European Archives of Oto-Rhino-Laryngology
https://doi.org/10.1007/s00405-024-08951-z

HEAD AND NECK

The utility and reliability of a deep learning algorithm as a diagnosis support tool in head & neck non-melanoma skin malignancies

Alfonso Medela[1] · Alberto Sabater[1] · Ignacio Hernández Montilla[1] · Taig MacCarthy[2] · Andy Aguilar[2] · Carlos Miguel Chiesa-Estomba[3,4,5,6]

Received: 16 May 2024 / Accepted: 27 August 2024

Abstract

Objective The incidence of non-melanoma skin cancers, encompassing basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC), is on the rise globally and new methods to improve skin malignancy diagnosis are necessary. This study aims to assess the performance of a CE-certified medical device as a diagnosis support tool in a head & neck (H&N) outpatient clinic, specifically focusing on the classification of three key diagnostics: BCC, cSCC, and non-malignant lesions (such as Actinic Cheilitis, Actinic Keratosis, and Seborrheic Keratosis).

Methods a prospective, longitudinal, non-randomized study was designed to evaluate the performance of a deep learning-based method as a diagnosis tool in a group of patients referred to the head & neck clinic for suspicious skin lesions.

Results 135 patients were included, 92 (68.1%) were male and 43 (31.9%) were female. The median age was 71 years +/- 9 (Min: 56/Max: 91). Of those, 108 were malignant pathologies (54 basal cell carcinoma and 54 squamous cell carcinoma) and 27 benign pathologies (14 seborrheic keratoses, 2 actinic keratoses, and 11 actinic cheilitis). Of special significance is the remarkable performance of the medical device in identifying malignant lesions (basal cell carcinoma and squamous cell carcinoma) within the top-5 most likely diagnoses in above 90% of cases, underscoring its potential utility for early diagnosis and treatment.

Conclusion In this study, the effectiveness of deep learning methods, with a particular focus on vision transformers, as a diagnostic aid for H&N cutaneous non-melanoma skin cancers was demonstrated, highlighting its potential value for early detection and treatment of non-melanoma skin cancers. In this vein, further research is needed in the future to elucidate the role of this technology, because of its potential in the primary care clinic, dermatology, and head & neck surgery clinic as well as in patients with suspicious lesions, as a self-exploration tool.

Keywords Skin cancer · Deep learning · Basal cell carcinoma · Squamous cell carcinoma · CADx system

Introduction

The incidence of non-melanoma skin cancers, encompassing basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC), is on the rise globally, representing the predominant malignancy in Western nations. Annually, the United States alone records approximately 3,500,000 cases of non-melanoma skin cancer [1]. BCC ranks as the most prevalent skin cancer, with its incidence steadily increasing due to demographic aging and pervasive sun exposure, constituting 50% of all cancers in the United States [2]. cSCC follows as the second most common cutaneous malignancy, with an estimated incidence ranging from 15 to 35 cases per 100,000 inhabitants annually [3, 4], also associated with

✉ Carlos Miguel Chiesa-Estomba
chiesaestomba86@gmail.com

1 Medical Data Science, Legit.Health, Bilbao, Spain

2 Clinical Endpoint Innovation, Legit.Health, Bilbao, Spain

3 Department of Otorhinolaryngology-Head & Neck Surgery, Osakidetza, Donostia University Hospital, San Sebastian 20014, Spain

4 Bioguipuzkoa Health Research Institute, San Sebastian 20014, Spain

5 Faculty of Medicine, Deusto University, Bilbao, Spain

6 Servicio de Otorrinolaringología – Cirugía de Cabeza y Cuello, Hospital Universitario Donostia, Paseo Dr. Begiristain #1. CP, San Sebastian- Donosti 20014, España

Published online: 06 September 2024

chronic sun exposure, advancing age, and predisposing genodermatoses [3, 5].

While mortality rates from BCC and cSCC remain low, patients afflicted with BCC can endure significant morbidity, particularly if the tumor remains untreated over an extended period [6]. Moreover, high-risk subtypes of cSCC exist [7], elevating the likelihood of lymph node metastasis development and displaying a short disease-specific survival with local recurrence.

Traditionally, dermatologists and other medical specialists have relied on visual inspection or dermoscopy as the primary methods for diagnosing skin malignancies. However, recent advancements in machine learning technology, particularly in deep learning techniques, have spurred research across various medical domains.

Initial publications on dermatological computer-aided classification [8–10] lacked the generalization capacity of medical practitioners due to limited data and a focus on standardized tasks such as dermoscopy [11–13] or histological image classification [14–17]. However, Esteva et al. [18] in 2017 demonstrated the efficacy of deep learning in dermatology, applying artificial intelligence (AI) to both common skin conditions and specific cancers using a single convolutional neural network (CNN) trained on general skin lesion classification. Subsequent studies explored CNN applications in melanoma skin cancer [19–21]. However, there remains a dearth of specific research on non-melanoma skin cancer.

This study aims to assess the performance of a CE-certified medical device as a diagnosis support tool in a head & neck (H&N) outpatient clinic, specifically focusing on the identification of three key diagnostics: BCC, cSCC, and non-malignant lesions (such as Actinic Cheilitis, Actinic Keratosis, and Seborrheic Keratosis).

Materials & methods

During the study period, two datasets were established. First, the clinically collected H&N dataset, which is used to compare and test the performance of the evaluated medical device in a real clinical scenario. Second, the MDD dataset, which allows the medical device's evaluation in images from diverse public and private sources.

Prospective study and H&N dataset

After obtaining approval from the Ethics Committee of our Centre, we conducted a prospective, longitudinal, non-randomized study between June 2021 and December 2021 comprising a group of patients referred to the H&N clinic for suspicious skin lesions. The study aimed to analyze the performance of an AI diagnosis support tool in a real clinical setting for identifying various skin malignancies. Additionally, it sought to evaluate the tool's ability to assist healthcare practitioners in improving the accuracy of clinical assessments, expediting clinical inspections, and optimizing clinical triage.

For this purpose, the study involved collecting images of BCC (54 images) and cSCC (54 images), along with images of benign skin pathologies commonly found in the H&N regions (11 images of actinic cheilitis, 2 images of actinic keratosis, and 14 images of seborrheic keratosis). All the included patient images underwent histological confirmation biopsy to analyze the AI-driven diagnosis support tool's ability to discriminate between benign and malignant pathologies that often share visual characteristics. The H&N dataset used for this study comprises these images after undergoing manual cropping to the captured lesion and was managed exclusively by the senior author.

The inclusion criteria for the study comprised patients aged 18 and above with clinically and histologically confirmed skin diseases relevant to this study. Patients were excluded if they did not have histological confirmation, if they underwent non-surgical treatment, had an immunosuppressive status, or if the final histology did not correspond to those considered in the study. All research activities adhered to relevant guidelines and regulations. Informed consent to use clinical data for research purposes was obtained from all patients prior to surgical procedures. The images were randomly selected from daily clinical practice to reflect a real clinical scenario as accurately as possible. Pictures were captured using an Apple iPhone 13© (Cupertino, California, USA), with a dual camera of 12 megapixels in regular mode. All the pictures were taken at 10 cm with regular light from the clinic, without mobile phone zoom or flash. Trained pathologists performed histopathological reviews on all cases after surgical excision or biopsy. As a result of this study, we have built the H&N dataset with the images compiled from head and neck regions.

Medical device and MDD dataset

Images from this study have been analyzed by a CE-certified medical device (Legit.Health - Bilbao, Spain). This is a software-only medical device that leverages computer vision algorithms to process dermatology-related smartphone and dermatoscopy images. Its principal function is to provide a wide range of clinical data from the analyzed images to aid healthcare practitioners and organizations in their clinical decision-making process, thus enhancing the efficiency and accuracy of care delivery. Of special relevance for this study, the diagnosis-support functionality of the medical device was employed, which encompasses

a vision transformer [22] (ViT-base) trained end-to-end to distinguish 239 different skin diseases and families of skin pathologies. The medical device's dataset (MDD), implied in its training and evaluation consists of a compilation of 200,000 + images from both public and private dermatological sources, containing both clinical and dermatoscopic images. Some of these images were labeled by dermatologists and/or verified by pathologists, while others were biopsy-confirmed. This MDD dataset was split into train, validation, and test sets. The train and validation sets were involved in the deep learning training procedure, while the test set was used for the external validation of the learned models. This study includes the performance analysis of the medical device's diagnosis support tool when using the subset of this test dataset that consists of 3,821 images related to the 5 skin pathologies present in the H&N dataset: 18 images of actinic cheilitis, 728 images of actinic keratosis, 952 images of BCC, 1760 images of seborrheic keratosis, and 363 images of cSCC.

Evaluation metrics

Regarding evaluation metrics for the diagnosis support assessment, we include the top-k accuracy metrics (top-1, top-3, top-5 accuracy) and the top-5 class-wise sensitivity and specificity. Where TP and TN stand for the number of correctly identified positive and negative samples, and FP and TN refer to the number of wrongly identified positive and negative samples. In the case of the top-5 class-wise sensitivity and specificity, we consider TN the samples whose top-5 predictions that are above a minimum threshold of 0.01 do not include the class that is being evaluated. The use of a threshold allows for a better understanding of the estimated diagnosis by suppressing unlikely pathologies within the top-5.

$$Topk\ Accuracy\ =\ \frac{TP+TN}{TP+FP+TN+FN},$$

$$top5\ class\ sensitivity = \frac{TP}{TP\ +\ FN},$$

$$top5\ class\ specificity\ =\ \frac{TN}{TN\ +\ FP}$$

Regarding the evaluation metrics for the malignancy estimation assessment, we compute the ROC curve and the AUC-ROC score. The ROC curve illustrates the relationship between the true positive rate (TPR, or sensitivity) and the false positive rate (FPR, or 1-specificity) when varying the value of a malignancy threshold, and the AUC-ROC score represents the integration or Area Under the Curve (AUC) of the ROC curve. Additionally, we also compute the binary sensitivity and specificity of this malignancy estimation assessment.

$$TPR\ =\ \frac{TP}{TP\ +\ FN},$$

$$FPR\ =\ \frac{TP}{TN\ +\ FP}$$

Results

135 patients were included, 92 (68.1%) were male and 43 (31.9%) were female. The median age was 71 years +/- 9 (Min: 56/Max: 91). Of those, 108 were malignant pathologies (54 basal cell carcinoma and 54 squamous cell carcinoma) and 27 benign pathologies (14 seborrheic keratosis, 2 actinic keratosis, and 11 actinic cheilitis). See Fig. 1 to visualize the different lesions involved in the study.

Based on the diagnostic performance results presented in Table 1, we can observe that the medical device's diagnosis tool presents higher overall performance on the H&N than in the MDD dataset, especially for the top-3 and top-5 metrics. This discrepancy can be attributed to the nature of the datasets: the H&N dataset is meticulously curated, with images carefully captured and cropped to focus on lesions, whereas the MDD dataset comprises images compiled from diverse sources. Of special significance to this study is the remarkable performance of the medical device in identifying malignant lesions (basal cell carcinoma and squamous cell carcinoma) within the top-5 most likely diagnoses in above 90% of cases, underscoring its potential utility for early diagnosis and treatment. Similarly, benign lesions also exhibit exceptional performance, consistently being identified correctly within the top-5 most likely pathologies.

When considering top-5 class-wise sensitivity (see Table 2), the H&N dataset exhibits nearly perfect values, while for the MDD dataset, these values are lower. This H&N high performance highlights the consistent identification of all lesions within the top-5 most likely diagnoses. Differently, the top-5 class-wise specificity (see Table 2) of the H&N dataset is lower compared to that of the MDD dataset. This lower specificity arises because H&N images often share visual characteristics present in multiple pathologies. For instance, there are similarities between seborrheic keratosis and squamous cell carcinoma lesions, resulting in the simultaneous inclusion of both pathologies within the top-5 most likely diagnoses when analyzing these lesions. While this increases diagnosis sensitivity, it diminishes specificity.

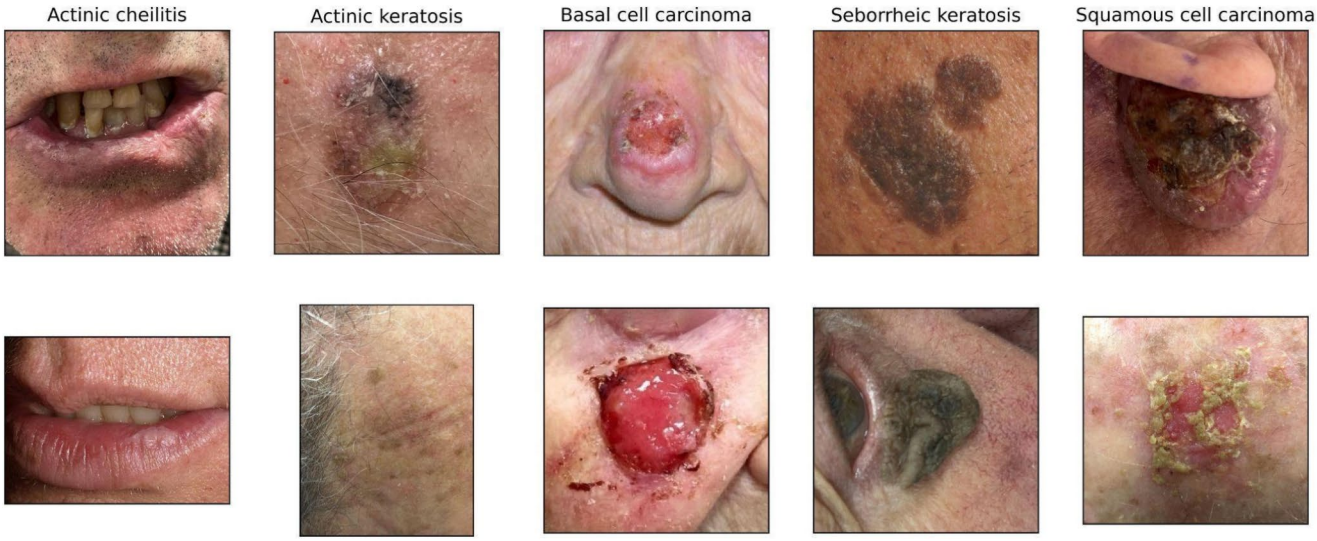

Fig. 1 Samples of the 5 different lesions

Table 1 Diagnosis performance on the H&N dataset

	H&N			MDD		
	Top-1 Accuracy	Top-3 Accuracy	Top-5 Accuracy	Top-1 Accuracy	Top-3 Accuracy	Top-5 Accuracy
Actinic cheilitis	72.7	100.0	100.0	77.8	94.4	94.4
Actinic keratosis	50.0	50.0	100.0	66.1	86.1	91.6
Basal cell carcinoma	77.8	94.4	98.1	72.0	90.4	94.0
Seborrheic keratosis	100.0	100.0	100.0	72.1	91.0	94.8
Squamous cell carcinoma	53.7	87.0	90.7	29.5	55.1	66.9
Bening	85.2	96.3	100.0	70.4	89.6	93.9
Malignant	65.7	90.7	94.4	60.2	80.7	86.5
Total	69.6	91.9	95.6	66.9	86.5	91.3

Table 2 Diagnosis top-5 sensitivity and specificity on the H&N dataset

	H&N		MDD	
	Top-5 Sensitivity	Top-5 Specificity	Top-5 Sensitivity	Top-5 Specificity
Actinic cheilitis	1.0	0.95	0.83	1.0
Actinic keratosis	1.0	0.76	0.85	0.94
Basal cell carcinoma	0.96	0.56	0.88	0.93
Seborrheic keratosis	1.0	0.91	0.89	0.89
Squamous cell carcinoma	0.85	0.58	0.54	0.98

Malignancy analysis

The results depicted in Figs. 2 and 3 illustrate that the medical device achieves an excellent balance between sensitivity and specificity across both the H&N and MDD datasets. The ROC and sensitivity/specificity curves provide a visual representation of the diagnostic tool's effectiveness in distinguishing between malignant and benign lesions. This effectiveness is further supported by the calculated AUC ROC scores of 0.93 for the H&N dataset and 0.92 for the MDD dataset, indicating strong inference capabilities and its suitability as a diagnostic-support tool.

Discussion

In our study, we showcase the effectiveness of deep learning methods, with a particular focus on vision transformers, as a diagnostic aid for H&N cutaneous non-melanoma skin cancers. The medical device evaluated demonstrated a high accuracy rate of over 90% in identifying malignant lesions like BCC and cSCC, highlighting its potential value for early detection and treatment of non-melanoma skin cancers.

Non-melanoma skin cancers typically manifest as slow-growing masses with a low likelihood of spreading regionally or distantly [23, 24]. However, they can lead to increased morbidity due to potential disfigurement, especially when located on the face or head. Beyond the physical implications, these cancers can also impact a patient's self-image and functional status following ablative surgery. Therefore,

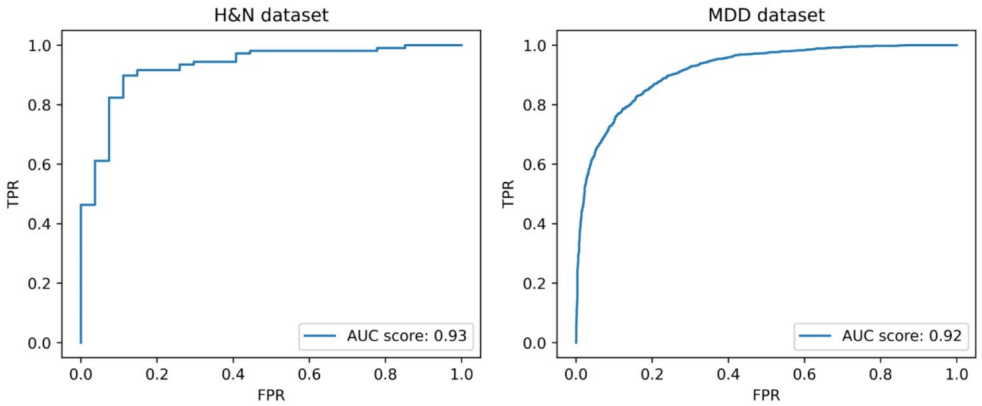

Fig. 2 AUC ROC evaluation for the malignancy estimation. Left: evaluation on the L&N dataset. Right: evaluation of the MDD dataset

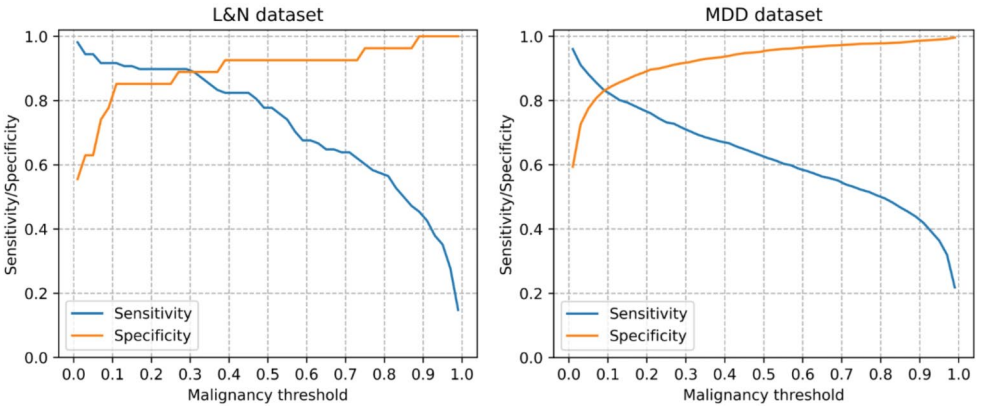

Fig. 3 Sensitivity and specificity curves. Left: evaluation on the L&N dataset. Right: evaluation of the MDD dataset

early detection is crucial not only to enhance the patient's quality of life but also to avoid unnecessary morbidity.

With the rising incidence of non-melanoma skin cancer, there's a growing demand for dependable methods to detect these malignancies earlier. This aims to enhance diagnostic accuracy and assist doctors in timely treatment interventions. The surge in interest in Artificial Intelligence (AI) has led to a wealth of data in recent years, emphasizing the significance of AI tools in treatment decision-making [25–27]. Additionally, the widespread adoption of telemedicine across various medical and surgical fields, accelerated by the COVID-19 pandemic, underscores the importance of virtual patient interactions [28]. This represents a new frontier in modern medicine, necessitating reliable methods to support diagnosis work-up and suggest potential treatment options.

Among various AI techniques, deep learning [29] stands out as a primary image analysis method, as we demonstrated, making the application of these techniques relevant for skin-malignancies screening and diagnosis. However, to reliably utilize deep learning technology for screening, the diagnostic sensitivity should be at least equivalent to that of trained experts in the field.

Moreover, considering the rapid advancements in mobile smartphone technology over recent decades, we now have access to powerful high-definition cameras with enhanced processing capabilities right in the hands of doctors and patients. During our study, the senior authors took all the clinical pictures with the same mobile phone, in the same light conditions, something that could positively influence the results. Altogether, this technological progress paves the way for the development of deep-learning-based smartphone applications for skin cancer screening. These apps can be utilized in outpatient clinics and for remote patient consultations, showcasing the vast potential of this technology.

Furthermore, there is ample evidence in current literature supporting the efficacy of these methods. Given the shortage of dermatologists or skin specialists in many countries, along with challenges in accessing specialized clinics or the physical distance faced by patients in rural areas from healthcare providers, where these smartphone-based solutions can be particularly beneficial. They offer a promising solution to bridge the gap in access to timely and accurate skin cancer screening and diagnosis [30].

In previous studies focusing on the use of deep learning for non-melanoma skin cancer diagnosis, Liu et al. published research on a deep learning system aimed at establishing a differential diagnosis for skin conditions. Their system demonstrated non-inferiority to specialists and outperformed non-specialists when analyzing various inputs like skin photographs, demographic information, and medical history. The authors achieved an accuracy rate of

66% in identifying the most probable diagnosis, which was comparable to dermatologists but higher than primary care providers (44%) and nurse practitioners (40%) [31]. Meanwhile, Ameri et al., Jaisakthi et al., and Serrano et al. proposed frameworks for classification and explored the role of dermoscopy combined with deep learning in detecting non-melanoma skin cancer [32–34]. By contrast, a quantitative review by Sharma et al. assessed the effectiveness of deep learning technology in diagnosing non-melanoma skin cancer. Their findings revealed no significant difference in sensitivity or specificity between dermatologists using dermoscopy and those utilizing machine learning techniques [35].

Another study published by Du-Harpur et al. explored how minor alterations to images could lead to errors in deep learning analysis. The researchers aimed to compare the accuracy of identifying malignant melanoma versus benign nevus. The authors recognized that even slight changes in color balance could significantly affect melanoma diagnosis, leading to a 235% increase in false negatives. Additionally, rotating images by 45º and 180º resulted in an 11% rise in false negatives, even though the images had been randomly rotated during the network's training. Interestingly, dermatologists showed no notable difference in accuracy when presented with these modified images compared to the original ones [36]. Analyzing the risk of image alternation, focusing on the misdiagnosis of melanoma, Winkler et al. in a study found that skin markings using blue surgical ink markers led to an increase in false positives of approximately 40% [37]. Something that we need to consider, when we obtain images for analysis or if we use pre-operative marked images for analysis.

In our study, we evaluated a CE-certified medical device that includes a ViT model trained to diagnose 239 different skin pathologies, including H&N non-melanoma skin cancer, utilizing a comprehensive dataset of over 200,000 images. Trying to improve our results, in comparison with previous studies, we took great care to standardize the image acquisition process across all cases, ensuring consistent camera settings and lighting conditions. Additionally, we avoided the use of pens on the skin and minimized image rotation to closely emulate real-world clinical scenarios. Our meticulous approach has yielded promising results, demonstrating a significant improvement in both sensitivity and specificity compared with previously published data. These outcomes strongly support our hypothesis regarding the potential utility of deep learning-based technology in enhancing the diagnosis and management of H&N non-melanoma skin cancer.

Overall, our study underscores the value of leveraging advanced machine learning techniques and standardized imaging protocols in dermatology. This research opens new avenues for improving diagnostic accuracy and patient care in the field of skin cancer detection.

Therefore, our findings showcase specialist-level sensitivity in the diagnosis of Non-Melanoma Skin Cancer (NMSC) using deep learning techniques, underscoring its potential as a valuable tool in screening efforts, helping to triage and diagnose cases more rapidly. However, when we check all the data available in the literature, this brings to light a common challenge associated with deep-learning-based models: their generalizability. This refers to the model's ability to perform effectively when analyzing new, previously unseen data. To mitigate this issue, it's crucial that training datasets are robust, diverse, large, and representative of the real-world scenarios in which the model will be deployed. Ensuring a comprehensive dataset can enhance the model's ability to generalize its findings to new cases accurately. Furthermore, it's imperative to subject these models to rigorous review and validation before their widespread clinical implementation. Identifying and addressing such weaknesses early on is essential to minimize the risk of potential harm to patients and to optimize the model's performance in clinical settings.

Finally, we need to highlight some limitations beyond the small sample size and the lack of comparison with human dermatologist. Upon visually inspecting the underperforming images, we found common issues like defocusing, which could affect the medical device's performance, as well as the need to test the performance of the CE-certified medical device with different mobile phones or cameras. Although the medical device achieves outstanding performance for both the diagnosis and malignancy evaluation, it is crucial to standardize and enhance the image-capturing process in clinical settings. This involves implementing robust protocols and tools for assessing image quality [38], which will greatly impact the results.

Conclusion

This study showcases the effectiveness of deep learning methods, with a particular focus on vision transformers, as a diagnostic aid for H&N cutaneous non-melanoma skin cancers. Results from this study highlight the potential value of the evaluated AI-driven medical device for early detection and treatment of non-melanoma skin cancers. In this vein, further research is needed to better elucidate the role of this technology, given its promising outcomes in real-world scenarios such as primary care clinics, dermatology, and head & neck surgery clinics, as well as in patients with suspicious lesions, where it serves as a self-exploration tool. Moreover, more research is required to explore the capabilities of these

AI-driven methods in diagnosing and evaluating further skin-related pathologies.

Acknowledgements We are grateful for the support provided by the doctors who filled out the questionnaire and made this study possible.

Author contributions A.M. and C.M.C conceived and conducted the experiment(s), A.S. analysed the experiment(s) results and wrote the manuscript. All authors reviewed the manuscript.

Funding None.

Data availability The majority of datasets used in this study are not publicly available due to specific institutional requirements governing privacy protection. Validation data for this study were made available to Editorial Board Members and referees at the time of submission for the purposes of evaluating the manuscript. The public datasets can be downloaded from different sources under author's consent.

Declarations

Competing interests The authors declare no competing interests.

Additional information Correspondence and requests for materials should be addressed to A.M.

References

1. Karia PS, Jambusaria-Pahlajani A, Harrington DP, Murphy GF, Qureshi AA, Schmults CD (2014) Evaluation of American Joint Committee on Cancer, International Union Against Cancer, and Brigham and women's Hospital tumor staging for cutaneous squamous cell carcinoma. J Clin Oncol 32(4):327
2. Apalla Z, Nashan D, Weller RB, Castellsagué X (2017) Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches. Dermatology Therapy 7:5–19
3. Rowe DE, Carroll RJ, Day CL Jr (1992) Prognostic factors for local recurrence, metastasis, and survival rates in squamous cell carcinoma of thse skin, ear, and lip: implications for treatment modality selection. J Am Acad Dermatol 26(6):976–990
4. Burton KA, Ashack KA, Khachemoune A (2016) Cutaneous squamous cell carcinoma: a review of high-risk and metastatic disease. Am J Clin Dermatol 17:491–450
5. Gray DT, Suman VJ, Su WD, Clay RP, Harmsen WS, Roenigk RK (1997) Trends in the population-based incidence of squamous cell carcinoma of the skin first diagnosed between 1984 and 1992. Arch Dermatol 133(6):735–740
6. Miller DL, Weinstock MA (1994) Nonmelanoma skin cancer in the United States: incidence. J Am Acad Dermatol 30(5):774–778
7. Motaparthi K, Kapil JP, Velazquez EF (2017) Cutaneous squamous cell carcinoma: review of the eighth edition of the American Joint Committee on cancer staging guidelines, prognostic factors, and histopathologic variants. Adv Anat Pathol 24(4):171–194
8. Neelapu R, Devi GL, Rao KS (2018) Deep learning based conventional neural network architecture for medical image classification. Traitement Du Signal 35(2):169
9. Rosado B, Menzies S, Harbauer A, Pehamberger H, Wolff K, Binder M, Kittler H (2003) Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. Arch Dermatol 139(3):361–367
10. Burroni M, Corona R, Dell'Eva G, Sera F, Bono R, Puddu P, Rubegni P (2004) Melanoma computer-aided diagnosis: reliability and feasibility study. Clin Cancer Res 10(6):1881–1886
11. Kittler H, Pehamberger H, Wolff K, Binder M J. T. I. O. (2002). Diagnostic accuracy of dermoscopy. Lancet Oncol, 3(3), 159–165
12. Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR (2015), October Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In International workshop on machine learning in medical imaging (pp. 118–126). Cham: Springer International Publishing
13. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1605.01397
14. Binder M, Kittler H, Seeber A, Steiner A, Pehamberger H, Wolff K (1998) Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network. Melanoma Res 8(3):261–266
15. Hoffmann K (1997) In: Altmeyer P, Stücker M (eds) Skin cancer and UV radiation. Springer, Berlin, pp 219–226
16. WH JR CLARK (1989) Model predicting survival in stage I melanoma based on tumor progression. J Natl Cancer Inst 81:1983–1904
17. Schindewolf T, Stolz W, Albert R, Abmayr W, Harms H (1993) Classification of melanocytic lesions with color and texture analysis using digital image processing. Anal Quant Cytol Histol 15(1):1–11
18. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
19. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Zalaudek I (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 29(8):1836–1842
20. Marchetti MA, Codella NC, Dusza SW, Gutman DA, Helba B, Kalloo A, International Skin Imaging Collaboration (2018) …. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. Journal of the American Academy of Dermatology, 78(2), 270–277
21. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Utikal JS (2019) Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer 119:11–17
22. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Houlsby N (2020) An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929
23. Lomas ALBJ, Leonardi-Bee J, Bath-Hextall FJBJ (2012) A systematic review of worldwide incidence of nonmelanoma skin cancer. Br J Dermatol 166(5):1069–1080
24. Samarasinghe V, Madan V (2012) Nonmelanoma skin cancer. J Cutan Aesthetic Surg 5(1):3–10
25. Loftus TJ, Shickel B, Ozrazgat-Baslanti T, Ren Y, Glicksberg BS, Cao J, Bihorac A (2022) Artificial intelligence-enabled decision support in nephrology. Nat Rev Nephrol 18(7):452–465
26. Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G, Xia H (2019) Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 25(3):433–438
27. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N (2021) Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 13:1–17

28. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF (2020) Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim care Community Health 11:2150132720980612
29. LeCun Y, Bengio Y, Hinton G (2015) Deep Learn Nat 521(7553):436–444
30. Asbeck SM, Imo BU, Okobi OE, Dorcé-Medard J (2023) The dermatologic care needs of a Rural Community in South Florida. Int J Environ Res Public Health 20(4):3071
31. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, Coz D (2020) A deep learning system for differential diagnosis of skin diseases. Nat Med 26(6):900–908
32. Ameri A (2020) A deep learning approach to skin cancer detection in dermoscopy images. J Biomedical Phys Eng 10(6):801
33. SM J, Aravindan PM, C., Appavu R (2023) Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimedia Tools Appl 82(10):15763–15778
34. Serrano C, Lazo M, Serrano A, Toledo-Pastrana T, Barros-Tornay R, Acha B (2022) Clinically inspired skin lesion classification through the detection of dermoscopic criteria for basal cell carcinoma. J Imaging 8(7):197
35. Sharma AN, Shwe S, Mesinkovska NA (2022) Current state of machine learning for non-melanoma skin cancer. Arch Dermatol Res 314(4):325–327
36. Du-Harpur X, Arthurs C, Ganier C, Woolf R, Laftah Z, Lakhan M, Lynch MD (2021) Clinically relevant vulnerabilities of deep machine learning systems for skin cancer diagnosis. J Invest Dermatol 141(4):916
37. Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Haenssle HA (2019) Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA Dermatology 155(10):1135–1141
38. Montilla IH, Carthy M, Aguilar T, A., Medela A (2023) Dermatology Image Quality Assessment (DIQA): Artificial intelligence to ensure the clinical utility of images for remote consultations and clinical trials. J Am Acad Dermatol 88(4):927–928