VIA MEDICA

PRACA ORYGINALNA

Forum Dermatologicum
2020, tom 6, nr 2, 40–42
DOI: 10.5603/FD.a2020.0004

ISSN 2451–1501

Can convolutional neural networks outperform clinicians in the detection of melanoma on dermoscopy images?

Dominika Kwiatkowska[1], Piotr Kluska[2], Adam Reich[1]

[1]*Department of Dermatology, Institute for Medical Sciencies, University of Rzeszow*
[2]*AXA XL Wroclaw*

ABSTRACT

Introduction: Artificial intelligence is widely used in various fields of medicine. It also has great potential for being used in the assessment of dermoscopy images.

This study aimed to evaluate whether a convolutional neural network model could match dermatologists' accuracy in the assessment of dermoscopic pictures.

Material and methods: For this research we used HAM10000 training dataset, that was extracted from "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection". All skin lesions were classified to one of the following group: (1) malignant melanoma, (2) melanocytic nevus, (3) basal cell carcinoma, (4) actinic keratosis/Bowen's disease, (5) benign keratosis, (6) dermatofibroma, and (7) vascular lesion. From the dataset, we have randomly extracted 104 images from all classes of lesions to create the online test presented to 14 dermatologists who were asked to classify each lesion out of 104 dermoscopic pictures to the groups mentioned above. Next, the ResNeXt model was evaluated on the same dataset.

Results: Dermatologists achieved better sensitivity than ResNeXt in malignant melanoma differentiation. However, precision and F1 score of ResNeXt were higher in comparison to dermatologists. Moreover, CNN was more precise and sensitive to other skin lesion types.

Conclusions: This research has shown that computer vision aided dermoscopy can be a supportive tool that could help physicians in the screening of patients for malignant melanoma.

Forum Derm. 2020; 6, 2: 40–42

Key words: articial intelligence, deep learning, dermoscopy, malignant melanoma

INTRODUCTION

Deep learning models can learn the complex representation of the data without manual extraction of features. Nowadays, artificial intelligence is widely used in many different fields. In medicine and healthcare deep learning models were successfully applied in the classification of MRI and X-ray imaging [1, 2], ocular imaging (diabetic retinopathy, age-related macular degeneration) [3, 4] or detecting cancer in pathology pictures [5]. It is also great potential for their use in dermatology, *e.g.* to analyze dermoscopy images [6]. Offering faster differentiation of malignant melanoma and other cutaneous malignancies from benign pigmented lesions, deep learning models could be of help in routine clinical practice. Recent studies have shown that state of the art deep learning models could even outperform dermatologists in malignant melanoma detection [7]. The authors' previously conducted research showed that deep learning models had achieved quite good precision in malignant melanoma detection within digitalized dermoscopy images [8]. This research aimed to evaluate whether the authors' best convolutional neural network model could match dermatologists' accuracy.

MATERIAL AND METHODS

Dataset

For this research, we used HAM10000 training dataset, that was extracted from "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" [9]. The authors of this dataset provided supplementary data about the origin of the lesions with a unique identifier. All skin lesions were classified either by histopathological examination, confocal microscopy, follow-up examination or experts' consensus. Within the dataset authors distinguished seven types of lesions: (1) malignant melanoma, (2) melanocytic nevus,

Adres do korespondencji:
Prof. Adam Reich, Department of Dermatology, Institute for Medical Sciencies, University of Rzeszow, Szopena 2, 35–055 Rzeszów, Poland
e-mail: adi_medicalis@go2.pl

(3) basal cell carcinoma, (4) actinic keratosis/Bowen's disease (intraepithelial carcinoma), (5) benign keratosis (solar lentigo/seborrheic keratosis/lichen planus-like keratosis), (6) dermatofibroma, and (7) vascular lesion (including angioma, angiokeratoma, pyogenic granuloma, microvenular haemangioma, angioma serpiginosum, port-wine stain, lymphangioma circumscriptum, targetoid haemosiderotic haemangioma, Kaposi's sarcoma, and angiosarcoma). From the dataset, we have randomly extracted 104 images from all classes of lesions to create the online anonymous test.

The test was presented to 14 dermatologists, with different level of clinical and dermoscopic expertise (they have on average 4.42 ± 3.98 years of professional experience), who were asked to classify each lesion out of 104 dermoscopic pictures to the groups mentioned above. All images were demonstrated on the computer screen in the same order to each participant. There was no time restriction for the assessment performed by the dermatologists. After providing the answer and before entering the next image, the physician was informed about the correct allocation of the lesion.

Next, the ResNeXt model which was found previously to be the most accurate in the differentiation of melanoma from other lesions on dermoscopic pictures [8] was evaluated on the same dataset and the results were compared to the mean scoring achieved by all dermatologists.

Model training and evaluation

In this study, we utilized ResNeXt convolutional neural network (CNN) trained on ImageNet dataset [10, 11]. However, we replaced the last output layer with randomly initialized fully connected layer with 7 output nodes with the softmax activation function. As a cost function, we selected weighted cross-entropy. The weights were calculated based on the inverse cardinality of class in the training dataset. Adam was selected as an optimization algorithm [12]. The model was trained up to 20 epochs during which standard data augmentation techniques were applied: random cropping, random rotation, and normalization. The training was stopped once we could observe overfitting on the validation dataset. To compare ResNeXt with dermatologists we have computed the following metrics on the test dataset: precision (1), sensitivity (2), F1 score (3), and specificity (4).

$$\text{Precision} = TP/(TP + FP) \quad (1)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (2)$$

$$F1 = 2 \times [1/(1/\text{Precision}) + 1/(1/\text{Sensitivity}) = \\ = 2TP/(2TP + FP + FN) \quad (3)$$

$$\text{Specificity} = TN/(TP + FN) \quad (4)$$

TP — true positives; TN — true negatives; FP — false positives; FN — false negatives

RESULTS

As seen in Table 1, dermatologists achieved better sensitivity than ResNeXt in malignant melanoma differentiation. However, precision and F1 score of ResNeXt are higher in comparison to dermatologists. Moreover, CNN was more precise and sensitive to other skin lesion types. These data may indicate that dermatologists took precautions actions — classifying benign lesions as malignant — during the examination.

DISCUSSION

Dermoscopy is a non-invasive skin examination technique, which significantly improves the diagnosis of various skin lesions. Different structures visible in enlarged images provide valuable information for proper diagnosis. Unfortunately, this brilliant technique has also limitations. One of the biggest ones is the personal experience of physicians. Moreover, specialist dermatological advice is not widely available, thus, many times examination must be performed by general practitioners who may not be familiar with dermoscopy and skin tumour diagnosis. Undoubtedly, public health would benefit from malignant melanoma prevention and fast detection at early tumour stages. Having this in mind we may suggest, that machine learning could be a helpful tool during the daily clinical routine to help physicians to gather the supportive opinion about the possible diagnosis of pigmented skin lesions. We do believe, that even experienced dermatologists may benefit from the assistance of CNN during diagnosing suspicious melanocytic nevi, although it has to be mentioned, that CNN cannot replace a well-skilled physician.

A recent study performed by Tschandl et al. [13] has shown that state of the art deep learning models may outperform dermatologists. The study compared 139 deep learning models with 511 human readers (283 board-certified dermatologists, 118 dermatology residents, 83 general practitioners) from 63 countries. This group was also divided by years of experience, where 27 of the respondents were experts with more than 10 years of experience. The authors' best model published in the aforementioned study were better than the reader group [13]. Here, an improved version of the authors' CNNs is presented. Nevertheless, it should be noted that the settings of those studies do not reflect the environment of the real dermoscopy examination in which dermatologists can evaluate skin lesions at different levels of zoom and angles. Furthermore, they usually have access to a patient's clinical meta-data (medical history, age, sex, location), which further may facilitate the proper diagnosis. Also, routinely, the entire patient's skin is evaluated during such examination, which creates the opportunity to compare the pattern of different lesions

Table 1. ResNeXt and mean (± standard deviation) dermatologists performance metrics on a test set 104

Skin Lesion	Precision	Sensitivity	F1 score	Specificity
Dermatologists				
Malignant melanoma	0.54 ± 0.18	0.67 ± 0.23	0.57 ± 0.17	0.91 ± 0.06
Melanocytic nevus	0.92 ± 0.06	0.85 ± 0.1	0.88 ± 0.06	0.86 ± 0.12
Basal cell carcinoma	0.71 ± 0.21	0.77 ± 0.3	0.72 ± 0.24	0.98 ± 0.02
Actinic keratosis/Bowen's disease	0.62 ± 0.29	0.71 ± 0.3	0.64 ± 0.27	0.99 ± 0.01
Benign keratosis	0.7 ± 0.16	0.66 ± 0.22	0.67 + 0.18	0.96 ± 0.02
Dermatofibroma	0.64 ± 0.29	0.86 ± 0.29	0.72 ± 0.28	0.99 ± 0.01
Vascular lesion	0.78 ± 0.32	0.68 ± 0.31	0.7 ± 0.28	1.0 ± 0.01
ResNeXt				
Malignant melanoma	0.86	0.5	0.63	0.99
Melanocytic nevus	0.97	0.99	0.98	0.95
Basal cell carcinoma	0.86	1.00	0.92	0.99
Actinic keratosis/Bowen's disease	1.00	1.00	1.00	1.00
Benign keratosis	0.73	0.92	0.81	0.96
Dermatofibroma	1.00	1.00	1.00	1.00
Vascular lesion	1.00	1.00	1.00	1.00

and distinguish "the ugly duckling" from other, normally looking, nevi [14].

Despite finding the results of interest, the authors' study may contain some limitations including a small number of participating dermatologists and a relatively short period of professional experience. Furthermore, the participants of the study also demonstrate differences in the level of dermoscopy training. In future research, one should evaluate CNNs in comparison to a more homogeneous group, with the same level of experience.

CONCLUSIONS

To summarize, this research has shown that computer vision aided dermoscopy can be a supportive tool that could help physicians in the screening of patients for malignant melanoma.

Acknowledgement

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. We would like also to thank the International Skin Imaging Collaboration: Melanoma Project (https://www.isic-archive.com) for publicly available images of skin lesions that were used in this research.

REFERENCES

1. Bacchi S, Zerner T, Dongas J, et al. Deep learning in the detection of high-grade glioma recurrence using multiple MRI sequences: A pilot study. Journal of Clinical Neuroscience. 2019; 70: 11–13, doi: 10.1016/j.jocn.2019.10.003.
2. Badgeley M, Zech J, Oakden-Rayner L, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. npj Digital Medicine. 2019; 2(1), doi: 10.1038/s41746-019-0105-1.
3. Arcadu F, Benmansour F, Maunz A, et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. npj Digital Medicine. 2019; 2(1), doi: 10.1038/s41746-019-0172-3.
4. Lee C, Baughman D, Lee A. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. Ophthalmology Retina. 2017; 1(4): 322–327, doi: 10.1016/j.oret.2016.12.009.
5. Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. Computational and Structural Biotechnology Journal. 2018; 16: 34–42, doi: 10.1016/j.csbj.2018.01.001.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553): 436–444, doi: 10.1038/nature14539.
7. Haenssle HA, Fink C, Schneiderbauer R, 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. 2018; 29(8): 1836–1842, doi: 10.1093/annonc/mdy166.
8. Kwiatkowska D, Kluska P, Reich A. Convolutional neural networks for the detection of malignant melanoma in dermoscopy images. Adv Dermatol Alergol 2020 – in press.
9. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018; 5(1), doi: 10.1038/sdata.2018.161.
10. Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009, doi: 10.1109/cvpr.2009.5206848.
11. Xie S, Girshick R, Dollar P, et al. Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, doi: 10.1109/cvpr.2017.634.
12. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv Prepr arXiv1412. 6980; 2014.
13. Tschandl P, Codella N, Akay B, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. The Lancet Oncology. 2019; 20(7): 938–947, doi: 10.1016/s1470-2045(19)30333-x.
14. Gaudy-Marqueste C, Wazaefi Y, Bruneu Y, et al. Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection. JAMA Dermatology. 2017; 153(4): 279, doi: 10.1001/jamadermatol.2016.5500.