Recent advances in deep learning applied for skin cancer detection

Anonymous Author(s)
Affiliation
Address
email

Abstract

Skin cancer is a major public health problem around the world. Its early detection is very important to increase patient prognostics. However, the lack of qualified professionals and medical instruments is a significant issue in this field. For this reason, over the past few years, deep learning models applied to automated skin cancer detection has become a trend. In this paper, we present an overview of the recent advances reported in this field as well as a discussion about the challenges and opportunities for improvement in the current models. In addition, we also present some important aspects regarding the use of these models in smartphones and indicate future directions we believe the field will take.

1 Introduction

Skin cancer is the most common cancer worldwide. The World Health Organization (WHO) estimates that one in every three cancers diagnosed is a skin cancer [1]. In countries such as USA, Canada, and Australia, the number of people diagnosed with skin cancer has been increasing at a fairly constant rate over the past decades [2, 3, 4]. The deadliest type of skin cancer is the melanoma and its early detection greatly improves the prognosis of patients [5]. Nonetheless, there is a lack of medical instruments and qualified professionals to assist the population, especially in rural areas [6] and in economically emerging countries [7]. In this sense, over the past decades, different computer-aided diagnosis (CAD) systems have been proposed to tackle skin cancer detection. These systems are mostly based on traditional computer vision algorithms to extract various features, such as shape, color, and texture in order to feed a classifier [8, 9, 10, 11, 12]. Recently, machine learning techniques became a trend to handle this task. Deep learning models, in particular, Convolutional Neural Networks (CNN), have been achieving remarkable results in this field. Yu et al. [13] presented a very deep CNN and a set of schemes to learn under limited training data. Esteva et al. [14] used a pre-trained CNN model to train more than 120 thousand images and achieve a dermatologist-level diagnostic. Haenssle et al. [15] and Brinker et al. [16] presented CNN models that have shown competitive or outperformed the dermatologists. Other efforts have been made using deep learning to detect skin cancer, such as ensemble of models [17, 18], feature aggregation of different models [19], among others [20, 21, 22].

The recent progress achieved by the machine learning methodologies has been leading to the accession of smartphone-based applications as a tool to handle the lack of dermatoscopes[1] available to dermatologists and general practitioners. According to the Ericsson mobile report [23], there are around 7.9 billion smartphones around the world. Thereby, a CAD system embedded in smartphones seems to be a low-cost approach to tackle this problem. However, even though this technology has the potential to be widely used in dermatology, there are important aspects that must be addressed

[1] a medical instrument that allows the visualization of the subsurface structures of the skin revealing lesion details in colors and textures

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

such as target users and how to present the system predictions. In addition, there are important ethical concerns regarding patient confidentiality, informed consent, transparency of data ownership, and data privacy protection [24].

Since the impact of machine learning in dermatology will increase in the next few years, the goal of this paper is to critically review the latest advances in this field as well as to reflect about the challenges and aspects that need to improve. To this end, first, we present the main methodologies and results reported for the task. Then, we provide a discussion about general limitations regarding the machine learning methods and about the smartphone application issues. Lastly, we conclude this paper with our perspectives about this field for the future.

2 Automated skin cancer detection

2.1 Recent advances

The automated skin cancer detection is a challenging task due to the variability of skin lesions in the dermatology field. The recent advances reported for this task have been showing that deep learning is the most successful machine learning technique addressed to the problem. In this sense, the International Skin Imaging Collaboration (ISIC) has been playing an important role by maintaining the ISIC Archive, an international repository of dermoscopic skin images, which includes, skin diseases and skin cancer [25]. This archive have been proving data for different deep learning methodologies such as the ones proposed by Yu et al. [13], Codella et. al. [17], Haenssle et al. [15], and Briker et al. [16]. Currently, the ISIC archive contains 25,331 images for training and 8,238 test available for research purposes.

While developing approaches using the ISIC archive is important, it constrains its use for dermoscopic images. It means this system cannot be used, for example, in smartphone apps, except if the device has a special dermoscope attached to it. In this context, it is necessary to expand the models to also handle clinical images. However, for this case, there is no large public archive available like ISIC. Thereby, Han et al. [20] combined clinical images from 5 repositories, public and private, in order to differentiate benign and malignant cutaneous tumors. Nonetheless, a breakthrough work was presented by Esteva et al. [14] in which the authors collected 129,450 clinical images and trained a convolutional neural network (CNN) that achieved a dermatologist level in the benign/malignant identification. Unfortunately, this dataset is private and it is not available for the research community.

Another trend in this field is to adopt an ensemble of deep models instead of a single method. The main goal of this method is to make the predictions more effective and reliable. Codella et al. [17] employed an ensemble of different deep models, including deep residual networks and convolutional neural networks (CNNs), in order to detect malignant melanomas, the deadliest type of skin cancer. Similarly, Gessert et al. [26] adopted several types of CNN architectures in order to classify 7 different types of skin diseases. In general, the ensemble of models have been achieving landmark results, in particular for ISIC archive [27].

In Table 1, we summarize all previously mentioned methods and their main contributions. It is important to note that all those models use only images to output their diagnostics. In fact, dermatologists do not trust only on the image screening, they also use the patient clinical information in order to provide a more reliable diagnostic. Pieces of information such as the patient's age, sex, ethnicity, if the lesion hurts or itches, among many others, are relevant clues towards a better prediction [28]. Thence, another breakthrough work has been recently proposed by Google Health researches in which they developed a deep learning system that is able to combine one or more images with the patient metadata in order to classify 26 skin conditions [29]. The addition of metadata provided a 4-5% consistent improvement in their model. They also report a result that is on par with U.S. board-certified dermatologists. Nonetheless, the authors indicate that is necessary to investigate prospectively the clinical impact of using this tool in actual clinical workflows.

To conclude this section, it is worth noting the recent work developed by Faes et al. [31]. In this work, the authors, who do not have any experience with algorithm development, used the Google Cloud AutoML to design several deep learning models for medical images, including skin cancer. They use a partition of the ISIC archive and reported a result that is comparable to other elementary classification tasks in this section. For one hand, it is a democratization of deep learning techniques. However, it also raise some questions about ethical principles when using these automated models.

Table 1: A summary of the recent deep learning models proposed to skin cancer detection

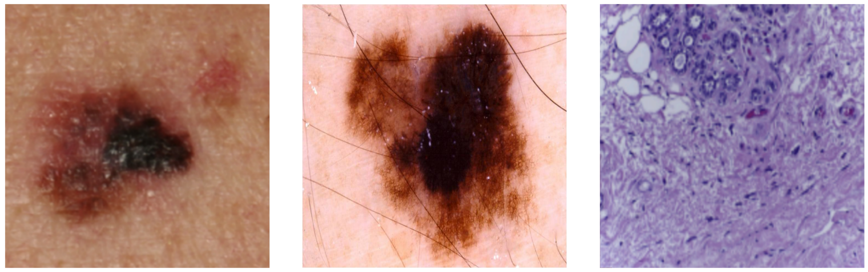
Ref.	Objective	Model	Main findings
[13]	Diagnose melanoma and non-melanoma using dermoscopic image	A two-stage framework composed of a fully convolutional residual network (FCRN) and a Deep Residual Network (DRN)	It was one of the first deep learning models applied to skin cancer detection and experimental results demonstrate the significant performance gains of the proposed framework compared to handcrafted feature models
[15]	Diagnose melanomas and nevus using dermoscopic images	Inception v4 CNN model	The authors compared the model performance to a group of 58 dermatologists using 100 images in the test set. The model AUC was greater than the average AUC of the dermatologists
[30]	Diagnose melanomas and nevus using dermoscopic images	ResNet50 CNN model	The authors compared the model to a group of 157 dermatologists using 100 images. The model outperformed 136 of them in terms of average specificity and sensitivity
[20]	Diagnose benign and malignant cutaneous tumors among 12 types of skin diseases using clinical images	ResNet-152 CNN model	The results achieved by the model was comparable to the performance of 16 dermatologists. The authors also affirm it is necessary to collect images with a broader range of ages and ethnicities in order to improve the model
[14]	Diagnose 757 types of skin diseases using clinical images	GoogleNet Inception v3 CNN model	The model achieved performance on par with 21 dermatologists considering the binary classification of the most common and the deadliest cases of skin cancer
[17]	Diagnose melanoma and non-melanoma using dermoscopic images	An ensemble composed of DRNs, CNNs and Fully CNNs	The ensemble was compared to the average of 8 dermatologists on a subset of 100 test images, and provided a higher accuracy and specificity, and an equivalent sensitivity
[26]	Diagnose 7 different types of skin diseases using dermoscopic images	An ensemble composed of ResNets, Densenets and Senets	The authors presented a new strategy based on a vast amount of unscaled image crops to generate final predictions. This approach outperforms most of the current models proposed for the ISIC archive

2.2 Challenges and opportunities

The models and results summarized in the previous section indicate the potential of CAD systems based on deep learning models applied for skin cancer detection. Nonetheless, there are several concerns that must be addressed in order to improve those systems. In this context, the goal of this section is to present a discussion about these concerns as well as indicate challenges and opportunities in this field.

2.2.1 Dataset, bias and uncertainty

It is known that to apply deep learning approaches it is necessary a large amount of data. However, collecting medical data, in particular from skin cancer, is a challenge task. Therefore, one of the main concerns of applying deep learning for this task is the lack of training data [20, 13]. As stated before, the ISIC archive is very important to tackle this issue. However, the number of samples available is still insufficient and very imbalanced among the classes. In order to tackle these issues, several approaches have been proposing such as transfer learning, data augmentation, up/downsampling and



(a) Clinical (b) Dermoscopic (c) Histopathological

Figure 1: The difference between the clinical [20], dermoscopic [25] and histopathological [34] images of a skin cancer

weighted loss [32, 33]. Nevertheless, there is still room for improvement and approaches to learn with limited data and based on weak supervision seem to be good choices to deal with it.

It is also important to note that the lack of open clinical data is a limiting factor for this task. As shown in Figure 1, dermoscopic and clinical images present significant differences related to the level of details available in each image. For this reason, reuse a model trained using only dermoscopic images to predict clinical images is not possible. The previously described works that deal with clinical data either combined some small datasets [20] or have access a private ones [14, 29]. In this sense, a concerted effort is needed in order to build a clinical image archive such as ISIC. Furthermore, it is important to include, along with the images, the patient information (metadata). As Liu et al. [29] has shown, the use of metadata may help the deep learning systems deal with the lack of a large number of images.

Another challenge regarding the skin cancer detection is to understand the current bias that distort the performance of the models. Bissoto et al. [35] carried out a study that suggests spurious correlations guiding the models. Moreover, some datasets, such as the used by Liu et al. [29], contain just a few samples of skin types IV and V [28], which contribute to the bias. All these points must be considered in order to deploy a model that is able to detect skin cancer for a more diverse group of people.

Beyond the bias, the patient metadata may contain uncertain information. Pieces of information such as family cancer history, if the lesion is painful or itching, among many others, are surrounded by uncertainty. Currently, the models do not take it into account, but it is an issue that should be addressed in the future.

2.2.2 Presenting the predicted diagnosis

Currently, the most common way the models provide the diagnostic is choosing the label that produces the highest probability. Some models also provide a ranking or a threshold for suspicious lesions [20, 29]. However, how can a clinician interpret a low probability assigned to a melanoma? In fact, they require more explanations than only the model's predictions [36]. Instead of focus only on the final accuracy, we need to improve how we present the results to the users. In this context, it is very important to determine the target user. Dermatologists, general practitioners, medical students, or even patients, have different levels of knowledge, hence, different needs.

In general, a clinician is interested in CAD systems that support their diagnostic by presenting insights and visual explanations of the features used by a model in classification process [36]. They want to know why the model is selecting such label. In this sense, we need to also focus on models that are able to output not only the labels probabilities but the pattern analysis as well. Kawahara and Hamarneh [37] proposed a model to detect dermoscopic feature classification, but it needs to be improved and extended to clinical data. In Figure 2 is depicted an example of the 7-point checklist, an algorithm based on pattern analysis commonly used to dermatologists to detect skin cancer [38]. As we can note, the expert is able to identify known patterns in the image in order to determine the final diagnosis. While it is a very challenging task, it should be the ultimate goal of a CAD system employed to skin cancer detection.

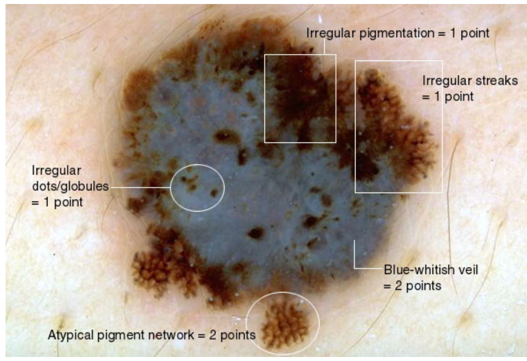


Figure 2: An example of the pattern analysis according to the 7-point checklist [39]

3 Skin cancer detection using smartphones

Due to the recent progress achieved by the CAD systems for skin cancer detection, currently, there are several smartphone-based applications that aims to deal with this task. As stated before, embed a skin cancer detection in a smartphone is a low-cost approach to tackle the lack of dermatoscopes in remote places. It is clear that this technology has the potential to impact positively on people's life. It may accelerate and help clinicians to provide a reliable diagnosis. However, developing such technology is not only deploy the model in a smartphone, there are important ethical aspects that must be addressed. The amount of those apps available for general users has drawn the attention of different researchers that claim several issues regarding its use. Kassianos et al. [40] carried out a study that identified 40 smartphone apps available to detect or prevent melanoma by nonspecialist users. Half of them enabled patients to capture and store images of their skin lesions either for review by a dermatologist or for self-monitoring. Chao et al. [24] conducted a similar study and concluded that only a few apps have involved the input of dermatologists. In addition, most of them do not provide a disclosure of authorship and credentials. As such, the application should make it clear how it handles user data. It must ensure patient confidentiality as well as let them know what the application does with their data after the processing. It may sound obvious, but as Chaos et al. [24] have shown, researchers/developers are not respecting that.

Beyond the problems regarding to patient confidentiality and privacy, the lack of regulation for those apps may result in harm to the patient or mislead them with an incorrect diagnostic. Let us consider a hypothetical situation of a false negative for melanoma to a given user. It may delay their treatment and, in the worst scenario, it may lead them to death. This is a serious problem that we, machine learning researchers, need to confront. First of all, it is quite important the opinion of dermatologists to improve the effectiveness of this technology. Then, those applications must be exhaustively tested before deployed. Lastly, in our opinion, they should not be allowed to general users before certification of a board of experts. To this end, it is necessary regulation and we need to advocate for this.

To conclude, in addition to the challenges described in the previous section, in particular, the target users and the way to present the results, there is an important technological issue about deploying deep learning models in smartphones that should be discussed. The main use of this kind of applications will be in remote places such as rural areas. In this sense, it is expected no internet access in those places. However, the current apps do not process the data inside the smartphone, but in a server, which demands internet. There are some fair reasons for this characteristic: the classification is based on more than one model, i.e., an ensemble; the models are computationally expensive, which demands better hardware than the ones usually found in smartphones; and the model's weights are large files, which may not fit in the smartphone memory. In summary, this is an important aspect that we could not find any discussion about it. In our opinion, this may lead to the development of lighter models in order to deal with it.

Figure 3: An example of the VQA problem applied to skin cancer detection

4 Final considerations and future directions

The recent advances in deep learning models for skin cancer detection have been showing the potential of this technique to deal with this task. Nonetheless, there are some limitations and important aspects that need to be addressed. In this paper, we presented a discussion about the state-of-the-art approaches as well as the main challenges and opportunities related to the problem. Despite the remarkable results reported, we indicated that there are rooms for improvement, especially for the way the results should be presented. In this context, we believe that in the future this task needs to be addressed as a variant of the visual and question answering (VQA) problem [41]. In Figure 3 is illustrated an example of the VQA problem applied to skin cancer detection. The main goal is to allow clinicians to make questions about the lesion in order to understand the predicted diagnosis outputted by the model. This approach is in accordance with the interest of the clinicians, which we described in Section 2.2.2. It is clear that addressing the skin cancer detection as a VQA problem increases the difficulty of the problem. However, it is an efficient way toward the goal of delivering a more useful tool for doctors.

Another aspect we believe will become a trend in the near future is the use of three types of skin cancer images: clinical, dermoscopic and histopathological. As we can see in Figure 1, each image presents different characteristics, which may help to correlate features in order to improve the predicted diagnostic. In addition, the CAD systems will be able to act from the clinical diagnostic to the biopsy, which makes it more desirable and useful. To conclude, regarding the deployment of deep models in smartphones, as noticed earlier, the use of lighter models is necessary in order to make the apps available in remote places. In this context, investigating better ways to improve the transfer learning and considering not only the image but also the patient metadata are important aspect to be explored in the future.

References

[1] World Health Organization (WHO). How common is the skin cancer?, 2019. `https://www.who.int/uv/faq/skincancer/en/index1.html` [Last accessed: 15 September 2019].

[2] Cancer Council Australia. Understanding skin cancer - a guide for people with cancer, their families and friends, 2018. `https://www.cancer.org.au/about-cancer/types-of-cancer/skin-cancer.html` [Last accessed: 15 September 2019].

[3] Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian cancer statistics 2014 - special topic: Skin cancers, 2014. `https://www.cancer.ca/statistics.` [Last accessed: 15 September 2019].

[4] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[5] Anthony F Jerant, Jennifer T Johnson, Catherine Demastes Sheridan, and Timothy J Caffrey. Early detection and treatment of skin cancer. *American family physician*, 62(2), 2000.

[6] Hao Feng, Juliana Berk-Krauss, Paula W Feng, and Jennifer A Stein. Comparison of dermatologist density between urban and rural counties in the united states. *JAMA dermatology*, 154(11):1265–1271, 2018.

[7] Richard M Scheffler, Jenny X Liu, Yohannes Kinfu, and Mario R Dal Poz. Forecasting the global shortage of physicians: an economic-and needs-based approach. *Bulletin of the World Health Organization*, 86:516–523B, 2008.

[8] Fikret Ercal, Anurag Chawla, William V Stoecker, Hsi-Chieh Lee, and Randy H Moss. Neural network diagnosis of malignant melanoma from color images. *IEEE Transactions on Biomedical Engineering*, 41(9):837–845, 1994.

[9] M Emre Celebi, Hassan A Kingravi, Bakhtiyar Uddin, Hitoshi Iyatomi, Y Alp Aslandogan, William V Stoecker, and Randy H Moss. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6):362–373, 2007.

[10] Paul Wighton, Tim K Lee, Harvey Lui, David I McLean, and M Stella Atkins. Generalizing common tasks in automated skin lesion diagnosis. *IEEE Transactions on Information Technology in Biomedicine*, 15(4):622–629, 2011.

[11] Ilias Maglogiannis and Konstantinos K Delibasis. Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy. *Computer Methods and Programs in Biomedicine*, 118(2):124–133, 2015.

[12] Catarina Barata, M Emre Celebi, and Jorge S Marques. Improving dermoscopy image classification using color constancy. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1146–1152, 2014.

[13] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*, 36(4):994–1004, 2017.

[14] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

[15] Holger A Haenssle, Christine Fink, R Schneiderbauer, Ferdinand Toberer, Timo Buhl, A Blum, A Kalloo, A Ben Hadj Hassen, L Thomas, A Enk, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 29(8):1836–1842, 2018.

[16] Titus J Brinker, Achim Hekler, Alexander H Enk, Joachim Klode, Axel Hauschild, Carola Berking, Bastian Schilling, Sebastian Haferkamp, Dirk Schadendorf, Tim Holland-Letz, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113:47–54, 2019.

[17] Noel CF Codella, Q-B Nguyen, Sharath Pankanti, DA Gutman, Brian Helba, AC Halpern, and John R Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4):5–1, 2017.

[18] Balazs Harangi. Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics*, 86:25–32, 2018.

[19] Zhen Yu, Xudong Jiang, Feng Zhou, Jing Qin, Dong Ni, Siping Chen, Baiying Lei, and Tianfu Wang. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. *IEEE Transactions on Biomedical Engineering*, 66(4):1006–1016, 2019.

[20] Seung Seog Han, Myoung Shin Kim, Woohyung Lim, Gyeong Hun Park, Ilwoo Park, and Sung Eun Chang. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7):1529–1538, 2018.

[21] Mohamed Attia, Mohamed Hossny, Saeid Nahavandi, and Anousha Yazdabadi. Skin melanoma segmentation using recurrent and convolutional neural networks. *IEEE 14th International Symposium on Biomedical Imaging*, pages 292–296, 2017.

[22] Nudrat Nida, Aun Irtaza, Ali Javed, Muhammad Haroon Yousaf, and Muhammad Tariq Mahmood. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering. *International journal of medical informatics*, 124:37–48, 2019.

[23] Ericsson Inc. Ericsson mobility report, 2019. `https://www.ericsson.com/en/mobility-report/reports/june-2019`. [Last accessed: 15 September 2019].

[24] Elizabeth Chao, Chelsea K Meenan, and Laura K Ferris. Smartphone-based applications for skin monitoring and melanoma detection. *Dermatologic clinics*, 35(4):551–557, 2017.

[25] International Skin Imaging Collaboration (ISIC). Isic archive, 2019. `https://www.isic-archive.com`. [Last accessed: 15 September 2019].

[26] Nils Gessert, Thilo Sentker, Frederic Madesta, Rüdiger Schmitz, Helge Kniep, Ivo Baltruschat, René Werner, and Alexander Schlaefer. Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting. *arXiv preprint arXiv:1808.01694*, 2018.

[27] Fábio Perez, Sandra Avila, and Eduardo Valle. Solo or ensemble? choosing a cnn architecture for melanoma classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[28] Klaus Wolff, Richard Allen Johnson, Arturo P. Saavedra, and Ellen K. Roh. *Fitzpatrick's color atlas and synopsis of clinical dermatology*. McGraw-Hill Education, New York, USA, 8 edition, 2017.

[29] Yuan Liu, Ayush Jain, Clara Eng, David H Way, Kang Lee, Peggy Bui, Kimberly Kanada, Guilherme de Oliveira Marinho, Jessica Gallegos, Sara Gabriele, et al. A deep learning system for differential diagnosis of skin diseases. *arXiv preprint arXiv:1909.05382*, 2019.

[30] Titus Josef Brinker, Achim Hekler, Jochen Sven Utikal, Niels Grabe, Dirk Schadendorf, Joachim Klode, Carola Berking, Theresa Steeb, Alexander H Enk, and Christof von Kalle. Skin cancer classification using convolutional neural networks: systematic review. *Journal of Medical Internet Research*, 20(10):e11936, 2018.

[31] Livia Faes, Siegfried K Wagner, Dun Jack Fu, Xiaoxuan Liu, Edward Korot, Joseph R Ledsam, Trevor Back, Reena Chopra, Nikolas Pontikos, Christoph Kern, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *The Lancet Digital Health*, 1(5):e232–e242, 2019.

[32] Cristina Nader Vasconcelos and Bárbara Nader Vasconcelos. Experiments using deep learning for dermoscopy image analysis. *Pattern Recognition Letters*, 2017. . In press.

[33] Fábio Perez, Cristina Vasconcelos, Sandra Avila, and Eduardo Valle. Data augmentation for skin lesion analysis. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pages 303–311. Springer, 2018.

[34] Irene Fondón, Auxiliadora Sarmiento, Ana Isabel García, María Silvestre, Catarina Eloy, António Polónia, and Paulo Aguiar. Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Computers in biology and medicine*, 96:41–51, 2018.

[35] Alceu Bissoto, Michel Fornaciali, Eduardo Valle, and Sandra Avila. (de) constructing bias on skin lesion datasets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[36] George A Zakhem, Catherine C Motosko, and Roger S Ho. How should artificial intelligence screen for skin cancer and deliver diagnostic predictions to patients? *JAMA dermatology*, 154(12):1383–1384, 2018.

[37] Jeremy Kawahara and Ghassan Hamarneh. Fully convolutional neural networks to detect clinical dermoscopic features. *IEEE journal of biomedical and health informatics*, 23(2):578–585, 2018.

[38] Giuseppe Argenziano, Gabriella Fabbrocini, Paolo Carli, Vincenzo De Giorgi, Elena Sammarco, and Mario Delfino. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, 134(12):1563–1570, 1998.

[39] Dermoscopy. 7-point checklist, 2019. `http://www.dermoscopy.org/consensus/2d.asp` [Last accessed: 17 September 2019].

[40] AP Kassianos, JD Emery, P Murchie, and Fiona Mary Walter. Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review. *British Journal of Dermatology*, 172(6):1507–1518, 2015.

[41] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 2425–2433, 2015.