Artificial Intelligence in Medicine 60 (2014) 13–26

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists

Maciel Zortea[a,*], Thomas R. Schopf[b], Kevin Thon[b], Marc Geilhufe[a], Kristian Hindberg[a], Herbert Kirchesch[c], Kajsa Møllersen[b], Jörn Schulz[a], Stein Olav Skrøvseth[b], Fred Godtliebsen[a]

[a] Department of Mathematics and Statistics, University of Tromsø, 9037 Tromsø, Norway
[b] Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, 9038 Tromsø, Norway
[c] Dermatology Office, Venloer Straße 107, 50259 Pulheim, Germany

ARTICLE INFO

Article history:
Received 1 July 2013
Received in revised form
28 November 2013
Accepted 29 November 2013

Keywords:
Computer-aided diagnosis
Supervised classification
Skin cancer detection
Melanoma
Pigmented skin lesions
Dermoscopy

ABSTRACT

Background: It is often difficult to differentiate early melanomas from benign melanocytic nevi even by expert dermatologists, and the task is even more challenging for primary care physicians untrained in dermatology and dermoscopy. A computer system can provide an objective and quantitative evaluation of skin lesions, reducing subjectivity in the diagnosis.
Objective: Our objective is to make a low-cost computer aided diagnostic tool applicable in primary care based on a consumer grade camera with attached dermatoscope, and compare its performance to that of experienced dermatologists.
Methods and materials: We propose several new image-derived features computed from automatically segmented dermoscopic pictures. These are related to the asymmetry, color, border, geometry, and texture of skin lesions. The diagnostic accuracy of the system is compared with that of three dermatologists.
Results: With a data set of 206 skin lesions, 169 benign and 37 melanomas, the classifier was able to provide competitive sensitivity (86%) and specificity (52%) scores compared with the sensitivity (85%) and specificity (48%) of the most accurate dermatologist using only dermoscopic images.
Conclusion: We show that simple statistical classifiers can be trained to provide a recommendation on whether a pigmented skin lesion requires biopsy to exclude skin cancer with a performance that is comparable to and exceeds that of experienced dermatologists.

1. Introduction

The incidence and mortality rates of melanoma in Caucasian populations have been increasing for many decades [1]. Though only accounting for one tenth of the new cases of skin cancer, melanoma are associated with more than 90% of the skin cancer deaths [2]. If detected at an early stage, the prognosis for the patient is excellent because the patient can be cured by simple excision of the tumor. However, early diagnosis is very challenging as melanomas are easily confused with benign skin lesions.

Dermoscopy is a method that allows doctors to examine structures in the skin that are not visible to the naked eye. When practiced by experts, dermoscopy improves the diagnostic accuracy of pigmented skin lesions (PSL) [3–5]. Several methods have been developed to help clinicians interpret the structures revealed through dermoscopy [6]. Well known algorithms include the ABCD rule of dermoscopy [7], the Menzies method [8], the Three-point checklist [9], the 7-point checklist [10], the CASH algorithm for dermoscopy [11], the Chaos and Clues algorithm [12], the BLINCK algorithm [13], and Pattern Analysis [14]. However, intensive and time consuming training is required to become an expert in dermoscopy. Furthermore, dermoscopy has its limitations, especially in the diagnosis of early melanoma [15]. In the early stages of the disease it may look like a common mole. Often there are no specific dermatoscopic features of melanoma, or the features appear subtle and are easily overlooked.

There have been efforts to develop computer programs to diagnose melanoma based on lesion images. Roughly, these studies follow intuitive steps in a standard pattern recognition processing chain: (a) image segmentation to separate the lesion area from the background skin, (b) extraction of image features for classification purposes, and (c) final classification using statistical methods.

* Corresponding author. Tel.: +47 45683608.
E-mail addresses: mzortea@gmail.com, maciel.zortea@uit.no (M. Zortea).

0933-3657/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artmed.2013.11.006

A wide range of ideas have been utilized in these three steps; see Korotkov and Garcia [16] for an overview and categorizations. Reporting sensitivity and specificity, Rosado et al. [17] presented a thorough overview of state-of-the-art methods at the time, and Dolianitis et al. [18] compared the diagnostic accuracy of four dermoscopy algorithms in the hands of 61 relatively inexpert medical practitioners in Australia. Day and Barbour [19] attempted to reproduce algorithmically the perceptions of dermatologists as to whether a lesion should be excised or not. Comparing results between different systems is difficult because results are very sensitive to the data set used for validation, and a major problem is the lack of publicly available databases of dermoscopic images. For a fair and representative comparison, a data set with a large number of examples of all types of lesions and all types of features expected to be encountered in clinical practice should be made available.

Following this, an interesting research question is:

> Assume identical information is made available to both computers and doctors for the same set of skin lesion images. Then, how does the accuracy of the computer system compare with the accuracy of the doctors?

An answer to the above question would make it easier to objectively assess the performance of any new or existing methods, and would provide an indication of how difficult the lesion images in the data sets used in the experiments were to diagnose. In a data set with a clear distinction between classes, high accuracy would be no surprise. Despite this being a conceptually rather simple experiment to conduct, the study could be demanding because it would require substantial effort by dermatologists to evaluate a large number of lesion images. Also, a more difficult question to answer is whether the data set is sufficiently representative. To be so, it needs to approximate the variability of cases found in a true clinical setting, including the prior information regarding the occurrence of each type of lesion.

Several studies have been reported where the diagnostic accuracy of a computer system is directly compared with human diagnosis. The diagnostic accuracy of the computer systems is generally not significantly different from that of human experts (for an overview, see Rosado et al. [17]). Most studies tend to compare the performance of their system exclusively with histopathological diagnosis, leaving it an open question how difficult the lesions are to diagnose by dermatologists. In the present study, in addition to the histopathological results, we compare the results of the computer system with those of three dermatologists to provide an indication of how challenging our dataset is to either type of analysis.

Korotkov and Garcia [16] recently listed 10 commercial computer-aided diagnosis (CAD) systems for the diagnosis of melanoma based on dermoscopy. As a rule they use powerful and dedicated video cameras. The cost related to the acquisition material and proprietary technologies are likely substantial barriers to the systems gaining widespread popularity among physicians [20]. Perrinaud et al. [21] reported on an independent clinical evaluation of some of these systems, and they found little evidence that such systems benefit dermatologists. Also, current limitations of state-of-the-art CAD systems motivate the development of new algorithms for analysis of skin lesions, and low cost data acquisition tools (e.g., digital cameras and dermatoscopes) are becoming commonly available. Following an approach that should be practical and intuitive to dermatologists, the images considered in this study are acquired by means of a consumer-grade digital camera with a dermatoscope attached. This simple image acquisition setup has been previously discussed, for instance in Gewirtzman and Braun [22], and has been used in the visual comparison system of Baldi et al. [23].

In the following sections we report on our experiments with a simple system we built using off-the-shelf equipment for fully automatic detection of melanomas. Our main contributions include the development of novel image features with the potential to handle morphological structures in dermoscopic images acquired with low-cost, off-the-shelf equipment. Also, we give an indication on how challenging the dataset is by asking three dermatologists to evaluate the same set of images evaluated by the computer system.

The remainder of the paper is organized as follows. Section 2 provides a description of the data, the image segmentation method, feature extraction and selection, and classification. Section 3 describes two experiments and reports on the findings. Section 4 provides a discussion, and conclusions are drawn in Section 5.

2. Materials and methods

2.1. Data

Dermoscopic images of 206 pigmented skin lesions were acquired using a portable dermatoscope (Dermlite Pro II HR, 3Gen LLC, CA, USA) attached to a consumer-grade digital camera (Canon G10, Canon Inc., Tokyo, Japan). Images were acquired at two locations, a private practice clinic in Germany by author H.K., and at the Department of Dermatology at the University Hospital of North Norway, in Tromsø, Norway by author T.R.S. 113 images were obtained consecutively at the private clinic between December 2009 and January 2010 from all patients requiring biopsy or excision of a pigmented skin lesion because of diagnostic uncertainty. In addition, we added 93 images photographed at both sites between December 2009 and December 2010. Of these, 60 images represented benign common lesions not requiring biopsy or excision and 33 images of melanomas. A total number of 206 lesion images were decided on because this number appeared realistic regarding the workloads of the three independent dermatologists participating in the evaluation of this study. All images have the same fixed values for aperture width, ISO value, focus distance, and focal length. The resulting raw images of size 4432×3326 pixels were raw-converted to.png images using a fixed white balance. Due to the presence of the dermatoscope, only a circular area with a diameter of 3326 pixels contains image information, which corresponds to about 14 mm on the skin. The typical dark circular fading pattern at the border of the image area due to the geometry of the dermatoscope is discarded. The resulting image is downsampled to 1650×1650 pixels to ease computations, resulting in a spatial resolution of approximately 3000 ppi, with a color depth of eight bits per channel.

The set of images was printed on high quality 178 mm $\times$ 178 mm paper sheets with a resolution higher than 200 dpi. The printed images were given to three dermatologists familiar with dermoscopy and who were not otherwise involved in the data collection. Two doctors were board certified dermatologists while the third doctor was an experienced resident doctor. They were asked to provide, for each case, (a) the probable diagnostic class (on a visual analog scale indicating benign, suspicious, or malignant), and (b) an indication regarding whether they would recommend excision of the skin lesion. For the purpose of our experiments, we will compare the outcome of the computer method with the doctors' excision recommendations (yes, no). No additional information was provided to the dermatologists (patient's gender, age, etc.). The participants noted their answers on the sheets and had no time constraints. In Table 1 the characteristics of the lesions used in the study are summarized. Notably, the Breslow depth is less than 1 mm in all cases except three, where a Breslow depth of <1 mm indicates early stage melanoma. Pigmented Bowen's disease and