Computer Methods and Programs in Biomedicine 195 (2020) 105631

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Computer-aided classification of suspicious pigmented lesions using wide-field images

Judith S. Birkenfeld[a,b,c,d,2,*], Jason M. Tucker-Schwartz[b,1], Luis R. Soenksen[a,b,f,g,h,1], José A. Avilés-Izquierdo[e], Berta Marti-Fuster[a,b,c,2]

[a] Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
[b] MIT linQ, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
[c] Brigham and Women's Hospital - Harvard Medical School, 75 Francis St, Boston, MA 02115, United States
[d] Massachusetts General Hospital - Harvard Medical School, 55 Fruit St, Boston, MA 02114, United States
[e] Department of Dermatology, Hospital General Universitario Gregorio Marañón, Calle del Dr. Esquerdo 46, 28007 Madrid, Spain
[f] Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
[g] Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Cir, Boston, MA 02115, USA
[h] Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 20 March 2020
Accepted 21 June 2020

Keywords:
Machine learning
Suspicious pigmented lesions
Computer-aided classification
Melanoma
Wide-field images

ABSTRACT

Background and objective: Early identification of melanoma is conducted through whole-body visual examinations to detect suspicious pigmented lesions, a situation that fluctuates in accuracy depending on the experience and time of the examiner. Computer-aided diagnosis tools for skin lesions are typically trained using pre-selected single-lesion images, taken under controlled conditions, which limits their use in wide-field scenes. Here, we propose a computer-aided classifier system with such input conditions to aid in the rapid identification of suspicious pigmented lesions at the primary care level.

Methods: 133 patients with a multitude of skin lesions were recruited for this study. All lesions were examined by a board-certified dermatologist and classified into "suspicious" and "non-suspicious". A new clinical database was acquired and created by taking Wide-Field images of all major body parts with a consumer-grade camera under natural illumination condition and with a consistent source of image variability. 3–8 images were acquired per patient on different sites of the body, and a total of 1759 pigmented lesions were extracted. A machine learning classifier was optimized and build into a computer aided classification system to binary classify each lesion using a suspiciousness score.

Results: In a testing set, our computer-aided classification system achieved a sensitivity of 100% for suspicious pigmented lesions that were later confirmed by dermoscopy examination ("SPL_A") and 83.2% for suspicious pigmented lesions that were not confirmed after examination ("SPL_B"). Sensitivity for non-suspicious lesions was 72.1%, and accuracy was 75.9%. With these results we defined a suspiciousness score that is aligned with common macro-screening (naked eye) practices.

Conclusions: This work demonstrates that wide-field photography combined with computer-aided classification systems can distinguish suspicious from non-suspicious pigmented lesions, and might be effective to assess the severity of a suspicious pigmented lesions. We believe this approach could be useful to support skin screenings at a population-level.

1. Introduction

Cutaneous melanoma is responsible for over 75% of skin cancer deaths [1,2]. In 2020, an estimated 100,350 patients will be diagnosed with melanoma, and 6850 patients are expected to die of melanoma in the U.S. [3]. However, the prognosis is excellent for localized disease and primary tumors which are 1.00 mm or

* Corresponding author at: MIT linQ, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA.
E-mail address: jbirkenf@mit.edu (J.S. Birkenfeld).
[1] Both the authors contributed equally to this work.
[2] These authors share senior authorship equally.

https://doi.org/10.1016/j.cmpb.2020.105631

less in thickness. The 5-year survival rate in these patients is more than 90% [1]. For later tumor stages III and IV where the tumor has spread to nearby and distant tissues, the survival rate drops to 62.6% and 16.1% [4], respectively, with a 20-fold increase in treatment costs [5]. Hence, early detection is key to reducing melanoma mortality and lowering treatment costs.

Pilot screening programs have shown encouraging results [6,7], and a recent literature review of international melanoma experts states the need to develop methods to identify precursor lesions at high risk for progression to melanoma [2]. The screening for these precursor lesions, to which we will refer to as suspicious pigmented lesions (SPLs) is known as macro-screening for outlier lesions [8]. For this purpose, characteristics of a lesion including asymmetry, border irregularities, uneven distribution of colors, lesion diameter, and changes over time (known as ABCDE criteria [9]) are taken into account, in combination with, patient history, the number of nevi, and recognition of the "ugly duckling" lesion (outlier lesion on the patient's mole types) [10]. Although the current approach of patient screening is effective, there are less than 15 dermatology visits per 100 persons annually in the USA [11], leaving a large population without expert opportunistic identification of these lesions. Unlike dermatologists, internists and family practitioners see around 330 million combined patient visits in the USA [11], putting primary care physicians (PCPs) in a prime position for opportunistic identification of melanoma. However, the majority of PCPs are currently not trained for skin examination [12], and some studies assert they have insufficient diagnostic and referral accuracy [13–15,44]. The ability to expand melanoma screening to a broader population could aid in identifying melanoma early and positively impacting patient outcomes.

Computer-aided diagnosis systems have been developed to identify malignant characteristics from images of single lesions, but only few are commercially available [16]. The purpose of such systems is to help dermatologists with individual lesion diagnosis and to assist non-dermatologists with accurate and fast referrals. However, to maximize sensitivity and avoid missing melanomas, the full spectrum of lesions on a patient should be assessed, which would require a significant amount of time when using the approach of validating every single lesion on a patients' body with these single-lesion imaging and analysis systems. Due to the time constraints in primary care practice, this could limit the practical use of computer-based screening in a non-expert setting, at least in a cost effective manner.

It may be possible to acquire more digital data per patient in a rapid manner for computer-based pre-screening systems by using wide-field digital images. Wide-field images can be quickly and easily acquired with a user-friendly commercial camera. Furthermore, these wider field images can include a multitude of skin lesions within one image, since the area imaged may be increased by more than tenfold in comparison to images that are taken with a dermoscope. The manual collection of a number of these images in sequence could allow for greater or even complete coverage of a patient's body for pre-screening in a simple and cost-effective way. However, the image quality is fundamentally different from dermoscopy images, since there is minimal illumination, magnification, or polarized control. Computer-based pre-screening systems that use these types of images could be able to do both, lesion recognition (identifying individual lesions on the image) and lesion analysis to determine which of the identified individual lesions require further investigation (i.e. are suspicious).

In this paper, the goal is to demonstrate that individual lesion analysis is possible for lesion-images that have been taken from wide-field photographs using a computer aided classification (CAC) system. The uniqueness of this approach is that (1) the developed computer aided classification system was optimized to work with lesion-images from wide-field photographs (lower resolution than previous approaches), (2) the said system gives lesions a suspiciousness score, and classifies suspicious and non-suspicious lesions, not the traditional melanoma vs non-melanoma, (3) training of the system explicitly included an important source of image variability due to acquisition considerations of the database, and (4) it allows for low-cost image acquisition instruments like digital photographic cameras and mobile phones.

Within this manuscript, our contributions are as follows:

- We have optimized a machine learning classification algorithm to distinguish suspicious (SPLs) from non-suspicious (non-SPLs) pigmented lesions
- We have acquired images of major body parts of 133 patients in a new clinical database of images, which included 1759 skin lesions
- We have established a suspiciousness score, which is aligned with common macro-screening (naked eye) practice.

2. Methods

Images from 133 patients were acquired with a consumer-grade camera, and a computer-aided classification (CAC) system to distinguish SPLs from non-SPLs was developed. Here we distinguish the term CAC from the more common term computer-aided diagnosis (CAD) as the intent of the present work is not to generate a tool for diagnosis but rather an assistive system for referral. A machine learning classifier, as part of the CAC system, was trained with the acquired database of skin lesions.

2.1. Patient population

All subjects were high risk patients (skin types I to III) recruited at the Department of Dermatology at the Hospital Gregorio Marañón, Madrid, Spain. Inclusion criteria for participants were signed informed consent, age of 18 years or older, and assured mental integrity. Exclusion criteria were marks on the subject's body that would prevent full anonymization (e.g. tattoos). Images were acquired by a melanoma expert with >15 years of experience. The dermatologist did a full body examination using his standard of care tools including dermoscopy. Lesions were classified into three different classes: 1) SPL, class A (SPL_A), if suspicious clinical features were detected by naked eye examination and confirmed by dermoscopy examination, 2) SPL, class B (SPL_B) if suspicious clinical features were found by using naked eye examination, but not confirmed after examination with a dermoscope, 3) (all other) non-suspicious lesions (non-SPL). The manual score for the lesions were 1 for SPL_A, 2 for SPL_B, and 0 for non-SPL. SPL_A lesions were tracked over time or biopsied by the melanoma expert as part of his standard of care. SPL_B lesions were recommended to be tracked over time by a non-expert dermatologist.

2.2. Ethics approval

This study was reviewed and approved by the Ethical Committee from Hospital General Universitario Gregorio Marañón (Madrid, Spain) and COUHES (Committee on the Use of Humans as Experimental Subjects) from Massachusetts Institute of Technology (Cambridge, MA, USA).

2.3. Image acquisition

All images were acquired with an Olympus E-420 (10 Mega Pixels) camera, in RAW format, at a distance of 0.2 m from the patient and anonymized before processing. A sample image is shown in Fig 1A. A card with an integrated ruler was held within the image as a reference point for measuring image pixel size and lesion size. Illumination was not controlled during image acquisition.

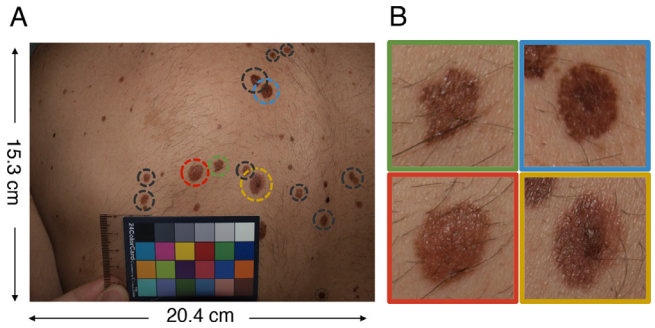

Fig. 1. A: Wide-field image (20.4 × 15.3 cm - 3648 × 2736 pixels) of a patient's back acquired from approximately 0.2 m distance. All lesions encircled (all colors) were analyzed. B: Enlarged view of example lesions from A. Enlarged lesions in B.

Potential illumination artifacts were corrected during image processing (see section "Image Pre-Processing"). At the predetermined distance, an image acquired in this study has the pixel size of 67 µm (measured using a positive USAF 1951 Resolution Test Target, Thorlabs). This resolution compares to eye resolution at the same distance at reading/viewing distance of (~58–72 µm 0.2–0.25 m), which is also the approximate distance in which lesions are examined by the expert during naked-eye examination.

2.4. Image pre-processing

Images were pre-processed by using a semi-automated image processing scheme. RAW images were first converted to TIFF format with camera-specific software (OLYMPUS Viewer 3). Single lesions with a diameter >3 mm were manually cropped from the original images and saved individually (Fig 1B). The cropping process can be automated, but for the purpose of this study, which aims to investigate practicality of wide-field images for single-lesion analysis, the automatization was not a priority. Special care was taken that lesions which appeared on more than one wide-field image were only added in the single-lesion data base once (see also supplement, part 3). The diameter threshold selected in this study for manually cropping was supported by the clinical evidence that most melanomas (> 94%) are bigger than 3 mm [17]. Within the image resolution previously described and assuming a circular lesion, a 3 mm diameter lesion in our study contained around 1575 pixels. All images (i.e., each mole, Fig 1B) were processed by using a set of automated image pre-processing algorithms including illumination correction (Otsu's method) and lesion border segmentation (Fig 2). Otsu's method was used for automatic thresholding a grayscale version of the lesion image into foreground (lesion) and background (skin). Once a mask was found, a morphological opening operation with a disk-shaped structuring element spanning one fifth of the total lesion image size was applied to remove foreground objects. A Gaussian blur was applied to this processed imaged to reduce high frequency components of the background while maintaining the non-uniform illumination information. The lightness value in HSV space of the opened image after Gaussian blurring was subtracted pixel by pixel from the original color image, correcting for non-uniform illumination.

A shading attenuation method previously described by Cavalcanti et al. [18] was used to correct inhomogeneous illumination artifacts. A previously described automatic lesion segmentation algorithm by Celebi et al. was used to segment the lesion border of each pigmented lesion (PL) [19] and separate the lesion from surrounding skin. Because the automatic segmentation algorithm included a blob-detection approach, only the centered blob was considered as the lesion.

2.5. Image feature extraction

After pre-processing, 399 features that measure specific components of the ABCD criteria [9] of suspicious lesion identification (not including E – Evolution) were extracted from each pre-processed lesion image and used to classify the lesion as suspicious or non-suspicious (see also supplement, section 2). Extracted features were mostly variations of features previously used to validate a computer-aided melanoma diagnostic method for dermoscopy images [20,21,25]. The 399 features are distributed between six categories, namely, lesion asymmetry ($n = 91$), border ($n = 192$), color ($n = 48$), texture ($n = 66$), pixel size ($n = 1$), and lesion area ($n = 1$). Subsequently we briefly describe how the features of the first four categories (i.e. the ABCD features) were obtained.

2.5.1. Asymmetry

A total of 91 features were calculated to extract characteristics describing asymmetry of the lesion. Ten intensity (V channel) thresholds (5 to 230, in steps of 25) were applied to the lesion image [21] to generate a set of 10 new image regions. Lesion pixels whose intensities were higher than the defined threshold were set to zero. Non-lesion pixels (skin) were also set to zero. For each of the 10 new image regions, the features calculated were: area ratio to the original lesion area, circularity, center of gravity differences between new and original lesion areas, standard deviation of the distribution and skewness of the distribution. Furthermore, an asymmetry index [20] was calculated in both directions, major and minor axes, for each of these 10 regions.

This index was defined as the difference between the areas of two halves of the lesion. The circularity of the original lesion was added as an additional feature.

2.5.2. Border

A total of 192 features were obtained to describe the border of each pigmented lesion. As proposed by Iyatomi et al. [21], each lesion area was divided into eight equi-angle regions. A window was centered on the lesion border to analyze pixels inside and outside of the lesion border. The features calculated for each lesion were the average and standard deviation of the color intensity ratio and gradient over the equi-angle regions for various window sizes. These features were calculated in each of the red, blue and green (RBG) channels, and also luminance and chrominance (Cb and Cr) channels.

2.5.3. Color

A total of 48 color-based features were calculated. For this, a peripheral region of the lesion was defined as the region inside the border that has an area equal to 30% of the tumor area [21]. The average, standard deviation and skewness in both RGB and HSV (Hue, Saturation and Value) color spaces were calculated for the following: complete lesion area, peripheral region of lesion, and (complete) area surrounding the lesion. The difference of these features was taken between the pigmented lesion and surrounding skin, as well as the peripheral region and surrounding skin. In addition, the number of unique colors in the peripheral and lesion regions in both RGB and HSV color spaces and the number of unique colors after quantizing the values in each image channel to 8 and 16 levels were included.

2.5.4. Differential texture

In order to quantify lesion textures, 66 features were extracted. Co-occurrence matrices were calculated for each image of the pigmented lesion [21]. Co-occurrence matrix distances were set to 11 different values between 1/2 and 1/64 the length of the major axis of the lesion. Energy, entropy and correlation for each of the 11 co-occurrence matrices were calculated. The final features

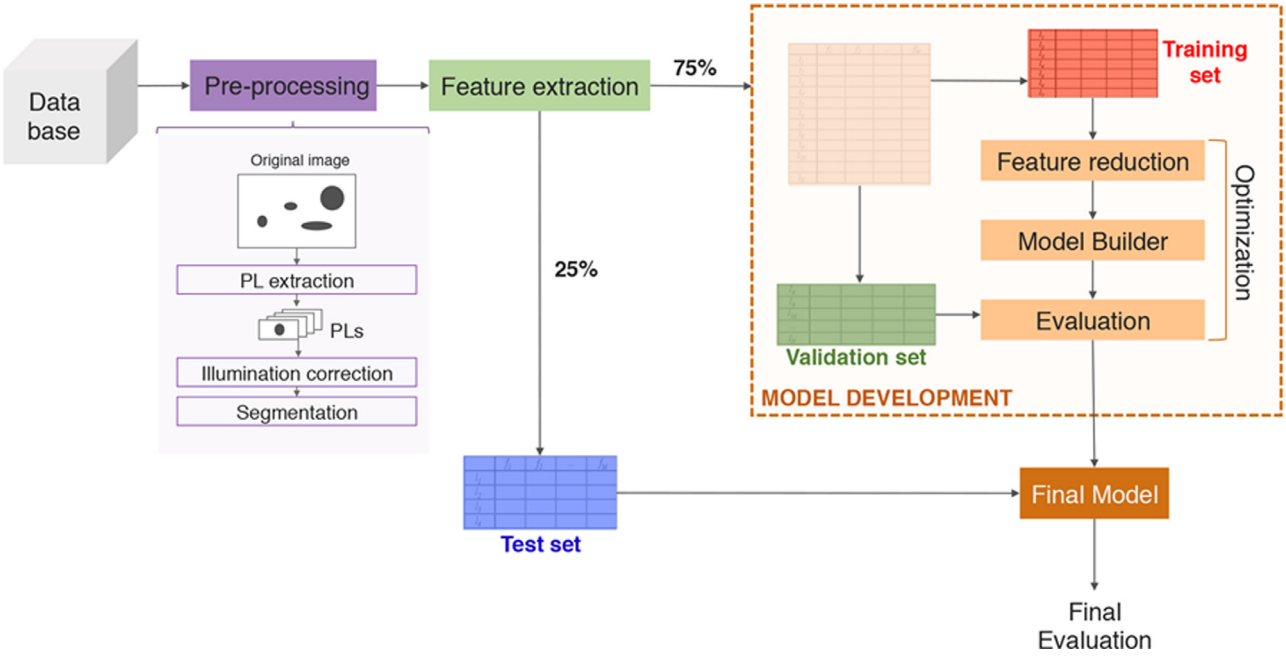

Fig. 2. Processing workflow to build and test an image-processing model to distinguish between SPLs and non-SPLs. Image pre-processing includes extraction of individual lesions from the original images, correction of illumination artifacts and image segmentation to extract the lesion's contour. Image pre-processing is followed by feature extraction. Once all information is extracted, data is divided into two groups: training (75%) and test (25%). Model development (orange dashed box) is a standard procedure for validating machine learning algorithms [39]. For an objective evaluation of the algorithm, the final model is applied to the separate test set.

used for the model were the average and standard deviation of energy, entropy, and correlation over four directions (0°, 45°, 90°, 135°). Hereby, the major axis length and orientation were calculated for each lesion, and the image was rotated so that the major axis would be at 0°. From this as a starting point 45°, 90° and 135° directions were generated and textural features extracted.

2.6. Classifier training and validation

To differentiate suspicious (SPL_A and SPL_B) from non-suspicious lesions, a machine learning classifier was built as part of the CAC System, using Logistic Regression on ABCD extracted features to create a simple, but effective classifier. The 1759 lesions were split into a training (n = 1187) and a test set (n = 572) (see supplement, section 4, for a comparison of splitting the data by lesion and by patient ID). The training set consisted of 46 SPL_As, and 1141 non-SPLs (approx. 75% of SPL_As and non-SPLs). SPL_Bs were not included in the training set. The reasoning behind this approach is based on the results of previous experiments that have shown that the classifier achieved better performance if clear examples of both lesion types were used (see supplement, section 1). The test set consisted of 181 suspicious lesions, (9 SPL_As, 172 SPL_Bs) and 391 non-SPLs (approx. 25% of SPL_As and non-SPLs and 100% of SPL_Bs).

We combined two methods to select features with high impact and reduce the dimensionality of our data, namely 1) univariate feature selection (ANOVA), that removes all but the k highest scoring features, and 2) principal component analysis (PCA). Due to the unbalanced nature of our data (46 vs. 1141 training samples) we used a learning method that internally had the choice to correct that unbalanced scenario. Namely, we used a logistic regression method implemented in the open source Scikit-learn Python package [22] which has a parameter that allows balancing the data (L_2 penalized logistic regression with class balancing).

Prior to input into the model, we standardized all feature data by scaling each feature distribution to have zero mean and unit variance. Moreover, features with no variance (i.e., variance $<$ 10–5) were not used to generate the model.

For the hyper-parameter optimization, we applied a random search method [23] with 105 iterations. The optimized parameters were the number of top features selected in the k-best reduction method (k), the number of PCA components (n_components), and the inverse of regularization strength (C) for the logistic regression algorithm. The ranges explored were k $\in$ [1, 399], n_components $\in$ [0, 399], and C $\in$ [10^{-4},10^{4}]. A cross-validation method of 10 folds was used and the area under the curve (AUC) was the maximized score during the optimization process. The classifier output for each lesion was set to be a suspiciousness score value between 1 (for all SPLs) and 0 (for non-SPLs).

After optimizing our classifier with a 10-fold cross validation, receiver operating characteristic (ROC) curves were generated by using a leave-one-out cross validation strategy. Thus, for every single lesion in the training set (1187 images, 46 SPL_A), the system was trained with all images but one. The trained classifier was then used to get the suspiciousness score of the lesion which was excluded during training (therefore, the ROC curve which is shown is a cross-validation ROC curve).

2.7. Classifier testing

The performance of the classifier was evaluated by using the test set images. True positive and false positive rates of the classifier in the test set were calculated by using the classifier score threshold that achieved 95% true positive in the training-validation process. Scikit-learn Python package was used to optimize and test our model [22]. All algorithms were run on a Mac Pro (2 $\times$ 2.4 GHz 6-Core Intel Xeon) station.

3. Results

133 patients (61 female, 72 male), ages 19 years to 76 years old (median age 43) participated in the study. 3–8 images were acquired per patient on different sites of the body: back, torso,

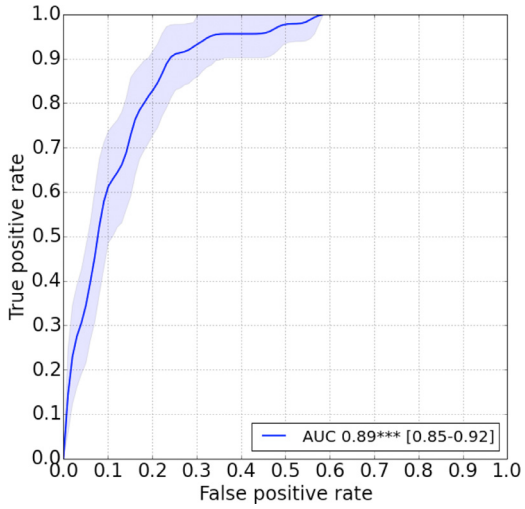

Fig. 3. Validation ROC curve. The ROC curve demonstrates the ability of the algorithm to distinguish SPLs from non-SPLs when compared to an expert dermatologist. The area under the curve (AUC) was 0.89 with a 95% confidence interval of 0.85–0.92 (light blue shading area). The cross-validation (leave-one-out) used 1141 non-SPLs and 46 SPLAs.

Table 1
Overview over cac system performance for test set when compared to the gold standard.

Lesion	Total	TP	Sensitivity
SPL	181	152	84.0%
SPL_A	9	9	100%
SPL_B	172	143	83.2%
Lesion	Total	TN	Specificity
non-SPL	391	282	72.1%

TP: True Positive; TN: True Negative; SPL: Suspicious Pigmented Lesion.

Table 2
Overall confusion matrix.

$n = 572$	Actual: YES	Actual: NO	
Predicted: YES	TP = 152 $TP_{SPL\_A = 9}$ $TP_{SPL\_B = 143}$	FP = 109	261
Predicted: NO	FN = 29 $FN_{SPL\_A = 0}$ $FN_{SPL\_B = 29}$	TN = 282	311
	181	391	572

TP: True Positive; TN: True Negative; SPL: Suspicious Pigmented Lesion; FN: False Negatives; FP: False Positives.

legs, arms and other. A total of 1759 pigmented lesions were extracted from the images (back: 1582, torso: 192, legs: 20, arms: 38, other: 4), with an average of 13 lesions per patient. The database included patients with both, family and personal history of melanoma (11/133 and 25/133, respectively, no overlap). Of the total number of 1759 pigmented lesions, 55 were marked as suspicious, class A by the expert dermatologist. Nine of the 55 lesions were biopsied. After histopathology the diagnosis of the biopsied lesions were: 3 melanomas, 2 basal cell carcinomas, 2 dysplastic nevi and 2 common nevi. The remaining 46 SPL_A were indicated in the patient file for regular tracking by the melanoma expert. 172 lesions were marked as suspicious, class B.

The optimization process previously described (see Methods section) showed that the combination of feature reduction parameters that achieved higher performance was using 4 of the color-related features ($k = 4$) and the first 114 PCA components (n_components = 114). The four selected features with the highest k-score were the number of colors present in the lesion and the peripheral region in the HSV color space, after quantizing the values in each image to 8 and 16 levels. The first 114 PCA components explained the 98% of the variance of our data. The optimization process found that the optimal value for the inverse regularization strength for the logistic regression was $C = 0.003$ (see also supplement, section 2: Feature Reduction 2).

Using the investigated features extracted from each image, a single value was regressed to perform a binary classification. The validation ROC curve presented in Fig. 3, shows the expected sensitivity and specificity of our model as the classification threshold is varied. It was generated by using a leave-one-out cross-validation approach with the training set (1141 non-SPLs and 46 SPL_As) to demonstrate the performance of the optimized classifier in differentiation between suspicious and non-suspicious lesions. The false positive and true positive rate was determined by comparing the classification output of the algorithm to our gold standard (as set by the melanoma expert). From 46 SPL_As, 44 were correctly classified in the training-validation set. The area under the curve was 0.89 with a 95% confidence interval of 0.85–0.92. A threshold, $th_{score} = 0.46$, was found to achieve the lowest false positive rate while still achieving our clinical design criteria of 95% true positive rate (TPR). Achieving such TPR is an essential aspect of our value proposition, and is primary reason why we selected $th_{score} = 0.46$ for our feature-based regression classifier.

The threshold identified in the above described cross-validation was used to predict the suspiciousness of the lesions in the test set by making a binary decision: All lesions with a value lower than the threshold were considered non-suspicious and those above it were considered as suspicious. Table 1 gives an overview over the CAC system performance for the test set when compared to the gold standard. Out of the 181 suspicious lesions in the test set, all suspicious lesions, class A (9) and 83.2% of suspicious lesions, class B (143) were identified correctly. From 391 non-SPLs in the test set, 282 were correctly classified as non-suspicious. These findings result in 75.9% accuracy, defining accuracy as the ratio between the sum of all true positives and true negatives, divided by the total number of lesions in the testing set. In Table 2, we present the confusion matrix of the proposed system to give an overview of performance analysis.

Fig 4 (left) shows the classifier output suspiciousness scores for all lesions in the test set categorized in three groups: SPL_A, SPL_B, and non-SPL. We found that there is a statistically significant difference ($p < 0.001$) between non-SPL suspiciousness scores and SPL suspiciousness scores, and between the suspiciousness score of SPL_A and SPL_B. Fig 4 (right) shows an image example for each pigmented lesion category with its corresponding classifier score (bottom-right) and its corresponding calculated area in mm2 (top-right). The median suspiciousness score was highest for the SPL_A than for the SPL_B, in agreement to the dermatologist opinion.

4. Discussion

Traditionally, the focus of computer aided systems in dermatology has been directed towards (non-invasive) diagnosis, i.e. the ability to classify selected lesions into pathology proven malignant vs. pathology proven benign. These approaches have been primarily developed to aid non-melanoma experts in reducing the number of unnecessary biopsies. However, before selecting any lesion for diagnosis and possible biopsy, it is clinical standard to perform a macro-screening on a patient's body. Indeed, macro-screening is the first step in melanoma care, and the focus is directed towards finding any suspicious lesions on the patient's body which need

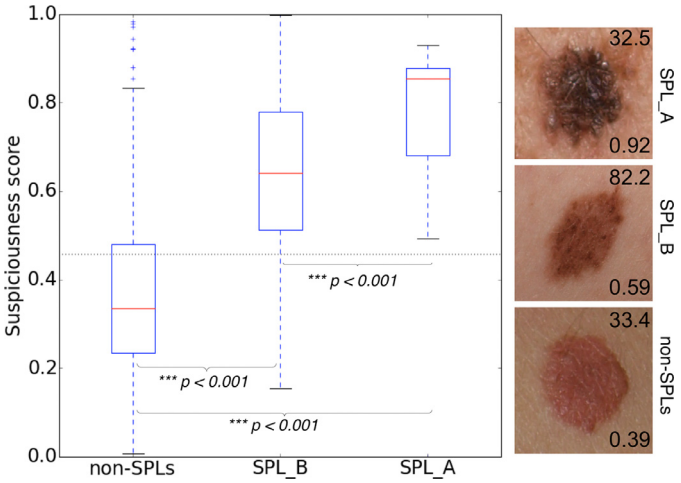

Fig. 4. Median suspiciousness score for three PL types (non-SPLs, SPL_B and SPL_A) in the test set. The boxed area is 25th to 75th percentile; the center line is median value. Whiskers indicate first and fourth quartiles. The horizontal dotted line indicates the selected threshold of the classifier after optimization (thscore= 0.46). ***There is a statistically significant difference between all groups ($p<0.001$). B: Example of lesions classified as SPL_A (top), SPL_B (middle) or non-SPLs (bottom) by the expert opinion. Numbers on the top right indicate the area of the lesion in mm2. Numbers on bottom right corner indicate scores obtained by the classifier.

further diagnosis. Untrained observers easily miss a suspicious lesion when macro-screening [13-15], therefore, putting the patient at risk for an undetected melanoma.

In this study we were interested in developing computational solutions for effective macro-screening for suspicious skin lesions to aid with lesion-selection and effective referral to a specialist. We have conducted a clinical study with high risk patients (skin type I to III) in which we 1) acquired images of all major body parts of the patients with a commercial camera for wide-field images of the body, and 2) classified all lesions into suspicious and non-suspicious, and 3) developed and optimized a computer aided classification system to work with lesion images that were taken from the wide-field images.

Most previously reported dermatological databases for pigmented lesions include single-lesion images that had been directly and carefully acquired in close proximity to patient skin. Hereby, each image within the database is usually from a different patient. It is unlikely that the resolution and quality of such images will be representative of those acquired in a situation of melanoma screening at the primary care level, where devices and imaging conditions are prone to vary. Our intent in acquiring wide-field images and then cropping individual lesions for this study was that for the first time this important source of image variability was explicitly included in the acquisition considerations of our database while still preserving interpretable image classification constraints used in previously described suspicious pigmented lesion CAD systems which perform their analysis lesion by lesion. Our results suggest that our classifier can distinguish between suspicious and non-suspicious skin lesions using images with variabilities in lightning, pixel size, and position within the image. Furthermore, our results indicate that the severity of a suspicious lesion could be distinguished by using our approach. To our knowledge, this is the first time these types of images, which represent image data from real-life scenarios in which devices, imaging conditions, and experience of handling are prone to vary, has been used for skin lesion classification.

For the classifier, we have aimed to train for two general classes, namely nSPL and SPL to improve the training dataset size. We have shown accuracy for the full test set, as well as for two distinct bins of suspiciousness, which might be able to give a better metric of accuracy in different presentations of SPL (see supplement, section 2). However, one limitation of our study was that examination with SPL_A as a class alone would not give representative results due to the low number of images in this class.

The advantage of our approach is the ability to image and assess large skin areas with multiple lesions using images that can be quickly acquired with an easy-to-use standard consumer grade camera. In the future, this could lead to a simple review and flagging of suspicious lesions on a patient.

Computer-aided systems have been previously proposed by several authors to non-invasively diagnose melanoma [24] in both dermoscopy [21,25–28] and non-dermoscopy (standard consumer grade camera) [18,29–31] images of individual skin lesions. For dermoscopy images, sensitivity and specificity of published computer-aided methods are between 85.9–97.4% and 44.2–97.87% [21,25–28], respectively whereas images taken with standard consumer-grade cameras may achieve a sensitivity and specificity between 77 and 100% and 68–97.78% [18,29–31], respectively.

More recently, deep learning approaches have also been explored on non-invasive diagnosis of skin lesions [32,33,38,40–43] and have shown their capabilities on classifying skin lesions with high accuracy (see Table 3). One of the advantages of using deep learning is to potentially remove the pre-processing steps required in traditional approaches (e.g. avoid lesion segmentation). It is important to mention that all of the above-mentioned approaches are based on single lesion analysis, and the algorithms described in the literature attempt to classify lesions into malignant and benign. That makes a direct comparison to our approach difficult, especially since histopathology can only be known for some of the lesions used in this manuscript (it would be impossible to have histopathology on all of a subject's lesions because this would have meant removing approx. 12 lesions per patient). Table 3 provides an overview with specifications of above-mentioned methods, in an attempt to compare published work to our approach. Keeping in mind that our aim is computer-aided macro-screening (as opposed to single lesion analysis), Table III shows that the performance of our classifier is well within the range of the literature. It also suggests that our system could be used as a first and/or complementary step on the way to better lesion management by identifying the suspicious lesions on the patient's body for further diagnosis. It could also be implemented within already existing tracking methods, as shown for example in [34]. While our results for detecting true positives are promising, for practical macro-screening applications, improved classification of true negatives would reduce unnecessary referrals and further burden on the dermatologists and the healthcare system. There are several ways to improve our current results for a potential use in macro screening. One possible and simple way would be to add individual patient information from medical records such as patient history (e.g. family/personal history of melanoma), patient background (e.g. sex, age), and other relevant information (e.g. evolution, location of lesion) as features. This type of information could be scored for suspiciousness and used as additional input data for the classifier. Indeed, previous publications have demonstrated the merit of adding patient information [30]. In the case of this study, not all information was available to us, therefore, it was not implemented. Another approach which has a less straight forward solution is the implementation of the ugly duckling sign within the classifier. In fact, this study was deliberately designed so that the images of the patients show multiple pigmented lesions on a large area of the patient's body for future testing of the feasibility of outlier detection methods.

Further advancement of our system would require an automated lesion detection method to identify individual pigmented lesions in wide-field images for all skin types. For this study, all lesions of high risk patients were manually cropped to assess the

Table 3
Comparison of proposed approach with selected methods from literature.

Total no of images	Extracted Standard Clinical Features	Classification	Sensitivity	Specificity	Accuracy	Ref
Dermoscopy Images (malignant vs benign)						
1258	Color, Symmetry, Border, texture	ANN	85.9%	86%	–	21
564	Shape, Color, texture	SVM	93.33%	92.34%	–	25
	Asymmetry, Border, Color, Diameter	Bayes	76.47%	85.11%	80.61%[a]	26
206	Asymmetry, Color, Border, Geometry, Texture	QDA	86%	52%	63.3%[b]	27
		LDA	80.5%	62.3%	63.3% [b]	
		CART	63%	63%	68.3% [b]	
173	Geometry, Border gradient, Color, Texture	Decision forest (1000 decision trees)	97.4%	44.2%		28
100	NA	CNN	74.1%[c] 87.5%	86.5% 60.0%[c]	81.33%	40
1276 1300	NA	CNN	86.4 ± 3.5%[d] 85.1 ± 2.2%[d]	85.5 ± 3. 2[d]		41
379	NA	CNN	78.66%	81.3 ± 2.9[d] 79.74%		43
Non-dermoscopy Images (malignant vs benign)						
152	Asymmetry, Border irregularity, Color variation, Differential structures	KNN	96.26%	97.78%	96.71%	18
		KNN-DT	96.26%	97.78%	96.71%	
152	Asymmetry, Border irregularity, Color variation, Differential structures & melanin variation feature	KNN	100%[e]	97%[e]	99%[e]	29
152	Asymmetry, Border Color, Texture	CFS w/ LMT	94.39%	68.89%	86.84%	30
		CFS w/ Adaboost & LMT	89.72%	75.56%	85.53%	
769	Shape, Texture, Color	SVM	85.63%	87.65%	90.64%	31
1942[f]	NA	CNN			72.1 ± 0.9%	33
100	NA	CNN[g]	89.4%[h]	68.2% (mean)		42
Wide-field non-dermoscopy Images (suspicious vs non suspicious)						
227	Asymmetry, Border, Color, Differential texture	Logistic Regression	SPL: 84.0% SPL_A: 100% SPL_B: 83.2%	Non-SPL: 72.1%	75.9%	Our results

[a] classification rate.
[b] correct rate score.
[c] adapted to mean sensitivity and specificity value of 157 dermatologists for a comparison of a head-to-head classification task.
[d] Average of classification of 12 common skin diseases, using Asan and Edinburgh dataset (upper and lower row, respectively).
[e] best results from discrimination results in 50 trials.
[f] clinical (mainly) and dermoscopy images.
[g] trained exclusively with dermoscopic images.
[h] adapted to mean sensitivity achieved by 154 dermatologists the results were compared to.SVM Support Vector Machines; ANN ArtificialNeuralNetwork; KNN K-nearest neighbors; KNN-DT KNN followed by a Decision Tree classifier; CFS w/ LMT Correlation-Based Feature Selector with Logistic Model Tree (LMT); QDA Quadratic discriminant Analysis; LDA Linear Discriminant Analysis; CART classification trees; CNN Convolutional Neural Network.

utility of our type of images in the lesion classification component. To better understand future improvements regarding the system's specificity, we show two examples of lesions that were misclassified due to the limitations of Otsu's segmentation method (Fig. 5). The non-SPL in Fig 5A is a congenital melanocytic nevus. It was classified as suspicious due to the grown excess terminal hair which resulted in a segmentation error. Fig 5B shows a misclassified halo nevus (classified as non-SPL by our classifier, but as SPL by the dermatologist).

These types of lesions are often classified as suspicious by clinicians and should be tracked, since the typical depigmented ring of the nevus can be a signal for an immune-mediated process resulting from damage or destruction of melanocytes [35,36]. In this case, the segmentation algorithm failed to recognize the depigmented halo, and classified only the pigmented lesion as a non-SPL. Possible ways to improve the segmentation process could include more non-lesion information from the surrounding area (i.e. cropping the lesion with larger field of view).

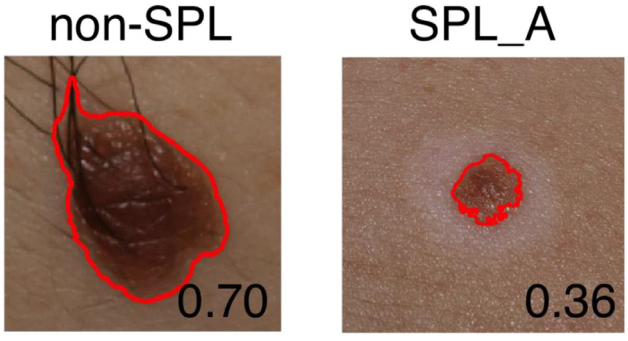

Fig. 5. Example of lesions misclassified by the algorithm in the training set. The numbers on bottom right corner indicate scores obtained by the classifier (suspicious >= 0.46).

It has been previously demonstrated that PCPs often misdiagnose skin lesions as malignant. One study found that 44.5% of in-

vestigated skin lesions that where reason for a referral actually appeared to be benign [37]. Hence, patients are often referred unnecessarily to the already overcrowded dermatology practice, with literature suggesting that 44.5–52.1% of referrals from PCPs to be unnecessary [37,44]. This leads not only to unnecessary biopsies but also patient anxiety. On the other hand, an alarming number of additional SPLs are not recognized by the PCPs. From a population consisting of 734 patients, 234 additional lesions were found on different body parts of 146 patients, of which 123 were premalignant and 111 malignant [37]. Another study [44] calculated the overall diagnostic accuracy of PCPs for skin cancer and precancer to 34.2% (14.8%for skin cancer, and 45.5% for precancer)In the attempt to lower the current high referral rate and increase the number of patients identified with early-stage melanoma, wide-field photography might be feasible for quick screening of patients for suspicious pigmented lesions, thereby, assisting the non-expert in correct SPL identification and referral decision, and leaving recommendation for biopsy and final diagnosis in the hands of the dermatologist.

Declaration of Competing Interest

None.

Acknowledgment

All authors thank the *M*+Visión IDEA3 faculty panel for their guidance and advice in developing this project. We thank Asif Naseem, PhD, for his involvement during the initial phase of this project in the M+Visión fellowship. Additionally, we would like to thank Raul San Jose Estepar, PhD, and Luca Giancardo, PhD for their assistance during model development, and our collaborators at Universidad Rey Juan Carlos, Norberto Malpica, PhD and Elena Viaña for their helpful contributions in data analysis. We especially thank all participating patients. The patient study was reviewed and approved by the Ethical Committee from Hospital General Universitario Gregorio Marañón (Madrid, Spain) and COUHES (Committee on the Use of Humans asv Experimental Subjects) from Massachusetts Institute of Technology (Cambridge, MA, USA). The authors have no conflict of interest to declare.

This work was supported through the EU FP7-PEOPLE-2011-COFUND Program within the M+Visión Project of Fundación Madri+d (Comunidad de Madrid) (JSB, JTS, LRS, BMF), and the DOE training grant DE-SC0008430 (JSB, BMF).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2020.105631.

References

[1] American Cancer Society, Cancer facts & figures, The Society (2008).
[2] G. Merlino, The state of melanoma: challenges and opportunities, Pigment Cell Melanoma Res. 29 (4) (2016) 404–416.
[3] R.L. Siegel, K.D. Miller, and A. Jemal., "Cancer statistics, 2020," CA Cancer J. Clin. vol. 70, no 1, 7–30, 2020
[4] N. Howlader, Surveillance, Epidemiology, and End Results Program, Bethesda: National Cancer Institute, 2011.
[5] G.P. Guy, S.R. Machlin, D.U. Ekwueme, K.R. Yabroff, Prevalence and Costs of Skin Cancer Treatment in the US, 2002– 2006 and 2007– 2011, Am. J. Prev. Med. 48 (2) (2015) 183–187.
[6] A. Waldmann, Skin cancer screening participation and impact on melanoma incidence in Germany–an observational study on incidence trends in regions with and without population-based screening, Br. J. Cancer 106 (5) (2012) 970–974.
[7] C. Curiel-Lewandrowski, S.C. Chen, S.M. Swetter, Melanoma prevention working group-pigmented skin lesion sub-committee. Screening and prevention measures for melanoma: is there a survival advantage? Curr. Oncol. Rep. 14 (5) (2012) 458–467.
[8] A. Scope, A. Marghoob, The “ugly duckling” sign: an early melanoma recognition tool for clinicians and the public, Melanoma Lett. 25 (2007) 1–2.
[9] N.R. Abbasi, Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria, JAMA 292 (22) (2004) 2771–2776.
[10] J. Grob, J. Bonerandi, The'ugly duckling'sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening, Arch. Dermatol. 134 (1) (1998) 103.
[11] C.-J. Hsiao, D.K. Cherry, P.C. Beatty, E.A. Rechtsteiner, National ambulatory medical care survey: 2007 summary, Natl. Health Stat. Rep. 27 (2010) 1–32.
[12] E. Wise, Rates of skin cancer screening and prevention counseling by US medical residents, Arch. Dermatol. 145 (10) (2009) 1131–1136.
[13] P.H. Youl, P.D. Baade, M. Janda, C.B. Del Mar, D.C. Whiteman, J.F. Aitken, Diagnosing skin cancer in primary care: how do mainstream general practitioners compare with primary care skin cancer clinic doctors? Med. J. Aust. 187 (4) (2007) 215.
[14] G. Argenziano, Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer, J. Clin. Oncol. 24 (12) (2006) 1877–1882.
[15] S. Menzies, Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial, Br. J. Dermatol. 161 (6) (2009) 1270–1277.
[16] A.G. Goodson, D. Grossman, Strategies for early melanoma detection: Approaches to the patient with nevi, J. Am. Acad. Dermatol. 60 (5) (2009) 719–735.
[17] N.R. Abbasi, Utility of lesion diameter in the clinical diagnosis of cutaneous melanoma, Arch. Dermatol. 144 (4) (2008) 469–474.
[18] P.G. Cavalcanti, J. Scharcanski, Automated prescreening of pigmented skin lesions using standard cameras, Comput. Med. Imaging Graph. 35 (6) (2011) 481–491.
[19] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, G. Schaefer, Lesion border detection in dermoscopy images using ensembles of thresholding methods, Skin Res. Technol. 19 (1) (2013) e252–e258.
[20] R. Garnavi, M. Aldeen, J. Bailey, Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis, IEEE Trans. Inf. Technol. Biomed. 16 (6) (2012) 1239–1252.
[21] H. Iyatomi, An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm, Comput. Med. Imaging Graph. 32 (7) (2008) 566–579.
[22] F. Pedregosa, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830 Oct.
[23] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (1) (2012) 281–305.
[24] M.E. Celebi, N. Codella, A. Halpern, Dermoscopy image analysis: overview and future directions, IEEE J. Biomed. Health Inform. 23 (2) (2019) 474–478.
[25] M.E. Celebi, A methodological approach to the classification of dermoscopy images, Comput. Med. Imaging Graph. 31 (6) (2007) 362–373.
[26] D. Ruiz, V. Berenguer, A. Soriano, B. SáNchez, A decision support system for the diagnosis of melanoma: a comparative approach, Expert Syst. Appl. 38 (12) (2011) 15217–15223.
[27] M. Zortea, Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists, Artif. Intell. Med. 60 (1) (2014) 13–26.
[28] L.K. Ferris, Computer-aided classification of melanocytic lesions using dermoscopic images, J. Am. Acad. Dermatol. 73 (5) (2015) 769–776.
[29] P.G. Cavalcanti, J. Scharcanski, G.V. Baranoski, A two-stage approach for discriminating melanocytic skin lesions using standard cameras, Expert Syst. Appl. 40 (10) (2013) 4054–4064.
[30] J.F. Alcón, Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis, Sel. Top. Signal Process. IEEE J. 3 (1) (2009) 14–25.
[31] W.-Y. Chang, Computer-aided diagnosis of skin lesions using conventional digital photography: a reliability and feasibility study, PLoS One 8 (11) (2013) e76212.
[32] L. Yu, H. Chen, Q. Dou, J. Qin, P.A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2016) 994–1004.
[33] A. Esteva, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[34] H. Mirzaalian, T.K. Lee, G. Hamarneh, Skin lesion tracking using structured graphical models, Med. Image Anal. 27 (2016) 84–92.
[35] A. Patrizi, M. Bentivogli, B. Raone, A. Dondi, M. Tabanelli, I. Neri, Association of halo nevus/i and vitiligo in childhood: a retrospective observational study, J. Eur. Acad. Dermatol. Venereol. 27 (2) (2013) e148–e152.
[36] P.M. Huynh, R. Lazova, J.L. Bolognia, Unusual halo nevi–darkening rather than lightening of the central nevus, Dermatology 202 (4) (2001) 324–327.
[37] M.C. Van Rijsingen, S.C. Hanssen, J.M. Groenewoud, G.J. Van Der Wilt, M.-J.P. Gerritsen, Referrals by general practitioners for suspicious skin lesions: the urgency of training, Acta Derm. Venereol. 94 (2) (2014) 138–141.
[38] S. Pathan, K.G. Prabhu, P.C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review, Biomed. Signal Process. Control 39 (2018) 237–262.
[39] K. Korotkov, R. Garcia, "Computerized analysis of pigmented skin lesions: a review, Artif. Intell. Med. 56 (2012) 69–90.
[40] T.J. Brinker, Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task, Eur. J. Cancer 113 (2019) 47–54.

[41] S.S. Han, M.S. Kim, W. Lim, G.H. Park, S.E. Chang, Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm, J. Invest. Dermatol. 138 (2018) 1529–1538.
[42] T.J. Brinker, et al., A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task, Eur. J. Cancer 111 (2019) 148–154.
[43] A.R. Lopez, X. Giro-i-Nieto, J. Burdick, O. Marques, Skin lesion classification from dermoscopic images using deep learning techniques, in: Proceedings of the 13th IASTED International Conference on Biomedical Engineering (BioMed), 2017, pp. 49–54.
[44] S.M. Swetter, J. Chang, A.R. Shaub, M.A. Weinstock, E.T. Lewis, S.M. Asch, Primary care–based skin cancer screening in a Veterans affairs health care system, JAMA Dermatol. 153 (2017) 797–801.