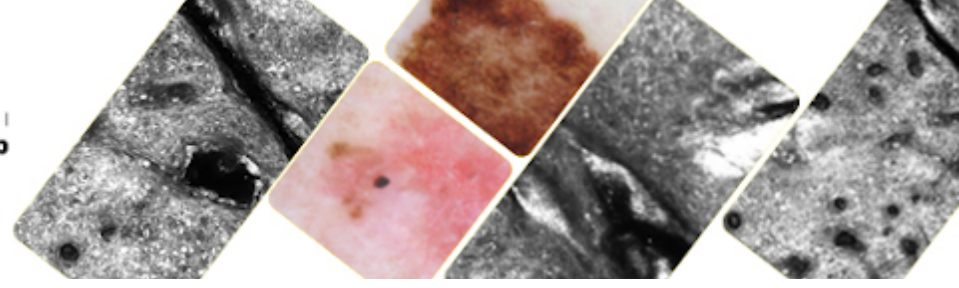
International Confocal Group



# DEEP LEARNING FOR FACIAL LESIONS' DIFFERENTIAL DIAGNOSIS UNDER REFLECTANCE CONFOCAL MICROSCOPY

Camila Scharf, Kivanc Kose, Marco Spadafora, Caterina Longo, Giuseppe Argenziano, Elvira Moscarella

Università degli Studi della Campania *Luigi Vanvitelli*

## INTRODUCTION

Reflectance confocal microscopy (RCM) is a high-resolution, noninvasive imaging technique that uses a diode laser (830nm) for cellular visualization of the epidermis and dermal structures, producing images based on skin structure refraction differences[1]. It offers horizontal sections with an 8x8 mm field and has sensitivity rates of 68% to 99% and specificity of 60% to 99%. Routine use in oncological centers has led to a 60% reduction in unnecessary biopsies.2

A main limitation of RCM is the extensive training needed to gain diagnostic expertise, as interpreting RCM images requires skills beyond traditional histopathology or dermoscopy. Current training resources are mainly short courses, with few formal programs or academic services available.

In this context, deep learning could help guide clinicians, especially those less experienced with RCM. Machine learning algorithms may provide a more quantitative and objective diagnostic approach, akin to previous efforts that improved diagnostic accuracy through automated image analysis of dermoscopy and dermatopathology slides[3]. These tools can generate scores reflecting malignant potential without needing clinician input.

## OBJECTIVES

This study aims to develop and assess an AI-based segmentation model applied to RCM mosaics of facial lesions.

## MATERIALS AND METHODS

This was a retrospective and prospective study conducted in collaboration between the Dermatology Unit of the University of Campania Luigi Vanvitelli (Naples, Italy) and the Memorial Sloan Kettering Cancer Center (New York, NY and Hauppauge, NY), evaluating all RCM cases perfomed from 2012 to 2023. The dataset consisted of 3,056 RCM mosaics from 855 patients, each mosaic corresponding to a unique skin lesion, collected at the University of Campania and the University of Modena and Reggio Emilia. Areas of suspicion in the mosaics were labeled at the pixel level by two expert readers using Seg3D, with unlabelled areas considered benign. The mosaics varied in size from 7000 × 8000 to 12,000 × 12,000 pixels (14–36 mm²) and represented common facial lesions in clinical practice.

Two AI models were developed, a tile-wise classification model and a tile-wise object detection model using Detectron. EfficientNet and Swin Transformer models were trained for image classification.

## RESULTS

The evaluation of the RCM cases included 3056 RCM mosaics, obtained from 855 different patients, being 430 females and 425 males with a mean age of 62 years.

Cases included 159 Actinic Keratosis (AK), 5 Angioma, 294 Atypical Melanocitary lesion (LMA), 128 Basal Cell Carcinoma (BCC), 8 Blue Nevus, 5 Bowen's Disease, 2 Keratoacanthoma, 43 Melanocytic Nevus (including junctional, compound and dermal), 47 Lichen Planus Like Keratosis (LPLK), 9 Squamous Cell Carcinoma (SCC), 68 Seborrheic Keratosis (SK), 80 Solar Lentigo (SL), 14 Spitz Nevus and 3 Trichoepitheliomas.

All mosaics including atypical honeycomb, atypical cobblestone, dendritic cells, roundish or plumb-bright cells, pagetoid spread, canalicular vessels, dark shadows, clefting or disarrange of the dermal epidermal junction where labeled, as it exemplified in figures 1 and 2.

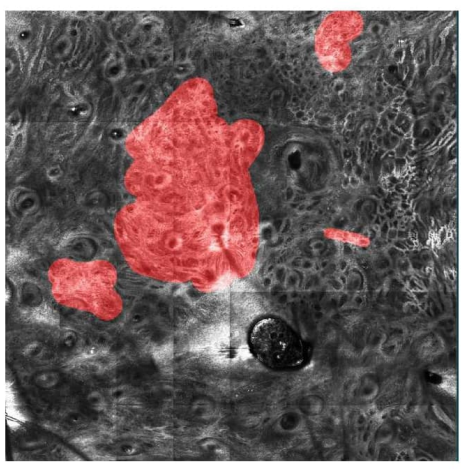


FIGURE 1: label of a LMA (in red)

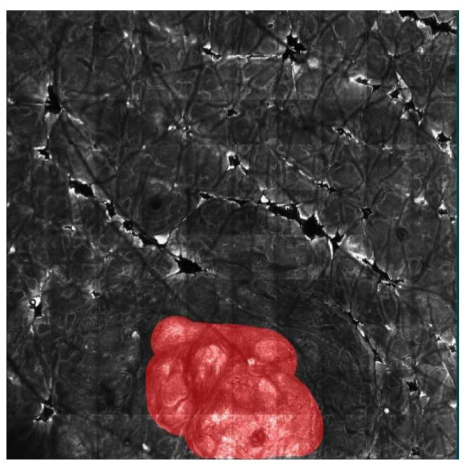


FIGURE 2: label of a BCC (in red)

Classification model performance

The Swin transformer base model trained on the tile-level labels achieved 0.775 AUROC on the validation set. We first threshold the tile wise model output probabilities to binarize the prediction mask and filter out the mosaics where the suspicious areas cover less than %1 of the whole mosaic. As presented in Figure 3, the model achieved an AUROC of 0.832. For example, at the threshold level of 0.63, sensitivity is 0.83 and specificity is 0.71

Semantic segmentation performance

We also ran the Detectron segmentation model in a sliding window fashion and merged the tile-wise results into mosaics and obtain a mosaic segmentation mask. As presented in Figure 4, the model achieved an AUROC of 0.79. At the sensitivity level of 0.8, the specificity of the model is 0.6.

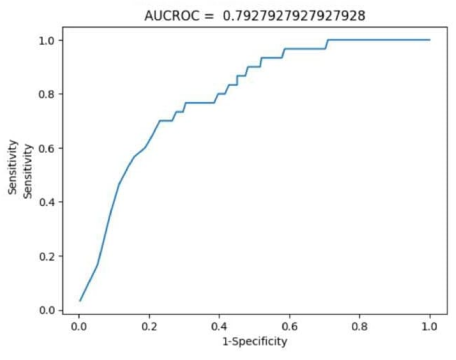



FIGURE 3: AUROC for the tile wise model

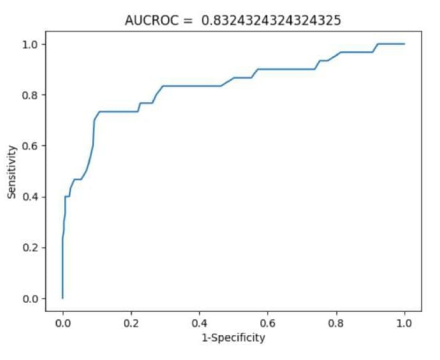



FIGURE 4: AUROC for the Detectron segmentation model

## CONCLUSIONS

This research yielded encouraging preliminary results, suggesting that with enhancements in both annotation quality and model refinement, there is potential for our approach to rival that of expert RCM reader or at least reach a comparable level of proficiency with human experts.

1. Rajadhyaksha M, Grossman M, Esterowitz D, Webb RH, Anderson RR. In vivo confocal scanning laser microscopy of human skin: melanin provides strong contrast. J Invest Dermatol. 1995 Jun;104(6):946-52.
2. Witkowski AM, Łudzik J, Arginelli F, Bassoli S, Benati E, Casari A, De Carvalho N, De Pace B, Farnetani F, Losi A, Manfredini M, Reggiani C, Malvehy J, Pellacani G. Improving diagnostic sensitivity of combined dermoscopy and reflectance confocal microscopy imaging through double reader concordance evaluation in telemedicine settings: A retrospective study of 1000 equivocal cases. PLoS One. 2017 Nov 9;12(11):e0187748.
3. Kurugol S, Kose K, Park B, Dy JG, Brooks DH, Rajadhyaksha M. Automated delineation of dermal-epidermal junction in reflectance confocal microscopy image stacks of human skin. J Invest Dermatol. 2015 Mar;135(3):710-717.