, *et al.*, 2012):

$$\mu = \frac{\sum_x^m = \sum_y^n = [I_u(x, y) - \bar{I}_u][I_j(x, y) - \bar{I}_j]}{(2m+1)(2n+1)\sigma_{I_u}\sigma_{I_j}}, \quad (2)$$

$$\mu = \frac{\sum_x^m = \sum_y^n = [I_j(x, y) - \overline{I_u, I_j}]}{(2m+1)(2n+1)\sigma_{I_u}\sigma_{I_j}},$$

where  $\bar{I}_u, \bar{I}_j$  are the mean grey-level values of the images and  $\sigma_{I_u}, \sigma_{I_j}$  are the standard deviations of the grey-level values of the images. The input to the cross-correlation algorithm comprised single-mole images (*see* Fig. 2). The  $\mu$  parameter was then calculated for all possible image pairs that included the unknown image sample and one of the verified database's image samples. The most similar images were considered as those having the largest  $\mu$  with the unknown image sample.

#### C. Content based image retrieval (FC)

Another perspective for investigating image similarity focuses on image content (features). The content-based image retrieval algorithm utilized in this study is the fuzzy c-means clustering algorithm (Jain, *et al.*, 1988), that was designed to partition image features into two clusters/classes: those characterizing benign cases, and those describing melanoma cases. Image features comprised 72 measurements related to the mole's morphology (10), grey-level histogram (4), texture (38), and colour (20) (Loukas, *et al.*, 2013, Ninos, *et al.*, 2013). The fuzzy c-means algorithm operates following an iterative procedure during which each image feature set (representing a unique image sample from the available database) gets a fuzzy allocation to a cluster according to distance metric criteria. The algorithm iteratively facilitates for minimization of the objective function (Eq. 3)

$$J(M, C) = \sum_{i=1}^N \sum_{j=1}^C m_{ij}^w \|x_i - c_j\|^2, \quad (3)$$

to provide a solution for the membership function matrix  $M$  and cluster centre matrix  $C$ , where  $m_{ij}^w$  is the degree of membership of  $x_i$  feature-vector in the cluster  $j$ ,  $\|x_i - c_j\|$  is the Euclidean distance between  $j$ - $th$  cluster centre and  $i$ - $th$  feature-vector, and  $w \in [1, \infty]$

is the fuzzy exponent, which determines the degree of fuzziness. The algorithm converges when  $|m_{ij}^{k+1} - m_{ij}^k| < \varepsilon$ , where  $\varepsilon \in (0,1)$  is a termination criterion and  $k$  is the iteration steps. Then an unknown feature-vector (from a “new” image sample) is assigned to the cluster with the minimum distance from its centroid.

The features were normalized to zero mean and unit standard deviation (Theodoridis, *et al.*, 2003). In order to avoid overfitting a feature selection methodology was followed by ranking features in descending order using a class separability criterion that was based on the Wilcoxon test and the correlation between features (Theodoridis, *et al.*, 2003). Following, only a part (one-third of the smallest class in the database, twelve features for our samples) of the ranked features were selected for further analysis by the FC algorithm.

## ENSEMBLE SCHEME

The different and complementary information that was assessed using the above mentioned three methods was combined through a majority vote rule in order to:

- provide the  $n$  most similar images that the majority of these three algorithms decides and
- classify the unknown image case as ‘benign’, if the majority of the  $n$  similar images emerges from the ‘benign mole category’, or as ‘melanoma’, if the majority of the  $n$  similar images emerges from the ‘malignant melanoma mole category’.

The majority vote rule is given by (Kittler, *et al.*, 1998):

$$D_c(X) = \sum_{i=1}^R d_{c,i}(X), \quad (4)$$

where  $c$  is the class (benign/melanoma),  $X$  is the unknown image sample,  $i = 1,2,3$  is the odd number of methods involved in the majority vote scheme,  $d_{c,i}$  is the binary decision value (0,1), 0 corresponds to melanoma and 1 to benign classes. Thus, if  $D_1(X) > D_0(X)$ , the unknown image-mole is categorized as benign, otherwise is categorized as melanoma.

## PERFORMANCE EVALUATION

- Performance of MI and COR methods in identifying single-mole images against their-self following rotation at different angles

Each image from the available database was rotated at eight (8) different angles, from  $-20^\circ$  to  $+20^\circ$  with a  $5^\circ$  step. Subsequently, each rotated image was

inputted to the MI and COR algorithms, which were asked to return the most similar image from the same database (including the original, un-rotated version of the inputted image). If the algorithms returned as the most similar image the un-rotated version of the inputted image, then a successful retrieval was considered, otherwise as unsuccessful. In this way, it was possible to determine the robustness of the MI and COR algorithms when images are slightly rotated. The FC algorithm is rotation invariant since it depends only on features extracted from segmented moles. The features that were included in this study are rotational invariant.

- Performance of the proposed ensemble scheme using a leave-one-out data splitting approach

In order to evaluate the performance of the proposed ensemble scheme, the following methodology was utilized: each image sample from the available verified database (benign or malignant) was tested against the remaining database (alike to the leave one out method (Theodoridis, *et al.*, 2003)). Then, the  $n$  most similar images to the unknown sample were retrieved along with their corresponding labels (benign, malignant). If the majority of the  $n$  most similar images were benign cases, then the unknown image was classified as benign, whereas if the majority of the most similar images were melanoma cases, then the unknown image was classified as melanoma. Based on the above classification, a truth table was constructed in order to evaluate and compare the performance of each single algorithm tested (mutual information, cross-correlation, fuzzy c-means) against the ensemble majority vote scheme. Moreover, the above evaluation process was repeated by changing the number of  $n$  most similar images from 1 to 19. The evaluation of the performance was based on five different metrics (scores 1-5) that are derived from the truth table, namely the accuracy (score 1), the sensitivity (score 2), the specificity (score 3), the diagnostic accuracy (score 4) and the Cohen-k (score 5) where:

$$Score1 (Accuracy) = \frac{TP + TN}{TP + TN + FP + FN}, \quad (5)$$

$$Score2 (Sensitivity) = \frac{TP}{TP + FN}, \quad (6)$$

$$Score3 (Specificity) = \frac{TN}{TN + FP}, \quad (7)$$

$$Score4 (Diagnostic accuracy) = \frac{TP}{TP + FP + FN}, \quad (8)$$

where  $TP$  is the number of true positive cases,  $TN$  is the number of true negative cases,  $FP$  is the number of false positive cases and  $FN$  is the number of false negative cases.

$$Score5(Choen-k) = \frac{n \sum_{i=1}^k M_{ii} - \sum_{i=1}^k M_{i.} M_{.i}}{n^2 - \sum_{i=1}^k M_{i.} M_{.i}} \quad (9)$$

where  $M$  is the confusion matrix,  $M_{.i}$  is the sum of elements of  $i$ -th column of  $M$  and  $M_{i.}$  is the sum of elements of  $i$ -th row of  $M$ .

### C. Performance of the proposed ensemble scheme using an external cross-validation data splitting approach

Moreover, an external cross-validation (ECV) splitting of the data was also performed, in order to get less biased estimates than the leave-one-out splitting. Data were randomly split into two subsets, each comprising 50% of all available images. Image samples from the first subset (testing data) were considered as unknown. Then, the algorithm was asked to retrieve the  $n$  most similar, to the testing cases, images by searching only the second dataset (template data), which was considered as having known labels. This process was repeated ten (10) times and the final estimate of the evaluation performance was computed as the average of all classification performances obtained for each different repetition. The above analysis was performed separately for each different  $n$  number of similar images. In this way, we considered that a less biased estimate might be obtained than using the leave-one-out method (Ambrose, *et al.*, 2002).

## RESULTS

Regarding the performance of the MI and COR algorithms in identifying single-mole images that have been rotated at different angles, results are summarized in Fig. 3, which illustrates a good performance with 84.2%-98.5% detection accuracy.

Regarding the performance of the proposed content-based image retrieval classification scheme for the leave-one-out data splitting, results are summarized in Fig. 4 for each single algorithm (mutual information,

cross-correlation and fuzzy c-means) and the ensemble scheme for different number of  $n$  similar images ( $n = 1:19$ ) and for each of the five different performance evaluation metrics described in the previous paragraphs (scores 1-5). For a small number of similar images (up to 3) the fuzzy c-means outperformed the mutual information and cross-correlation algorithms for all metrics. The cross-correlation method became the most effective algorithm for more than 3 similar images. The mutual information algorithm presented the best specificity, independently of the number of similar images investigated. The ensemble scheme proved the most accurate, outperforming each single algorithm tested for all metrics.

Regarding the performance of the proposed content-based image retrieval classification scheme for the external cross-validation data splitting, the ensemble scheme resulted in optimal performances for smaller numbers of similar images (see Fig. 5 and Table 1). The increase in the prediction accuracy with the majority vote scheme may be justified by the fact that the proposed methods (MI, FC and COR) combined complementary information.

Moreover, and for comparison reasons, the SVM algorithm (El-Naqa, *et al.*, 2004) was also tested, as an alternative to our method, and led to  $78 \pm 5\%$  overall accuracy (with various kernels) using the ECV method.

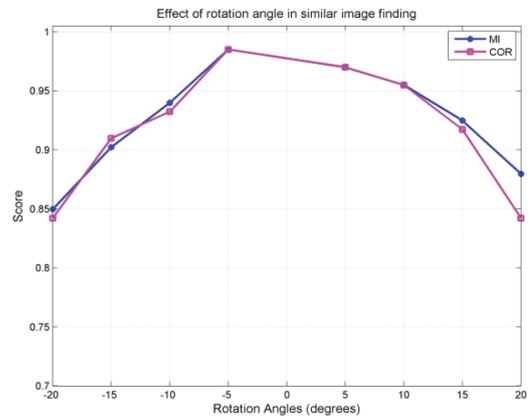


Fig. 3. The dependence of the MI and COR in detecting a single-mole image that has been rotated at different angles (MI: mutual information, COR: cross-correlation).

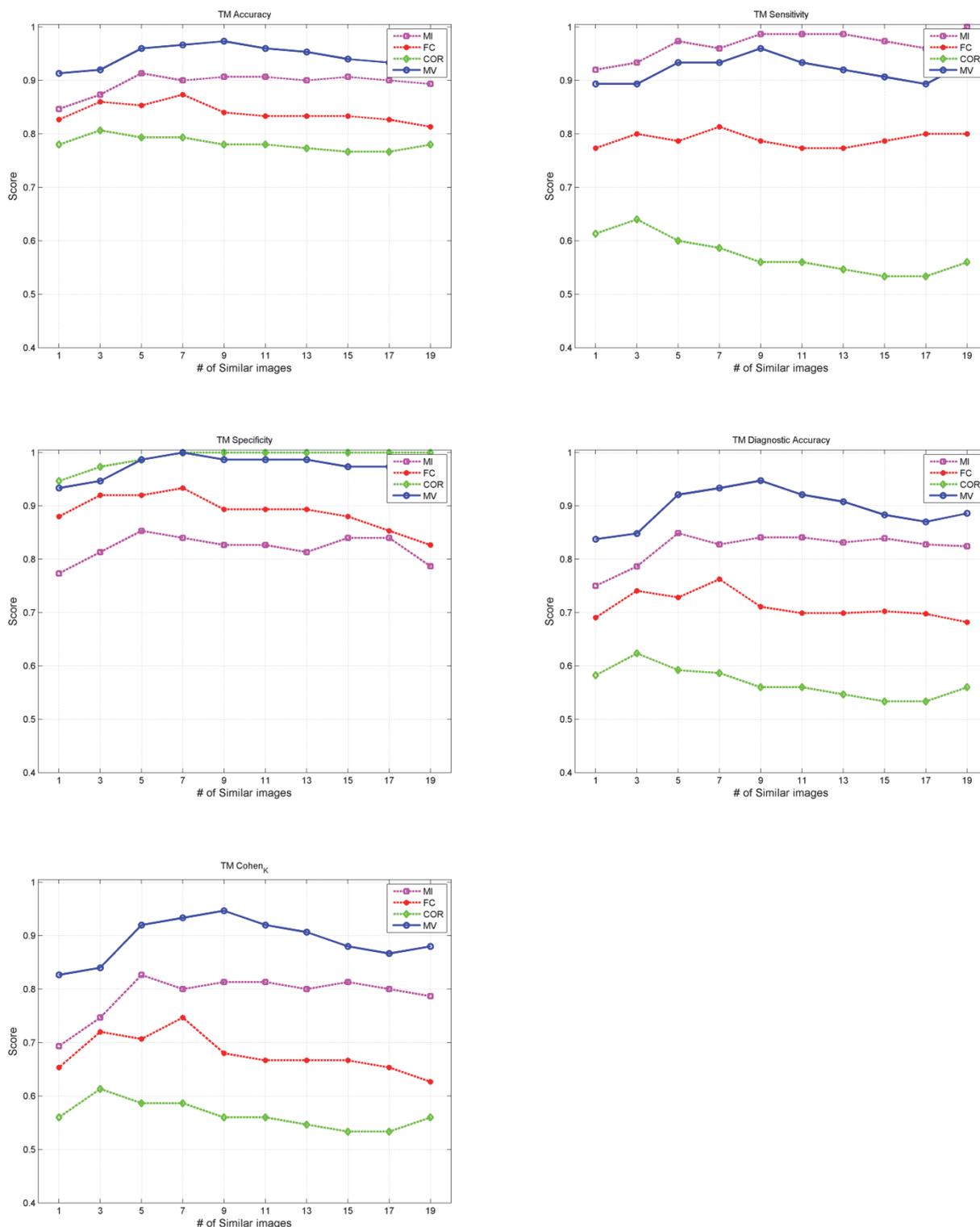


Fig. 4. Each plot corresponds to the score of each metric (accuracy, sensitivity, specificity, diagnostic accuracy and Cohen-k) for the three methods (MI: mutual information, FC: features clustering, COR: cross-correlation) and majority vote rule (MV), when the Leave One Out method was implemented.

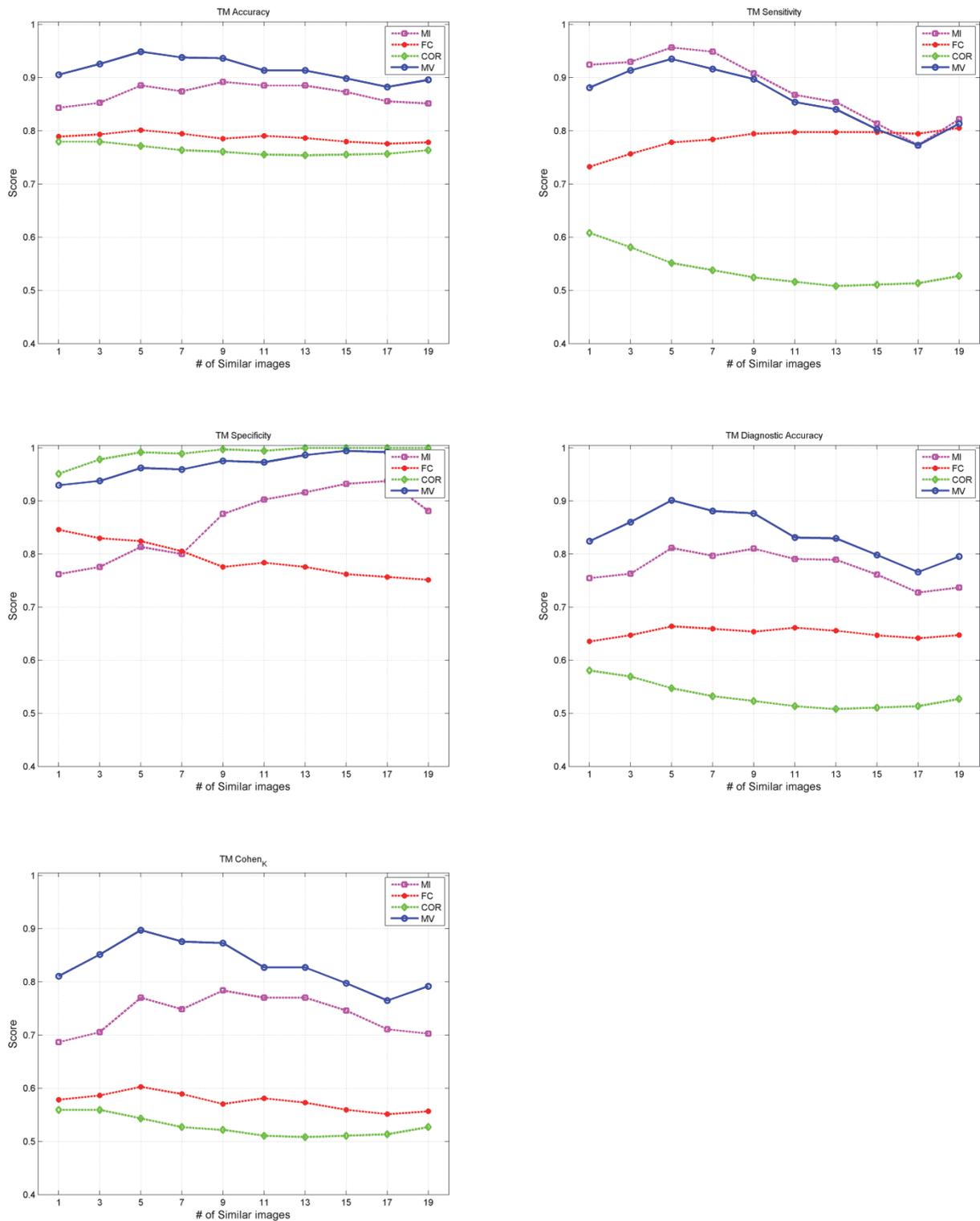


Fig. 5. Each plot corresponds to the mean score of each metric (accuracy, sensitivity, specificity, diagnostic accuracy and Cohen-k) for the three methods (MI: mutual information, FC: features clustering, COR: cross-correlation) and majority vote rule (MV), when the external cross-validation method was employed for ten repetitions.

Table 1. Best average performances of MI, FC, COR and MV regarding the five metrics for the 10 generated datasets, and the corresponding number of similar images required.

	MI	FC	COR	MV
	MnV $\pm$ Std	MnV $\pm$ Std	MnV $\pm$ Std	MnV $\pm$ Std
<b>Accuracy</b>	89.2 $\pm$ 3.8	80.1 $\pm$ 4.0	78.0 $\pm$ 3.2	94.9 $\pm$ 1.5
# of similar images required	9	5	3	5
<b>Sensitivity</b>	95.7 $\pm$ 2.6	80.5 $\pm$ 8.8	60.8 $\pm$ 5.4	93.5 $\pm$ 3.4
# of similar images required	5	19	1	5
<b>Specificity</b>	93.8 $\pm$ 9.5	84.6 $\pm$ 8.7	100.0 $\pm$ 0.0	99.5 $\pm$ 1.7
# of similar images required	17	1	13	15
<b>Diagnostic Accuracy</b>	0.81 $\pm$ 0.05	0.66 $\pm$ 0.05	0.58 $\pm$ 0.06	0.90 $\pm$ 0.03
# of similar images required	9	5	1	5
<b>Cohen k</b>	0.78 $\pm$ 0.08	0.60 $\pm$ 0.08	0.56 $\pm$ 0.04	0.90 $\pm$ 0.03
# of similar images required	9	5	3	5

## DISCUSSION

In this study, an ensemble template matching and content-based image retrieval scheme were designed for assisting physicians towards early detection of melanomas and alerting patients towards the urgency for a physician visit. The proposed system may assist the expert physician by a/ providing the most similar, to the examined case, images from a known database of skin mole and melanoma images and b/ providing an automated second opinion consultation regarding the nature of the examined skin lesion.

Moreover, the proposed scheme can be of assistance to the patient by providing consultations regarding solely the necessity for evaluation of the examined skin mole by an expert physician.

The ensemble scheme was constructed using three complementary approaches, the MI, COR, and the FC. These algorithms sought for similarities from a different point of view, involving the image's entropy, the image's cross-correlation and the specific morphological, textural, and color characteristics of each investigated mole, which were in total 72 features. Entropy was used to investigate the organization of the textural information in the image. Cross-correlation was used to investigate the texture correlation between different image patterns. Melanomas have been found to exhibit elaborated textural patterns, which can be encoded by means of the spatial distribution of the various colors and intensities of the mole pixels, thus, the MI and COR algorithms may be used to capture these diagnostic meaningful differences. More-

over, with the FC algorithm it was possible to investigate the morphology, texture and colour properties of the examined moles and relate these properties with patterns appearing in melanoma cases, providing, in this way, a complementary perspective of the examined image mole signatures.

Although these three algorithms, when operating in a standalone mode, provided average performances in the five different metrics tested (*i.e.*, accuracy MI 91.3%, COR 79.3%, FC 85.3%), when combined under the majority vote scheme the performances were boosted up (*i.e.*, accuracy MV 96.0%) when tested using the leave-one-out method. The increased performance might be explained by the complementarity of the nature of the information that each distinct algorithm offered to the ensemble scheme. Regarding the external cross-validation data splitting, the MV method outperformed all other methods in terms of accuracy with 94.9  $\pm$  1.5% (MI 89.2  $\pm$  3.8%, COR 78.0  $\pm$  3.2% and FC 78.0  $\pm$  3.2%).

Considering the fact that the database utilized in this study comprised extraction from multiple dermatological atlases that contain publicly available images, the high performance of the proposed scheme may justify its effectiveness to detect melanoma signatures on plain photography images, despite the digitization equipment, the angle of photography, the lighting conditions etc., under the premise that the photographs have sufficient diagnostic quality.

A lot of research efforts have been previously presented for melanoma detection based on dermoscopy images or normal digital camera images. Two

main categories of studies may be identified. The first category consists of efforts focusing on statistical pattern recognition, whereas the second category comprises efforts focusing on template matching and content-based image retrieval. Regarding the first category, representative studies may be found in (Cavalcanti, *et al.*, 2013), which proposed a k-nearest neighbor (k-NN) classifier using 52 features extracted based on the ABCD rule with 99.3% overall accuracy, in Jaleel *et al.* (2012), which proposed an artificial neural network (ANN) classifier with 100% prediction accuracy and in Ruiz *et al.* (2011), which proposed an ensemble pattern recognition scheme combining three distinct classifiers, the k-NN, the Bayesian and the ANN, with accuracy 87.76%. Regarding the second category, representative studies can be found in Ballerini *et al.* (2010; 2013), which proposed a content-based image retrieval system investigating textural and color features, in Maragoudakis and Maglogiannis (2011), which proposed an ontology structure model based on features extracted from skin lesion images based on agglomerative clustering and distance criteria and in Chen *et al.* (2016), which is a recent study proposing a content-based image retrieval system that identified melanomas on plain photography images with performances exceeding 90% for all metrics tested.

In terms of classification effectiveness, a direct comparison of the proposed ensemble scheme with previous studies is difficult to be performed due to differences in the data sets and differences in evaluation algorithms utilized. Many previous studies have presented very high prediction rates, such as 100% in Ruiz *et al.* (2011); however, such prediction rates were obtained by testing the constructed classification models using internal evaluation approaches, that have been shown to give optimistically biased estimates (Ambrose, *et al.*, 2002). These estimates may be indicative of the model's performances on the training data; however, these estimates are far from being representative of the effectiveness of the model to new, unseen data. In this study, we have attempted to approximate the performance of the proposed model in new, unseen data by using an external cross-validation approach, which enabled us to approximate the generalization prediction rate of the proposed scheme ( $94.9 \pm 1.5\%$ ). If one wanted to select a single optimum number of similar images, we would have to optimize our system based on one of the five performance evaluation criteria that we have utilized (*i.e.*, accuracy, sensitivity, specificity, and diagnostic accuracy or Cohen k). Using the accuracy as the performance evaluation criterion, the external cross-

validation method indicated that the optimum number of similar images is 5, with 94.9% performance with the Majority Vote scheme. Moreover, another significant difference of the proposed study against previous studies is that the proposed ensemble scheme tested images originating from different dermatological atlases with great generalization potential for all criteria tested. Finally, another difference of the proposed study against the previous studies is that the template matching and content-based image retrieval scheme is used not only to retrieve the most similar, to the examined case, images, but also to characterize the nature of the unknown case using a combination of three different algorithms, which, to the best of our knowledge, is for the first time investigated.

In terms of clinical effectiveness, the proposed scheme offers the possibility to both patients and physicians to exploit consultations that will guide them towards more accurate decisions. The patient may use the proposed scheme as a second opinion consultation during the self-skin examination process by photographing with a standard consumer smartphone or other type of digital camera and requesting from the proposed scheme to assess the urgency for a potential visit.

In order to render our database less dependent upon the smartphone camera technology, we used the following approaches: a/ although the size of the mole has a significant importance in diagnosing melanoma, this feature was not used since the size of the mole not only depends on the magnification of the photograph, but also depends on the distance of the camera from the mole. Thus, our database does not rely on either the magnification or the distance of the camera from the mole, b/ we use the mean shift algorithm (Fukunaga, 1975), which is very effective in flattening the image's texture, thus, we can correct for different levels of illumination, c/ we use mainly features of texture in our algorithms. These features are less depended on the technology of the smartphone camera and the viewing angle, than features of size and shape, d/ we use the DullRazor algorithm (Lee, *et al.*, 1997) to eliminate hair pixels and smooth the image, reducing overall noise levels and facilitating the subsequent step of segmentation.

When the proposed scheme identifies that the most similar images are retrieved from the melanoma category, then the consultation will be towards an urgent physician visit. In this way, the probability for early stage detection of melanoma will potentially increase, since the patient may visit the expert physician soon enough. On the other hand, the physician

may also benefit by the proposed scheme by means of:

- a. second opinion consultations regarding the nature of the examined moles,
- b. retrieval of the most similar images from a verified melanoma cases data source and
- c. distance monitoring of patients.

In this way, potential diagnostic misinterpretations might be reduced and the overall patient management might be improved.

This study is part of the MARK1 project. The MARK1 application may capture an image, assign the image to a special dermatologist and give the dermatologist a series of image processing and decision support services in order to conclude regarding the administration of the case. More information may be found at: <http://mark1-project.eu/>.

### ACKNOWLEDGMENTS

Research activities of this work have been carried out within the context of the project “MARK1- A decision support system for the early detection of malignant melanoma” with ref. Number ISR\_3233, under the bilateral cooperation between Greece & Israel action 2013-2015 that has been co-funded by the European union and the General Secretariat for Research & Technology, Ministry of Education, Research and Religious Affairs of the Hellenic Republic.

### REFERENCES

- Abbasi NR, Shaw HM, Rigel DS, Friedman RJ, McCarthy WH, Osman I, et al. (2004). Early diagnosis of cutaneous melanoma: Revisiting the ABCD criteria. *JAMA* 292:2771–6.
- Abikhair MR, Mahar PD, Cachia AR, Kelly JW (2014). Liability in the context of misdiagnosis of melanoma in australia. *MED J Australia* 200:119–21.
- Ambroise C, McLachlan G J (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *PNAS* 99:6562–6.
- Asgarizadeh M, Pourghassem H, Shahgholian G, Robust object tracking using regional mutual information and normalized cross correlation, (2012), Proceedings - 4th International Conference on Computational Intelligence and Communication Networks, CICN 2012, Mathura, Uttar Pradesh, India, 411–5.
- Balch CM, et al. (2001). Long-term results of a prospective surgical trial comparing 2 cm vs. 4 cm excision margins for 740 patients with 1-4 mm melanomas. *Ann Surg Oncol* 8:101–8.
- Ballerini L, Fisher R B, Aldridge B, Rees J (2013). A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. In: Celebi EM, Schaefer G, eds. *Color medical image analysis*, Dordrecht: Springer Netherlands, 63–86.
- Ballerini L, Li X, Fisher RB, Rees J (2010). A query-by-example content-based image retrieval system of non-melanoma skin lesions. In: Caputo B, Müller H, Syeda-Mahmood T, Duncan JS, Wang F, Kalpathy-Cramer J, eds. *Medical content-based retrieval for clinical decision support: First miccai international workshop, mcbr-cds 2009*, london, uk, september 20, 2009, revised selected papers, Berlin, Heidelberg: Springer Berlin Heidelberg, 31–88.
- Carli P, De Giorgi V, Nardini P, Mannone F, Palli D, Giannotti B (2002). Melanoma detection rate and concordance between self-skin examination and clinical evaluation in patients attending a pigmented lesion clinic in italy. *Br J Dermatol* 146:261–6.
- Cavalcanti PG, Scharcanski J, Baranoski GVG (2013). A two-stage approach for discriminating melanocytic skin lesions using standard cameras. *Expert Sys Appl* 40:4054–64.
- Chen RH, Snorrason M, Enger SM, Mostafa E, Ko JM, Aoki V, Bowling J (2016). Validation of a skin-lesion image-matching algorithm based on computer vision technology. *Telemed J E Health* 22:45–50.
- Cohn-Cedermark G, et al. (2000). Long term results of a randomized study by the swedish melanoma study group on 2-cm versus 5-cm resection margins for patients with cutaneous melanoma with a tumor thickness of 0.8-2.0 mm. *Cancer* 89:1495–501.
- El-Naqa I, Yang Y, Galatsanos NP, Nishikawa RM, Wernick MN (2004). A similarity learning approach to content-based image retrieval: Application to digital mammography. *IEEE T Med Imaging* 23:1233–44.
- Field LM (1994). Clinical misdiagnosis of melanoma as well as squamous cell carcinoma masquerading as seborrheic keratosis. *J Dermatol Surg Oncol* 20:222.
- Fukunaga KLDH (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE T Inform Theory* 21:32–40.
- Gaidhane VH, Hote YV, Singh V (2012). An efficient similarity measure technique for medical image registration. *Sadhana - Academy Proceedings in Engineering Sciences* 37:709–21.
- Gonzalez RC, Woods RE (2002). *Digital image processing*, NY: Addison-Wesley Pub, 518–28.
- Grant-Kels JM, Bason ET, Grin CM (1999). The misdiagnosis of malignant melanoma. *J Am Acad Dermatol* 40:539–48.
- Jain AK, Dubes RC (1988). *Algorithms for clustering data*, Prentice-Hall Inc.
- Jaleel JA, Salim S, Aswin RB (2012). Artificial neural network based detection of skin cancer. *IJAREEIE* 1:200–05.
- Kassianos AP, Emery JD, Murchie P, Walter FM (2015). Smartphone applications for melanoma detection by

- community, patient and generalist clinician users: A review. *Br J Dermatol* 172:1507–18.
- Kittler J, Hatef M, Duin RPW, Matas J (1998). On combining classifiers. *IEEE T Pattern Anal* 20:226–39.
- Leachman SA, et al. (2016). Methods of melanoma detection. *Cancer Treat Res* 167:51–105.
- Lee T, Ng V, Gallagher R, Coldman A, McLean D (1997). Dullrazor: A software approach to hair removal from images. *Comput Biol Med* 27:533–43.
- Li CH, Lee CK (1993). Minimum cross entropy thresholding. *Pattern Recog* 26:617–25.
- Lorentzen H F, Weismann K, Grønhøj Larsen F (2001). Structural asymmetry as a dermatoscopic indicator of malignant melanoma - a latent class analysis of sensitivity and classification errors. *Melanoma Res* 11:495–501.
- Loukas C, Kostopoulos S, Tanoglidi A, Glotsos D, Sfikas C, Cavouras D (2013). Breast cancer characterization based on image classification of tissue sections visualized under low magnification. *Comp Math Methods Med* 2013:7 pages.
- Lucas R, McMichael T, Smith W, Armstrong B (2006). Solar ultraviolet radiation: Global burden of disease from solar ultraviolet radiation In: A. Prüss-Üstün, H. Zeeb, C. Mathers, M. Repacholi, eds. *Environmental burden of disease series 13*, Geneva: World Health Organization.
- Maragoudakis M, Maglogiannis I (2011). A medical ontology for intelligent web-based skin lesions image retrieval. *Health Informatics J* 17:140–57.
- Mazurowski MA, Lo JY, Harrawood BP, Tourassi GD (2011). Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis. *J Biomed Inform* 44:815–23.
- Ming ME (2000). The histopathologic misdiagnosis of melanoma: Sources and consequences of "false positives" and "false negatives". *J Am Acad Dermatol* 43:704–6.
- Ninos K, et al. (2013). Computer-based image analysis system designed to differentiate between low-grade and high-grade laryngeal cancer cases. *Anal Quant Cytol* 35:261–72.
- Paddock LE, Lu SE, Bandera EV, Rhoads GG, Fine J, Paine S, et al. (2016). Skin self-examination and long-term melanoma survival. *Melanoma Res*: in press.
- Pfahlberg AB, Gefeller O (2008). Errors in assessing risk factors for melanoma: Lack of reproducibility is the minor problem. *Melanoma Res* 18:300–1.
- Rastrelli M, Tropea S, Rossi CR, Alaibac M (2014). Melanoma: Epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo* 28:1005–11.
- Ringborg U, et al. (1996). Resection margins of 2 versus 5 cm for cutaneous malignant melanoma with a tumor thickness of 0.8 to 2.0 mm: Randomized study by the Swedish melanoma study group. *Cancer* 77:1809–14.
- Robson Y, Blackford S, Roberts D (2012). Caution in melanoma risk analysis with smartphone application technology. *Br J Dermatol* 167:703–4.
- Ruiz D, Berenguer V, Soriano A, Sanchez B (2011). A decision support system for the diagnosis of melanoma: A comparative approach. *Expert Syst Appl* 38:15217–23.
- Russakoff DB, Tomasi C, Rohlfing T, Maurer Jr CR (2004). Image similarity using mutual information of regions. *3023:596–607*.
- Schein O, Westreich M, Shalom A (2009). Effect of dermoscopy on diagnostic accuracy of pigmented skin lesions emphasizing malignant melanoma. *Harefuah* 148:820–3.
- Stoecker WV, Rader RK, Halpern A (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection: Representative lesion sets and the role for adjunctive technologies. *JAMA Dermatol* 149:884.
- Stringa M (1988). Misdiagnosis of choroidal melanoma. *Panminerva Med* 30:89–92.
- Tenenhaus A, Nkengne A, Horn JF, Serruys C, Giron A, Fertil B (2010). Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions. *Skin Res Technol* 16:85–97.
- Theodoridis S, Koutroumbas K (2003). *Pattern recognition*, San Diego:Elsevier.
- Vañó-Galván S, Paoli J, Ríos-Buceta L, Jaén P (2015). Skin self-examination using smartphone photography to improve the early diagnosis of melanoma. *Actas Dermosifiliogr* 106:75–7.
- Veierod MB, Parr CL, Lund E, Hjartaker A (2009). Response: Errors in assessing risk factors for melanoma. *Melanoma Res* 19:61.
- Veronesi U, Cascinelli N (1991). Narrow excision (1-cm margin). A safe procedure for thin cutaneous melanoma. *Arch Surg* 126:438–41.
- Wolf JA, Moreau JF, Akilov O, Patton T, English JC, 3rd, Ho J, Ferris LK (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol* 149:422–6.
- Yan Y, Huang X, Zheng Y, Xu W (2011). An efficient template matching between rotated mono- or multi-sensor images, MIPPR 2011: Parallel Processing of Images and Optimization and Medical Imaging Processing, Guilin, China, 80050M-80050M-9.
- Zagrouba E, Barhoumi W (2004). A preliminary approach for the automated recognition of malignant melanoma. *Image Anal Stereol* 23:121–35.
- Zhang S, Gao F, Wan D (2010). Effect of misdiagnosis on the prognosis of anorectal malignant melanoma. *J Cancer Res Clin Oncol* 136:1401–5.