(Article begins on next page)

An Annotation Tool for
a Digital Library System of Epidermal Data

Fabrizio Balducci, Guido Borghi

Dipartimento di Ingegneria "Enzo Ferrari"
Università degli Studi di Modena e Reggio Emilia
Via Vivarelli 10, Modena MO 41125, Italy
{name.surname}@unimore.it

Abstract. Melanoma is one of the deadliest form of skin cancers so it becomes crucial the developing of automated systems that analyze and investigate epidermal images to early identify them also reducing unnecessary medical exams. A key element is the availability of user-friendly annotation tools that can be used by non-IT experts to produce well-annotated and high-quality medical data. In this work, we present an annotation tool to manually crate and annotate digital epidermal images, with the aim to extract meta-data (annotations, contour patterns and intersections, color information) stored and organized in an integrated digital library. This tool is obtained following rigid usability principles also based on doctors interviews and opinions. A preliminary but functional evaluation phase has been conducted with non-medical subjects by using questionnaires, in order to check the general usability and the efficacy of the proposed tool.

Keywords: annotation, epidermal images, skin, usability

1 Introduction

The interest of biomedical and computer vision communities in acquisition and analysis of epidermal images has been increased during the last decades: the possibility to automatically detect and classify early melanomas is investigated because it is one of the most common and danger skin cancer. Only considering United States, 100,000 new cases are diagnosed every year with over 9,000 deaths correlated [9, 8]. In this context new automated system for fast and accurate skin image acquisition and investigation, melanoma detection and classification are well accepted in the biomedical community, also considering that the diagnostic accuracy with trained clinicians is around 75-84% [15].

Traditionally, clinical experts manually categorize and examine printed medical images so it becomes a very time consuming task; to avoid this issue, advanced user-friendly annotation tools must be developed, in order to improve the digital libraries in biomedical and related fields in size and quality of annotated data that will be processed by machine learning and pattern recognition techniques. The usability of the tools, with also the assurance that their hardware

devices are medical-compliant, is a fundamental element due to user's lack of deep IT-skills.

This paper is organized as follows: Section 2 describes literature based on epidermal and, in particular, melanoma images while Section 3 proposes an architecture for the digital library with the implementation details of the annotation tool; Section 4 reports results obtained by the experimental evaluation and, finally, conclusions and future work are illustrated in Section 5.

2 Related Work

A complete and rich survey about lesion border detection is in [6]. The first step for skin inspection is the acquisition of digital images: the main techniques involve *Epiluminence Microscopy* (ELM, or dermoscopy), *Transmission Electron Microscopy* (TEM) and the acquisition through standard RGB cameras [19]. In last decades, standard video devices are commonly used for skin lesion inspection systems, in particular in the telemedicine field [28]. However, these solutions present some issues, like low camera spatial resolution (melanoma or other skin details can be very small) and distortions caused by the camera lenses. Moreover, variable illumination conditions can strongly deteriorate the quality of acquisitions [19, 3].

After the digital acquisition, the second step consists in the analysis and investigation of the epidermal images acquired: several works in literature have been proposed for automated epidermal image analysis, in order to support biomedical experts; most of them are based on Computer Vision approaches, typically combining low-level visual features representation, image processing techniques and machine learning and pattern recognition algorithms [8].

In [7] manually pre-segmented images, already cropped around the region of interest, have been used in conjunction with hand-coded and unsupervised features to achieve state-of-the-art results in melanoma recognition task, with a dataset of 2,000 dermoscopy images. Specifically, a combination of sparse coding, deep learning and Support Vector Machine (SVM) learning algorithms are exploited. In [21, 4] several machine learning classifiers, like SVMs and K-nearest neighbors (kNNs), based on color, edge and texture descriptors are investigated and compared. Learning approaches [25, 24] and deep learning techniques [10, 20, 14, 5] have been exploited in literature while a combination of hand-coded features extractors, sparse-coding methods, SVMs and deep learning techniques are used focusing on melanoma recognition and segmentation tasks in dermoscopy and dermatology domain [8].

Finally, in 2016 a new challenge, called *Skin Lesion Analysis toward Melanoma Detection* [12], has been presented: the aim is to use one of the most complete dataset of melanoma images, collected by the *International Skin Imaging Collaboration* (ISIC) and obtained with the aggregation dataset of dermoscopic images from multiple institutions, to test and evaluate the automated techniques for the diagnosis of melanomas; best scores in classification and segmentation are achieved using deep learning approaches [13, 17].
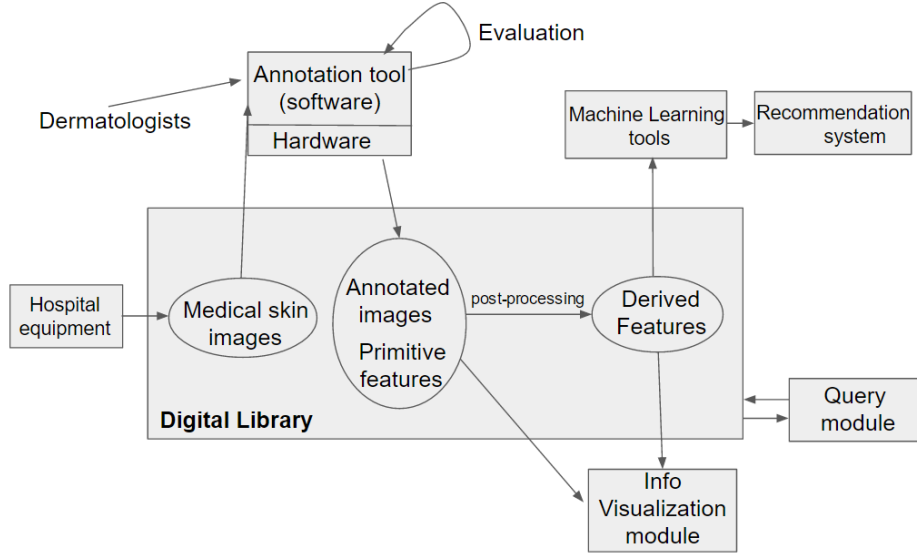
Fig. 1. The general architecture proposed for a biomedical digital library. The digital library and the annotation tool are the core of the presented method.

Before this challenge, the exploitation of deep learning techniques, like Convolutional Neural Networks (CNNs) was partially bounded by the limited size of datasets present in literature. Generally, the amount if training data and high quality annotations are key aspects for deep learning approaches [16].

3 Our proposal

This paper presents a new user-friendly annotation tool for epidermal images which must also be mobile oriented by running on smartphones and tablets. The annotation tool is the first step to create an heterogeneous integrated system, which architecture is depicted in Figure 1.

The system core is the Digital Library which stores and organizes the dataset acquired from hospital equipment: it currently consists of 436 medical skin images in standard JPEG format with a high spatial resolution of 4000x2664 or 3000x4000 pixels. The software has been developed to enable the annotation of images by domain experts, like dermatologists, and it is described in the following subsection: it permits to easily draw strokes with different colors (8 choices) and pen sizes (4 choices), where each color denotes a specific semantic given by dermatologists (this aspect will be addressed in future works).

The main goal of this tool is the retention of annotations in the form of a new enriched images also indexing each drawn stroke (in turn composed by points correspondent to image pixels). The organization of this data is made in form of specific folder and file names, composed together to incorporate and identify

Fig. 2. A skin image annotated with strokes of different colors and width (primitive feature).

the resources by using the annotator username and the path and the name of the original JPEG file.

An example of annotated skin image is depicted in Figure 2. The primitive meta-data directly extracted by the annotation tool are:

- a new color image, featuring all the draw strokes (primitive features)
- a binary (black and white) image for each color used to drawn the strokes (each image assembles the strokes of the same color)
- a text file with the features of each stroke (list of coordinate points, color code and pen size): this file is used to rebuilt and reload a saved annotation

After the annotation phase and the *primitive features* extraction and storing, it is necessary to perform a pre-processing phase before the extraction of *advanced features* to check two essential requirements for each annotated image:

1. remove thick hairs, because this artifacts influence the shape and contours extraction [29, 1]
2. assure that for each stroke and color all the drawn contours are closed [23], to preserve the consistency of extracted features

After this pre-processing phase new images are then created:

- a binary image for each color channel (RGB) used for the thick hair removal phase
- a new color image without thick hairs

Now it is possible to apply internal image processing functions and algorithms to extract *derived features* like contours, shapes, intersections, color features and numerical values, as shown in Figure 3.
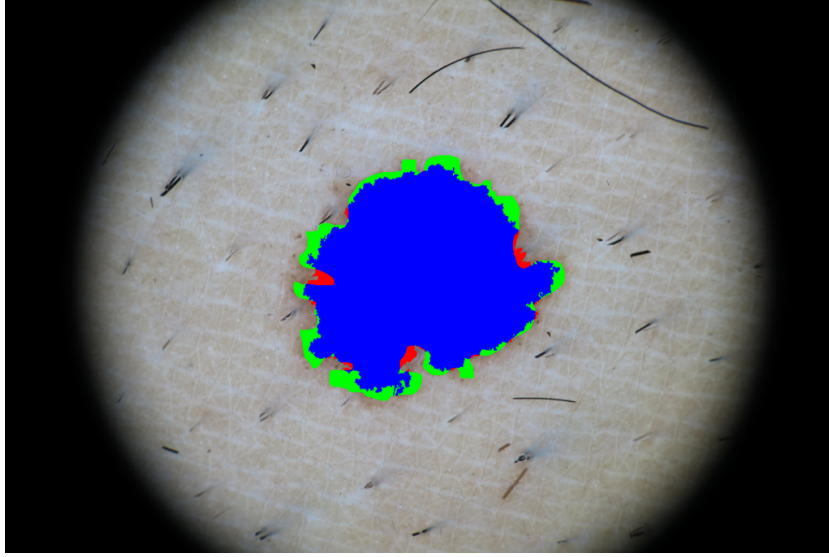
Fig. 3. An example of derived feature: the automatic area (red) made by the system and the intersection area (blue) obtained with the manual one (green) made by a dermatologist.

By using the new extracted features it will be possible to use *Machine Learning tools* to develop the *Recommendation module*. Image classification, to distinguish between benign and malignant (melanomas) lesions, could be carried out exploiting deep learning techniques; recently, this kind of approach reached best performances in many Computer Vision research tasks like image and video classification, objects detection and pose estimation.

Given an epidermal input image, a Convolutional Neural Network (CNN) could be used to produce the classification: a CNN is a neural model composed by sequences of convolutional layers that apply a set of filters (typically followed by Spatial Pooling layers) to down sample the input tensor in spatial dimensions, and by Fully Connected layers in which each neuron could be linked with all other neurons in the following layers. An activation function is exploited to introduce non-linearity inside the architecture [16]. Finally, a loss function represents the cost associated to the achievement of a final goal, like correct classification, regression and so on. The main issue of CNN is the huge amount of data needed in the training phase: with an unbalanced or small dataset they are prone to both over fitting and under fitting phenomena; to avoid this, data augmentation is often performed so an annotation tool to easily acquire and produce high-quality annotated biomedical images could be really useful and important.

Finally, in the architecture we considered an *Information Visualization module* for various and dynamic data visualization and a specific *Query module* to permit search facilities [2].
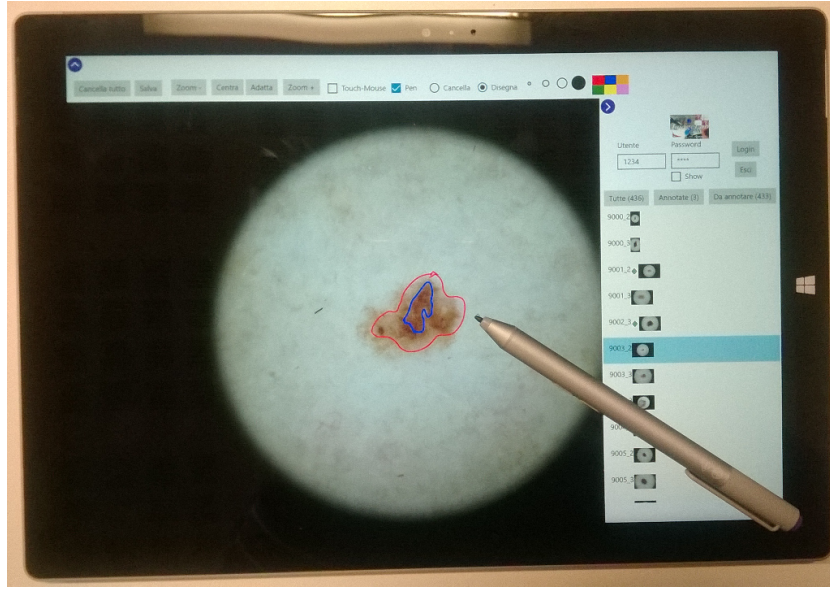
Fig. 4. The *Microsoft Surface Pro 3* is the hardware platform chosen for the annotation tool. In this way, portability and error-free annotations (thanks to the *Surface Pen*) are guaranteed.

An example of standard for handling, storing, printing, and transmitting information in medical imaging is the DICOM (Digital Imaging and Communications in Medicine, ISO standard 12052:2006) of which the National Electrical Manufacturers Association (NEMA) holds the copyright. This standard has been widely adopted by hospitals and is making inroads in smaller applications like dentists' and doctors' offices: it includes a file format definition and a network communications protocol which uses TCP/IP to communicate between systems and enables the integration of medical imaging devices like scanners, servers, workstations, printers and network hardware from multiple manufacturers.

The different devices come with DICOM Conformance Statements which clearly state which classes they support; a DICOM data object consists of attributes (such as name, ID, datetime) with a special one containing the pixel data: in this way the medical image contains for example the patient ID within the file so that the image can never be separated from this information (similarly to the JPEG format that embeds tags to describe an image).

3.1 Developing the Annotation Tool

We decided to use as hardware platform the *Microsoft Surface Pro 3*, depicted in Figure 4: this is a portable tablet device, powerful as a modern PC but less invasive, so it can be used in mobility into a medical environment like an hospital. Dermatologists can annotate images with the assurance that the only

strokes recognized by the touch screen are those which comes from the specific *Surface Pen*, avoiding unwanted strokes coming from touch gestures or oversight movements.

To acquire the necessary data for the digital library, we developed an annotation tool following the principles of usability and *Human-Computer Interaction* [27] bearing in mind the following principles:

1. final users are domain experts that could be unfamiliar with technical tool or data organization and analysis
2. physician are usually overworked so they do not have much time to skill themselves on externals tools
3. physician and specialists like dermatologists have peculiar working protocols and pipelines, so the tool must be non-invasive with the aim to not impact on their daily activities
4. the image annotation task must be as much as possible fast and user-friendly for a dermatologist, imitating what they would do naturally
5. the medical environment has peculiar safety and security requirements so, in addition to the software tool, the hardware introduced in this areas must be considered

The task of *annotate* with a precise technological pen on a glass screen is natural and intuitive and shows the characteristic of *Affordance* [11]: in the field of psychology the term includes all actions that are physically possible on an object or environment (like what a physician wants to performs on a real medical printed photo). In a general way, when the concept is applied to design and develop activities it refers to those action possibilities that the user is aware of, also considering the available or permissible peripherals that allow such interactions. The *perceived affordance* refers not only to the user physical capabilities but also on his goals, beliefs, and past experiences [22] so that an object (real or virtual) can naturally 'suggests' how to interact with it.

Another usability aspect is the *direct manipulation* allowed by the gestures on the touch screen like *to pitch* (zoom-in and zoom-out operations) and *to drag* (scrolling) which are alternatives to the button widgets provided by the user interface. Formally, Direct manipulation is an *interaction style* in which users act on displayed objects of interest using physical, incremental, reversible actions whose effects are immediately visible on the screen [26].

In his works Shneiderman identified several principles of this interaction style:

− continuous representation of the objects of interest: while performing an action user can see its effects on the state of the system
− physical actions instead of complex syntax: in contrast with command-line interfaces, actions are invoked physically via input peripherals, visual widges (button, menu,) and gestures.
− continuous feedback and reversible incremental actions: it results easy to validate each action and fix mistakes
− rapid learning: users learn by recognition instead of remember complex syntax commands
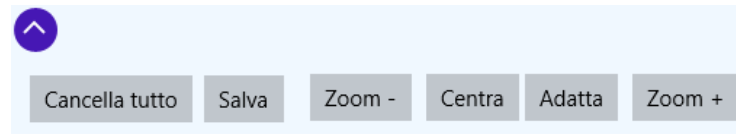
Fig. 5. The top panel of the annotation tool (part 1): here, user can save or delete the annotation and manipulate the image display.
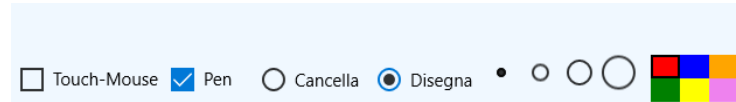

Fig. 6. The top panel of the annotation tool (part 2): here, user can set the interaction and input modalities and manage the stroke features.
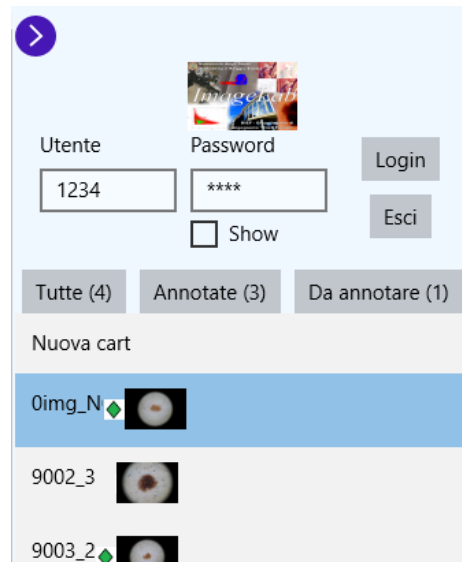
Fig. 7. The right panel of the annotation tool: it contains the login module, the list of images (with a thumbnail) and the tabs that dynamically filter them.

The interface is minimal but functional and presents two retractable panels at the top (separated in Fig. 5 and 6 and at the right (Fig. 7) part of the screen. The top panel contains a set of functionalities regarding image and annotation processing:

1. delete or store the annotation strokes
2. manipulate the image display (zoom in/out, center, scale adaptation)
3. select the input device (mouse/touch or pen)
4. manage the stroke features like color and width

Strongly Disagree			Indifferent			Strongly Agree
1	2	3	4	5	6	7

Fig. 8. The 7-degree *Likert* scale used for answer to a question.

The right panel presents two horizontal sections: the superior one permits the login and the exit from the application (exploiting the internal users database that also allows the meta-data organization previously described); the lower panel allows the image selection and loading by showing a list with a small preview of them (if an image has been annotated previously a green rhombus appears next to its name).

In the previous panel there are three tabs that have a dual functionality, in fact they show respectively the total number of images, the number of the annotated images and the number of the images that shall be annotated; moreover, when the user selects one tab, a *dynamic filter* updates the image list based on the selected preference.

4 Experimental setup

A preliminary evaluation was performed to test the annotation facilities and to understand if the user interface and its functionalities were easy to understand and perform; we used six non-medical students which, after a small brief on the capabilities and the target of this tool, were observed while carrying the annotation task on a subset of 25 images, encouraged to explore the tool.

After each experimentation phase, a questionnaire made by following Human Computer Interaction studies guidelines and consisting by 20 questions was given to the subjects. The evaluation for each question was given by a 7-degree ordered *Likert* scale as in Figure 8 and was asked to rate agreement with the statements, ranging from strongly disagree to strongly agree. Questions are divided into four sections concerning *Usefulness*, *Ease of Use*, *Ease of Learning* and *Satisfaction*.

In its work, Lund [18] states that users evaluate primarily using the dimensions of *Usefulness*, *Satisfaction*, and *Ease of Use* which are used to discriminate between interfaces. Partial correlations calculated using scales suggested that Ease of Use and Usefulness influence one another, such that improvements in Ease of Use improve ratings of Usefulness and vice-versa; while both drive Satisfaction, the Usefulness is relatively less important when the systems are internal systems that users are required to use; users are more variable in their Usefulness ratings when they have had only limited exposure to a product while, as expected from the literature, Satisfaction was strongly related to the usage (actual or predicted).

Questions and answers (the frequency of votes in *Likert* scale) are reported in Table 1: the majority of answers expresses positive agree as for question j

Table 1. Questions and Answers in Likert scale

	1	2	3	4	5	6	7
a) It gives me more control over the activities	0	0	0	0	0	3	3
b) It makes the things I want to accomplish easier	0	0	0	0	0	3	3
c) It saves me time when I use it	0	0	0	0	2	4	0
d) It meets my needs	0	0	0	1	0	4	1
e) It does everything I would expect it to d	0	0	0	0	2	1	3
f) It is easy to use	0	0	0	0	0	2	4
g) Using it is effortless	0	0	0	0	0	2	4
h) I can use it without written instructions	0	0	0	0	0	3	3
i) I don't notice any inconsistencies as I use it	0	0	0	0	2	3	1
j) Both occasional and regular users would like it	0	0	0	0	0	1	5
k) I can recover from mistakes quickly and easily	0	0	1	0	2	1	2
l) I learned to use it quickly	0	0	0	0	0	2	4
m) I easily remember how to use it	0	0	0	0	0	2	4
n) It is easy to learn to use it	0	0	0	0	0	3	3
o) I quickly became skillful with it	0	0	0	0	0	2	4
p) I am satisfied with it	0	0	0	0	2	4	0
q) I would recommend it to a friend	0	0	0	0	0	5	1
r) It is fun to use	0	0	0	0	1	3	2
s) It works the way I want it to work	0	0	0	0	1	3	2
t) I feel I need to have it	0	0	1	1	2	0	2

(Both occasional and regular users would like it) or question f (It is easy to use). Few low votes in interval [3-5] considering question t (I feel I need to have it) and question k (I can recover from mistakes quickly and easily).

In Figure 9 is shown a plot which summarizes and visualizes all the questionnaire results: it consists in an histogram featuring various dimensions:

1. X-axis: question code
2. Y-axis: the overall score in percentage, expresses the polarization towards a "strongly agreement" by considering a weighted vote for each given answer
3. Color: associated to a rating value of the *Likert* scale [1-7]; they go from blue palette (dislike mood) to the yellow (medium liking mood) and to the green palette (positive mood)
4. Number (into a bar): answer frequency, how many subjects answered to the question with the same rating
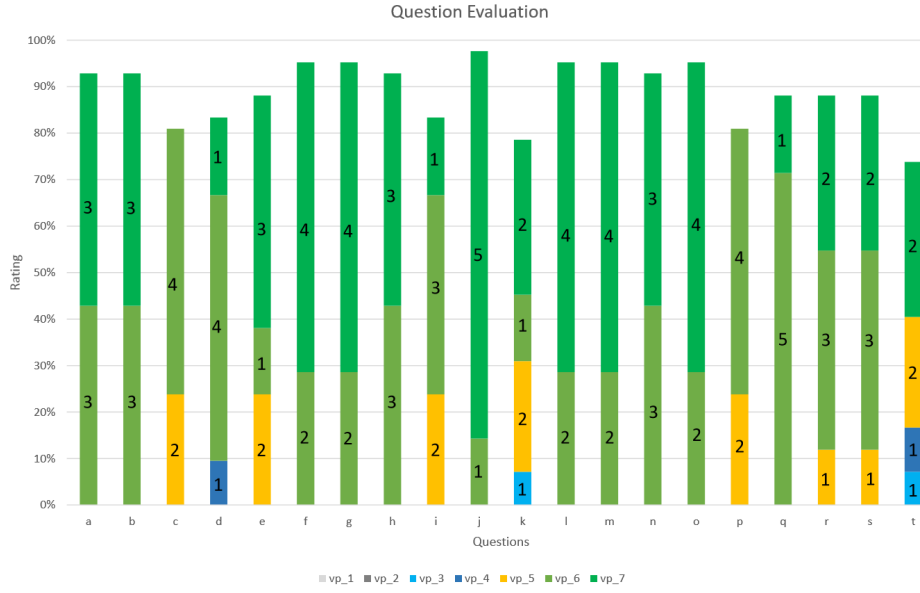
Fig. 9. The histogram about questionnaire answers.

5. Sub-bar: its width expresses the weight that have a specific answer on the overall score (100% would be all subjects scoring the highest vote of 7)

Observing the plot it becomes evident that, by considering the overall score in percentage, for all questions subjects expresses a positive liking mood (all bar height exceed 80%). By exploiting this visualization the differences between questions become evident, for example between question a and question c: in the second one two subjects answered value '5' (yellow color evidences that the value expresses a medium liking mood) and, since the sub-bar height is related to the liking mood, expressing such a value prevents that the global bar height could exceeds one with better scores.

5 Conclusions and Future work

This work is the first step towards the development of a more complex medical data system which achieves the entire pipeline from image gathering and annotation to the analysis and visualization of the (meta)data associated to them, with the aim to create and manage a biomedical Digital Library.

Regarding the development of the software tool presented here, the preliminary evaluation shows encouraging results both for user-centered approach and for the data retention and management. After observing common unskilled users, the future evaluation will include dermatologist in their work environment also taking into account their comments and observations.

With the aim to develop all of the proposed architecture modules, we will improve this approach by realizing the embedding of this heterogeneous information into the corresponding image file, considering also the specific module dedicated to the data query and search; raw and structured data will be analyzed with a wide variety of machine learning techniques such as Convolutional Neural Networks and SVM classifiers, with the aim to define patterns for the automatic detection and prevention of skin melanomas; in that cases, the system will be used as an agent to support dermatologists in their decisions.

The final step will be the evaluation of information visualization facilities by using the domain experts suggestions and also by the proposition of new and more effective techniques.

References

1. Dullrazor: A software approach to hair removal from images. Computers in Biology and Medicine 27(6), 533 – 543 (1997)
2. Ahlberg, C., Williamson, C., Shneiderman, B.: Dynamic queries for information exploration: An implementation and evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 619–626. CHI '92, ACM, New York, NY, USA (1992), `http://doi.acm.org/10.1145/142750.143054`
3. Ali, A.R.A., Deserno, T.M.: A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. In: SPIE Medical Imaging. pp. 83181I–83181I. International Society for Optics and Photonics (2012)
4. Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J.S.: Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal 8(3), 965–979 (2014)
5. Binder, M., Steiner, A., Schwarz, M., Knollmayer, S., Wolff, K., Pehamberger, H.: Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. British Journal of Dermatology 130(4), 460–465 (1994)
6. Celebi, M.E., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., Schaefer, G.: A state-of-the-art survey on lesion border detection in dermoscopy images. In: Dermoscopy Image Analysis, pp. 97–129. CRC Press (2015)
7. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In: International Workshop on Machine Learning in Medical Imaging. pp. 118–126. Springer (2015)
8. Codella, N.C.F., Nguyen, Q., Pankanti, S., Gutman, D., Helba, B., Halpern, A., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. CoRR abs/1610.04662 (2016), `http://arxiv.org/abs/1610.04662`
9. Diepgen, T., Mahler, V.: The epidemiology of skin cancer. British Journal of Dermatology 146(s61), 1–6 (2002)
10. Ercal, F., Chawla, A., Stoecker, W.V., Lee, H.C., Moss, R.H.: Neural network diagnosis of malignant melanoma from color images. IEEE Transactions on biomedical engineering 41(9), 837–845 (1994)
11. Gibson, J.J., Shaw, R.: Perceiving, acting, and knowing: Toward an ecological psychology. The Theory of Affordances pp. 67–82 (1977)

12. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1605.01397 (2016)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
14. Jaleel, J.A., Salim, S., Aswin, R.: Artificial neural network based detection of skin cancer. IJAREEIE 1, 200–205 (2012)
15. Kittler, H., Pehamberger, H., Wolff, K., Binder, M.: Diagnostic accuracy of dermoscopy. The lancet oncology 3(3), 159–165 (2002)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
18. Lund, A.M.: Measuring usability with the use questionnaire12 (2001)
19. Maglogiannis, I., Doukas, C.N.: Overview of advanced computer vision systems for skin lesions characterization. IEEE transactions on information technology in biomedicine 13(5), 721–733 (2009)
20. Marín, C., Alférez, G.H., Córdova, J., González, V.: Detection of melanoma through image recognition and artificial neural networks. In: World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada. pp. 832–835. Springer (2015)
21. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R., Rozeira, J.: Ph 2-a dermoscopic image database for research and benchmarking. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5437–5440. IEEE (2013)
22. Norman, D.A.: Affordance, conventions, and design. interactions 6(3), 38–43 (May 1999), http://doi.acm.org/10.1145/301153.301168
23. Pellacani, G., Grana, C., Seidenari, S.: Comparison between computer elaboration and clinical assessment of asymmetry and border cut-off in melanoma images. In: EXPERIMENTAL DERMATOLOGY. vol. 11, pp. 609–609. BLACKWELL MUNKSGAARD 35 NORRE SOGADE, PO BOX 2148, DK-1016 COPENHAGEN, DENMARK (2002)
24. Rastgoo, M., Morel, O., Marzani, F., Garcia, R.: Ensemble approach for differentiation of malignant melanoma. In: The International Conference on Quality Control by Artificial Vision 2015. pp. 953415–953415. International Society for Optics and Photonics (2015)
25. Schaefer, G., Krawczyk, B., Celebi, M.E., Iyatomi, H.: An ensemble classification approach for melanoma diagnosis. Memetic Computing 6(4), 233–240 (2014)
26. Shneiderman, B.: 1.1 direct manipulation: a step beyond programming languages. Sparks of innovation in human-computer interaction 17, 1993 (1993)
27. Shneiderman, B.: Designing the user interface: strategies for effective human-computer interaction. Pearson Education India (2010)
28. Singh, S., Stevenson, J., McGurty, D.: An evaluation of polaroid photographic imaging for cutaneous-lesion referrals to an outpatient clinic: a pilot study. British journal of plastic surgery 54(2), 140–143 (2001)
29. Zagrouba, E., Barhoumi, W.: A preliminary approach for the automated recognition of malignant melanoma. Image Analysis & Stereology 23(2), 121–135 (2011)