Dermatology Practical & Conceptual

Assessment of a Smartphone-Based Neural Network Application for the Risk Assessment of Skin Lesions under Real-World Conditions

Teresa Kränke[1], Philipp Efferl[1], Katharina Tripolt-Droschl[1], Rainer Hofmann-Wellenhof[1]

1 Department of Dermatology and Venereology, Medical University of Graz, Austria

Key words: Dermatology, Artificial intelligence, Convolutional neural networks, Skin lesion assessment, Skin cancer

Citation: Kränke T, Efferl P, Tripolt-Droschl K, Hofmann-Wellenhof R. Assessment of a Smartphone-Based Neural Network Application for the Risk Assessment of Skin Lesions under Real-World Conditions. *Dermatol Pract Concept*. 2025;15(3):5110. DOI: https://doi.org/10.5826/dpc.1503a5110

Accepted: February 5, 2025; **Published:** July 2025

Funding: None.

Competing Interests: Teresa Kränke was employed by the medical device company (medaia GmbH, Graz, Austria) after completion of the publication. Philipp Efferl was employed by the medical device company (medaia GmbH, Graz, Austria) during the study. This does not alter our adherence to Dermatology Practical & Conceptual policies on sharing data and materials. The remaining authors have no conflict of interest to declare.

Authorship: All authors have contributed significantly to this publication.

Corresponding Author: Teresa Kränke, MD, Department of Dermatology and Venereology Medical University of Graz, Auenbruggerplatz 8, Graz, Austria. ORCID ID: 0000-0001-8406-6063. E-Mail: teresa.kraenke@medunigraz.at

ABSTRACT

Introduction: The diagnostic performance of convolutional neural networks (CNNs) in diagnosing different types of skin cancer has been quite promising. Mobile phone applications with integrated artificial intelligence (AI) are an understudied area.

Objective: We evaluated the risk assessment of the SkinScreener (Medaia GmbH, Graz, Austria) AI-based algorithm in comparison with an expert panel of three dermatologists.

Methods: In this retrospective single-center study at the Department of Dermatology and Venereology in Graz, Austria. Photographs of lesions were taken by the users' mobile phone cameras. The algorithm allocated them to three risk classes. Blinded to AI´s results, the images were evaluated by three dermatologists—our reference standard. A consensus was defined as at least a two-thirds majority.

Results: A total of 1,428 skin lesions were included. In 902 lesions (63.16%), there was full agreement, and in 441 lesions (30.88%) a two-thirds majority was reached. Eighty-five lesions (5.69%) had to be discussed in a joint review process. The tested algorithm reached a sensitivity of 76.9% (95% CI: 71.7%–81.5%) and a specificity of 80.9% (95% CI: 78.5%–83.2%). Overall accuracy results were 77.2%.

Conclusions: Our results indicate that the tested mobile phone algorithm is a valuable tool for the correct risk classification of various skin lesions. As expected, its performance is worse than in a professional setting. Nonetheless, the use of these applications on mobile phones should raise awareness of skin cancer and encourage users to deal more intensively with preventive measures. In light of our results, these applications are also reliable for use by non-professionals.

Introduction

Skin cancer represents a significant burden on public health concerns. Melanoma is reported to be among the most rapidly rising forms of cancer worldwide, with a continuous increase in incidence rates, particularly in fair-skinned people. Despite advancements in early detection, melanoma is still the leading cause of skin cancer-related deaths [1,2]. Furthermore, the incidence of keratinocyte carcinomas (KC) is expected to double in Germany in the next few years [3]. Methods for the early detection of any kind of skin cancer (total body photography, (sequential digital) dermoscopy, reflectance confocal microscopy) has enormously improved and are well established in daily practice [4,5]. However, these techniques are used in a professional setting and require a high level of expertise.

There is growing evidence for the supplementary use of AI-based applications (mainly based on convolutional neural networks – CNN) in the dermatological practice. Several studies have proven that these CNNs are reliable tools concerning melanoma recognition. These neural networks, however, were mainly tested in the evaluation of high-quality images and focused on the detection of melanoma [6-15].

In the recent past, AI-based applications on mobile phones have become popular. The ubiquitous availability of mobile phones (about 6.3 billion users in 2021 worldwide) combined with technical improvements (e.g., image quality) has led to this phenomenon. In analogy to the use of AI-based tools in a professional setting, these mobile phone applications are able to detect various benign and malignant skin lesions and to provide a risk classification. The majority of these algorithms allocate the lesions to one of the three risk levels (low risk, medium risk, and high risk) by using pattern recognition software and provide further recommendations [16-22]. Current data, however, show that the performance of these mobile phone applications strongly depends on image-associated factors like image quality or scratching artefacts, which may lead to over- and underdiagnoses [18].

Recently, the diagnostic accuracy of the MDR IIa certified CE medical device App SkinScreener (Medaia GmbH, Graz, Austria) has been evaluated in a professional setting [23]. The integrated AI-based algorithm achieved a sensitivity of 96.4% (95% CI: 93.94–98.85) and a specificity of 94.85% (95% CI: 92.46–97.23) under clinical testing conditions. The authors concluded that the algorithms used in the study may have a positive impact on the healthcare system and reduce unnecessary visits and histological examinations.

Objective

The aim of the study presented herein was to investigate the risk assessment and diagnostic accuracy of the SkinScreener AI-based algorithm concerning the risk classification of various skin lesions in comparison with an expert panel of three dermatologists. Notably, images for evaluation were taken by l non-professionals, providing insights into the app's performance under real-world conditions for the first time.

Methods

Study Design

This was a retrospective single-center study at the Department of Dermatology and Venereology in Graz, Austria. The study was approved by the local ethics committee (approval number: 34-070 ex21/22 1508-2021) and was performed in accordance with the Declaration of Helsinki.

Based on previous publications and the functionality of the app [23], we defined three risk classes (green=low, yellow=medium, red=high) indicating the respective risk of a lesion being malignant and the subsequent approach, as follows:

- **Green:** Benign, no action needed
- **Yellow:** Suspicious, timely dermatological examination needed
- **Red:** Highly suspicious, immediate dermatological examination needed

The algorithm's risk classification was stated as correct if it matched at least the two-thirds majority of the dermatologists. As only images without any further participant information were gathered, histopathological examination was not performed. Hence, we chose the consensus of the dermatologists as reference standard.

Study Population

All participants were users of the CE-certified smartphone app SkinScreener. By using this app, they agreed to the processing of their data and the privacy policy. The users also agreed that their data may be processed for research, development, and market monitoring purposes. Inclusion criteria were age 18 or over and a skin type I-IV according to the Fitzpatrick scale.

Exclusion criteria were age under 18 and skin types V and VI as well as several image-associated factors that could influence an accurate assessment (Table 1).

The participants took images of the respective lesion with their mobile phone cameras.

Data were fully anonymized, and every participant got an internal ID. No further information (e.g., sex, age, development/duration of the lesion, location, medical record, or family history) were recorded due to data privacy policy. The recruitment phase was from the 1 June 2021 to the 1 August 2021.

Table 1. Listing of the Image-Associated Exclusion Criteria.

The user has skin type V (dark brown) or VI (darkest brown) according to Fitzpatrick
The lesion has low visual contrast to the surrounding skin area
The lesion is surrounded or covered by hair
The skin is sunburned
The lesion has previously been traumatized (excised/biopsied)
The surrounding skin is not intact (e.g., open wounds, ulcers, bleeding, irritation)
The lesion is located on or next to anatomical structures ("special sites") such as ear, eye, genitals, hair, mouth, nails, nose, or nipples
The lesion is very close to scars or tattoos or areas partly or fully covered with opaque or glittering substances like make-up or any kind of skin cream
The lesion is on mucosal surfaces like lips, in the mouth, or the genital region
The lesion is in a skinfold
The lesion is not on human skin
The scanned region is even partially covered by clothes
The lesion is not captured in focus

Procedure and Reader Study

The images for the evaluation process were randomly selected by using a special program written in Python to reduce the primary dataset of 12,766 images to a secondary dataset of 1,567. This was followed by a manual review by the second author to further exclude lesions not matching the inclusion criteria and to avoid a selection bias; 139 lesions were excluded after this additional review, resulting in a final dataset of 1,428 lesions.

The lesions were primarily evaluated by the algorithm and allocated to a risk group (green, yellow, or red). The images analyzed by AI were provided as JPG files with associated XML file for coding the image ID, risk class, diagnosis, and percentage estimate of diagnosis. This information was not visible to the dermatologists. The lesions were afterwards presented to the three dermatologists (TK, KTD, RHW) for allocation to a risk group. The lesion evaluations were made consensually. A consensus was defined as a two-thirds majority.

In the following scenarios, the corresponding images were evaluated in a joint review process:

- Lesions that were assessed with all three risk ratings (low-medium-high)
- Two of the three dermatologists rated a lesion as low risk, and the third dermatologist as high risk
- Two of the three dermatologists rated a lesion differently, and the third dermatologist did not give a rating

Algorithm

This application is a class IIa CE-marked and MDR approved medical device that has been already placed on the market. An image data set of 19,576 anonymized images was used for training (18,384) and testing (1,192) labeled with one of 47 distinct subcategories. Training was not only performed with images from the hospital archives but also with images that were collected from users of the app. The algorithm presented in this study employs a 2-step method for image classification: First, probabilities (ranging from 0.0% to 100%) for each label are calculated, followed by a risk assessment based on these probabilities. The risk assessment is given to the user of the app. The diagnosis as well as the probabilities for the respective label are stored in the background, which the developers shared with us as part of the study design. Detailed information about the algorithm is given in the Supplementary Material.

Statistical Analysis

The calculations were performed by using IBM SPSS Statistics 28. To quantify the risk assessment, accuracy, sensitivity, and specificity were calculated. As sensitivity and specificity are binary classifications, the risk groups "yellow" and "red" were summarized ("non-benign") and compared with the risk group "green" ("benign"). This approach was chosen due to the fact that the risk groups "yellow" and "red" need further attention, independently of their risk level.

The exact Clopper-Pearson 95% confidence interval (CI) was used to calculate the corresponding confidence intervals. Furthermore, we evaluated inter-rater variability using Cohen's Kappa, a metric for measuring agreement between two raters categorizing items on a nominal scale ("yes" or "no"). To assess inter-rater reliability among dermatologists, we compared pairs of dermatologists, resulting in three distinct Cohen-Kappa coefficients. These coefficients demonstrate the consistency of risk-assessments among multiple experts.

Detailed information on statistical analyses and sample size calculations are given in the Supplementary Material.

Results

A total of 1,428 lesions was included. Due to European data protection regulations for medical devices, no statement could be made about the exact number of participants nor their age and sex distribution. In 902 lesions (63.16%) there was full agreement among all three dermatologists, and in 441 lesions (30.88%) a two-thirds majority was reached. Following our definition, a consensus was found in 1,343 lesions (94.04%). Eighty-five lesions (5.69%) had to be discussed in a joint-review process as either no two-thirds majority was achieved or lesions were within under above-mentioned rules.

Risk Assessment and Diagnoses Made by the Algorithm

Seventy-eight lesions (5.46%) were allocated to the high-risk group, 372 lesions (26.06%) to the medium-risk group, and 978 lesion (68.48%) to the low-risk group.

Risk-Assessment and Diagnoses Made by the Dermatologists

The allocation of the lesions to the three risk classes done by the three dermatologists is shown in Figure 1.

Joint Review Process

The 85 lesions (5.69%) in the joint review process scattered as follows: 49 lesions (57.65%) were assessed with all three risk ratings (low-medium-high), 32 lesions (37.65%) were rated as low risk by two dermatologists and as being high risk by the third dermatologist, four lesions were rated differently by two dermatologists, and the third dermatologist did not give a rating.

During this process, 10 lesions were assessed as high risk, 19 as medium risk, and 56 as low risk. Figure 2 illustrates the lesion allocation to the three risk-classes before and after the joint review process.

Accuracy of the Algorithm

After the joint review process, statistical analyses of the classification groups low risk ("benign") versus medium and high risk ("non-benign") showed a sensitivity of 76.9% (95% CI: 71.7%–81.5%) and a specificity of 80.9% (95% CI: 78.5%–83.2%). Overall accuracy resulted in 77.2%.

Review Process with the Senior Dermatologist

After the joint review process, an additional review process done solely by the senior dermatologist (RHW) not blinded to the algorithm's assessments was added. This approach was chosen in order to elucidate potential causes of AI's incorrect classifications.

The following lesions were selected:

- Lesions that were classified as being low risk by the algorithm, whereas the consensus opinion was medium or high risk
- Lesions that were assessed as being high risk by the algorithm, whereas the consensus opinion was low risk

Low Risk by AI versus Medium/High Risk by Consensus

Seventy-one lesions were classified as low risk by AI. After review by the senior dermatologist, 22 (30.98%) of these were confirmed as correctly being classified as low risk. These 22 lesions included nine black or red nevi, one congenital nevus, and 12 nevi without specific features.

High Risk by AI versus Low Risk by the Consensus

Twenty-eight lesions were classified as high risk by the algorithm. After review, 10 (35.71%) of these were confirmed as correctly being classified as high risk. The remaining 18 lesions mostly showed features like scratching artefacts (n=4)

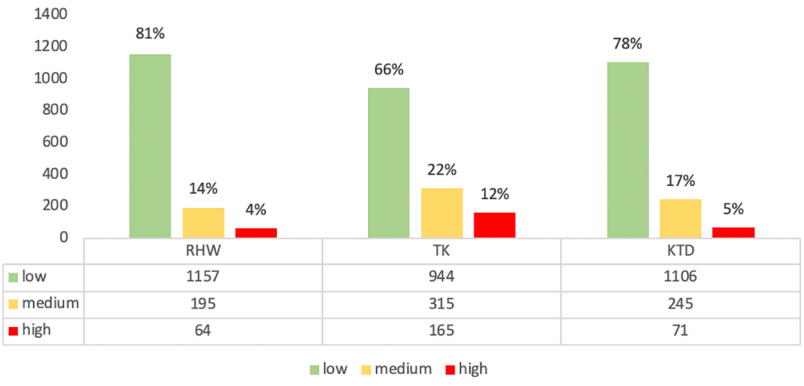

Figure 1. Bar diagram showing the allocation of the lesions to the respective risk classes done by the dermatologists.

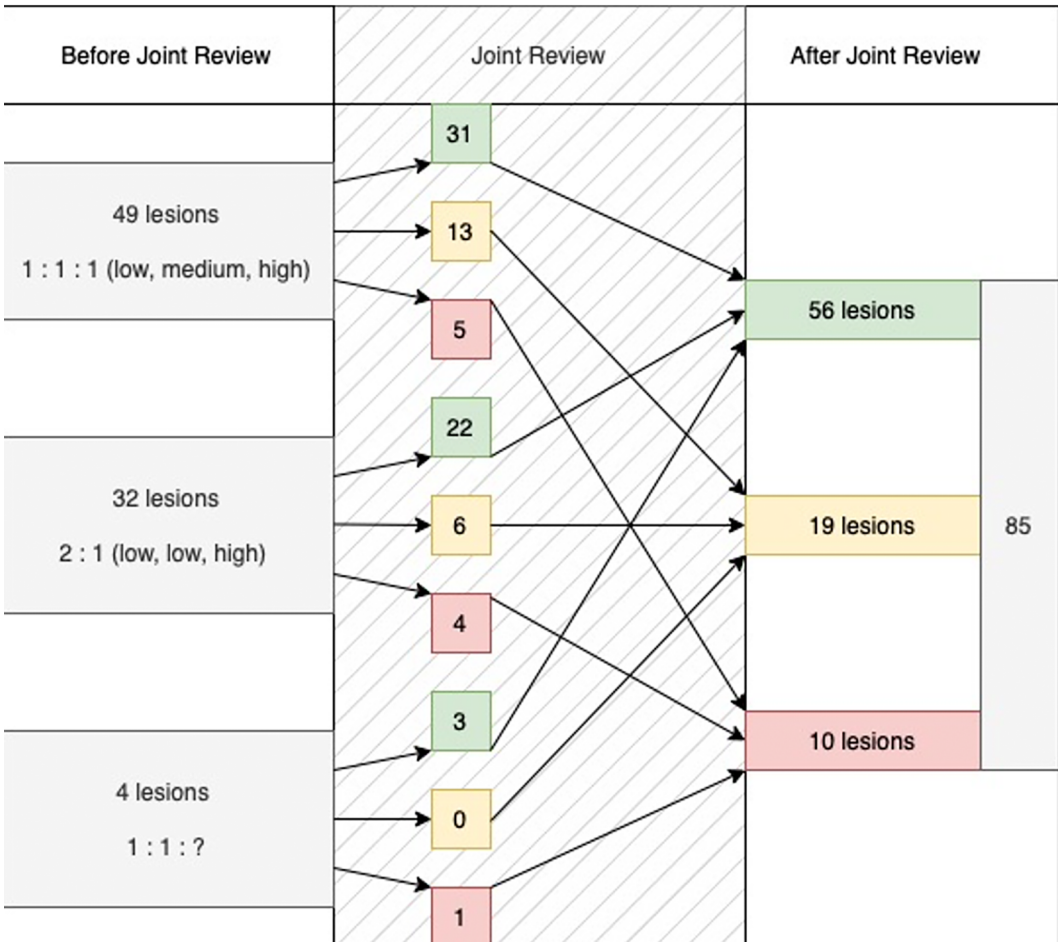

Figure 2. Graphical depiction of the lesion allocation to the three risk classes before and after the joint review process of the dermatologists.

or reddish parts (n=4). For the remaining lesions, no reason for the misclassification could be found.

Discussion

Our results indicate that the tested algorithm on a mobile phone application is a valuable tool for the correct risk classification of various skin lesions by non-professionals. The present study was conducted as a follow-up work of that by Kränke et al. [23] in terms of a post-market surveillance study. In this context it should be emphasized that the present study is the first to examine an AI-based algorithm in in a consumer application.

As expected, the sensitivity and specificity were lower, as in the prior study, which was performed in a professional setting. In detail, the algorithm showed a sensitivity of 76.9% (95% CI: 71.7%–81.5%) and a specificity of 80.9% (95% CI: 78.5%–83.2%) in the application by non-professionals, whereas both parameters were over 90% in a professional setting [23]. One explanation might be the image quality, as the images in the present study more often showed artefacts or were overexposed, leading to a significant decrease in sensitivity and specificity. Moreover, the algorithm´s evaluation and risk assessment in both studies were solely based on macroscopic images. However, numerous studies have shown the additional use of dermoscopy to be superior in correctly assessing benign and malignant skin lesions [16,24,25].

The reliance on macroscopic images represents a general limitation of these mobile phone applications. Dermoscopic close-up images, readily obtainable with magnification attachments for various mobile phones, could provide more detailed images, hence improving both sensitivity and specificity [26]. However, the implementation of these attachments would result in additional costs for the users and potential handling challenges. One development of the algorithm could be the automatic magnification of the uploaded images within the application. This would certainly increase the diagnostic accuracy. However, in addition to technical issues, further studies are needed in order to train the algorithm on these enlarged and detailed images.

One limitation of the prior study was the generalizability of the study population, as a highly selected population with mostly patients at high risk for developing any kind of skin cancer was chosen. Furthermore, the lesions examined were specifically selected by the dermatologists. The images in the

present study were randomly selected by the users, probably including a broader spectrum of cases including various age and risk groups. Although we do not have any demographic data, we assume that our current results are more generalizable. However, in our opinion, this application was primarily used by younger people, who showed little or no sign of chronic UV-damaged skin. The frequent occurrence of collision lesions and its poikilodermatous appearance hamper correct clinical assessment. If more users showed chronic UV-damaged skin, it could worsen the performance of the algorithm. Furthermore, the algorithm was solely trained and tested on lesions on Fitzpatrick skin types I-IV, underrepresenting individuals with darker skin (skin type V+VI). As no histopathological data were available due to the study design, we stated the consensus opinion of three dermatologists (a two-thirds majority was mandatory) as reference standard.

One study [27] described a combined reference standard of histology and clinical follow-up of benign lesions as reliably providing more generalizable results. In addition, a histopathological examination of every (including clearly benign) lesion is neither ethically justifiable, practical, nor cost-effective [28]. In the previous work [23], a clinical and dermoscopic digital follow-up of all non-excised "atypical nevi" was performed after six months in order to minimize this verification bias. Due to the study design, this procedure was not feasible in the current study.

The consensus of three dermatologists as reference standard provided valuable insights into possible future approaches. Following our definition, we reached a consensus in 94.04% cases, which is surely convincing. The lack of any histology is obviously a limitation. However, the initial review process, the following joint-review process, and the excellent agreement among the dermatologists make the chosen reference standard a robust one, in our opinion.

As dermoscopy is known as a bridge between the clinical and histopathological examination, it may be considered as a future addition to the reference standard in this kind of study [29].

Comparison of False Positives and False Negatives

As outlined above, 18 lesions were falsely classified by the algorithm as being high risk and 49 lesions as being low risk. Within the group of false positive-rated lesions, we identified user-induced irritations (e.g., scratching artefacts) and lesions with a high content of red color as the main reasons for the misclassification, meaning overrating by AI (Figure 3). These findings emphasize the need to refine AI models to better handle artifacts and different colors within a lesion (especially red). Furthermore, users should be made aware that scratched or traumatized lesions may lead to a false risk assessment. Considering the false negatives, some lesions were reclassified as benign upon reanalysis, while others were categorized as black or red nevi. The high pigmentation of these lesions likely contributed to the initial confusion among dermatologists. However, in the remaining cases, the reasons for the false negative rating by the algorithm are still unclear (see Figure 4).

The possible consequences of false positive and false negative ratings must be emphasized. False negative-assessed lesions can result in a delayed diagnosis and possibly lead to higher morbidity and mortality rates. False positive ratings worry the users and may result in unnecessary biopsies. As the primary goal is to "do no harm," these points need to be improved in further studies. In order to mitigate these mentioned risks, especially that of false-negative classifications, the following must be mentioned: Using this application does not replace regular follow-up visits to a dermatologist. Rather, it is intended as a supplementary tool for early skin cancer recognition and to encourage users to be more proactive with preventive measures (primary and secondary prophylaxis).

Conclusion

Our study shows promising results that AI-based mobile phone applications may be a good addition for the correct risk classification of various skin lesions. As expected, its performance is worse than in a professional setting, as indicated

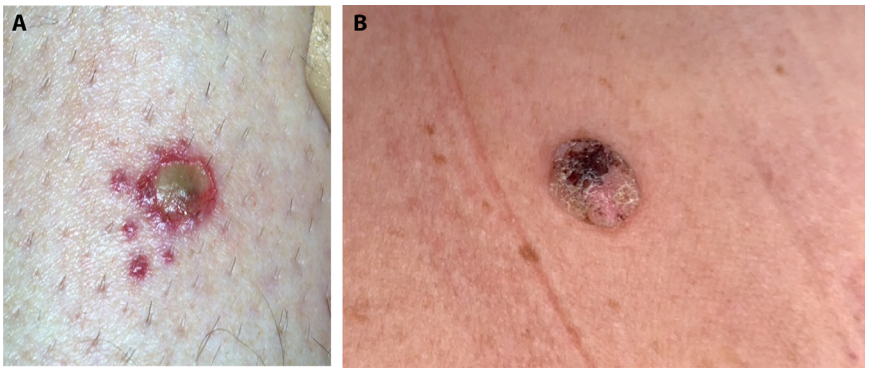

Figure 3. Two examples of falsely high-risk rated lesions by the algorithm due to a high content of (A) red color or (B) scratching artefacts.

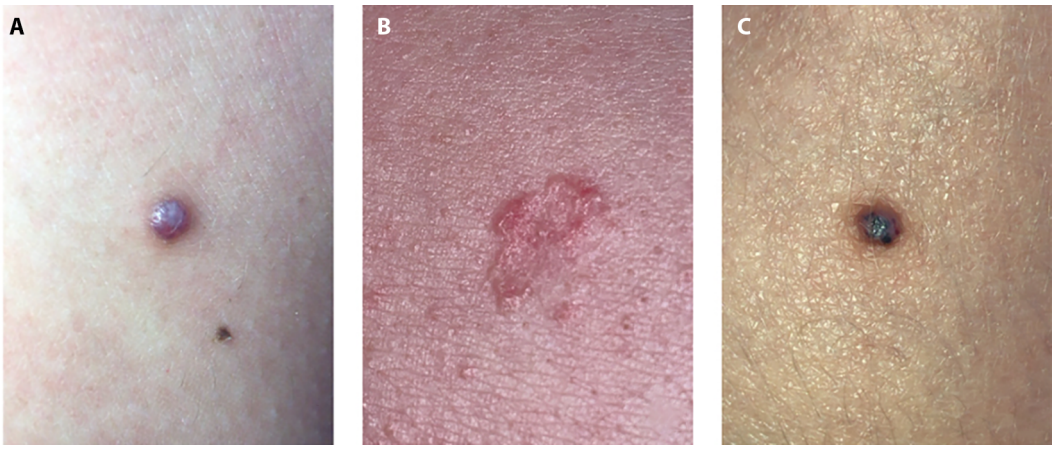

Figure 4. Three examples of false negative-assessed lesions by the algorithm.

by a lower sensitivity and specificity. Nonetheless, the use of these mobile phone applications should raise awareness of skin cancer and encourage users to be more proactive about prevention. Considering the algorithm's promising performance in our setting, it could be used as auxiliary tool in skin cancer recognition, especially in regions with a low density of dermatologists.

Due to our results and the permanent technical improvements of the algorithms, these applications can be reliably used by non-professionals.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin.* 2021 Jan;71(1):7–33. DOI: 10.3322/caac.21654. PMID: 33433946.
2. Nikolaou V, Stratigos AJ. Emerging trends in the epidemiology of melanoma. *Br J Dermatol.* 2014 Jan;170(1):11–9. DOI: 10.1111/bjd.12492. PMID: 23815297.
3. Hollestein L, Weinstock M, Le Roux E, Olsen C. More Than Many: How to Manage the Most Frequent Cancer? *J Invest Dermatol.* 2017 Sep;137(9):1823–6. DOI: 0.1016/j.jid.2017.06.017. PMID 28843292.
4. Petrie T, Samatham R, Witkowski AM, Esteva A, Leachman SA. Melanoma Early Detection: Big Data, Bigger Picture. *J Invest Dermatol.* 2019 Jan;139(1):25–30. DOI: 10.1016/j.jid.2018.06.187. PMID 30482597.
5. Dorrell DN, Strowd LC. Skin Cancer Detection Technology. *Dermatol Clin.* 2019 Oct;37(4):527–36. DOI: 10.1016/j.det.2019.05.010. PMID: 31466592.
6. Marka A, Carter JB, Toto E, Hassanpour S. Automated detection of nonmelanoma skin cancer using digital images: A systematic review. *BMC Med Imaging.* 2019;19(1):21. DOI: 10.1186/s12880-019-0307-7. PMID: 30819133.
7. Maron RC, Weichenthal M, Utikal JS et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer.* 2019;119:57-65. DOI: 10.1016/j.ejca.2019.06.013. PMID: 31419752.
8. Brinker TJ, Hekler A, Enk AH et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer.* 2019;111:148-154. DOI: 10.1016/j.ejca.2019.02.005. PMID: 30852421.
9. Brinker TJ, Hekler A, Enk AH et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer.* 2019;113:47-54. DOI: 10.1016/j.ejca.2019.04.001. PMID: 30981091.
10. Zhao XY, Wu X, Li FF et al. The Application of Deep Learning in the Risk Grading of Skin Tumors for Patients Using Clinical Images. *J Med Syst.* 2019;43(8):283. DOI: 10.1007/s10916-019-1414-2. PMID: 31300897.
11. Cui X, Wei R, Gong L et al. Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review. *J Am Acad Dermatol.* 2019;81(5):1176-1180. DOI: 10.1016/j.jaad.2019.06.042. PMID: 31255749.
12. Fujisawa Y, Otomo Y, Ogata Y et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol.* 2019;180(2):373-381. DOI: 10.1111/bjd.16924. PMID: 29953582.
13. Al-Masni MA, Kim DH, Kim TS. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Programs Biomed.* 2020;190:105351. DOI: 10.1016/j.cmpb.2020.105351. PMID: 32028084.
14. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *J Invest Dermatol.* 2018;138(7):1529-1538. DOI: 10.1016/j.jid.2018.01.028. PMID: 29428356.
15. Udrea A, Mitra GD, Costea D et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J Eur Acad Dermatol Venereol.* 2020;34(3):648-655. DOI: 10.1111/jdv.15935. PMID 31494983.
16. Liu Y, Jain A, Eng C et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med.* 2020 Jun;26(6):900–8. DOI: 10.1038/s41591-020-0842-3. PMID: 32424212.
17. Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017 Feb 2;542(7639):115–8. DOI: 10.1038/nature21056. PMID: 28117445.
18. Chao E, Meenan CK, Ferris LK. Smartphone-Based Applications for Skin Monitoring and Melanoma Detection. *Dermatol Clin.*

2017 Oct;35(4):551–7. DOI: 10.1016/j.det.2017.06.014. PMID: 28886812.
19. Wurm EM, Curchin CE, Soyer HP. Recent advances in diagnosing cutaneous melanomas. *F1000 Med Rep*. 2010;2:46. DOI: 10.3410/M2-46. PMID: 20948838.
20. Waweru AK, Ahmed K, Miao Y, Kawan P. Deep Learning in Skin Lesion Analysis Towards Cancer Detection. In: 2020 24th International Conference Information Visualisation (IV) [Internet]. Melbourne, Australia: IEEE; 2020 [cited 2023 Nov 29]. p. 740–5. https://ieeexplore.ieee.org/document/9373111/
21. Acharya P, Mathur M. Artificial intelligence in dermatology: the 'unsupervised' learning. *Br J Dermatol*. 2020 Jun;182(6): 1507–8. DOI: 10.1111/bjd.18933. PMID: 32030726.
22. Felmingham C, MacNamara S, Cranwell W et al. Improving Skin cancer Management with ARTificial Intelligence (SMARTI): protocol for a preintervention/postintervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a specialist dermatology setting. *BMJ Open*. 2022 Jan;12(1):e050203. DOI: 10.1136/bmjopen-2021-050203. PMID: 34983756.
23. Kränke T, Tripolt-Droschl K, Röd L et al. New AI-algorithms on smartphones to detect skin cancer in a clinical setting—A validation study. *PLOS ONE*. 2023 Feb 15;18(2):e0280670. DOI: 10.1371/journal.pone.0280670. PMID: 36791068.
24. Tschandl P, Rosendahl C, Akay BN et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol*. 2019 Jan 1;155(1):58. DOI: 10.1001/jamadermatol.2018.4378. PMID: 30484822.
25. Young AT, Vora NB, Cortez J et al. The role of technology in melanoma screening and diagnosis. *Pigment Cell Melanoma Res*. 2021 Mar;34(2):288–300. DOI: 10.1111/pcmr.12907. PMID: 32558281.
26. Jahn AS, Navarini AA, Cerminara SE, et al. Over-Detection of Melanoma-Suspect Lesions by a CE-Certified Smartphone App: Performance in Comparison to Dermatologists, 2D and 3D Convolutional Neural Networks in a Prospective Data Set of 1204 Pigmented Skin Lesions Involving Patients' Perception. *Cancers*. 2022 Aug 7;14(15):3829. DOI: 10.3390/cancers14153829. PMID: 35954491.
27. Freeman K, Dinnes J, Chuchu N et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ*. 2020 Feb 10; 368:m127. DOI: 10.1136/bmj.m127. PMID: 32041693.
28. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. *BMJ Evid-Based Med*. 2018 Apr;23(2):54–5. DOI: 10.1136/bmjebm-2018-110919. PMID 29595130.
29. Chen X, Lu Q, Chen C, Jiang G. Recent developments in dermoscopy for dermatology. *J Cosmet Dermatol*. 2021 Jun;20(6): 1611–7. DOI: 10.1111/jocd.13846. PMID: 33197276.