Cosmetic oncology: innocent mole or malignant melanoma? Subjective assessments, objective semiology and aided diagnosis

M. Burroni[1], U. Wollina[2], F. Arcangeli[3], R. Schwartz[4], R. Sadoughifar[5] and T. Lotti[6]

[1]Honorary Member of the Italian Society of Dermatology; [2]Department of Dermatology and Allergology, Schermistico, Klinikum Dresden, Academic Teaching Hospital of the Technical University, Dresden, Germany; [3]Prof. Arcangeli - Department of Dermatology and Venereology, University, of Rome G. Marconi, Rome, Italy; [4]Prof. Schwartz - Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, Ontario Tobacco Research Unit, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; [5]University of Rome G. Marconi, Rome, Italy, Department of Dermatology, University Medical Center Mainz, Mainz, Germany, Department of Dermatology, University Hospital Basel, Basel, Switzerland; [6]Institute of Dermatological Sciences, Florence, Italy; G. Marconi, Rome, Italy; [7]Department of Dermatology, University G. Marconi, Rome, Italy

Corresponding Author:

Roxanna Sadoughifar MD,
Institute of Dermatological Sciences,
Via Alfonso Lamarmora 29,
50129, Florence, Italy

Key words: *cosmetology, oncology, beauty moles, melanoma, diagnosis*

Abstract

Dermoscopy is a well established non invasive technique which allows the clinician to confirm his/her early diagnosis of so called "early melanoma". In several cases "early melanoma" is difficult to recognize versus innocent moles. In this position paper experts in the field of melanoma compare the personal diagnostic "eye-brain system" versus the dermoscopic imaging combined with the computerized pattern analysis of pigmented skin lesions. Digital image processing pattern recognition seems to be presently a reliable tool for aiding doctors in differentiating innocent moles versus the "early melanoma of the skin". The Medical Devices equipped with Artificial Intelligence should however rely on peculiar measurements extracted by the pattern analysis in order to provide the objective motivations of a diagnosis to clinicians and not simply based on image matching recognition. Experts need the motivations of a potential diagnostic aid.

Introduction

Beauty marks (also named beauty spots) are common spots usually located on the face and on the whole skin surface. Malignant melanoma can easily mimic pigmented beauty marks or non-pigmented beauty marks (achromatic cutaneous and mucous melanoma). In this position paper the authors suggest that the digital image processing algorithms represents the most reliable diagnostic technique for the first screening and the eventual subsequent detailed analysis. All cosmetic doctors should become familiar with the digital imaging processing methods oriented to the pigmented cutaneous lesions recognition and description.

Dermoscopy is a non-invasive analysis technique that allows the early detection of melanoma through the visual inspection of cutaneous subsurface structures of Pigmented Skin Lesions (PSL). This technique represents a tool to increase the clinical low sensitivity performed by the naked eye (1-5).

A skilled clinician is capable of recognising a huge number of features on the base of experience and known examples; his common sense leads him to the difficult process called "early diagnosis". Here we must underline the importance of the term "early" because the "late" diagnosis is not useful; this recognition could simply be performed by non-expert clinicians by means of the naked eye. The prominent role of experts is to provide the "knowledge" for the image interpretation and the rationalization of the thought process that leads from the image to the diagnosis. This knowledge, in the domain of Artificial Intelligence (AI), is called "Training Set" and similarly to the human brain a machine correlate the unknown cases to this given set. It is however important to understand if this set is based on the images, leaving a machine to do the stuff, or it is based on images and measurements based on peculiar features defined by experts in order to provide also a diagnostic motivation.

The human processing of the diagnosis begins with the patient's and lesion's history, the clinical and dermatoscopic inspection, proceeds with the brain and, of course, ends with a response to colleagues and/or patients. Moreover, we must consider that the diagnostic accuracy increases when the diagnosis is performed in-vivo (6) so we eventually should rely on real-time computerized analysis.

ABCD rule and dermoscopy

Dermatoscopy is widely used in the clinical diagnosis of PLS. A very important and often overlooked problem in the dermoscopic field is represented by the unambiguous definition of significant features, which is actually performed by expert clinicians. On one side, there are difficulties due to subjective perception's laws: each clinician has his own individual perception of shapes and colors so we must consider any "rigorous definition" a sort of "flexible tool". On the other side, there are semantic problems arising from certain primitives of language which are difficult to be standardized and described with unambiguous delineation (6, 8).

A well-known method arose many years ago with the ABCD paradigm developed to teach patients and students (and not clinicians) to observe their own lesions at home: still today Asymmetry, Irregular borders, Dark colors and large Dimensions help to suggest patients to go to the specialist for further examination. The ABCD rule in dermoscopy, an evolution of the traditional method, reports some limits and also when used by experts in dermoscopy may fail on small melanocytic lesions (7) and should be considered secondary for the pigment network analysis (12).

Expert clinicians do not use crispy rules: they look globally at the lesion and then observe, instantly, a huge number of detailed features following the dermoscopic criteria based upon analysis of textures, color shades, and colors clusters. An expert clinician is able to process in his brain a large amount of data and, following long experience and common sense, the diagnosis is generally performed with good sensitivity and specificity (6). The literature reports values among (80% - 95%) for sensibility and (70-85%) for specificity.

Artificial intelligence

The human perception can be surely improved by AI based filters (figure 1), based on histogram equalization and training set parameters, in order to enhance peculiar features. One study reported improved sensibility by 10% while keeping the same value of specificity (15). This diagnostic process along with the patient's anamnesis and the history of the analyzing lesions offers a reliable tool for the diagnosis.

Algorithms and rules based methods have been reported to be really useful for teaching and providing a sort of flow-chart to non experienced dermatologists (6). Using them as a diagnostic tool in a daily routine, however, we are slightly

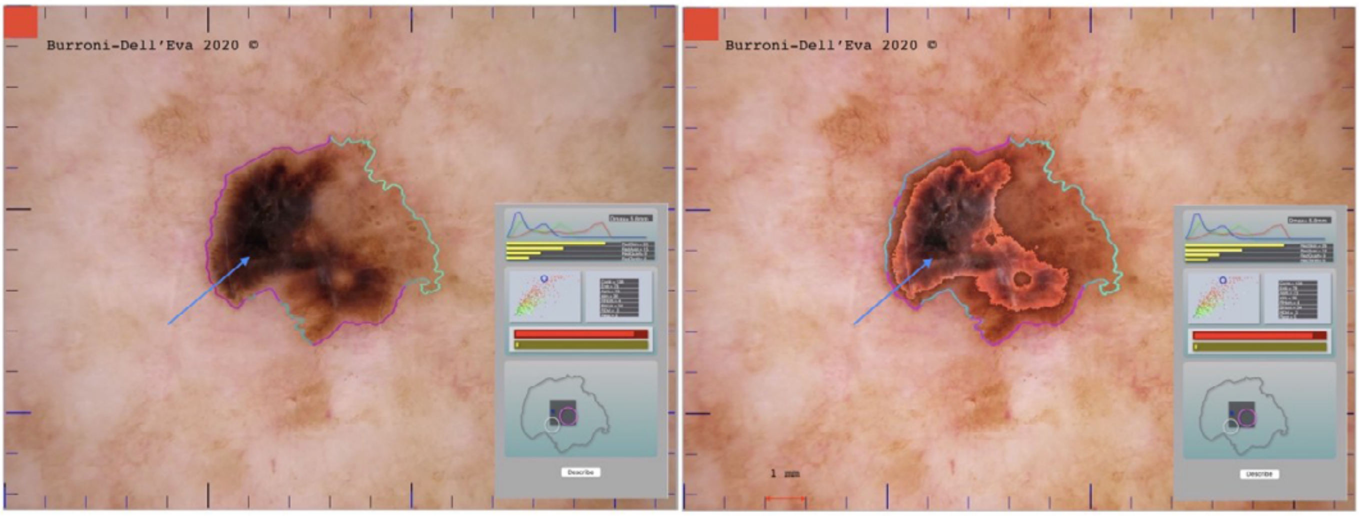

Fig. 1. *A screen-shot of the Islands Filter able to enhance the darker areas (blue arrow) and the related color shades in order to make them more visible to the clinician's perception.*

moving towards a mechanistic model that can be efficiently implemented by using computerized analysis. In addition, we have to take into strong consideration the evaluation of the intra-observer agreements of many visual features: the first variability in subjective definitions is due to the human perception's limits; the second variability is due to the difficult process of understanding the other's definitions caused by language standardization problems (6,8).

Some limitations of the subjectivity

The lack of standardization of terms and the objective limits in the perceptual ability of human beings leads to lower intraobserver agreements as reported in some studies (14, 17) based mainly on shared images and not on the in vivo inspection. The human brain tends to make rules on the basis of subjective impressions, sometimes based on unreproducible terms and tends to accept exceptions to rules basing on heuristic exclusions. The subjective assessment performed by the algorithmic methods is obviously subject to a highly reduced membership assignment continuing, at the same time, the reproducibility problems. An important task in the diagnosing process is the expenditure of time required for each "subjective" examination by rules: when a clinician observe a patient he performs instantly the selection of the lesions to be deeply examinated. This is a crucial task toward the needing of a real-time instrument as an aid specially in case many moles need to be examined. And, in case of (controversial) improvement of the diagnosis in the hand of non-expert clinicians, we maybe could use the common sense added to a computerized aided suggestion by means of reliable measurements of patterns starting from the traditional image processing methods (13) and their modern evolutions.

Progress by artificial intelligence

In order to overcome some of the aforementioned problems, focusing on dermatologists, some peculiar objective variables have been defined in order to evaluate both colors and color's clusters inside the lesions, i.e. "Islands of Colours" (9), experimenting later this new semiology providing a correlation with the clinical observation over atypical moles and early melanomas (10). The motivation of this path was suggested by the fact that medical devices should explain the reasons of a diagnosis through variables easy to be understood by clinicians. The perception laws and the cultural differences are creating a big gap not to be solved by the human eye-brain system. The main factor in the failure of some type of methods respect to the high accuracy revealed in the reality by the same experts who designed them is, that the PSL are natural objects with all the natural variations that life gives to things and simple descriptive rules create overlapping regions in the lesions classifications leading to "catchy" definitions instead of a complete and fluid complex diagnostic path. The human brain gives to a reconstructed image the right colors and the right meaning on the basis of previous experiences: this is a well-known aspect of the human perception called "visual memory law". The state of the art of the computer analysis permits not only the unambiguous mathematical definitions of terms, but mainly the automatic features evaluations leading to computerized aided suggestions.

It has been tested through a daily-routine based multicenter study that computerized aided differentiation of melanoma from other benign lesions is feasible and reliable (18). In the cited study two different departments were equipped with integrated systems (DB-Mips© Dell'Eva-Burroni), owning real-time AI engines: clinicians used the Systems in 4 years of daily routine inspecting the patients in-vivo. Subsequently the images and the related measurements, as automatically evaluated by the machine in the hands of clinicians, were collected and the results verified through a traditional statistical approach. The inevitable bias was caused by the necessity of exploring only the histology-proven lesions. It should be considered however that the analyzed lesions were eventually more difficult to be diagnosed respect to those not removed leading in the aforementioned study to a higher percentage of false positives respect to the reality and not viceversa. The study considered the flat pigmented lesions to be diagnosed by dermatoscopy. The "reliability study" indicated a mean sensitivity of 97% with a mean specificity of 75% of the integreted classifier and this result were confirmed by the statistical comparison of the same data. This approach took into strong consideration the real-clinical environment: it was not a subsequent processing of images but the analysis of those objective variables extracted by the machines during the examinations by clinicians in the routine.

The process of digital "image understanding", which refers to the techniques used to "teach" machines to recognize something, began many years ago by many research groups and continues with promising results even if the results may differ among different devices on the base of the kind of the adopted technology and the studied populations.

Artificial intelligence: pattern analysis and classifiers versus deep learning

The so called "AI" can be trained in several ways. We are reporting here the two currently used approaches.

One kind of Artificial Intelligence works by using the pattern analysis and proper classifiers based on peculiar measurements as reported in the aforementioned approach. The software evaluates from the examined lesions proper patterns defined by experts and then learn to recognise lesions from a traning-set built over these features (Figure 2). In this kind of approach clinicians know the objective motivations of a diagnosis based on pre-defined features of a mole. Using this technology doctors know the motivations of a machine's output. Moreover it is possible to perform statistical analysis in order to objectively describe different classes of pigmented lesions: the morphological features, when measured, can improve the understanding of the visual structure of the pigmented lesions.

The Deep Leaning approach, instead, is using a set of images leaving the machine to do the analysis by itself usually basing on Convolutional Neural Networks. Unlike the more traditional vision software, where programmers define objective rules, in deep-learning the algorithms find the rules by themselves but don't explain the decision. In recent years the use of the "Deep Learning", is actually rising many questions related to the lack of motivations of the diagnoses to clinicians.

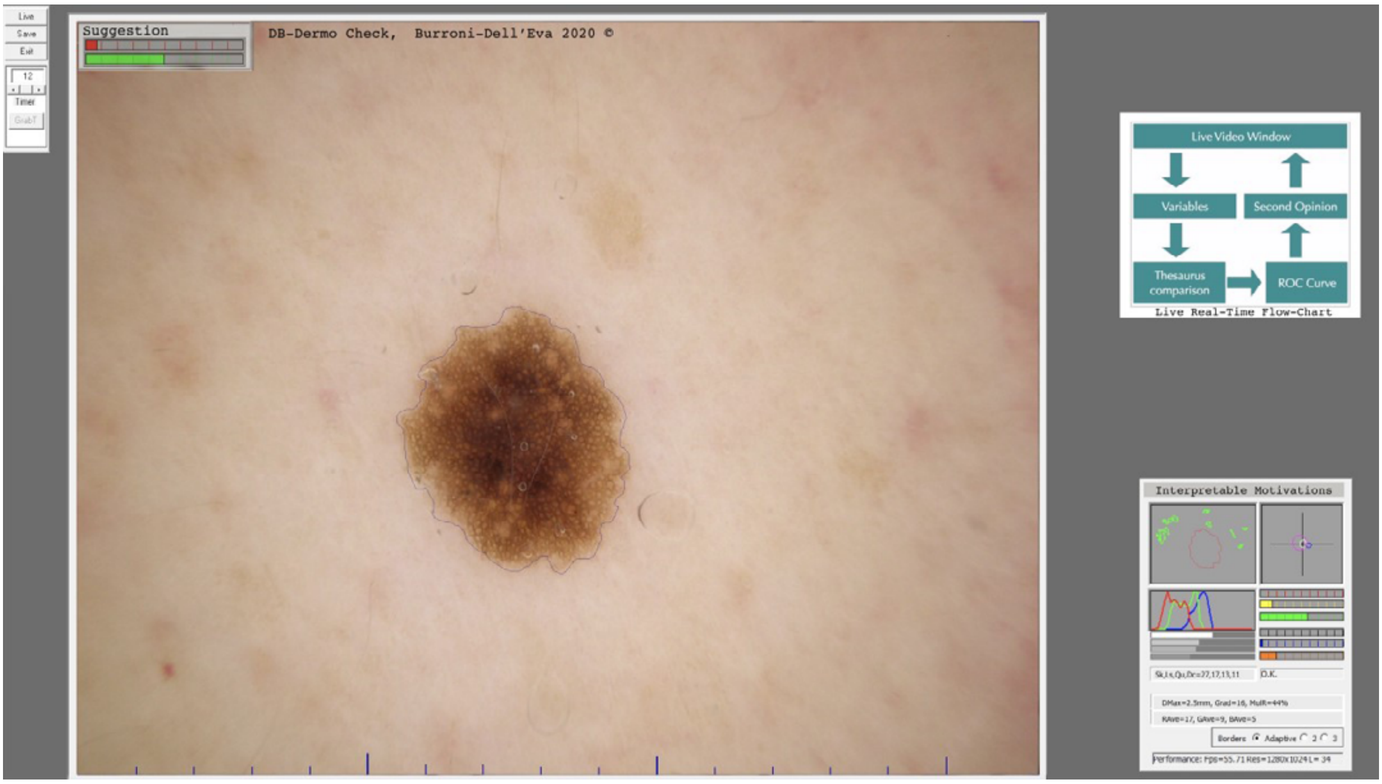

Fig. 2. *The image grabbing process of an Artificial Intelligence System of a Benign Naevus. The image's parameters are evaluated in real-time during the live inspection and compared with the training-set providing an aid (suggestion of risk probability). The peculiar parameters, used by the classifier, are shown on a proper window containing interpretable charts and measurements. The differential diagnosis during the live inspection is shown through horizontal bars in different colors. The Image and the related measurements are stored along with the Patient's database containing exaustive informations and total body photos.*

Cynthia Rudin, Professor of computer science at the Duke University, in a recent paper published on Nature, warns the use of Black-Boxes in "high stakes" fields (19). We should take into consideration the target and then decide which kind of aid is needed, providing interpretable results or not, on the base of the user's experience.

Discussion

It has been widely predicted that, over the next decade or two, biology will go completely quantitative. Medicine is following the same route. All modern measurement devices, from electrocardiograms to CAT scans or magnetic resonance images provide quantitative results and, unless we envision a future role of the clinician as a calculating machine, computerized analysis of results is the obvious consequence. Machines will never replace clinicians with their common sense, but clinicians may benefit by utilizing machines making complex calculations and providing interpretable motivations. The advantage in the daily routine could rise from using computers and peculiar measurements equipped with proper Patients database knowledge in the hands of dermatologists.

The use of the traditional pattern analysis method cooperating with a well-trained "objective based" machine seems to be actually a reliable tool. Dermatologists perform the diagnosis of a lesion using their experience, the touch, both patients and lesion's clinical history, then the naked eye inspection and finally a thorougly dermatoscopic examination. Recent reviews demonstrate how the in-vivo inspection of patients increase the diagnostic accuracy (6) and in the daily routine the need of real-time machine tools is almost evident. The actual debate refers also to the need of using interpretable models, in this case based on the pattern analysis and classifiers (18), instead of other models (19).

Conclusions

In our opinion clinicians need the diagnostic reasons explained, or to be clear, a motivation-based diagnosis performed on the same variables used to teach the training set. A modern device can be able to assist clinicians in selecting the suspicious lesions and then deliver a robust aid to them through a deeper analysis of the objective interpretable semiology. The keywords and the sense of this paper is objectivity, aid and interpretability in the hands of physicians.

Acknowledgements

Marco Burroni, a Biomedical Electronic Engineer, is actually involved into Melanoma Research and the related development, as author, of Medical Software and Devices oriented to the Early Diagnosis of melanoma and Telemedicine of Dermatological Deseases.

References

1. Kenet RO, Kang S, Kenet BJ, Fitzpatrick TB, Sober AJ, Barnhill RL. Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy. Grading protocol and atlas. Arch Dermatol 1993;129(2):157-74
2. Yadav S, Vossaert KA, Kopf AW, Silverman M, Grin-Jorgensen C. Histopathologic correlates of structures seen on dermoscopy (epiluminescence microscopy). Am J Dermatopathol. 1993;15(4):297-305.
3. Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, Bilek P, Braun-Falco O, Plewig G. The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions. J Am Acad Dermatol. 1994;30(4):551-9.
4. Binder M, Schwarz M, Winkler A, Steiner A, Kaider A, Wolff K, Pehamberger H. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Arch Dermatol. 1995;131(3):286-91.
5. Wolf IH, Smolle J, Soyer HP, Kerl H. Sensitivity in the clinical diagnosis of malignant melanoma. Melanoma Res. 1998;8(5):425-9.
6. Deeks J, Chuchu N, Ferrante di Ruffano L, et al. Cochrane Skin Cancer Diagnostic Test Accuracy Group. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. Cochrane Database Syst Review 2017 Dec.4;12(12):CD011902.
7. Pizzichetta MA, Talamini R, Piccolo D, Argenziano G, Pagnanelli G, Burgdorf T, Lombardi D, Trevisan G, Veronesi A, Carbone A, Soyer HP. The ABCD rule of dermatoscopy does not apply to small melanocytic skin lesions. Arch Dermatol. 2001;137(10):1376-8.
8. Stanganelli I, Burroni M, Rafanelli S, Bucchi L. Intraobserver agreement in interpretation of digital epiluminescence microscopy. J Am Acad Dermatol. 1995;33(4):584-9.
9. Burroni M. Understanding Digital Melanoma: Islands of Colors. Melanoma Res. 2001;11(Supp.1):22-3.
10. Andreassi L, Perotti R, Rubegni P, Burroni M, Cevenini G, Biagioli M, Taddeucci P, Dell'Eva G, Barbini P. Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology. Arch. Dermatol. 1999;135(12):1459-65.
11. Lorentzen H, Weismaan K, Secher L, Petersen CS, Larsen FG. The dermatoscopic ABCD rule does not improve diagnostic accuracy of malignant melanoma. Acta Derm Venereol. 1999;79:469-72.
12. Lorentzen H, Weismann K, Kenet RO, Secher L, Lar-sen FG. Comparison of dermatoscopic ABCD rule and risk stratification in the diagnosis of malignant melanoma. Acta Derm Venereol. 2000;80(2):122-6
13. Naranjan N.,Takanori F. Advances in Artificial Intelligence and Data Engineering. Springer, Ritsumeikan University Shiga Japan, 2017I SBN 978-981-15-3514-7.
14. Argenziano G, Soyer HP, Chimenti S, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. J Am Acad Dermatol. 2003;48(5):679-93.
15. Bauer P, Cristofolini P, Boi S, Burroni M, Dell'Eva G, Micciolo R, Cristofolini M. Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection. Melanoma Res. 2000;4:345-50.
16. Stang A, Garbe C, Autier P, Jöckel KH. The many unanswered questions related to the German skin cancer screening programme. Eur J Cancer. 2016;64:83-8.
17. Carrera C., Marchetti M., Dusza S., Argenziano G. et al. Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma: A Web-Based International Dermoscopy Society Study. JAMA Dermatol. 2016;152(7):798-806
18. Burroni M., Corona M., Dell'Eva G., Sera F. et al. Melanoma computer-aided diagnosis: reliability and feasibility study. Clin Cancer Res. 2004;10(6):1881-6.
19. Cyntia Rudin. Stop explayning black box machine learning models for high stakes decision and use interpretable models instead. Nature Machine Intelligence 2018;1:206-15.