Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assessing RCM's utility and impact of multimodal AI algorithm for diagnosing equivocal basal cell carcinoma

Jilliana MONNIER
Assistance Publique Hôpitaux de Marseille
Rabah Iguernaissi
Aix-Marseille Université
Nicholas Kurtansky
Memorial Sloan Kettering Cancer Center
Konstantinos Liopyris
Andreas Sygros Hospital
Ucalene Harris
Memorial Sloan Kettering Cancer Center
Babar Rao
Rutgers New Jersey Medical School
Marco Spadafora
University of Modena and Reggio Emilia
Sara Bassoli
University of Modena and Reggio Emilia
Cristina Rodriguez-Garcia
Skin inspection
Nathalie De Carvalho
Universidade Federal do Estado do Rio de Janeiro
Genevieve Ho
The University of Sydney
Gene Rubenstein
Dermatology and Laser Centre of Studio City
Banu Farabi
Icahn School of Medicine at Mount Sinai
Nicola Lippolis
University of Modena and Reggio Emilia
Attiya Haroon
Rutgers New Jersey Medical School
Beatrice Martinez Zugaib Abdalla

Faculdade de Medicina do ABC

Julia Curtis
University of Utah

Lilia Correa
Rutgers New Jersey Medical School

Caterina Longo
University of Modena and Reggio Emilia

Giovanni Pellacani
Sapienza University of Rome

Pascale Guitera
The University of Sydney

Djamal Merad
Aix-Marseille Université

Manu Jain
jainm@mskcc.org

Memorial Sloan Kettering Cancer Center

Article

Keywords: Basal Cell Carcinoma mimickers, Multimodal AI model, Reflectance Confocal Microscopy, Dermoscopy, Human reader study

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11503503/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title: Assessing RCM's utility and impact of multimodal AI algorithm for diagnosing equivocal basal cell carcinoma

Jilliana Monnier[1,2,3,18] jilliana.monnier@ap-hm.fr ,
Rabah Iguernaissi[2] rabah.iguernaissi@univ-amu.fr ,
Nicholas Kurtansky[4] kurtansn@mskcc.org ,
Konstantinos Liopyris[4,5] konstantinosliopyris@gmail.com ,
Ucalene Harris[4] ucaleneh@gmail.com ,
Babar Rao[6,7] babarrao@gmail.com ,
Marco Spadafora[8] spadafora.marco91@gmail.com ,
Sara Bassoli[8] sarabassoli79@gmail.com ,
Cristina Rodriguez-Garcia[9] dr.rodriguez@skininspection.co.uk ,
Nathalie De Carvalho[16] dranathaliedecarvalho@unirio.br,
Genevieve Ho[10] gen.ho7@gmail.com ,
Gene Rubenstein[11] grmd@sbcglobal.net ,
Banu Farabi[12] banufarabi91@gmail.com ,
Nicola Lippolis[8] lippolisnicola@yahoo.it ,
Attiya Haroon[6] attiya.haroon9@gmail.com ,
Beatrice Martinez Zugaib Abdalla[4,15] bmzabdalla@gmail.com ,
Julia Curtis[13] julia.curtis@hsc.utah.edu ,
Lilia Correa[6] liliacorrea2@gmail.com ,
Caterina Longo[8] longo.caterina@gmail.com ,
Giovanni Pellacani[17] Pellacani.giovanni@gmail.com ,
Pascale Guitera[10, 14] pascale.guitera@melanoma.org.au
Djamal Merad[2+] djamal.merad@univ-amu.fr
Manu Jain[4,7+*] jainm@mskcc.org

[1] Centre de Recherche en Cancérologie de Marseille, Inserm UMR1068, CNRS UMR7258, Aix-Marseille Université, France
[2] Aix-Marseille Univ, Université de Marseille, CNRS UMR 7020, Laboratoire d'Informatique et des Systèmes, Marseille, France
[3] Dermatology and skin cancer department, La Timone hospital, AP-HM, Marseille, France
[4] Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, USA
[5] Department of Dermatology-Venereology, Faculty of Medicine, National and Apodyteria University of Athens, Andreas Sygros Hospital, 16121 Athens, Greece
[6] Dermatology Department, Rutgers University, Robert Wood Johnson Medical Center, NY, USA
[7] Dermatology Department, Weil Cornell Medicine, New York, USA
[8] Department of Dermatology, University of Modena and Reggio Emilia, Modena, Italy
[9] Skin inspection, London, United Kingdom
[10] Melanoma Institute Australia, Faculty of Medicine and Health, University of Sydney, Australia
[11] Dermatology and Laser Centre of Studio City, Los Angeles, California, USA
[12] Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[13] University of Utah School of Medicine, Department of Dermatology, Salt Lake City, UT 84132, USA
[14] Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital Sydney, Australia
[15] Dermatology Department, Faculdade de Medicina do ABC, Brazil
[16] Dermatology Department, Federal University of the State of Rio de Janeiro, Rio de Janeiro, Brazil
[17] Dermatology Department, La Sapienza University in Rome, Italy
[18] Automated melanoma screening center of Marseille, La Conception hospital, AP-HM, Marseille, France

+co-shared senior authors
*Corresponding author

Corresponding author:
Manu Jain
Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York.
jainm@mskcc.org

Funding information: P30 CA008748/CA/NCI NIH HHS/United States

Conflicts of Interest: None declared.

Ethics statement: The patients in this manuscript have given oral informed consent to the publication of their case details.

Data availability statement: The dermoscopic images used to develop the dermoscopic model were obtained from the publicly available International Skin Imaging Collaboration (ISIC) 2019 dataset. Reflectance confocal microscopy (RCM) images were newly acquired at Memorial Sloan Kettering Cancer Center (MSKCC) during this study and were used for the multi-reader study and RCM model development. The RCM data and other materials supporting the findings of this study are available from the corresponding author upon reasonable request.

Manuscript word count: 4207 words/4500 words [excluding, abstract, methods, references, figures, tables]

Abstract word count: 150/150 words
References: 23
Figures: 4
Supplementary figures: 6
Tables: 3
Supplementary tables: 3

Keywords: Basal Cell Carcinoma mimickers, Multimodal AI model, Reflectance Confocal Microscopy, Dermoscopy, Human reader study

Abstract

Dermoscopy and reflectance confocal microscopy (RCM) are adjunct tools for diagnosing lesions equivocal for basal cell carcinoma (BCC). We performed a multi-reader study containing 249 equivocal lesions; diagnostic predictions and confidence scores were first evaluated using dermoscopy-alone and then again after adding RCM. We also built a fusion dermoscopy-RCM AI model and compared its accuracy with humans. RCM did not substantially improve diagnostic accuracy over dermoscopy (76.9%, 75.1% sensitivity; 53.5%, 53.7% specificity, respectively); however, confidence levels significantly rose from 20% to 48%. The fusion AI model yielded a high AUC of 0.83. It had similar sensitivity but a higher specificity (72.6%), especially compared to novices. Thus, RCM's real clinical value is in increasing clinician's confidence, which could aid management at the bedside. Furthermore, as RCM images are challenging to read by novices, a multimodal AI model could be of great assistance, leading to a wider integration of RCM in clinics.

INTRODUCTION

Basal cell carcinoma (BCC) is the most common skin cancer worldwide[1]. When diagnosed early, it can be treated surgically or non-surgically (imiquimod, photodynamic therapy or cryotherapy), especially for non-aggressive superficial BCC subtypes. Current diagnosis relies on visual examination, followed by dermoscopy for challenging lesions. Although dermoscopy has a high sensitivity and specificity (91.3% and 99%) for pigmented BCC, these are lower (84.3% and 73.2%) for non-pigmented lesions due to limited dermoscopic features, leading to unnecessary benign biopsies for histopathology confirmation[2]. Biopsy can cause bleeding, pain, scarring, and infection, making it undesirable for facial or lesions on sensitive sites and for patients with multiple lesions. Although histopathology is the gold-standard for diagnosis, it requires time consuming tissue processing, delaying management by weeks and even months.

Reflectance confocal microscopy (RCM) can noninvasively diagnose skin cancers, including BCCs, at cellular resolution and reduces the rate of benign biopsies[3, 4]. In a meta-analysis[5] 15 studies over 163 lesions show a high sensitivity (92%) and specificity (93%) for diagnosing BCC on RCM. However, this meta-analysis shows these studies have a high to uncertain risk of bias related to the inclusion of pigmented and non-equivocal BCC lesions, and inclusion of experts' interpretation[6]. Similar to dermoscopy, non-pigmented lesions remains challenging to diagnose on RCM due to lack melanin and appears as dark silhouettes[7, 8]. Furthermore, these studies do not reflect clinical workflow where a final diagnosis and management of a lesion relies on a sequentially integrated approach including clinical, dermoscopy, and RCM diagnosis. Recently, Longo et al showed that adding RCM to dermoscopy improved BCC diagnosis in suspicious lesions, increasing sensitivity and specificity from 93.2% and 51.7% to 97.8% and 86.8%, respectively[9]. However, RCM's added value remains unclear for equivocal lesions, which often undergo RCM evaluation in clinics. RCM acquired CPT billing codes in the US and is being used in clinics. However, grayscale and *enface* nature of the images requires an extensive training and a long learning curve[8, 10] posing a major roadblock in its wider adaptation. Computer-aided diagnosis based on artificial intelligence (AI) have been separately developed for diagnosing BCC on dermoscopy[11] and RCM images[12, 13] and shown a high potential for aiding beginners and comparable diagnostic performance with experts. However, to simulate clinical workflow, a multimodal AI model that combines dermoscopy and RCM is needed.

The goal of this study was to mimic clinical scenario in accessing the added value of RCM over dermoscopy in lesions equivocal for BCCs and to build a multi-modal AI model to reflect the human interpretation. First, we performed a multi-reader study including fifteen RCM readers (4 experts and 11 novices) and evaluated and compared their diagnostic accuracy on a combined dermoscopy-RCM approach over dermoscopy-alone. Additionally, we developed, for the first time, a multimodal dermoscopy-RCM AI algorithm for BCC diagnosis and compared its performance with the human readers'.

RESULTS

Demographic characteristics

The study population included a total of 204 Caucasian patients. The male-to-female sex ratio was 1.01:1 (103 males/ 101 females). The mean age was 61 years (range, 19-90 years). The patient sample was geographically diverse, including 136 patients from the United States; 58 from Europe (Italy) and 10 from Australia, (Table 3).

Lesion characteristics

A total of 249 clinically and dermoscopically challenging lesions equivocal for BCC were included. Majority, 128/249 (51%) of the lesions were papules; 105/249 (42%) were macules, nodules, patches, and plaques each accounted for 5/249 lesions (2%); and 1/249 (0.4%) was maculo-papular. Most lesions, 186/249 (75%), were non-pigmented. The rest of the lesions, 46/249 (18.5%), were pigmented and 17/249 (6.8%) were lightly pigmented. Lesions were predominantly located, 105/249 (42%), on the head and neck region, 95 lesions (38%) on the trunk, 26 lesions (10%) on the lower extremities, and 23 lesions (9%) on upper extremities. The medium diameter of the lesions was 6.46 mm [1; 100 mm], (Table 3).

Based on their anatomical location, 121 lesions were imaged using a hand-held RCM device (Vivascope 3000) and 128 lesions with a wide-probe device (Vivascope 1500). Some were imaged using both the devices. All the lesions were biopsied and had a histopathology diagnosis. We broadly classified lesions as BCC and not-BCC (benign or another malignant lesion such as squamous cell carcinoma and melanoma). Of 249 lesions, 144 (58%) were BCCs and 105/249 (42%) were not-BCC. For BCC, subtyping was available for 135/144 (93.75%). The most common subtypes were superficial (79/135; 58.5%) and nodular BCCs (70/135; 52%), and less common were infiltrative (16/135; 12%) and micro-nodular (6/135; 4.4%), (Table 3).

Multi-reader study

We performed the following evaluations to assess the utility of the combined RCM and dermoscopy multimodal approach over dermoscopy alone.

<u>Performance comparison of a multimodal approach over dermoscopy-alone for BCC diagnosis across all the readers:</u>

We evaluated the sensitivity and specificity for BCC diagnosis across all the readers, including experts and novices. Dermoscopy-alone and the multimodal approach had a similar sensitivity of 75.1% (95% CI: 71.4%-78.5%) and 76.9% (95% CI: 73.3%-80.1%); and a specificity of 53.7% (95%CI: 48.8%-58.5%) and 53.5% (95% CI: 48.6%-58.3%), respectively, for the diagnosis of BCC (Table 1). There was no significant improvement of a combined approach over dermoscopy-alone (*p-value*> 0.05).

<u>Performances comparison of a multimodal approach over dermoscopy-alone for BCC diagnosis between experts and non-experts</u>

When comparing dermoscopy-alone results separately for experts and non-experts, experts had a higher sensitivity of 77.6% (95% CI: 70.6%-83.4%) and a specificity of 59.5% (95% CI: 50.4%-68%) compared to 74.2% (95% CI: 69.7%-78.1%) and 51.4% (95% CI: 45.6%-56.4%) for non-experts. Likewise, the experts had a higher sensitivity of 80.1% (95% CI: 73.3%-85.6%) and 60.3% specificity (95% CI: 51.2%-68.8%) compared to 75.6% (95% CI: 71.2%-79.5%) and 50.7% (95% CI: 44.9%-56.4%) for non-experts with the multimodal approach.

For the expert readers, the multimodal approach resulted in an increased trend towards a higher sensitivity from 77.6% with dermoscopy-alone to 80.1%, without reaching a statistical significance (*p-value*>0.05). However, the specificity remained similar 60.3% with or without RCM (*p-value*>0.05), (Table 1, Table S1).

Sensitivity for BCC Detection and subtype classification across all readers (experts and non-experts)

We compared the sensitivity for BCC detection based on their histopathological subtype and sensitivity for correct BCC subtyping (superficial, nodular, micro-nodular, and infiltrative) between dermoscopy-alone and multimodal approach, (Table S2).

With dermoscopy-alone, highest sensitivity for BCC detection was for micro-nodular BCC (79.2%), followed by 78.7%. for nodular BCC. Slightly lower sensitivity of 76.1% was noted for superficial BCC and the lowest of 71.2% for infiltrative BCC. With addition of RCM (multi-modal approach), highest sensitivity of 81.9% was for nodular BCC, followed by 79.7% for infiltrative BCC, and 79.2% for micro-nodular BCC. The lowest sensitivity of 76.7% for superficial BCC. When assessing the utility of addition of RCM over dermoscopy-alone, we noticed a numerical increase in sensitivity for detecting nodular BCC from 78.7% with dermoscopy-alone to 81.9% with the multimodal approach. Similar trends were noticed for infiltrative BCC with improvement from 71.2% with dermoscopy-alone to 79.7% with multimodal approach. Despite these increasing trends, the addition of RCM did not result in a statistically significant increase in sensitivity for any BCC subtypes (*p-values* > 0.05), (Table S2).

When analyzing sensitivity of subtyping various BCCs, with the dermoscopy-alone approach, superficial BCC had the highest sensitivity of 49.8%, followed by nodular BCC (45.3%), and lower for infiltrative (27.1%) and 0% for micro-nodular BCC. With the multi-modal approach, highest subtyping sensitivity was for nodular BCC (49.1%), followed by 33% for superficial BCC and 28.8% for infiltrative BCC. Micro-nodular BCC had the least sensitivity. When assessing the utility of addition of RCM over dermoscopy-alone. We did not see any statistically significant improvement in subtyping nodular and infiltrative BCC with addition of RCM. However, the sensitivity for micro-nodular subtype showed a trend toward improvement with the addition of RCM compared with dermoscopy alone (16.7% versus 0%, respectively) *(p-value* = 0.062*)*, albeit based on a small sample size. On the contrary, subtyping sensitivity for sBCC decreased from 49.8% with dermoscopy-alone to 33% with the addition of RCM. This difference was statistically significant (*p-value*<0.001), (Table S2).

Confidence level

In addition to diagnostic accuracy assessments, we also evaluated confidence of readers for diagnosing BCC before and after addition of RCM over dermoscopy (Table 2). A statistically significant association between indicated confidence and imaging modality was observed across the overall sample, as well as among experts and nonexpert ratings, distinctly. There was an increase in confidence with the combined approach. The 100% confidence level was indicated 47.9% of the time with RCM as opposed to 19.5% with dermoscopy-only, and the impact was greater for non-experts (17.8% to 48.3%) than experts (23.8% to 46.9%).

Despite an overall increase in confidence level with the combined approach for all readers, we noticed some outliers. Confidence increased with RCM for all but 2 readers, one expert and one non-expert, for whom the percentage of decisions for which confidence increased was equivalent to the percentage of decisions for which confidence decreased. For most clinicians (7/11), confidence increased in more than 50% of their decisions. The detailed results of this change of confidence are illustrated in Figure 2 for all readers and the details about the improvements for individual lesions for each reader are illustrated in Figure S2. Figure S2 indicates, for each reader, the frequency table for

indicated confidence before and after considering RCM images. Most lesions are represented on the super-diagonal, indicating that confidence was higher after RCM as compared to dermoscopy-only.

We have also evaluated the relationship between improvements of confidence and improvement in terms of sensitivity and specificity for each reader. The results of this evaluation are shown in Figure S3 and Figure S4, respectively. The results obtained did not demonstrate an evident correlation between the reader-specific increasement in confidence of the diagnosis and the improvement in terms of diagnosis accuracy. These results may imply that clinicians become more confident about their initial diagnosis

Artificial intelligence model performances

In the second part of the study, we built three AI models - dermoscopy-alone, RCM-alone, and a multimodal fusion model combining dermoscopy and RCM to demonstrate the added value of incorporating RCM into a dermoscopy-based model. For this, several evaluations were conducted, starting with the dermoscopy model and then the multimodal model which involved fusing RCM and dermoscopy models.

AI Dermoscopy results

The AI dermoscopy model was tested on polarized dermoscopy images from the 249 lesions included in the multi-reader study. This model reached good performances with an AUC of 0.71 (ROC curve, Figure S1), a sensitivity of 71.3%, and a specificity of 61.3% (Table S3).

AI Multimodal results

Finally, to evaluate the contribution of the multimodal model over the dermoscopy-alone model, we fused the dermoscopy model prediction with RCM, lesion-level, prediction. This model achieved high performance reaching an AUC of 0.83 (ROC curve, Figure S1), a sensitivity of 77.6%, and a specificity of 72.6% (Table S3).

Comparing Multimodal results vs Dermoscopy-alone

We demonstrated the contribution of the multimodal model approach over AI dermoscopy model: the AUC with 0.12, the sensitivity with 6.5%, and the specificity with about 12% (Table S3).

Heatmaps results

We generated heatmaps to enhance the interpretability of the CNN predictions (14). The heatmaps highlighted the most relevant regions, within a given image on which, the AI model relied most heavily when for classifying the image as BCC or non-BCC. Figure 3 shows representative heatmaps of RCM images based on model predictions for true-positive (TP) (Images A), true-negative (TN) (Images B), false-positive (FP) (Images C), and false-negative (FN) (Images D) classifications for BCC.

RCM images (A) with a TP model prediction showed an excellent correspondence between the areas of high positive (BCC) attribution on the heatmap and the tumor nodules, which appeared as dark silhouettes on RCM. Similarly, RCM image (B) with a TN model prediction showed an overall excellent correspondence between the areas of high negative (not-BCC) attribution on the heatmap and the normal epidermis. However, within this image, there was an area of focal high-attribution located within a skin fold.

RCM images (C) represented a benign histologically diagnosed as lesion lichenoid keratosis and dermal scar, that was false predicted as BCC by the AI model. On heatmap, areas of high attribution were

observed within the dark shadows of the epidermis. Image D which had a FN prediction of BCC by the AI, was classified as a benign lesion. Interestingly, the heatmap showed areas of high attribution with in the hair follicles rather than within the BCC tumor nodules[12].

AI performance compared to human readers

AI dermoscopy model versus human results

The AI dermoscopy model, had a better specificity than non-expert readers (61.3% for AI vs 51.4% for non-experts) and equivalent to expert readers (61.3% for AI vs 59.5% for experts). However, its sensitivity of 71.3% was much lower than both categories of readers (75.6% for non-experts and 80.1% for experts) (Figure 4).

AI multimodal model versus human results

Similar to AI dermoscopy model, the multimodal fusion AI model had a higher specificity of 72.6% compared to both the expert (60.3%) and non-expert readers (50.7%). Unlike AI dermoscopy model, the fusion model achieved a slightly higher sensitivity of 77.6% compared to 75.6% for non-expert readers but with a slightly decreased sensitivity compared to 80.1% for expert readers. These performances are shown on Figure 4 illustrating the performance of the AI model and all readers.

DISCUSSION

RCM is being used in clinics as an adjunct to dermoscopy; however, the real clinical impact of adding RCM still needs to be explored. Furthermore, RCM images are challenging to read and requires a long learning curve—a major barrier in its widespread adaptation.

In this multi-center study, we performed a multi-reader analysis to assess the added value of RCM following dermoscopy triage of lesions equivocal for BCC amongst experts and novice readers. We also evaluated the impact of RCM addition on diagnostic confidence of the readers. Furthermore, for the first time, we developed a multimodal AI algorithm (combining dermoscopy and RCM) to mimic human reader bedside analysis and compared its performance with the readers. To be clinically relevant, we only included lesions equivocal for BCC, challenging to diagnose clinically and dermoscopically, as they usually undergo RCM analysis in clinics (Figure 1).

Our study showed that the addition of RCM lead to only a slight improvement in sensitivity (76.9%) over dermoscopy (75.1%); more for experts than for non-experts (2.5% vs 1.4%). However, specificity remained similar (53.5% after RCM vs 53.7% on dermoscopy) across all readers (60.3% for experts and 51.4% for non-experts) with both approaches. When assessing the utility of addition of RCM over dermoscopy-alone for detecting various subtypes of BCCs, we noticed a numerical increase in sensitivity for detecting nodular BCC (from 78.7% with dermoscopy-alone to 81.9% with the multimodal approach) and for infiltrative BCC with (71.2% with dermoscopy-alone to 79.7% with multimodal approach). But none of the results reached statistical significance (*p-values* > 0.05).

When analyzing sensitivity of subtyping various BCCs, with the dermoscopy-alone approach, superficial BCC had the highest sensitivity of 49.8%, followed by nodular BCC (45.3%), and lower for infiltrative (27.1%) and 0% for micro-nodular BCC. With the multi-modal approach, highest was for nodular BCC (49.1%), followed by 33% for superficial BCC and 28.8% for infiltrative BCC. Micro-nodular BCC had the least sensitivity.

When assessing the utility of addition of RCM over dermoscopy-alone for subtyping BCCs, for nodular and infiltrative BCC we saw slight improvement in sensitivity, but it was not statistically significant. However, the sensitivity for subtyping micro-nodular BCC showed a significant improvement *(p-value* = 0.062*)* with the addition of RCM (16.7% versus 0%, respectively). This observation is consistent with previous reports Jain et al[8] suggesting that superficial BCC may be particularly challenging to characterize by RCM. This is related to the *en face* nature of the RCM images where it is difficult to visualize the attachment of BCC cords and nodules with the overlying epidermis. Conversely, dermoscopy features of sBCC such as milky-pink color macule[15] and pigmented structures (leaf like structures) are easy to detect due to their superficial nature. Taken together, these findings suggest that the diagnostic value of RCM varies according to BCC subtype and may be particularly advantageous for the identification of deeper nodular or micronodular BCC, while offering less benefit for superficial lesions.

Longo et al[9] conducted a prospective study including dermoscopy and RCM. Compared to our study, they reported an overall higher sensitivity for both dermoscopy-alone approach (93.2%) and after addition of RCM (97.8%). However, similar to our study, they showed that addition of RCM led to only a slight improvement in sensitivity. They also reported lower specificity on dermoscopy-alone approach (51.7%) but a marked improvement in specificity (86.8%) with the addition of RCM. Overall lower results in our study could stem from the fact that we included only dermoscopically challenging lesions equivocal for BCC. Lower sensitivity could be due to majority of non-pigmented or lightly pigmented lesions, as they may be missed due to lack of defined patterns on dermoscopy and low or no melanin contrast on RCM. No subtyping analyzing was described on RCM. In fact, several studies

have shown that non-pigmented BCC appears as dark silhouette due to lack of melanin contrast on RCM[7, 8]. Lower specificity could be related to several factors including location of majority (76.2%) of the lesions in the head and neck region, where hair follicles could often be misinterpreted as BCC tumor nodules.

Although, RCM did not significantly improve diagnostic accuracy, it had a significant (*p-value*<0.001) impact on the diagnostic confidence across all the readers over dermoscopy from 19.5% to 47.9% (17.8% to 48.9% for non-experts; and 23.8% to 46.9% for experts). Increased confidence in diagnosis could aid in making appropriate management (surgical or non-surgical or follow-up) at the bedside, making RCM a valuable diagnostic tool in clinics.

For the first time, we developed a multimodal model AI model combining RCM and dermoscopy predictions to mimic clinical workflow. For this, we first built two separate CNN models one for dermoscopy and one for RCM and then fused their predictions. Dermoscopy model had an AUC of 0.71 (sensitivity 71.3%; specificity of 61.3%). Foahom et al[16] reported a dermoscopy model for BCC with a higher AUC (0.99) compared to our model. However, they used dermoscopy images of BCCs from the International Skin Imaging Collaboration (ISIC) 2018 dataset, which included BCC with obvious dermoscopy patterns. Lower AUC, in our study, could be due to the inclusion of only very difficult to diagnose BCC lesions on dermoscopy. In fact, due to the small number of lesions, we used ISIC 2019 dataset for training the model and kept the dermoscopy images from our study for testing only. This could have introduced differences related to image quality (device used) and may induce a bias about reproducibility, limiting generalization of AI models in real-world.

The RCM model, based on an EfficientNet-B0 architecture, demonstrated a moderate discriminative performance for BCC detection at the lesion level, achieving an AUC of 0.81 with a sensitivity ranging from 70.1% to 80%. This was lower than reported by our previous study by Campanella Gabriel et al.[12], which achieved an AUC of 0.883 using a convolutional neural network (CNN)-based algorithm for BCC detection in RCM images. This could be related to the differences in dataset in both studies as the previous study also included clinically obvious BCCs. Notably, while both approaches leverage deep learning for feature extraction from RCM images, the present model emphasizes computational efficiency through the use of a lightweight EfficientNet-B0 backbone, potentially offering advantages in scalability and deployment. Overall, these findings suggest that the RCM model achieves performance within the range of previously published CNN-based approaches while maintaining a balance between accuracy and model efficiency.

The multimodal model (at the lesion level) showed a higher accuracy than the dermoscopy AI model, with an AUC of 0.83 (sensitivity 77.6%; specificity of 72.6%) compared to an AUC of 0.71 (sensitivity 71.3%; specificity 61.3%) for the dermoscopy-alone model. Significant improvement in the results with the addition of RCM model could be related to the availability a larger number of RCM images (multiple images within a stack) in comparison to single dermoscopy image.

Lastly but more importantly, we compared the performance of dermoscopy-alone and multimodal AI models with the human readers. Both, dermoscopy and multimodal models, reached a better specificity (61.3%; 72.6%) compared to the non-expert readers (51.4%; 50.7%), demonstrating promising AI's ability to aid beginners during their practice and learning phase. Dermoscopy AI model had an equivalent specificity to expert readers (60.3%), but the multimodal AI model had better results (72.6%) compared to expert readers (60.3%). Improvement in specificity with the multimodal AI model was higher for both expert and non-experts could be related to better handling a large dataset of RCM with extensive information by AI than human. Higher specificity using multimodal approach will overcome the critical issue of false positive results on RCM and could reduce unnecessary benign

biopsies. However, sensitivity for both the models (71.3%; 77.6%) compared to experts (77.6%; 80.1%) and non-experts (74.2%; 75.6%) readers remained similar.

Our multimodal AI model could serve as a supportive tool to enhance human performance across different levels of expertise. For experts, it may serve as a decision-support tool by, streamlining complex analyses, reducing cognitive load, and providing second-opinion insights to improve accuracy and efficiency. For novices, it could provide structured guidance system and facilitate, learning through explainable AI tools, such as heatmaps, that highlight image regions influencing the model's predictions. In our study, for example, heatmaps showed high positive (BCC) attribution within tumor nodules that appeared as dark silhouette on RCM, a feature that can be diagnostically challenging for novice readers. In addition, interactive explanations could serve as an educational aid by allowing trainees to compare their own interpretations with the model's reasoning and expert-validated features. However, its role should be considered complementary rather than substitutive, with human oversight remaining essential to ensure appropriate interpretation and application of the model's outputs. This is exemplified in the heatmaps in our study, in which non-BCC regions, such as dark shadows in the epidermis, skin folds, and hair follicles, showed high positive (BCC) attribution. These regions could be further evaluated and appropriately interpreted by a trained human reader. For example, a hair follicle can be readily differentiated from a BCC tumor nodule by performing z-depth analysis, while a skin fold can be easily recognized on RCM based on its location and characteristic appearance on RCM.

In fact, some articles demonstrated that dermatologists' performance can be highly improved with AI models[17]. The main limitation of deep learning (CNN) is the lack of interpretation of the decision process and the conclusions, especially when these systems make some errors. Here, decision support tools such as GRAD-CAM heatmaps or occlusion maps[12] can guide the interpretation of RCM images, which could be beneficial to increase the confidence of BCC diagnosis. However, these tools are imprecise and needs further refinement, leaving decision predominantly based on AI prediction[18].

Limitations

This study has several limitations. First, its retrospective design may introduce selection and information biases, as the data were collected from existing records rather than through a controlled, prospective process. Additionally, the relatively small sample size reduces the statistical power of the analysis and may limit the generalizability of the results as the test dataset were from a specialized tertiary cancer center. Consequently, the findings should be interpreted with caution, and further studies with larger, prospective cohorts are warranted to confirm these results.

To conclude, these results demonstrated the possible use of AI models for BCC diagnosis assistant especially beginners regarding dermoscopy and RCM analyses, and to increase physician confidence to reduce the number of excisions of benign lesions. The improvement of confidence is a high mark of the value of RCM in this situation of very difficult to diagnose BCC.

METHODS ONLINE

Human reader study

Lesion selection and data collection:

This retrospective study involves participation of three tertiary health care centers including Memorial Sloan Kettering Cancer Center (MSKCC), New York; University of Modena and Reggio Emilia, Italy; and Melanoma institute, Australia. The primary study site was the Dermatology Department at MSKCC and conducted under an Institutional Review Boards (IRB # 17-078). All other participating sites transferred their data to MSKCC under a Data Transfer Agreement (DTA).

We included the lesions that had an equivocal diagnosis BCC on dermoscopy and had undergone a reflectance confocal microscopy (RCM) evaluation using either a wide-probe or handheld device; Vivascope 1500 or Vivascope 3000, (Caliber I.D., Rochester, NY) for confirmation. We only included lesions that were biopsied and had a histopathology diagnosis from MSKCC; University of Modena and Reggio Emilia; and from Sydney Melanoma Diagnostic Centre). For these lesions, we collected a complete set of images including at least one polarized dermoscopy image and a set of RCM images (i.e., a minimum of 1 mosaic/layer: stratum spinosum, basal layer, dermal-epidermal junction, and papillary dermis and at least one stack acquired). Clinical overview and close-up images were also included for each case for information, however the readers were asked not to record features of clinical images.

Selection of human readers:

We sent a survey to 15 RCM readers across the globe with at least one year of experience reading RCM images (Figure S6). A predetermined threshold for expertise was at least 10 years of self-reported experience, and the survey identified 4 experts and 11 novices.

Dataset compilation and sharing:

Each lesion was de-identified and assigned a randomized study number. Folders were created with these unique study numbers, each contained two subfolders: the first contained clinical overview, clinical close-up, and dermoscopy images and the second contained only RCM images. These folders were then uploaded on a data sharing software platform (BOX; company location) and were accessible to the readers.

Reader analysis:

Each reader was assigned a random set of cases to evaluate. All the lesions were seen by at least one expert to ensure that the entire dataset has an expert evaluation. The readers were directed to open the assigned study folders. They first opened the clinical/dermoscopy folder and analyzed clinical and dermoscopy images. They noted dermoscopy features for BCC (if any) and rendered a primary diagnosis based on the dermoscopic and clinical images. They also noted their level of diagnostic confidence level on a discrete ordinal scale (i.e., 0, 25%, 50%, 75%, and 100%). Next, the readers viewed the confocal images (stacks and mosaics) for the same lesion and noted any RCM features of BCC, if present. BCC terminology for RCM followed established terms in the literature[19] and dermoscopy (Figure S6). Readers provided their final diagnosis after reviewing confocal images. They also provided the impact of confocal images over clinical and dermoscopy alone and then graded their final diagnostic confidence level on the scale of 0 to 100.

All answers were recorded in an electronic survey (Google Forms) designed for the project (refer to google form below).

Artificial Intelligence development

We built two AI models to compare the utility of a combined RCM-dermoscopy approach over dermoscopy alone. The first model, referred to as "dermoscopy model" was based on dermoscopy images only, the second, referred to as a "RCM model" was based on RCM images only and then we developed a "multimodal model", which was based on a combined dermoscopy and RCM images.

AI Dermoscopy model

Data collections and annotation: we used polarized dermoscopy images for developing AI model. For this, we used 2 data sources: the International Skin Imaging Collaboration (ISIC) 2019 Challenge[20, 21, 22] training set was used for training and cases of the human reader study was used for testing. Training dataset includes 5373 dermoscopic images, which are split into 80% for training and 20% for validation. The dermoscopy images were classified at lesion level as "BCC" and "not-BCC" benign and malignant (see below).

Development algorithm: Training dataset includes 5373 dermoscopic images. 2178 dermoscopic images were of BCC, and not-BCC included 867 actinic keratosis, 500 benign keratosis, 200 dermatofibroma, 500 melanocytic nevus, 628 squamous cell carcinomas, 200 vascular lesion and 300 melanomas.

To develop this model, we used a pretrained EfficientNetb2[23] which was initially trained on ImageNet dataset. The classification layer was replaced by two fully-connected layers and a binary classification layer. We used Adam optimizer and 60 epoch to train the model. The backbone was frozen and only the new layers were trained. The images were resized for the EfficientNetB2 at 448x448 pixels. Then, we tested this model on an independent test dataset, which is the dermoscopy polarized images from the multi-reader study (249 lesions) 144 BCC and 105 not-BCC lesions.

RCM AI model

Dataset and annotation: RCM stacks (749) and 29 mosaics (transformed split into confocal images) and 22 lesions with confocal images from 249 lesions (used in the human reader study) were retrieved. Each image within a given stack was labeled as Basal cell carcinoma (BCC), Suspicious of Basal cell carcinoma, not-BCC (Squamous cell carcinoma, Melanoma and Benign (B), Normal (N), and Bad quality (BQ). A final stack-level label was also rendered. The labeling was performed by an expert RCM reader (MJ) who was aware of the histopathology diagnosis of these lesions. In case of uncertainty, another expert (PG) was consulted, and a consensus was made between the two readers. All images that were labeled as BQ were removed for the model development. In total, the model was trained on 35933 RCM images (n= 20,957 not-BCC and 14,946 BCC).

Development of the algorithm: We developed a RCM model using a pretrained EfficientNetb0[23] which was initially trained on ImageNet dataset. The classification layer was replaced by two fully-connected layers and a binary classification layer. We froze 60% of the parameters and used early stopping to control the convergence of the model. To train and evaluate this model we used a 5-fold cross validation procedure. We trained with 40 epochs, 40% finetuned parameters. The images were resized for the EfficientNetb0 to 448x448. This model gives the probability at image level from each stack of acquisition. These probabilities are then combined with a median to obtain the probabilities on the stacks and the probability of the full acquisition based on RCM imaging.

For each lesion, the half of images with less probability to be a BCC were weighted with 0, and the others half was used to calculate the probability on the lesion which is the median value. To choose the number of images weighted with 0 a grid search was used.

AI Multimodal model

Fusion of predictions of Dermoscopy and RCM models: To construct the multimodal AI model, we combined the RCM and the dermoscopy models developed above. We fused the predictions for each lesion to construct the multimodal model. This model gives the probability of each lesion based on dermoscopy and RCM lesion level probabilities. The fusion was done by a mean of the two probabilities (Figure S5).

Heatmaps generation

We generated heatmaps based on Grad-CAM[14] to gain insight into what features of an image are important for the CNN classification and to aid interpretation of the CNN predictions.

These visualizations rely on a color-coded attribution scale with areas color-coded in red strongly contributes to a BCC prediction area and those color-coded as blue corresponds to regions associated with benign lesions or healthy skin predictions.

Statistical analysis

Multi reader study:

We used McNemar's statistical test to compare the effect of RCM on human reader sensitivity and specificity (separately) in classifying equivocal BCC from BCC mimickers, using their paired predictions over dermoscopic/clinical images only and with the inclusion of RCM images. The significance of RCM on their indicated diagnostic confidence was evaluated using the Chi-Squared test. Level of significance was 0.05. This approach was applied to BCC overall as well as distinct BCC subtypes: superficial, nodular, micro-nodular, and infiltrative. We also analyzed changes in indicated confidence and changes in diagnostic accuracy with the inclusion of RCM, as well as their relationship to one another, using descriptive statistics.

AI vs HUMAN comparison methods.

We compared dermatologists' performances (experts and non-experts readers) with AI dermoscopy model and AI multimodal performances based on AUC, sensitivity and specificity, on the same test dataset.

For the evaluation of both readers and algorithms, we calculated the sensitivity and the specificity with confidence intervals at 95%.

Authors' contribution:

Conceptualization: MJ, JM, DM

Methodology: MJ, JM, DM, RI, NK

Data curation: MJ, KL, UH, RI

Formal analysis: NK, RI, JM, MJ

Multi-Readers study: BR, MS, SB, CR, ND, GH, GR, BF, NL, AH, BM, JC, LC, CL, GP, PG

Algorithms development: RI, DM

Writing – original draft: JM, MJ

Writing – review and editing: JM, MJ, NK, PG, RI

Supervision: MJ, DM

All authors reviewed and approved the final version of the manuscript.

REFERENCES

1. Wild CP, Weiderpass E, Stewart BW, éditeurs. World Cancer Report: Cancer research for cancer prevention [Internet]. Lyon (FR): International Agency for Research on Cancer; 2020 [cité 25 août 2026]. (IARC World Cancer Reports). Disponible sur: http://www.ncbi.nlm.nih.gov/books/NBK606505/ PubMed PMID: 39432694.

2. Reiter O, Mimouni I, Dusza S, Halpern AC, Leshem YA, Marghoob AA. Dermoscopic Features of Basal Cell Carcinoma and its Subtypes: A systematic Review. J Am Acad Dermatol. sept 2021;85(3):653-64. doi:10.1016/j.jaad.2019.11.008 PubMed PMID: 31706938; PubMed Central PMCID: PMC9366765.

3. Pastore LM, Zeidan R, Nazir Z, Manci R, Harris U, Marghoob AA, et al. Real time tele-reflectance confocal microscopy in clinical dermatologic practice: A prospective observational study. J Am Acad Dermatol. 1 sept 2024;91(3):537-9. doi:10.1016/j.jaad.2024.05.005 PubMed PMID: 38729432.

4. Pellacani G, Farnetani F, Ciardo S, Chester J, Kaleci S, Mazzoni L, et al. Effect of Reflectance Confocal Microscopy for Suspect Lesions on Diagnostic Accuracy in Melanoma. JAMA Dermatol. juill 2022;158(7):754-61. doi:10.1001/jamadermatol.2022.1570 PubMed PMID: 35648432; PubMed Central PMCID: PMC9161119.

5. Lupu M, Popa IM, Voiculescu VM, Caruntu A, Caruntu C. A Systematic Review and Meta-Analysis of the Accuracy of in Vivo Reflectance Confocal Microscopy for the Diagnosis of Primary Basal Cell Carcinoma. J Clin Med. 13 sept 2019;8(9):1462. doi:10.3390/jcm8091462 PubMed PMID: 31540342; PubMed Central PMCID: PMC6780971.

6. Dinnes J, Deeks JJ, Chuchu N, Saleh D, Bayliss SE, Takwoingi Y, et al. Reflectance confocal microscopy for diagnosing keratinocyte skin cancers in adults. Cochrane Database Syst Rev. 4 déc 2018;2018(12):CD013191. doi:10.1002/14651858.CD013191 PubMed PMID: 30521687; PubMed Central PMCID: PMC6516892.

7. Pogorzelska-Antkowiak A, Grzegorczyn Sł, Corneli P, Szepietowski JC. A Comparative Study of Pigmented and Non-pigmented Basal Cell Carcinoma in Reflectance Confocal Microscopy. In Vivo. 2021;35(1):423-7. doi:10.21873/invivo.12274 PubMed PMID: 33402492; PubMed Central PMCID: PMC7880731.

8. Jain M, Pulijal SV, Rajadhyaksha M, Halpern AC, Gonzalez S. Evaluation of Bedside Diagnostic Accuracy, Learning Curve, and Challenges for a Novice Reflectance Confocal Microscopy Reader for Skin Cancer Detection In Vivo. JAMA Dermatol. 1 août 2018;154(8):962-5. doi:10.1001/jamadermatol.2018.1668 PubMed PMID: 29998289; PubMed Central PMCID: PMC6143027.

9. Longo C, Guida S, Mirra M, Pampena R, Ciardo S, Bassoli S, et al. Dermatoscopy and reflectance confocal microscopy for basal cell carcinoma diagnosis and diagnosis prediction score: A prospective and multicenter study on 1005 lesions. J Am Acad Dermatol. mai 2024;90(5):994-1001. doi:10.1016/j.jaad.2024.01.035 PubMed PMID: 38296197.

10. Navarrete-Dechent C, Rajadhyaksha M, Marghoob AA, Nehal KS, Halpern AC, Rossi AM, et al. Continuing Medical Education-accredited reflectance confocal microscopy course in the USA and its impact on learning and clinical integration: report from a single centre. Clin Exp Dermatol. 1 déc 2023;49(4):408-9. doi:10.1093/ced/llad422 PubMed PMID: 38039138; PubMed Central PMCID: PMC10957219.

11. Liu H, Shang G, Shan Q. Deep Learning Algorithms in the Diagnosis of Basal Cell Carcinoma Using Dermatoscopy: Systematic Review and Meta-Analysis. J Med Internet Res. 3 oct 2025;27:e73541. doi:10.2196/73541 PubMed PMID: 41043135; PubMed Central PMCID: PMC12534767.

12. Campanella G, Navarrete-Dechent C, Liopyris K, Monnier J, Aleissa S, Minhas B, et al. Deep Learning for Basal Cell Carcinoma Detection for Reflectance Confocal Microscopy. J Invest Dermatol. janv 2022;142(1):97-103. doi:10.1016/j.jid.2021.06.015 PubMed PMID: 34265329; PubMed Central PMCID: PMC9338423.

13. Shavlokhova V, Vollmer M, Gholam P, Saravi B, Vollmer A, Hoffmann J, et al. Deep Learning on Basal Cell Carcinoma In Vivo Reflectance Confocal Microscopy Data. J Pers Med. 8 sept 2022;12(9):1471. doi:10.3390/jpm12091471 PubMed PMID: 36143256; PubMed Central PMCID: PMC9506260.

14. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. févr 2020;128(2):336-59. doi:10.1007/s11263-019-01228-7

15. Emiroglu N, Cengiz FP, Kemeriz F. The relation between dermoscopy and histopathology of basal cell carcinoma. An Bras Dermatol. 2015;90(3):351-6. doi:10.1590/abd1806-4841.20153446 PubMed PMID: 26131865; PubMed Central PMCID: PMC4516090.

16. Foahom Gouabou AC, Iguernaissi R, Damoiseaux JL, Moudafi A, Merad D. End-to-End Decoupled Training: A Robust Deep Learning Method for Long-Tailed Classification of Dermoscopic Images for Skin Lesion Classification. Electronics. janv 2022;11(20):3275. doi:10.3390/electronics11203275

17. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. août 2020;26(8):1229-34. doi:10.1038/s41591-020-0942-0 PubMed PMID: 32572267.

18. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. mai 2019;1(5):206-15. doi:10.1038/s42256-019-0048-x

19. Navarrete-Dechent C, DeRosa AP, Longo C, Liopyris K, Oliviero M, Rabinovitz H, et al. Reflectance confocal microscopy terminology glossary for nonmelanocytic skin lesions: A systematic review. J Am Acad Dermatol. mai 2019;80(5):1414-1427.e3. doi:10.1016/j.jaad.2018.12.007 PubMed PMID: 30529706; PubMed Central PMCID: PMC6830574.

20. Hernández-Pérez C, Combalia M, Podlipnik S, Codella NCF, Rotemberg V, Halpern AC, et al. BCN20000: Dermoscopic Lesions in the Wild. Sci Data. 17 juin 2024;11(1):641. doi:10.1038/s41597-024-03387-w PubMed PMID: 38886204; PubMed Central PMCID: PMC11183228.

21. Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) [Internet]. 2018 [cité 25 oct 2026]. p. 168-72. Disponible sur: https://ieeexplore.ieee.org/document/8363547 doi:10.1109/ISBI.2018.8363547

22. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 14 août 2018;5(1):180161. doi:10.1038/sdata.2018.161

23. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning [Internet]. PMLR; 2019 [cité 25 août 2026]. p. 6105-14. Disponible sur: https://proceedings.mlr.press/v97/tan19a.html

FIGURES

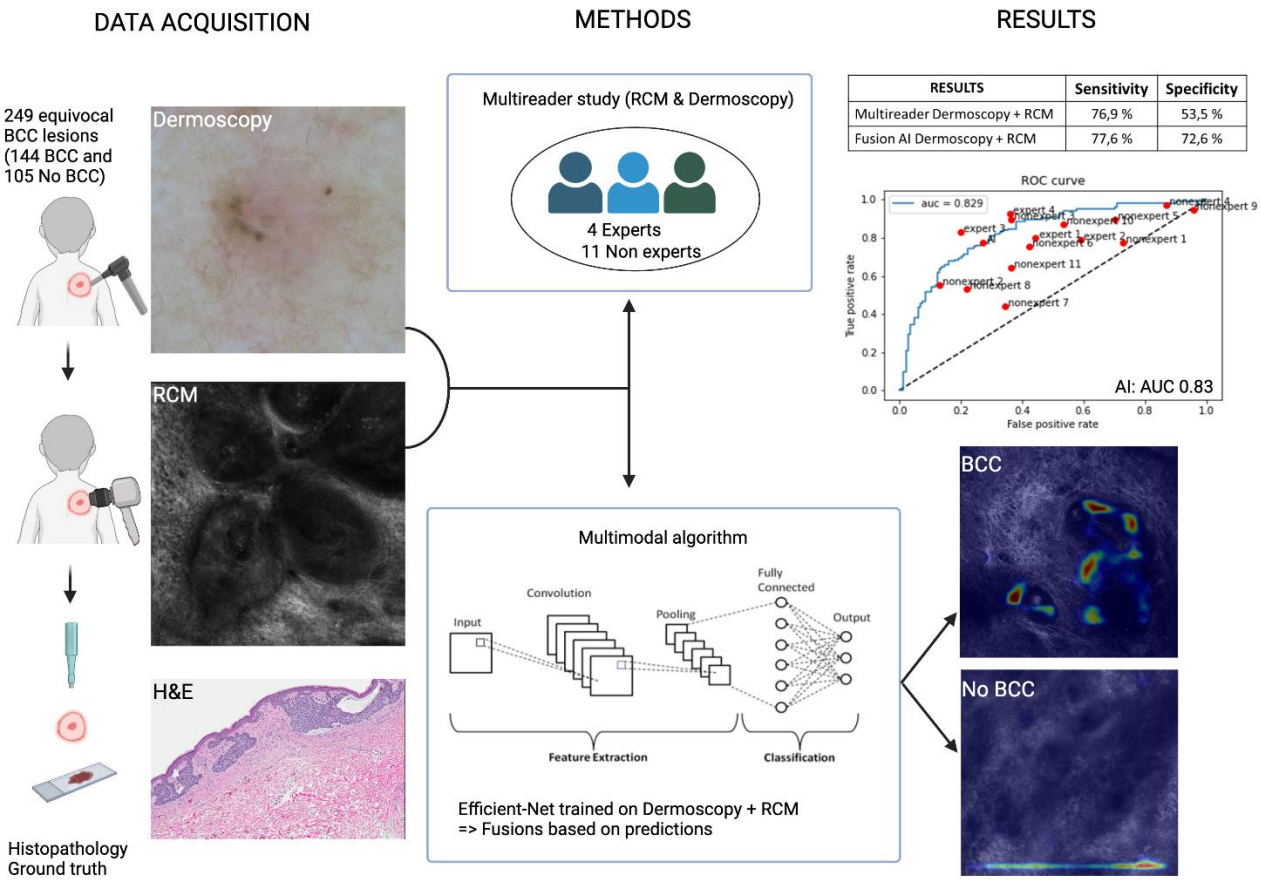


Figure 1: Graphical abstract

Graphical abstract represents an overview of data acquisition, methods, and results for the 249 lesions (144 BCC and 105 which were not a BCC) included in the study. Left column shows image acquisition process including a dermoscopic image and followed by a reflectance confocal microscopic image and histopathology as a ground-truth. Middle column represents the "Methods" including a multi reader study to assess the readers' performances on dermoscopy and the confocal, and development of an efficient-net AI fusion (dermoscopy + RCM) model. Right column shows the "Results" of the multi reader study and comparison with the AI fusion (dermoscopy + RCM) model. Heatmaps were generated to increase the interpretability of the prediction.

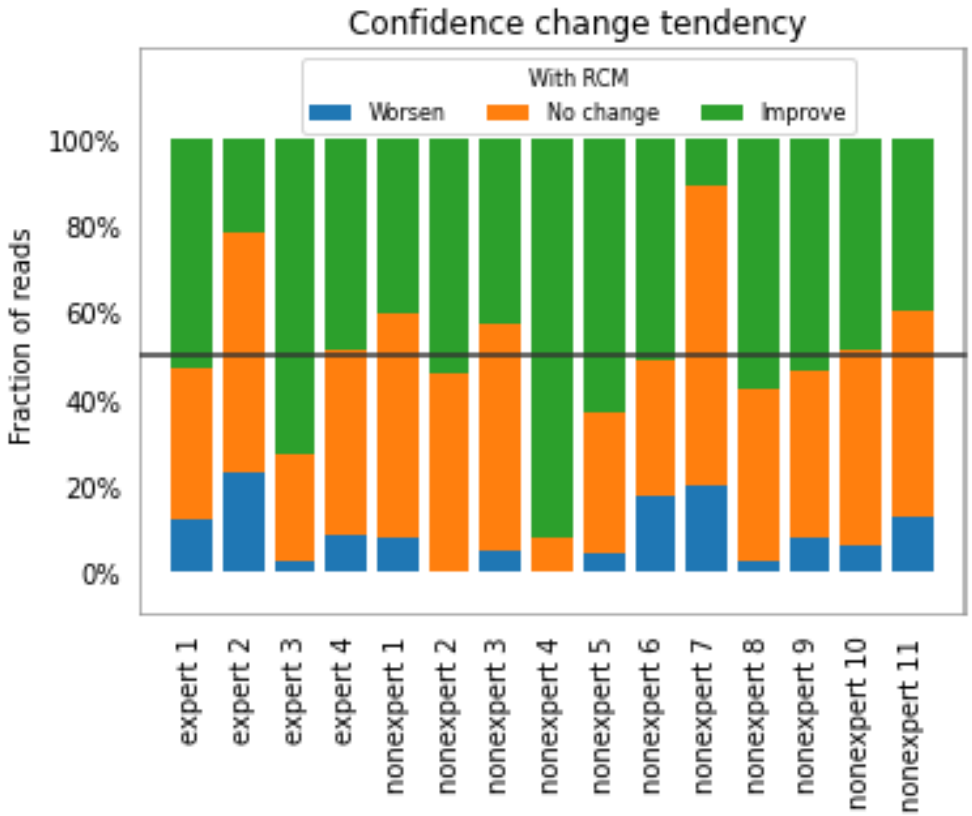
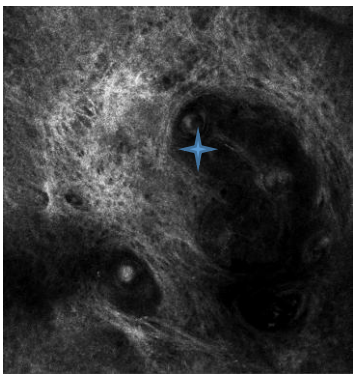
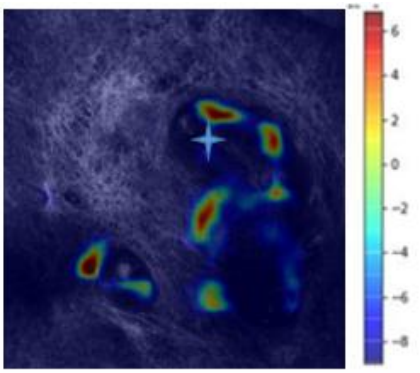
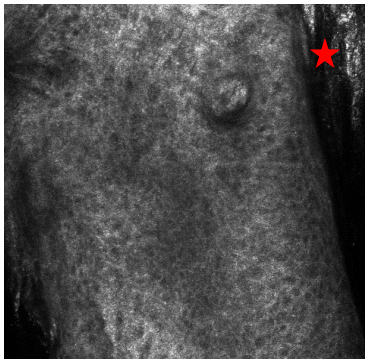
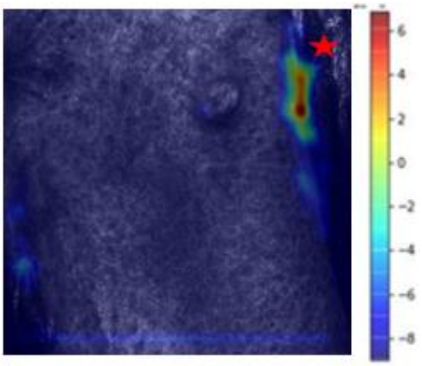
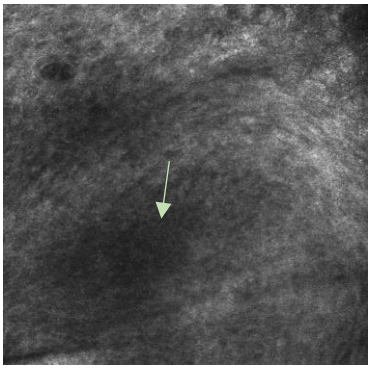
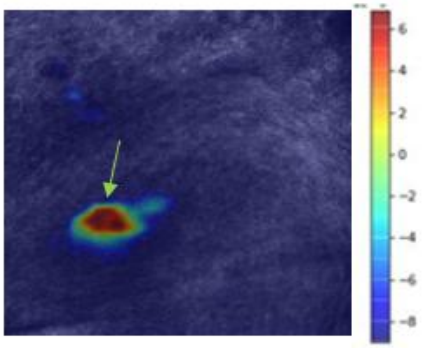
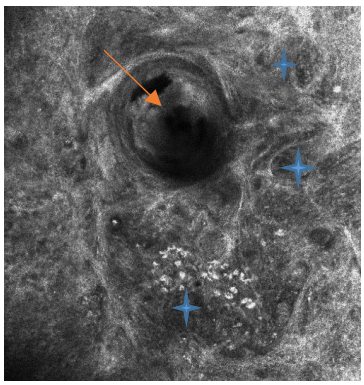
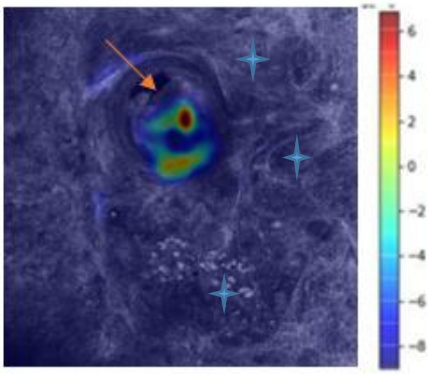


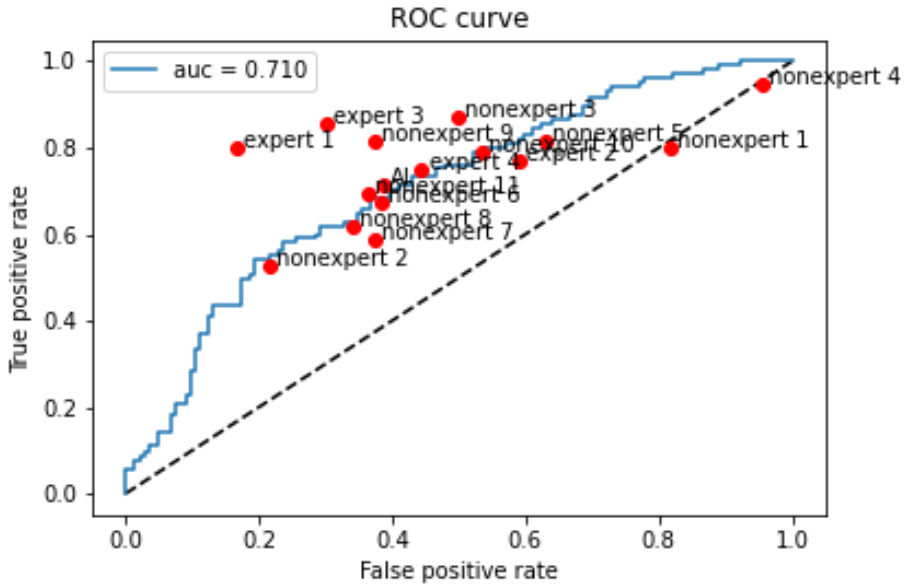
Figure 2: Confidence tendency with RCM for each reader.

RCM Images	Heatmaps Images
A) TP 	TP 
B) TN 	TN 
C) FP 	FP 
D) FN 	FN 

760
761

Figure 3: RCM images with their corresponding heatmaps: (A) RCM image with a true positive AI prediction shows BCC dark tumor nodules (blue stars) and high attribution area (red area) is seen within the nodules. (B) RCM image with a true negative AI prediction shows normal skin epidermis with regular honeycomb pattern which corresponds to low-attribution area (blue area) on the heatmap. On the upper-right corner of the image, a focal area of high-attribution (red area) is visible which corresponds to the skin fold (red stars) in the RCM image. (C) RCM image of a lichenoid keratosis and dermal scar with a false positive AI prediction for BCC shows dark shadow in the epidermis (green arrow), these areas had high-attribution (red area) for BCC in the heatmap. (D) RCM image of a nodular BCC with a false negative AI prediction shows BCC tumor nodules (blue stars) and a central hair follicle (red arrow). High-attribution area (red area) on the heatmap corresponds to the hair follicles.

A



B

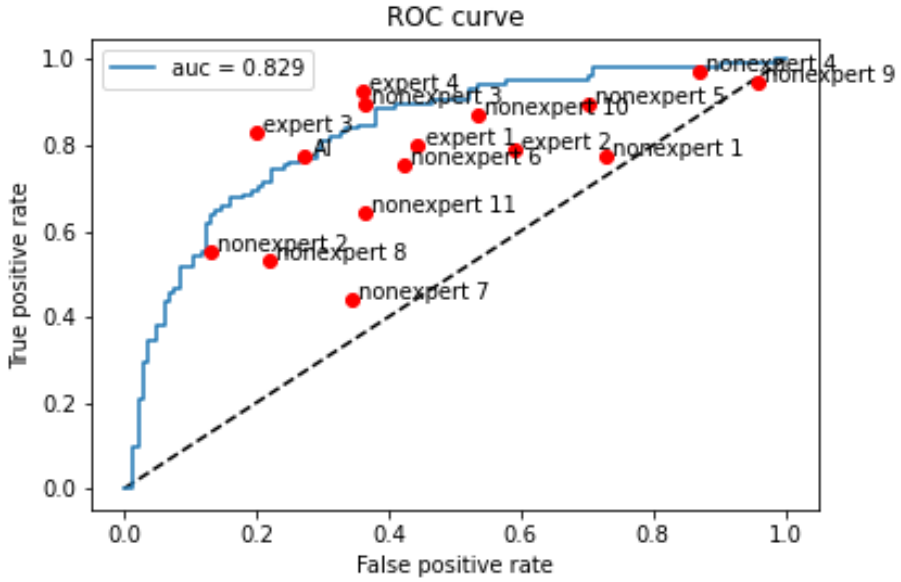


Figure 4: A. AUC Roc curve comparing dermoscopy AI model performances with all the 15 readers. B. AUC Roc curve comparing the AI fusion model (dermoscopy+RCM) performances with all the 15 readers.

TABLES

Reader class	Imaging modalities	Sensitivity	(95% CI)	p-value	specificity	(95% CI)	p-value
Overall	With RCM	76.9%	(73.3 – 80.1)	0.320	53.5%	(48.6 – 58.3)	0.919
	Dermoscopy	75.1%	(71.4 – 78.5)		53.7%	(48.8 – 58.5)	
Nonexperts	With RCM	75.6%	(71.2 – 79.5)	0.464	50.7%	(44.9 – 56.4)	0.807
	Dermoscopy	74.2%	(69.7 – 78.1)		51.4%	(45.6 – 57.1)	
Experts	With RCM	80.1%	(73.3 – 85.6)	0.499	60.3%	(51.2 – 68.8)	0.860
	Dermoscopy	77.6%	(70.6 – 83.4)		59.5%	(50.4 – 68.0)	

Table 1: Sensitivity and specificity for BCC across all the readers.

Reader experience	Imaging modalities	Indicated confidence					
		0%	25%	50%	75%	100%	p-value
Overall	With RCM	3.5%	6.3%	8.9%	33.3%	47.9%	<0.001
	Dermoscopy	1.8%	9.3%	30.5%	38.8%	19.5%	
Nonexperts	With RCM	3.8%	7.3%	8.8%	31.8%	48.3%	<0.001
	Dermoscopy	2.6%	9.8%	31.2%	38.6%	17.8%	
Experts	With RCM	2.5%	4.0%	9.4%	37.2%	46.9%	<0.001
	Dermoscopy	0.0%	7.9%	28.9%	39.4%	23.8%	

Table 2: Confidence with and without RCM.

Characteristic	Value / Count (n)	Percentage (%) / Range
Patient demographics (N = 204)		
Age (years)		
Mean [Range]	61	[19–90]
Sex		
Male	103	50.5%
Female	101	49.5%
Male-to-female ratio	*1.01:1*	—
Ethnicity		
Caucasian	204	100.0%
Geographic origin		
United States (MSKCC)	136	66.7%
Europe (Italy)	58	28.4%
Australia	10	4.9%
Lesion characteristics (N = 249)		
Clinical morphology		
Papule	128	51.4%
Macule	105	42.2%
Nodule	5	2.0%
Patch	5	2.0%
Plaque	5	2.0%
Maculopapular	1	0.4%
Non-pigmented lesions	**186**	**74.7%**
Pigmented lesions	**46**	**18.5%**

Characteristic	Value / Count (n)	Percentage (%) / Range
Lightly pigmented lesions	**17**	**6.8%**
Anatomical site		
Head and neck	105	42.2%
Trunk	95	38.2%
Lower extremities	26	10.4%
Upper extremities	23	9.2%
Lesion diameter (mm)		
Median [Range]	6.46	[1.0–100.0]
RCM imaging device used		
Handheld probe (VivaScope 3000)	121	48.6%
Wide-probe device (VivaScope 1500)	128	51.4%
Histopathological diagnosis		
Basal cell carcinoma (BCC)	144	57.8%
Non-BCC (benign or other malignancy)	105	42.2%
BCC histological subtyping (n = 135)*		
Superficial	79	58.5%
Nodular	70	51.9%
Infiltrative	16	11.9%
Micronodular	6	4.4%

Note: RCM, reflectance confocal microscopy; BCC, basal cell carcinoma.

*Histological subtyping was available for 135 of the 144 BCC lesions (93.75%). Subtype percentages exceed 100% due to the presence of mixed-pattern histological phenotypes.

Table 3: Demographic, clinical, and histopathological characteristics

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterialsJ.MonnieretalsubmissionequivocalBCC.docx