applied biosciences

MDPI

Review

Artificial Intelligence in Reflectance Confocal Microscopy for Cutaneous Melanoma Computer-Assisted Detection: A Literature Review of Related Applications

Luana Conte [1,2,3,†], Angela Filoni [4,†], Luca Schinzari [4], Ester Sofia Congedo [4], Lucia Pietroleonardo [4], Rocco Rizzo [3,5,6], Ugo De Giorgi [7], Donato Cascio [1,2], Giorgio De Nunzio [3,6,8,*,‡] and Maurizio Congedo [4,‡]

1 Department of Physics and Chemistry "E. Segrè", University of Palermo, 90128 Palermo, Italy; luana.conte@unipa.it (L.C.); donato.cascio@unipa.it (D.C.)
2 Laboratori Nazionali del Sud, National Institute for Nuclear Physics, 95123 Catania, Italy
3 Advanced Data Analysis in Medicine (ADAM), Laboratory of Interdisciplinary Research Applied to Medicine (DReAM), Local Health Authority (ASL) of Lecce, 73100 Lecce, Italy; rocco.rizzo1@unisalento.it
4 UOSD Dermatology and Allergology Unit, Vito Fazzi Hospital Lecce, 73100 Lecce, Italy; filonidermatologo@gmail.com (A.F.); luca.schinzari93@gmail.com (L.S.); estercongedo@gmail.com (E.S.C.); pietroleonardolucia@yahoo.it (L.P.); mauriziocongedo@gmail.com (M.C.)
5 Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy
6 National Institute for Nuclear Physics, Branch of Lecce, 73100 Lecce, Italy
7 Department of Experimental Medicine, University of Salento, 73100 Lecce, Italy; ugo.degiorgi@unisalento.it
8 Laboratory of Biomedical Physics and Environment, Department of Mathematics and Physics "E. De Giorgi", University of Salento, 73100 Lecce, Italy
* Correspondence: giorgio.denunzio@unisalento.it
† Co-First Author: Luana Conte and Angela Filoni.
‡ Co-Last Author: Giorgio De Nunzio and Maurizio Congedo.

Abstract

Cutaneous melanoma is one of the most aggressive skin cancers, and early diagnosis remains essential to reduce mortality. Reflectance Confocal Microscopy (RCM) provides non-invasive, quasi-histological images of the epidermis, dermoepidermal junction (DEJ), and dermis, enabling real-time assessment of melanocytic lesions. However, interpretation still relies on expert visual evaluation, which is time-consuming and subjective. In this context, Artificial Intelligence (AI) and Computer-Assisted Detection (CAD) systems are emerging as valuable tools to improve diagnostic accuracy and reproducibility. This review summarizes research on AI applications in RCM imaging for melanoma, focusing on three major areas: delineation of skin strata, segmentation of tissues and morphological patterns, and classification of benign versus malignant lesions. Early approaches included Bayesian classifiers, wavelet-based decision trees, and logistic regression, while recent studies have employed support vector machines, random forests, and increasingly deep learning architectures such as convolutional and recurrent neural networks. The results demonstrate encouraging accuracy in DEJ localization, the segmentation of diagnostically relevant patterns, and the discrimination of melanoma from benign nevi. We distinguish the maturity of dermoscopy-based AI (AUC (ROC) > 0.80 on large multicenter cohorts) from the still-exploratory evidence for RCM-based AI. Nonetheless, current studies are often limited by small datasets, heterogeneous protocols, and a lack of multicenter validation. Overall, progress in AI applied to RCM supports the development of CAD systems that could assist clinicians during acquisition and diagnosis, reducing unnecessary biopsies and improving early melanoma detection. Future work should address standardization, dataset expansion, and the integration of advanced AI methods to move closer to clinical implementation.

Academic Editor: Nikolaos Kourkoumelis

Received: 30 October 2025
Revised: 4 January 2026
Accepted: 19 February 2026
Published: 9 March 2026

Keywords: melanoma; reflectance confocal microscopy; artificial intelligence; neural networks; deep learning

1. Introduction

Cutaneous melanoma (MM) is one of the most aggressive skin cancers, with a high metastatic potential. According to GLOBOCAN, melanoma of the skin accounted for an estimated 331,722 new cases and 58,667 deaths in 2022 globally. Early diagnosis is essential to reduce mortality, and modern imaging techniques allow for increasingly accurate and detailed data [1]. These include dermatoscopy, digital epiluminescence, high-frequency ultrasound, optical coherence tomography, and confocal laser scanning microscopy (CLSM), available in both reflectance and fluorescence (FLSM) modes [2].

Non-invasive techniques are designed to support specialists in the real-time assessment of melanoma at an early and more easily treatable stage, minimizing unnecessary surgical biopsies and associated scarring in the case of benign lesions [3]. They also enable lesion monitoring through digital image storage. Among these methodologies, reflectance confocal microscopy (RCM) is the only one capable of showing the structure of the skin at the microcellular level, thus representing a potential alternative to traditional histological examination [3].

Despite this, diagnostic confirmation is often completed by surgical removal of the lesion for biopsy, an invasive, painful, and scarring procedure that requires several days of waiting for a report [3]. Furthermore, the reading of RCM images depends largely on the operator's experience, as the analysis remains predominantly qualitative. This non-invasive approach nevertheless allows for direct in vivo observation with a resolution almost comparable to that of histology of the epidermis, dermal–epidermal junction (DEJ), and superficial dermis, producing grayscale images in which the pixel value reflects the refractive index of different types of tissues and cellular components [2].

RCM is now an established practice that has demonstrated a significant increase in diagnostic sensitivity and specificity compared to simple clinical evaluation with the naked eye. In meta-analyses, RCM shows high diagnostic accuracy for cutaneous melanoma, with pooled per-lesion sensitivity around 92–93% and a specificity of around 70–76%. Moreover, comparative analyses suggest that RCM can improve diagnostic specificity compared with dermoscopy in equivocal lesions, thereby reducing unnecessary excisions [4]. Patients and healthcare professionals are showing a strong interest in standardized, reproducible, and automated methodologies; for this reason, a gradual transition towards computer-aided diagnosis (CAD) systems is emerging, made possible by image digitization. CAD software is increasingly being adopted for the detection and discrimination of various abnormalities obtained through various diagnostic techniques.

AI is based on the development of machines capable of learning and autonomously reproducing cognitive functions similar to those of the human brain. Patel et al. [5] compared studies between dermatologists and algorithms applied to dermoscopic images; AI-based systems have shown ROC values above 80% in melanoma detection. This study systematically analyzed the state of the art of AI-based techniques applied to dermatological diagnostics, with particular reference to the use of RCM, evaluating accuracy, sensitivity, and specificity. All studies that directly compared the performance of AI-based techniques with that of dermatologists reported superior or comparable results for AI methodologies in improving melanoma detection. In particular, studies evaluating RCM showed an average accuracy value of 82.72%. Overall, the results suggest that AI-based techniques have high potential for improving diagnostic accuracy and clinical outcomes, primarily

through the early identification of melanoma. RCM has been shown to be a method capable of increasing diagnostic sensitivity and specificity in most skin cancers, supporting the indication for biopsy in suspicious lesions, and contributing to the accurate mapping of tumor margins in large lesions, facilitating correct surgical excision. In this context, AI could allow for the more accurate delineation of the DEJ in RCM images, while optimizing the time required for the diagnostic process. However, further studies are needed to address current limitations and ensure the reliability and clinical validity of these approaches. Studies directly comparing RCM with AI applications are still limited, as these technologies are at a relatively early stage of development. Further research is therefore needed to evaluate the generalizability of AI-based techniques across heterogeneous populations and different skin phototypes, as well as to improve the standardization of image acquisition and processing protocols. Most of the studies included focused mainly on subjects with fair-skin phototypes, so the extension of the results to patients with darker skin or other skin variants is currently limited and represents an area for further investigation.

In the past few years, developments in AI models with increasingly high representation power have allowed for significant progress in many fields, such as science, engineering, medicine, nursing, finance [6–15]. In the case of skin cancers, and especially of melanoma, the use of these techniques combined with newer technologies, such as RCM [15], represents an important challenge to ensure early diagnosis and reduce the risk of death. Scientific studies dealing with automatic methods for RCM image analysis are still few in number. Luck et al. [16] were among the first to introduce automated RCM image processing procedures, proposing a Bayesian classification algorithm for the identification of various tissues. Subsequently, Kurugol et al. [17] developed a semi-automatic technique for determining the dermal–epidermal junction (DEJ). Later, Koller et al. [18] introduced a classification approach based on wavelet decision trees with the aim of discriminating between benign and malignant melanocytic skin lesions in RCM images.

In another contribution, Hames et al. [19] developed a logistic regression-based classifier capable of automatically segmenting the different skin layers in RCM scans. Ghanta et al., in 2017 [20], implemented a complex mathematical shape model in order to accurately define the depth of DEJ in RCM acquisitions. In Bozkurt et al. [21], an algorithm based on deep neural networks was developed to segment DEJ in RCM images.

Based on these observations, our paper aims to survey and review articles reporting Artificial Intelligence and Image Processing applications directly or indirectly relevant to the Computer-Assisted Detection and Diagnosis of cutaneous melanoma in images acquired by RCM. Three main challenges were considered. As melanoma originates from the DEJ, some papers concerning the localization of the junction in a stack of RCM images were examined. Tissues (particularly those composing the DEJ) contain different varieties of cells and structures; therefore, works aiming at the segmentation of skin tissues were considered. Finally, specific articles on the discrimination between healthy and tumoral tissues were taken into account. All of these applications share the purpose of improving diagnostic accuracy during the automatic detection of skin pathologies with CAD systems. Beyond summarizing the technical and clinical aspects, this review critically appraises AI evidence for RCM in comparison with dermoscopy, focusing on dataset scale, external validation, and generalizability, and explicitly delineates the different maturity levels of dermoscopy vs. RCM-based AI evidence.

This article is a narrative review and was not designed as a systematic or scoping review. To describe how references were selected, we performed an iterative literature search in Scopus and PubMed from database inception to 7 April 2025. Searches were conducted using combinations of keywords related to “reflectance confocal microscopy”/RCM (and related terminology, such as CLSM, when used in the literature), “melanoma”/“melanocytic

lesion", and "artificial intelligence"/"machine learning"/"deep learning", and also including terms reflecting the three tasks discussed in this review (skin layer delineation/DEJ localization, tissue or pattern segmentation, and diagnostic classification). Peer-reviewed original studies dealing with automated or AI-based analysis of RCM images for cutaneous melanoma or melanocytic lesions were considered eligible; studies focusing on other imaging modalities without RCM, not involving automated/AI analysis or not related to melanocytic lesions, were excluded. The final selection was based on relevance to the scope of the review and on coverage of the main methodological families and CAD tasks. Additional relevant records were identified by screening the reference lists of included papers.

2. RCM Imaging Technique and Melanomas

The principle of confocal imaging was patented in 1957 by Marvin Minsky [22]. CLSM can be performed in either fluorescence or reflectance mode. The first has been used predominantly in experimental studies; the second relies on differences in the refractive indices of cellular structures. In many works, CLSM and RCM nomenclatures are used to denote the same technology. In this discussion, we prefer to use the RCM abbreviation.

RCM works with incident monochromatic light that penetrates the skin, highlighting small points inside the tissue. In the optical system, light reflected from the skin tissue is conveyed to a detection sensor. Before reaching this, it passes through a very small aperture. RCM scans have a high resolution with a horizontal resolution of between 0.5 and 1 μm, while the axial resolution reaches values between 3 and 5 μm, with a maximum signal penetration depth of between 150 and 200 μm [23,24].

Each individual two-dimensional image covers an area of 500 μm × 500 μm. However, it is possible to extend the scan to the entire lesion, up to an area of 8 mm × 8 mm, obtaining a series of images that form a continuous horizontal mosaic, known as VivaBlock. The device also allows for the automatic acquisition of deep images, each of the same size (500 μm × 500 μm), thus generating a three-dimensional reconstruction of the analyzed area, called VivaStack [3].

Mosaics obtained using RCM tend to cover the entire extent of the skin lesion and, where possible, one or two millimeters of adjacent healthy skin. As a result, each mosaic normally includes both pathological portions and portions of unaffected skin.

One of the main difficulties associated with RCM concerns insufficient image quality, which can occur when the technician fails to correctly identify the optimal skin level during scanning, or when visual distortions are generated by patient or device movement or by the presence of elements that obstruct visualization, such as microbubbles, debris, hair, or other interfering structures. Such distortions are referred as areas of no informational value [25].

Another important feature of confocal microscopy is the ability to record video sequences at a speed of 15–25 frames per second, allowing for biological processes to be observed [3].

RCM's ability to acquire mosaics and stacks in real time enables the noninvasive evaluation of a large area of tissue in vivo and provides substrates for dermoscopic patterns, corresponding to specific histopathological criteria [26]. This means that, in specific cases, RCM can avoid skin biopsy.

In the last decade, RCM has been used in clinical practice as a second-level examination after dermoscopic evaluation, improving early skin cancer diagnosis and reducing the number of unnecessary excisions due to surgical biopsies of benign lesions [27,28]. MM has been an early subject of systematic RCM analysis because melanocytic lesions can be depicted with great definition due to the clear endogenous contrast of melanin. For representative examples, see Figures 1 and 2 of [27], which illustrate in vivo confocal

microscopic features characteristic of melanomas and nevi, respectively. Clinically, MMs appear as dark, flat or slightly raised marks on the skin, with different colors. The contours of the lesion may be uneven, with irregularities, indentations, or small breaks in the demarcation line [3]. Images obtained using confocal microscopy, displayed in grayscale, allow elements that appear lighter to be identified, as they are made up of structures with a higher refractive index than the surrounding tissue; these areas are defined as backscattered zones. The phenomenon of backscattering depends on the refractive index (n) of the individual cellular components in relation to the medium in which they are found. Among the main skin elements that show a high capacity to reflect light are melanin and melanosomes, dermal fibers and keratin. These structures are particularly bright [29].

There are four main types of melanoma skin cancer, i.e., superficial spreading, nodular, lentigo maligna and acral lentiginous. Superficial spreading melanoma (SSM) is the most common and lethal form of skin cancer, with an estimated incidence of 70% of all cases [30]. It grows outward and spreads across the surface of the skin. Two distinct microscopic subtypes of SSM can be identify: pagetoid melanoma and solar melanoma [3]. Pagetoid melanoma normally occurs in adults with intermittent solar exposure and a high number of nevi on the skin [3]. It is characterized by the abundant growth of large intraepidermal cells, with prominent bright cytoplasm and dispersed melanin. Pagetoid melanoma, in dermoscopy, exhibits a reticular and non-specific pattern [31,32]. On RCM, melanocytes are visible as a large roundish shape, with prominent and bright cytoplasm, and they are normally present in the upper epidermis of DEJ, as single cells or arranged as nests [3]. Solar melanoma is frequent in patients with low nevi, in areas with high solar exposition. There is no evidence of this occurring in pre-existing nevi [3]. Solar melanoma is frequently indistinguishable from lentigo maligna, which usually occurs on the face of patients that live in the most sun-exposed areas.

Histologically, atypical melanocytes are visible in the superficial dermis, with lymphocytic infiltration [3]. In dermoscopy, it shows in regression zones with white depigmentation similar to scars. On RCM, atypical melanocytes appear with dendritic branches, forming branching structures in both epidermis and DEJ levels [3]. The normal structure is disarranged and pagetoid spread is formed by pleomorphic cells with elongated bodies, forming branches with variable morphology [3].

One of the first uses of confocal microscopy for malignant melanoma (MM) diagnosis was in 2001 by Langley et al. [33]. They studied forty pigmented skin lesions and verified the RCM's ability to identify distinct patterns and cytologic features, suggesting the possibility of a further improvement in the diagnostic accuracy of MM. After this work, researchers mainly focused on the differentiation of benign and malignant skin lesions, enabling the "real time" identification of melanocytic lesions, as described in the next sections [34–36]. In clinical practice, considering the different three skin levels in RCM (epidermis, dermal–epidermal junction (DEJ) and dermis), two aspects have to be considered: cytological and architectural features. The presence of pagetoid melanocytes is a reproducible and characteristic RCM criterion in several studies [37,38]. This feature presents as widespread, round, spindle, or dendritic atypical cells in epidermal layers. Another aspect of probable malignancy is the roughness of the superficial pigmented basal layer near the DEJ. For these reasons, the identification of skin strata and the subsequent study of the specific features represented two targets that can be used to identify useful AI algorithms for CAD systems.

In 2020, Kose et al. [25] trained the MED-Net model using a set of 117 RCM mosaics. These mosaics were labeled at the single-pixel level by two RCM experts, who reached a common agreement, with the aim of identifying portions without diagnostic relevance and measuring the model's performance in terms of sensitivity and specificity on a larger dataset. In addition, the segmentation output produced by MED-Net was used to assess how

frequently the non-informative areas in the RCM mosaics coincide with the pathological area of the tissue.

The MED-Net model is designed to learn visual structures at multiple levels of scale (i.e., different degrees of zoom and resolution), starting from a general view and progressively moving to more detailed levels. Semantic segmentation for each scale is performed by secondary networks, consisting of fully convolutional encoder–decoder neural architectures.

Since dermoscopic images and RCM mosaics are spatially aligned, the portion of tissue identified as lesion in dermoscopy corresponds directly to the same area in the RCM scan. This allows for the precise quantification of how much of the lesion is composed of areas lacking useful information. In this study, all regions where the image was unusable for diagnostic evaluation, either due to technical problems (poor lighting, signal saturation) or physical obstacles, were classified as "non-informative."

The analysis showed that, considering the entire mosaic, the average percentage of unusable pixels varied approximately between 4% and 17% (with a standard deviation between 6% and 23%). Limiting the measurement to the pathological area alone, the proportion of non-informative areas ranged from 3% to 15% (standard deviation 4–18%). The ability to view the map of obscured areas identified by the algorithm on the dermoscopic image plays a key role, as it allows for the verification of whether clinically relevant parts of the lesion were compromised during acquisition.

An automated evaluation system, capable of indicating, in real time, the quality of the mosaic and the position of unusable areas, could guide the technician to repeat the acquisition immediately.

In a subsequent phase, MED-Net was applied to RCM mosaics of the dermo-epidermal junction related to pigmented lesions, with the aim of measuring the recurrence of non-evaluable areas both on the entire mosaic and within the lesion. The data show that, when considering only the overall percentage of non-informative pixels in the image, the non-interpretable area present in the pathological area tends to be overestimated. Combining the visualization of the overlap with the overall percentage value, on the other hand, provides adequate objective information, avoiding the laborious process of the physician having to manually mark the boundaries of the lesion on the dermoscopy image.

The evidence obtained confirms that image processing using machine learning models is capable of reliably recognizing areas that do not provide clinically relevant information. The results also suggest that the use of an automated system for image quality assessment can play a significant role, both in defining standardized quality parameters and as an objective method for monitoring quality itself. This would allow for the introduction of measurable control during image recording, reducing the need to recall patients due to inadequate acquisitions and, during the analysis phase, supporting the estimation of the risk that the image obtained does not correctly represent the lesion examined.

Technical Specifications of RCM Imaging

Reflectance Confocal Microscopy (RCM) provides high-resolution, in vivo imaging of the skin using near-infrared monochromatic light and a confocal optical system. The technique enables the visualization of cellular and subcellular structures with contrast derived from variations in refractive index among tissue components.

Resolution and Penetration Depth

- Lateral (horizontal) resolution: approximately 0.5–1 μm, allowing for the detailed visualization of individual cells and microstructures.
- Axial (vertical) resolution: approximately 3–5 μm, enabling the accurate delineation of epidermal layers and the dermo–epidermal junction (DEJ).

- Maximum penetration depth: typically 150–200 µm, sufficient to image the epidermis, DEJ, and superficial dermis.

Field of View and Acquisition Modes

- Each single RCM image covers an area of 500 µm × 500 µm.
- For broader coverage, the system acquires mosaics (VivaBlock) by stitching adjacent fields, extending up to 8 mm × 8 mm, which can encompass the entire lesion and adjacent healthy skin.
- Vertical stacks (VivaStack) are obtained by sequential imaging at increasing depths, enabling a three-dimensional reconstruction of the scanned region.

Image Characteristics

- Images are displayed in grayscale, where pixel intensity reflects the refractive index of tissue components. Structures with high refractive indices—such as melanin, melanosomes, keratin, and dermal fibers—appear bright due to strong backscattering.
- Video acquisition at 15–25 frames per second is possible, allowing for the dynamic observation of biological processes.

Common Artifacts and Quality Considerations

- Image quality may be compromised by motion artifacts, poor focus, or physical obstructions (e.g., hair, microbubbles, debris).
- Non-informative regions should be identified and minimized during acquisition; automated quality-control algorithms are increasingly integrated into RCM workflows.

3. Artificial Intelligence in RCM Imaging

Machine learning and deep learning algorithms represent powerful tools for many research groups, who apply them to CAD systems in medical applications. Until the early 2000s, all CAD systems supporting melanoma diagnosis were based on dermoscopic images. In fact, dermoscopy is, at present, the routine method used by dermatologists for skin cancer screening. Although still not widespread, the introduction of RCM technology in clinical practice marks the beginning of a new era of dermo-oncology. In this context, AI algorithms will eventually become a helpful tool to identify the most relevant features for MM diagnosis and to combine them in simple algorithms useful for practical diagnostic purposes. Figure 1 describes the temporal and geographical trends in publications containing the keywords "Artificial Intelligence" and "Reflectance Confocal Microscopy", showing a marked increase in the last five years and a predominant contribution from the United States and Italy. Figure 1 does not represent the set of papers included in this narrative review.

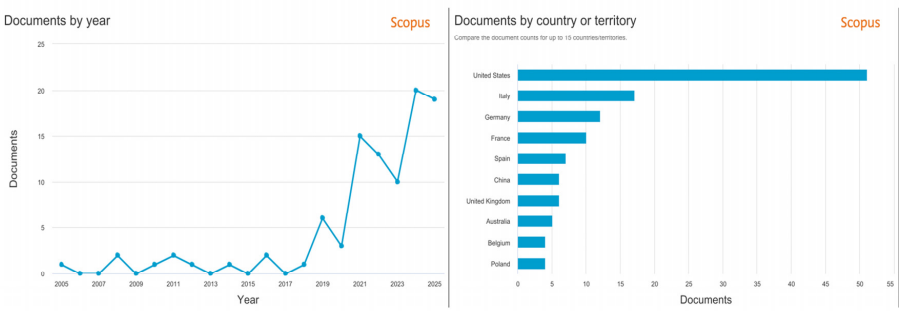

Figure 1. Trends in the scientific production of publications containing the keywords "Artificial Intelligence" and "Reflectance Confocal Microscopy". (**Left**) Annual distribution of documents from 2005 to 2025. (**Right**) Geographical distribution by country, showing the United States and Italy as the leading contributors (source: Scopus, accessed on 7 November 2025). The analysis is descriptive and intended to illustrate temporal and geographical trends in the topic.

In the next sections, we describe three principal applications of RCM that, in combination with different AI methods, might help clinicians in daily routine.

3.1. AI for Delineation of Skin Strata

Identifying the boundary between epidermis and dermis in RCM scans remains a technically demanding task. This interface is of particular clinical relevance, as many melanocytic malignancies arise and evolve precisely at this level. A reliable and consistent delineation of the DEJ can therefore support early detection, diagnostic assessment, and evaluation of lesion progression. Moreover, determining the DEJ position is a prerequisite for several downstream automated analyses in RCM, making its accurate localization a key step for developing clinically useful computer-assisted tools. For these reasons, automated and semi-automated algorithms were developed by many research groups in order to detect and classify skin strata, and particularly the location of the DEJ.

Building on the previous achievements by the same group [17,39,40], one of the first in-depth studies for 3D DEJ boundary delineation in fair skin was presented in 2011 by Kurugol and colleagues [40]. They developed a semiautomated algorithm for DEJ localization in RCM stacks from fair-skin types (60 slices per stack). Expert-labeled boundaries were used for evaluation. Performance was reported as mean error $\approx$ 7–12 µm, with >80–90% of boundaries within 15 µm of expert markings and epidermis/dermis misclassification < 10%, using a hybrid sequence segmentation and LS-SVM approach. Their goal was to reproduce the clinician's visual expertise, highlighting texture and contrast differences between layers of the epidermis and dermis, in order to locate the DEJ.

They focused exclusively on participants with fair skin, a group characterized by low melanin content and therefore by limited contrast at the DEJ—an aspect that makes boundary detection particularly challenging. Their method combined concepts from three different domains—texture-based segmentation, pattern classification, and sequence segmentation—with the latter referring to the progressive "dynamic" changes observed along the vertical axis of the stack. Building on these elements, the authors introduced a semi-automated hybrid algorithm that integrates sequence segmentation with machine-learning classification. The approach divides z-stacks into homogeneous regions by modeling the transitions between skin layers, and then assigns each region to epidermis, dermis, or the intermediate DEJ zone using a diverse set of texture descriptors (gray-level co-occurrence matrix features, statistical measures, wavelet-derived coefficients, log-Gabor features, and radial spectral characteristics). After dimensionality reduction, two support vector machine classifiers [41] were trained: one to distinguish epidermis from non-epidermis and the other to separate dermis from non-dermis. Following the fusion of classification outputs and a smoothing step, the system achieved solid performance, with epidermis/dermis misclassification rates below 10% and an average deviation of approximately 8.5 µm from expert-annotated boundaries. The algorithm was subsequently evaluated on a larger dataset comprising 24 stacks (15 from fair skin and 9 from dark skin) [42], confirming its robustness. Misclassification rates remained under 10% (using 25 $\times$ 25 µm tiles), and the mean distance from expert-defined boundaries was ~6.4 µm for fair skin and ~5.3 µm for dark skin. In this study, the authors also assessed inter-expert agreement. In 2015, the same research group [43] expanded their work by developing a version specifically tailored to delineate skin strata in dark-skin image stacks. In these cases, the high melanin content produces strong contrast, so the algorithm primarily relies on detecting intensity variations along depth. The authors computed the median intensity for each tile and generated a Gaussian-smoothed median-intensity profile as a function of depth. In most stacks, single peaks in the profiles were observed, corresponding to the basal cell layer. Having found regions that clearly containing basal cells, the authors could then calculate

texture templates for them, which were then used to locate the correct depth of basal cells in profiles with multiple peaks. Finally, the DEJ can be spotted by examining each profile below the selected peaks. The authors also suggested that the further development of their algorithms might guide the assessment of abnormal morphological features at the DEJ. In 2017, Ghanta et al. [20] tackled the problem of 3D reconstruction/segmentation of the RCM stack proposed by Kurugol and colleagues [39,43]. They developed an advanced framework aimed at improving boundary delineation by integrating a mathematical shape prior modeled through a marked spatial Poisson process, which captures the characteristic "hills and valleys" along the DEJ contour. A marked Poisson process represents a specific class of marked point processes with several practical advantages, and it has previously been applied to the detection and segmentation of irregularly shaped structures in microscopy data (see, e.g., [44]). The algorithm introduced in [20] is designed as a general-purpose method, allowing for its adaptation to various imaging problems through the selection of an appropriate shape prior informed by domain expertise. When used for DEJ localization, the method achieves a mean error of 5.41 ± 3.94 µm for dark skin and 12.1 ± 7.0 µm for fair skin, outperforming the model by Kurugol et al. [39,43], which reported average errors of 7.9 ± 6.4 µm and 8.3 ± 5.8 µm, respectively.

Within the broader field of DEJ delineation, another line of research focuses on classifying individual slices within an RCM stack. An entirely unsupervised, texton-based strategy is described in [45]. In this approach, each image in the stack is represented by a texton histogram generated using a filter bank, which is then projected into a lower-dimensional space via Principal Component Analysis. The five skin layers—stratum corneum, stratum granulosum, stratum spinosum, stratum basale, and regions of the papillary dermis—are subsequently identified using k-means clustering. Although the authors report a strong correspondence between predicted labels and ground truth, the reliability of this finding is limited by the small and suboptimal evaluation dataset.

Hames et al. [19] expanded the segmentation/classification paradigm by assigning anatomical strata at the pixel level rather than at the image-slice level. Their method employs a bag-of-features representation derived from normalized and whitened small patches extracted from RCM mosaics, combined with hierarchical clustering techniques [46,47] and spherical k-means [48]. Depth stacks are then classified using conditional random fields and a structured support vector machine, achieving the correct classification for 85.7% of test sections. Pixel-wise segmentation offers clear advantages over dermatologist-annotated section labels, providing higher spatial precision, reduced variability, and improved sensitivity to age-related changes in skin thickness. Notably, the automated per-pixel method can detect both thickening of the stratum corneum and flattening of the DEJ.

In 2016, Kaur et al. [49] introduced the first Deep Learning (DL) approach for the automatic identification of skin layers in RCM stacks. Their hybrid model uses a Convolutional Neural Network that processes traditional texton-based feature vectors—similar to those in [45]—instead of raw 2D images. They classify six strata, adding an "outside epidermis" category not included in earlier studies. This DL model surpasses previous state-of-the-art methods [19], achieving a test accuracy of 81.73%. When the CNN is replaced with a support vector machine, accuracy drops to 0.75. In 2017, Bozkurt et al. [21] proposed a more sophisticated architecture combining deep convolutional and recurrent neural networks to classify skin layers within RCM stacks. The Toeplitz attention model was evaluated on 504 RCM stacks annotated by expert consensus. Patient-level separation was enforced. Image-wise classification accuracy reached 88.17%, with DEJ sensitivity $\approx 95.98\%$ and specificity $\approx 90.48\%$ [21].

Another work in this field is that of Robic et al. [50]. They introduce an approach for estimating the DEJ depth by integrating texture-based descriptors with the three-

dimensional spatial relationships among pixels in an RCM stack, enabling segmentation at the pixel level. Their framework relies on a random forest classifier combined with a 3D Conditional Random Field [51], which enforces spatial consistency in the predicted labels and incorporates the biologically plausible constraints of skin architecture. According to the reported evaluation, the method performs on par with leading techniques in the field, achieving a classification accuracy of 86%. Table 1 summarizes the main features of the papers reviewed in this paragraph.

Overall, DEJ and skin-layer delineation is a foundational step for RCM-based CAD and current AI approaches show promising accuracy in controlled settings. However, performance remains sensitive to acquisition variability, annotation heterogeneity, and limited external validation. Larger multicenter datasets, standardized ground truth definitions, and patient-level evaluation protocols are needed to ensure robust generalization.

Table 1. Main features of the papers reviewed in Section 3.1.

Reference	Input Data	Skin Site/Lesion Type	Dataset Size and Class Balance	Ground Truth Definition	Model Family	Validation Strategy	Performance Metrics
Kurugol 2011 [40]	Z-Stacks	Fair skin (low contrast)	Four data stacks from different subjects.	Expert-labeled boundaries	Classical ML: hybrid sequence segmentation + locally Smooth SVM (texture, wavelet features)	Two scenarios: same-stack and cross-stack training/testing. 6-fold cv.	Misclassification rate < 10%; Mean dist. from GT ~8.5 µm
Kurugol 2012 [42]	Z-Stacks	Fair and dark skin	24 stacks (15 fair, 9 dark)	Expert-labeled (inter-expert consistency evaluated)	Classical ML: SVM (same as [40]), using 170 texture features (GLCM, wavelets, Gabor, and spectral)	Comparison of automatic boundaries against manual GT	Mean dist.: ~6.4 µm (fair), ~5.3 µm (dark). Epidermis/dermis misclassification <10% for fair skin.
Kurugol 2015 [43]	Z-Stacks	Fair and dark skin	30 stacks (15 fair, 15 dark)	Expert-labeled by consensus (at least two expert readers)	Classical ML hybrid algorithms: intensity-based peak detection for dark skin, sequential segmentation combined with LS-SVM for fair skin	Comparison with [20]	Mean error: 7.9 ± 6.4 µm (dark), 8.3 ± 5.8 µm to 7.6 ± 5.6 µm (fair); classification accuracy 87–89% for dark skin.
Ghanta 2017 [20]	Z-Stacks	Fair and dark skin	10 RCM stacks (5 from dark skin and 5 from fair skin).	Manual segmentation (labels) of the DEJ boundary provided by an expert reader.	Math Model: an unsupervised generative Bayesian framework using a Marked Spatial Poisson Process (shape prior fitting hills/valleys). Dark skin: raw intensity values are used; for fair skin, texture features are extracted.	Comparing the automatically detected 3D DEJ boundary against GT	Mean error: 5.41 ± 3.94 µm (dark), 12.1 ± 7.0 µm (fair).
Somoza 2014 [45]	Z-Stacks	Healthy human skin	Small evaluation dataset (three stacks); five strata classes	Expert visual inspection and manual labeling of the stacks.	Unsupervised ML: Texton histogram + PCA + K-means clustering	Comparison of the automated five-layer separation against expert-graded GT	High correlation and promising results reported (qualitative)
Hames 2015 [19]	Z-Stacks	Healthy human skin from the dorsal (sun-exposed) and volar (less sun-exposed) forearms	308 depth stacks from 54 volunteers	Weak labels provided by an experienced dermatologist. Four strata: stratum corneum, viable epidermis, DEJ, and papillary dermis	Classical ML: Bag-of-features + hierarchical/spherical K-means + CRF + structured SVM	A held-out test set of 18 participants (stratified by age)	Section classification accuracy: 85.7%. Age-related DEJ thinning and stratum corneum thickening detected.

Table 1. *Cont.*

Reference	Input Data	Skin Site/Lesion Type	Dataset Size and Class Balance	Ground Truth Definition	Model Family	Validation Strategy	Performance Metrics
Kaur 2016 [49]	Z-Stacks	Normal human skin	15 stacks (1500 images).	Manual labeling by experts, six classes: stratum corneum, granular, spinous, basal, DEJ, dermis.	Deep Learning: Hybrid DNN against standard CNN (input: texton features, not raw images)	Five-fold cross-validation + test set	Test accuracy: 81.73% Sensitivity 71.74%, Specificity 96.20%, Precision 72.36%, F-score 0.71.
Bozkurt 2017 [21]	Z-Stacks	Normal, benign melanocytic, diseased skin	504 stacks (large dataset)	Manual labeling of each image by experts.	Deep Learning: CNN + RNN (Recurrent Neural Network) + Toeplitz vs. Global Attention models	Patient-wise hold-out cv (training, validation, and testing sets of 245, 61, and 198 stacks respectively).	Accuracy: 87.97% (base), 88.17% (with attention)
Robic 2019 [50]	Z-Stacks	Normal cheek skin, targeting chronological aging.	23 stacks	Pixel-level manual labeling of three layers: epidermis, an "uncertain area" containing the DEJ, and dermis	Classical ML: Random Forest + 3D Conditional Random Field (CRF). Texture features: statistics, power spectrum, GLCM, Gabor filters, Laplacian variance	Performance comparison of 3D against 2D CRF and standalone RF baselines, with subject-wise 10-fold cv	Acc.: 90%, 54%, 75%; Sens.: 90%, 68%, 93%; Spec.: 96%, 92%, 95% (resp. epidermis, uncertain/DEJ/dermis)

3.2. AI for Skin Layer and DEJ Delineation, and Tissue/Pattern Segmentation

In the previous section, we reported the main automatic and semi-automatic systems representing the current state of the art for the identification of skin layers. These works mainly focus on RCM studies of healthy skin, with the exception of Bozkurt [21,52,53], who also included various types of lesions in the dataset.

In this section, we aim to describe the principal methods at the state-of-the-art for the segmentation/classification of skin mosaic acquired with the RCM technique for melanocytic lesions. This topic is of fundamental importance because it represents the starting point for all subsequent algorithms for the discrimination of skin cancers and, in particular, melanoma detection.

The first work that connects the study of confocal images of melanoma with DL for strata segmentation and classification is by Gareau et al. [54].

The study focuses on superficial spreading melanoma (SSM), employing a high-resolution image-analysis framework that leverages two hallmark RCM criteria: the presence of pagetoid melanocytes within the epidermis and architectural disruption at the DEJ. Using multiple stacks acquired from melanocytic lesions and nevi, the method estimates the surface Dsps(x,y), defined as the voxel with the highest reflectance along the z-axis for each x–y coordinate. This allows for the automatic quantification and localization of the most superficial pigmented layer—Dsps(x,y)—which corresponds to pagetoid melanocytes in the epidermis and pigmented basal cells near the DEJ. Because melanin concentration changes markedly when transitioning from the epidermis to the basal layer, the DEJ can be approximated, outlined, and visualized through the Dsps(x,y) surface.

The automated annotation of RCM images for segmentation tasks has only rarely been reported, likely due to the scarcity of labeled datasets. In 2016, Kose et al. [55] introduced a machine-learning approach capable of distinguishing six common morphological patterns in RCM images (background, meshwork, ring, clod, mixed, and aspecific). Their workflow involved dividing 20 suspicious RCM mosaics into 0.5 mm $\times$ 0.5 mm tiles. Noise reduction was achieved using an isotropic Gaussian low-pass filter, while an adaptive histogram equalization enhanced reflectance contrast. For each tile, texture was characterized using SURF descriptors, which were encoded into histograms. A support vector machine classifier was then trained on a subset of annotated mosaics, treating each tile as an independent sample. The reported performance ranged from 55 to 81% sensitivity and 81–89% specificity across the six morphological categories.

The first deep learning (DL) contribution in this domain was presented by Kose et al. in 2017 [56]. They developed a DL-based classifier to differentiate the morphological patterns observed in RCM mosaics of 20 melanocytic lesions at the DEJ. Three CNN-based strategies were evaluated for identifying the six established pattern types: background, meshwork, ring, clod, mixed, and aspecific. In the first experiment, a CNN pre-trained on natural images was fine-tuned for the RCM task. In the second, the authors inserted an additional fully connected layer before the output and retrained only the final two layers. In the third approach, the convolutional layers of the CNN were used solely as feature extractors; the resulting descriptors were encoded using a bag-of-words model and classified with a support vector machine. Bozkurt et al. [53] extended DL implementation using a multiresolution "nested encoder–decoder" CNN. They aimed to identify the six morphological patterns with three U-Net [57] subnetworks nested together, with mosaics of original size and a downsampling of ½ and ¼. The dataset is composed of 56 RCM mosaics from DEJ, collected from suspicious melanoma lesions. The 3-UNet model achieves an overall accuracy of 73%, which is almost 4% better than its closest competitor.

In 2020, Kose et al. [25] conducted a study based on a set of 117 RCM mosaics obtained from melanocytic skin lesions, recorded at the dermo–epidermal junction. The aim was

to classify these mosaics into six categories considered clinically relevant. Four of these classes represent different cellular configurations that are observable under the microscope—arranged in a ring, with a reticular structure, organized in clusters, or lacking a defined pattern—while the last two categories were added to distinguish both areas of healthy skin and regions where image quality is compromised by technical artifacts. In this study, MED-Net was used, an automated semantic segmentation method. MED-Net was trained and tested on 117 mosaics using 5-fold cross-validation, with pixel-wise annotation by two RCM experts who reached consensus. Performance was quantified by sensitivity (82%), specificity (93%), and Dice coefficient (0.78). The model was further applied to an independent dataset of 372 mosaics for quality assessment. The results showed that MED-net modestly outperforms the other networks (DNN-SegNet-Deeplab-UNet) in terms of sensitivity; however, the comparison between models highlights variations based on the different categories analyzed. In terms of specificity, all four networks show comparable performance, both when looking at the overall average value and when considering individual classes. A distinctive feature of MED-Net is its tendency to extend the labels assigned to an area to immediately adjacent areas that have not yet been annotated, producing more uniform segmentation maps than the other solutions examined. In addition, the model proved particularly effective in recognizing the absence of specific cellular patterns, with a limited number of false positives.

In 2021, D'Alonzo et al. [58] introduced a weak supervision machine learning method to distinguish two types of regions in RCM mosaics: 'benign' and 'non-specific'. The weakly supervised approach was trained on 157 mosaics (one per patient) with a patient-level split (70/10/20%). Pixel-wise labels were used only for evaluation, while training relied on patch-level labels. Performance reached AUC = 0.969 and Dice coefficient = 0.778 for segmentation, with patch-level classification accuracy $\approx$ 88.4%.

The evidence shows that the use of techniques based on CAM can help make the diagnostic process more efficient and, from a clinical perspective, potentially reduce melanoma-related mortality, as it allows for the timely identification and reporting of all suspicious lesions that require biopsy or further specialist examination. It is plausible that, in the future, this type of approach will be integrated into structured risk assessment systems, thereby increasing the reliability and/or safety of diagnoses even when the operator does not have advanced dermatological expertise.

In 2023, Mandal et al. [59] demonstrated that it is possible to achieve high accuracy in the automatic classification of Lentigo Maligna melanoma and atypical intraepidermal melanocytic lesions using stacks of RCM images, whose diagnostic validity was already confirmed by clinical evaluation and biopsy analysis. The study included 110 patients and 517 biopsy-confirmed RCM stacks. Patient-level splitting was applied during 5-fold cross-validation to avoid leakage. The best-performing DenseNet169 model achieved a test accuracy of 0.80, with the ROC curves and AUC reported. The research team compared several pre-trained convolutional neural networks (CNN) and also created a lighter CNN model by combining the features extracted from the networks with traditional machine learning classifiers. The study included 110 patients: pigmented lesions considered suspicious according to clinical criteria were analyzed with a portable RCM device (Vivascope 3000), and subsequently, the portions of tissue recognized as atypical were biopsied and subjected to histopathological diagnosis for final confirmation.

The model aimed to deliver a binary classification that could differentiate between intraepidermal melanocytic proliferation and Lentigo Maligna Melanoma samples with a test accuracy of 0.80.

The key characteristics of the studies discussed in this section are summarized in Table 2.

Table 2. Main features of the papers reviewed in Section 3.2.

Reference	Input Data	Skin Site/Lesion Type	Dataset Size and Class Balance	Ground Truth Definition	Model Family	Validation Strategy	Performance Metrics
Gareau 2010 [54]	Stacks	Superficial Spreading Melanoma (SSM), and benign nevi (including junctional lentigenous and congenital nevi)	Ten skin sites in total: five unequivocal SSMs and five nevi	Histopathology and review by an expert confocal pathologist to confirm unequivocal morphology (pagetoid melanocytes [PM], DEJ disarray)	Surface fitting to isolate the most superficial pigmented surface (D_SPS) based on reflectivity, and quantify malignant features.	Quantitative comparison of metrics (PM count and DEJ roughness) between the SSM and nevi groups by statistical significance tests.	PMs identified in all five SSMs; none in the nevi. DEJ roughness significantly higher in SSM (11.7 ± 3.7) compared to nevi (5.5 ± 1.0)
Kose 2016 [55]	RCM mosaics divided into localized processing areas	Melanocytic skin lesions, specifically focusing on patterns at the DEJ	Twenty mosaics, with six classes (background, meshwork, ring, clod, mixed and aspecific.)	Manual labeling of each image by experts	Classical ML: SURF descriptors (texture) + SVM	Hold out	Sensitivity: 55–81%; Specificity: 81–89% (class dependent)
Kose 2017 [56]	Mosaics (at DEJ) divided into smaller localized "tiles"	Melanocytic lesions	Twenty mosaics, with six classes (background, meshwork, ring, clod, mixed, and aspecific)	Manual labeling of the six morphological patterns by two expert users.	DL: CNN Fine-tuning vs. CNN + fully connected layer and partial retraining vs. CNN as Feature Extractor + BoW + SVM	Comparison of three CNN strategies against each other and against [55]	Results comparable to [55], with sensitivity from 55% to 81% and specificity from 81% to 89% (class-dependent)
Bozkurt 2018 [53]	Downsampled RCM mosaics. Processing performed using 256×256 pixel tiles with 75% overlap.	Melanoma-suspicious melanocytic lesions. Six morphological patterns at the DEJ: non-lesion/background, artifact, meshwork, ring, nested (clod), and Aspecific	Fifty-six mosaics. Class imbalance handled using coefficients in the loss function to weight under-represented classes.	Consensus labeling by two expert readers. Partial labels: experts annotated only representative regions of the large mosaics, rather than every pixel.	MUnet, an encoder–decoder CNN composed of multiple nested U-Net sub-networks. Trainable on partially labeled data.	Hold-out cv (46 mosaics for training and 10 for testing)	Sensitivity: from 50.52% to 93.86%. Specificity: from 90.38% to 97.88% More coherent segmentation compared to baseline models.
Kose 2020 [25]	RCM mosaics (at various skin layers) and coregistered dermoscopy images	Atypical pigmented lesions (including melanocytic neoplasms)	A total of 117 mosaics from seven clinics (training/validation) and a test set of 372 coregistered RCM-dermoscopic image pairs (from five clinics); six classes (included Healthy and Artifacts)	Pixel-level manual labeling (semantic segmentation) performed by two RCM experts who reached consensus	Deep Learning: MED-Net Multiscale Encoder–Decoder Network, a L CNN.	Comparison vs. DNN, SegNet, DeepLab, U-Net	Modest sensitivity improvement; few false positives. High performance in delineating uninformative areas: Sensitivity: 82%. Specificity: 93%.

Table 2. *Cont.*

Reference	Input Data	Skin Site/Lesion Type	Dataset Size and Class Balance	Ground Truth Definition	Model Family	Validation Strategy	Performance Metrics
D'Alonzo 2021 [58]	RCM mosaics collected at the dermal-epidermal junction (DEJ). Images processed as small patches	Pigmented lesions, "benign" (regular architectural patterns) and "aspecific" (disrupted architecture for melanoma, injury, or inflammation).	A total of 157 mosaics (1 per patient)	Weak labels provided by expert readers. Uses patch-level labels (identifying whether a specific FOV contains a certain pattern).	Weakly Supervised ML: Patch-based + CAM (Class Activation Maps)	Training (70%), validation (10%), and test (20%) sets.	Area Under the Curve (AUC) of 0.969 and Dice coefficient of 0.778
Mandal 2023 [59]	Stacks (Vivascope 3000). LZP to project 3D stacks into 2D images while preserving diagnostic info.	Sun-exposed skin. Lentigo Maligna vs. Atypical Intraepidermal Melanocytic Proliferation.	A total of 110 patients, 517 RCM stacks	Histopathology (Biopsy)	Hybrid DL: Pre-trained CNNs + Lightweight CNN w/ML classifiers (SVM/KNN)	k-fold cross-validation + test	Validation AUC of 0.88 and accuracy of 81% for distinguishing LM from AIMP. Accuracy on test: 80%

AI methods for tissue and morphological pattern segmentation have progressed from handcrafted descriptors to deep learning models that can learn clinically meaningful micro-architectural cues. The main bottlenecks are the scarcity and inconsistency of pixel-level annotations, variability across body sites and devices, and the presence of non-informative regions and artifacts. Future work should prioritize standardized annotation schemes, robust quality-control components, and validation across diverse clinical settings.

3.3. AI for Lesion-Level Diagnostic Classification

Finally, the last but the most important challenge for the discrimination of benign common nevi from malignant melanoma one is the implementation of an automatic system that is able to do this starting from the most informative skin stratus, such as DEJ. One of the first studies on automated RCM image analysis systems is by Wiltgen et al. [60]. Their aim was to check the discrimination power of texture features for the automatic identification of significant diagnostic regions in CLSM views of malignant melanoma (50 benign common nevi vs. 50 malignant melanoma). To do this, they used Tissue Counter Analysis [61], subdivided in a feature extraction step (in the spatial and frequency domains) from squared subregions of the entire mosaic, a classification step with the Classification and Regression Trees procedure [62], and a relocation final step in order to visualize the classification results for single subregions in a complete overview. Features based on the wavelet transform provide the best results, with 96.0% correctly classified elements from benign common nevi and 97.0% from malignant melanoma. The same research group, in 2009 [63], proposed a CAD system that analyses each RCM image, dividing it into square elements of an equal size of 256 $\times$ 256 pixels. Following the Tissue Counter Analysis procedure, several function classifiers are tested, such as linear logistic regression model, sequential minimal optimization algorithm for support vector classification and multilayer perceptron, but the best result came from the Classification and Regression Trees procedure algorithm. In 2010, Koller et al. [18] worked on the same task, implementing a machine learning algorithm based on image analysis features with wavelet transform and the Classification and Regression Trees procedure. For all 209 skin tumors (51 melanoma and 158 benign nevi), they collected 16.269 RCM images, classified into the stratum corneum, the granulosum/spinosum layers and the dermo-epidermal junction zone. Using all data, without differentiation of strata, the model provided a correct classification of all 24 melanoma in the test set and of 23.53% of the benign melanocytic skin lesions (area under the curve of 0.638). A separate analysis of the different skin layers shows that for the DEJ images in the test set, 47.66% of the melanoma and 62.67% of the nevi images were correctly categorized.

Starting from the same dataset used in [60], their research group proposed, in 2016 [64], the newest application to distinguish malignant melanoma from benign melanocytic lesions based on CNN. They consider all RCM images of skin lesions in layers of various depths (not only DEJ, as in previous studies). The LeNet-5 convolutional neural network [65] was implemented and trained, achieving 93% accuracy on the unseen test set.

In 2019, Wodzinski et al. [66] designed their melanoma recognition algorithm, based on fine-tuning a ResNet and pre-trained on ImageNet [67], which was tested on the RCM mosaics of 429 subjects with MM, basal cell carcinoma and naevus. They used the entire downsampled mosaics with a 10-fold cross-validation method, achieving an overall accuracy of 0.82 $\pm$ 0.02.

Table 3 provides an overview of the main features of the papers reviewed above.

Table 3. Main features of the papers reviewed in Section 3.3.

Reference	Input Data	Skin Site/Lesion Type	Dataset Size and Class Balance	Ground Truth Definition	Model Family	Validation Strategy	Performance Metrics
Wiltgen 2008 [60]	Mosaics (subdivided into square subregions for feature extraction)	Malignant Melanoma vs. Benign Common Nevi	A total of 100 lesions (50 melanoma, 50 nevi), and a total of 6897 images	Images from the DEJ and labeled based on visual comparison of typical cellular and architectural morphological structures.	Classical ML: tissue counter analysis + wavelet features + CART (Classification and Regression Trees)	Relocation/Visualization step	Correct classification: 96.0% nevi, 97.0% melanoma
Wiltgen 2011 [63]	Mosaics	Malignant Melanoma vs. Benign Nevi	Same dataset as [60] (50 vs. 50), with a total of 6897 images	Diagnosis (implied histopathology)	Classical ML on wavelet features: Bayes-, tree-, rule-, function (numeric)-, and lazy-classifiers. Best: CART	Hold-out	CART identified as best-performing and transparent (accuracy similar to [60])
Koller 2010 [18]	Single RCM images (from stacks)	Melanoma vs. Benign Nevi	A total of 209 tumors (16,269 images) from 178 patients: 51 melanoma (4669 images); 158 Nevi (11,600 images);	Histopathological assessment for all 51 melanomas and 67 of the nevi; the remaining 91 nevi were diagnosed based on clinical/dermoscopic criteria.	Classical ML: wavelet features + CART	Hold-out with a test set	Test set: 55.68% of melanoma images were classified as malignant, while 46.71% of benign nevi images were also classified as malignant.
Wiltgen 2016 [64]	Images from various skin layers. For the DL approach, images were resized to 64 × 64.	Malignant Melanoma vs. Benign Nevi	Same dataset as [60] (50 vs. 50), with a total of 6897 images	Histopathology	Multires analysis with wavelet trans. + CART, and CNN (LeNet-5 architecture)	Hold-out with a test set; dropout to control overfitting	Accuracy: 97% during validation; 81% on the test set
Wodzinski 2019 [66]	Entire mosaic (downsampled to 1024 × 1024 or 2048 × 2048 for processing)	MM, Basal Cell Carcinoma (BCC), Nevi	A total of 429 subjects (110 BCC, 160 MM, and 160 NE)	Histopathology	Deep Learning: ResNet (Pre-trained on ImageNet, Fine-tuned)	10-fold cv + test	Test set accuracy 87%; specific F1-scores: 0.91 BCC, 0.80 MM, 0.86 NE.

Lesion-level diagnostic classification from RCM has shown clear potential, but most studies are still limited by small, single-center cohorts and heterogeneous reference standards, which can inflate performance and reduce reproducibility. Evidence would be strengthened by external testing on independent cohorts, explicit patient-level separation, and the reporting of clinically relevant metrics (including sensitivity/specificity and calibration). Ultimately, clinically deployable systems will likely require interpretable outputs and integration within real-world workflows, potentially alongside complementary modalities.

Figure 2 illustrates a conceptual AI-assisted diagnostic pipeline that synthesizes the reviewed literature into a coherent clinical workflow. The central pathway depicts the current state-of-the-art approach, which operates sequentially: it begins with an automated Quality Control step to filter out artifacts and non-informative regions immediately after acquisition. This is followed by Navigation and Localization algorithms that automatically identify the DEJ within the image stack, ensuring that analysis focuses on the correct histological level. Subsequently, Tissue Segmentation models isolate specific diagnostic criteria, such as pagetoid cells or architectural disarray, which serve as inputs for the final Lesion Diagnosis step to classify the lesion as benign or pathological. Parallel to this sequential process, the dashed lines on the right represent the Future Outlook towards end-to-end deep learning and multimodal integration. Unlike the multi-step approach, these next-generation models aim to process raw data directly into diagnostic risk scores, potentially learning latent features without the need for explicit intermediate segmentation, thereby streamlining the decision support system.

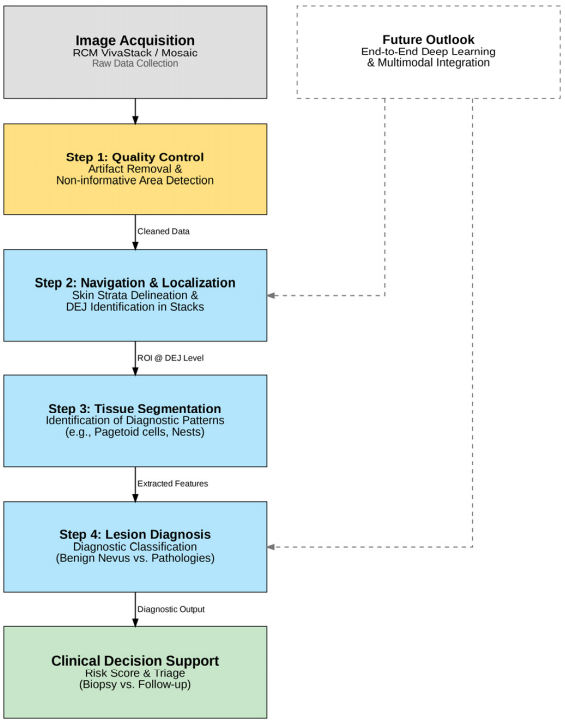

Figure 2. Conceptual AI-assisted RCM diagnostic pipeline. The diagram maps the AI methods reviewed in this study onto a sequential clinical workflow. Step 1 (Quality Control) ensures the validity of the acquisition by detecting artifacts or non-informative areas. Step 2 (Navigation) automates the localization of the dermal–epidermal junction (DEJ). Step 3 (Segmentation) isolates clinically relevant morphological patterns. Step 4 (Diagnosis) provides a classification output. The dashed section (Future Outlook) represents the potential for next-generation end-to-end models to integrate these components into a unified decision support system.

3.4. Limitations

AI-based methods also have inherent limitations that should be considered their before clinical translation. Model performance depends strongly on the quality, representativeness, and labeling reliability of the training data, and can be influenced by class imbalance, selection bias, and annotation variability (e.g., single-reader versus consensus, or different reference standards). Deep models are prone to overfitting when trained on limited cohorts, and the results can be overly optimistic if patient-level separation is not strictly enforced or if subtle data leakage occurs. Importantly, the domain shift across acquisition protocols, body sites, devices, and patient populations may lead to performance degradation when algorithms are deployed outside the setting in which they were developed. These limitations highlight the need for standardized acquisition and annotation practices, the transparent reporting of evaluation design, and robust external prospective validation across centers, devices, and diverse skin types prior to routine clinical use.

3.5. Critical Appraisal and Comparison with Dermoscopy

The consolidation of studies across skin strata delineation, tissue segmentation, and diagnostic classification (Tables 1–3) reveals distinct evolutionary trends in the application of AI to Reflectance Confocal Microscopy. By analyzing these works, several key shifts in model architecture, input data complexity, and validation rigor become apparent. For example, the most prominent trend is the transition from classical ML pipelines to DL architectures (and, consequently, from hand-crafted features to end-to-end learning). Another important aspect is the evolution of input data from patches to the 3D context. Finally, ground truth definitions move from expert consensus toward histopathological validation. Several other important aspects deserve attention.

In what follows, external validation denotes the evaluation on independent, patient-level separated cohorts, ideally multicenter cohorts, with rigorous leakage control. Artificial Intelligence (AI) for dermoscopy frequently reports an AUC (ROC) > 0.80 for melanoma detection, typically supported by large, multicenter datasets and external validation, which enhances generalizability (e.g., systematic reviews such as that of Patel et al., 2023 [5]). A summary of the key differences between dermoscopy-based and RCM-based AI approaches is provided in Table 4. By contrast, the current evidence for RCM-based AI is limited in scale and validation: most studies rely on small, single-center cohorts and heterogeneous acquisition/annotation protocols, with scarce external testing. As a result, conclusions from dermoscopy cannot be directly translated to RCM without modality-specific validation and workflow considerations.

Table 4. Dermoscopy- vs. RCM-based AI for melanoma: dataset scale, validation, performance, labeling, domain shift, and transferability (References as numbered in the manuscript).

Aspect	Dermoscopy-Based AI	RCM-Based AI
Dataset scale and provenance	Large cohorts; often multicenter; public datasets and benchmarks available.	Small cohorts; typically single-center; limited public data; heterogeneity across studies.
External validation	Frequent (independent test sets; multicenter external testing).	Limited/rare; studies mainly report internal cross-validation or single-center splits.

Table 4. *Cont.*

Aspect	Dermoscopy-Based AI	RCM-Based AI
Typical performance	AUC > 0.80 for melanoma detection (systematic reviews) [5].	Heterogeneous: pattern/strata tasks ≈ 0.73–0.88; lesion-level classification ≈ 0.80 in controlled settings (e.g., Kose, Bozkurt, D'Alonzo, Mandal, Wodzinski) [25,53,55,56,58,59,66].
Labeling and ground truth	Image-level labels with consensus; increasing use of curated datasets.	Pixel-/patch-level labels are often weak or single-reader; variable standards across tasks (DEJ, strata, patterns).
Domain shift sensitivity	Moderate; mitigated by dataset size/variety.	High: performance varies across devices, body sites, lesion types, and acquisition quality.
Transferability	Higher (due to scale and external validation).	Limited at present; conclusions from dermoscopy do not automatically generalize to RCM.
Representative studies	Patel et al., 2023 (systematic review; dermoscopy AUC > 0.80) [5].	Kose 2016/2017/2020; Bozkurt 2018; D'Alonzo 2021; Mandal 2023; Wodzinski 2019 [25,53,55,56,58,59,66].

In terms of dataset size and design, dermoscopy studies often include thousands of lesions from multiple centers, whereas RCM datasets are modest and single-center: examples include Kose et al.'s pattern studies on 20 mosaics (DEJ, six patterns) and 56 mosaics (nested U Net) [53,55,56], Kose et al.'s 2020 quality assessment on 117 mosaics [25], D'Alonzo et al.'s 2021 study of weakly supervised segmentation on 157 mosaics (one mosaic per patient) [58], Mandal et al., 2023, who performed a lesion-level classification on 110 patients (LM vs. atypical intraepidermal lesions) [59], and Wodzinski et al., 2019, who performed a CNN classification of 429 subjects (mixed diagnoses) [66]. Even where accuracy figures are encouraging (e.g., strata classification ≈ 0.82–0.88; DEJ segmentation improvements; lesion level accuracy ≈ 0.80 in a controlled setting), the limited external validation and single-center sampling constrain generalizability and raise concerns about potential overfitting.

Methodologically, RCM AI faces labeling heterogeneity (pixel level vs. section level, single reader vs. consensus), patient-level separation issues, and domain shift across devices, body sites, and populations, all of which can degrade performance outside the development environment. These factors, highlighted in the prior RCM literature and in our Limitations section, underscore the need for curated multi-institutional datasets, standardized ground truth at the pixel and patient levels, transparent reporting (strict patient level separation; leakage control), and prospective external validation across centers and devices. Practical clinical deployment will likely benefit from quality-control modules (to detect non-informative regions/artifacts during acquisition) and from multimodal strategies (e.g., combining RCM + dermoscopy) to improve robustness and translation.

4. Conclusions

Progress in AI pushes towards the implementation of modern and effective systems to support diagnosis in the biomedical field. If, until now, numerous models have been proposed for dermoscopic images, with the advent and the increasing diffusion of new

imaging systems such as RCM in the hospital setting, it is necessary to develop new algorithms that put together confocal microscopy applied to the skin lesions with CAD systems. We have seen that, for many research groups, the target points currently focus on the identification of the skin layers and, in particular, of the DEJ, as an informative layer, as they are more involved in the proliferation of malignant cells. These systems exploit transversal RCM acquisitions, mostly of healthy skin, in order to define a three-dimensional profile useful to the clinician. Another branch of studies, instead, aims to segment the DEJ into six main characteristic patterns of DEJ skin level. The identification of any aspecific tissue, characterized by disarray and the presence of melanocytic cells, is the starting point for a research work that aimed to classify the RCM of benign nevi with respect to malignant melanomas. This represents the last step and the most important; however, there are still few works focused on this aspect.

While dermoscopy-based AI frequently reports AUC > 0.80 in large, multi-center cohorts with external validation, RCM-based AI remains exploratory, with small, single-center datasets, non-standardized annotations, and limited generalizability. This distinction is crucial for correctly positioning the field: dermoscopy benchmarks do not readily transfer to RCM. The clinical translation of RCM AI will depend on curated multi-institutional datasets, harmonized ground truth, prospective external validation across devices and diverse skin types, and practical workflow integration (including acquisition quality modules). Multimodal strategies combining RCM + dermoscopy may further enhance robustness (see Section 3.4 and Table 4 for comparative evidence).

This summary of the state of the art in the field of RCM and AI aims to lay the foundations for the development of a CAD system to help the clinician during the acquisition and diagnosis of any malignant lesion. The first clinical difficulty, in fact, lies in the identification of the DEJ during a routine RCM acquisition. Once this layer has been identified, a segmentation step focused on the type of tissue can certainly facilitate the clinician in identifying any atypical patterns. From here, finally, classification systems of this aspecific pattern can be useful to discern the various types of lesions that can be detected.

Future work should prioritize multi-institutional curated datasets, standardized ground truth and reporting, and prospective evaluations. Multimodal systems integrating RCM with dermoscopy and/or OCT may further improve robustness. Crucially, algorithms should be validated across ethnically diverse skin types and across different RCM devices and settings, alongside practical implementation requirements (workflow integration, clinician acceptance, and regulatory considerations), before routine clinical adoption.

Author Contributions: Conceptualization, L.C. and G.D.N.; methodology, L.C., A.F., L.S. and E.S.C.; software, L.C.; validation, L.C., A.F., L.S., E.S.C. and L.P.; formal analysis, L.C.; investigation, L.C., A.F., L.S., E.S.C., L.P. and R.R.; resources, L.C., A.F., L.S., E.S.C., L.P. and R.R.; data curation, L.C., A.F., L.P., R.R., U.D.G., D.C., G.D.N. and M.C.; writing—original draft preparation, L.C., A.F., L.S. and G.D.N.; writing—review and editing, L.C., A.F., L.P., U.D.G., D.C., G.D.N. and M.C.; visualization, L.C., L.P., U.D.G., D.C. and G.D.N.; supervision, L.P., R.R., U.D.G., D.C., G.D.N. and M.C.; project administration, L.C., L.P., U.D.G., D.C., G.D.N. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ulrich, M.; Lange-Asschenfeldt, S.; Gonzalez, S. The Use of Reflectance Confocal Microscopy for Monitoring Response to Therapy of Skin Malignancies. *Dermatol. Pract. Concept.* **2012**, *2*, 43–52. [CrossRef]
2. Calzavara-Pinton, P.; Longo, C.; Venturini, M.; Sala, R.; Pellacani, G. Reflectance Confocal Microscopy for in Vivo Skin Imaging. *Photochem. Photobiol.* **2008**, *84*, 1421–1430. [CrossRef]
3. Hoffmann-Wellenhof, R.; Pellacani, G.; Malvehy, J.; Soyer, H. *Reflectance Confocal Microscopy for Skin Diseases*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
4. Pezzini, C.; Kaleci, S.; Chester, F.; Farnetani, F.; Longo, C.; Pellacani, G. Reflectance Confocal Microscopy Diagnostic Accuracy for Malignant Melanoma in Different Clinical Settings: Systematic Review and Meta-Analysis. *J. Eur. Acad. Dermatol. Venereol.* **2020**, *34*, 2268–2279. [CrossRef]
5. Patel, R.H.; Foltz, E.A.; Witkowski, A.; Ludzik, J. Analysis of Artificial Intelligence-Based Approaches Applied to Non-Invasive Imaging for Early Detection of Melanoma: A Systematic Review. *Cancers* **2023**, *15*, 4694. [CrossRef]
6. Caldo, D.; Bologna, S.; Conte, L.; Amin, M.S.; Anselma, L.; Basile, V.; Hossain, M.M.; Mazzei, A.; Heritier, P.; Ferracini, R.; et al. Machine Learning Algorithms Distinguish Discrete Digital Emotional Fingerprints for Web Pages Related to Back Pain. *Sci. Rep.* **2023**, *13*, 4654. [CrossRef]
7. Vitale, E.; Lupo, R.; Calabrò, A.; Cornacchia, M.; Conte, L.; Marchisio, D.; Caldararo, C.; Carvello, M.; Carriero, M.C. Mapping Potential Risk Factors in Developing Burnout Syndrome between Physicians and Registered Nurses Suffering from an Aggression in Italian Emergency Departments. *J. Psychopathol.* **2021**, *27*, 148–155. [CrossRef]
8. De Nunzio, G.; Conte, L.; Lupo, R.; Vitale, E.; Calabrò, A.; Ercolani, M.; Carvello, M.; Arigliani, M.; Toraldo, D.M.; De Benedetto, L. A New Berlin Questionnaire Simplified by Machine Learning Techniques in a Population of Italian Healthcare Workers to Highlight the Suspicion of Obstructive Sleep Apnea. *Front. Med.* **2022**, *9*, 866822. [CrossRef]
9. Conte, L.; Lupo, R.; Lezzi, A.; Sciolti, S.; Rubbi, I.; Carvello, M.; Calabrò, A.; Botti, S.; Fanizzi, A.; Massafra, R.; et al. Breast Cancer Prevention Practices and Knowledge in Italian and Chinese Women in Italy: Clinical Checkups, Free NHS Screening Adherence, and Breast Self-Examination (BSE). *J. Cancer Educ.* **2024**, *40*, 30–43. [CrossRef] [PubMed]
10. Conte, L.; Rizzo, E.; Civino, E.; Tarantino, P.; De Nunzio, G.; De Matteis, E. Enhancing Breast Cancer Risk Prediction with Machine Learning: Integrating BMI, Smoking Habits, Hormonal Dynamics, and BRCA Gene Mutations—A Game-Changer Compared to Traditional Statistical Models? *Appl. Sci.* **2024**, *14*, 8474. [CrossRef]
11. Lupo, R.; Vitale, E.; Panzanaro, L.; Lezzi, A.; Lezzi, P.; Botti, S.; Rubbi, I.; Carvello, M.; Calabrò, A.; Puglia, A.; et al. Effects of Long COVID on Psycho-Physical Conditions in the Italian Population: A Statistical and Large Language Model Combined Description. *Eur. J. Investig. Heal. Psychol. Educ.* **2024**, *14*, 1153–1170. [CrossRef] [PubMed]
12. Conte, L.; Rizzo, E.; Grassi, T.; Bagordo, F.; De Matteis, E.; De Nunzio, G. Artificial Intelligence Techniques and Pedigree Charts in Oncogenetics: Towards an Experimental Multioutput Software System for Digitization and Risk Prediction. *Computation* **2024**, *12*, 47. [CrossRef]
13. Conte, L.; Caruso, G.; Philip, A.K.; Cucci, F.; De Nunzio, G.; Cascio, D.; Caffo, M. Artificial Intelligence-Assisted Drug and Biomarker Discovery for Glioblastoma: A Scoping Review of the Literature. *Cancers* **2025**, *17*, 571. [CrossRef]
14. Conte, L.; Amodeo, I.; De Nunzio, G.; Raffaeli, G.; Borzani, I.; Persico, N.; Griggio, A.; Como, G.; Cascio, D.; Colnaghi, M.; et al. Congenital Diaphragmatic Hernia: Automatic Lung and Liver MRI Segmentation with NnU-Net, Reproducibility of Pyradiomics Features, and a Machine Learning Application for the Classification of Liver Herniation. *Eur. J. Pediatr.* **2024**, *183*, 2285–2300. [CrossRef] [PubMed]
15. Cortesi, L.; Galli, G.R.; Domati, F.; Conte, L.; Manca, L.; Berio, M.A.; Toss, A.; Iannone, A.; Federico, M. Obesity in Postmenopausal Breast Cancer Patients: It is Time to Improve Actions for a Healthier Lifestyle. The Results of a Comparison Between Two Italian Regions with Different "Presumed" Lifestyles. *Front. Oncol.* **2021**, *11*, 769683. [CrossRef] [PubMed]
16. Luck, B.L.; Carlson, K.D.; Bovik, A.C.; Richards-Kortum, R.R. An Image Model and Segmentation Algorithm for Reflectance Confocal Images of in Vivo Cervical Tissue. *IEEE Trans. Image Process.* **2005**, *14*, 1265–1276. [CrossRef] [PubMed]
17. Kurugol, S.; Dy, J.G.; Rajadhyaksha, M.; Brooks, D.H. Localizing the Dermis/Epidermis Boundary in Reflectance Confocal Microscopy Images with a Hybrid Classification Algorithm. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2009; pp. 1322–1325.
18. Koller, S.; Wiltgen, M.; Ahlgrimm-Siess, V.; Weger, W.; Hofmann-Wellenhof, R.; Richtig, E.; Smolle, J.; Gerger, A. In Vivo Reflectance Confocal Microscopy: Automated Diagnostic Image Analysis of Melanocytic Skin Tumours. *J. Eur. Acad. Dermatol. Venereol.* **2011**, *25*, 554–558. [CrossRef]
19. Hames, S.C.; Ardigo, M.; Soyer, H.P.; Bradley, A.P.; Prow, T.W. Anatomical Skin Segmentation in Reflectance Confocal Microscopy with Weak Labels. In *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, SA, Australia, 23–25 November 2015*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2015; pp. 1–8.

20. Ghanta, S.; Jordan, M.I.; Kose, K.; Brooks, D.H.; Rajadhyaksha, M.; Dy, J.G. A Marked Poisson Process Driven Latent Shape Model for 3D Segmentation of Reflectance Confocal Microscopy Image Stacks of Human Skin. *IEEE Trans. Image Process.* **2017**, *26*, 172–184. [CrossRef]
21. Bozkurt, A.; Kose, K.; Coll-Font, J.; Alessi-Fox, C.; Brooks, D.H.; Dy, J.G.; Rajadhyaksha, M. Delineation of Skin Strata in Reflectance Confocal Microscopy Images Using Recurrent Convolutional Networks with Toeplitz Attention. *arXiv* **2017**, arXiv:1712.00192. [CrossRef]
22. Minsky, M. Memoir of Inventing the Confocal Scanning Microscope. *Scanning* **1988**, *10*, 128–138. [CrossRef]
23. Rajadhyaksha, M.; Marghoob, A.; Rossi, A.; Halpern, A.C.; Nehal, K.S. Reflectance Confocal Microscopy of Skin in Vivo: From Bench to Bedside. *Lasers Surg. Med.* **2017**, *49*, 7–19. [CrossRef]
24. Shahriari, N.; Grant-Kels, J.M.; Rabinovitz, H.; Oliviero, M.; Scope, A. In Vivo Reflectance Confocal Microscopy Image Interpretation for the Dermatopathologist. *J. Cutan. Pathol.* **2018**, *45*, 187–197. [CrossRef] [PubMed]
25. Kose, K.; Bozkurt, A.; Alessi-Fox, C.; Brooks, D.H.; Dy, J.G.; Rajadhyaksha, M.; Gill, M. Utilizing Machine Learning for Image Quality Assessment for Reflectance Confocal Microscopy. *J. Invest. Dermatol.* **2020**, *140*, 1214–1222. [CrossRef]
26. Pellacani, G.; Longo, C.; Malvehy, J.; Puig, S.; Carrera, C.; Segura, S.; Bassoli, S.; Seidenari, S. In Vivo Confocal Microscopic and Histopathologic Correlations of Dermoscopic Features in 202 Melanocytic Lesions. *Arch. Dermatol.* **2008**, *144*, 1597–1608. [CrossRef]
27. Pellacani, G.; Guitera, P.; Longo, C.; Avramidis, M.; Seidenari, S.; Menzies, S. The Impact of In Vivo Reflectance Confocal Microscopy for the Diagnostic Accuracy of Melanoma and Equivocal Melanocytic Lesions. *J. Invest. Dermatol.* **2007**, *127*, 2759–2765. [CrossRef]
28. Longo, C.; Zalaudek, I.; Argenziano, G.; Pellacani, G. New Directions in Dermatopathology. *Dermatol. Clin.* **2012**, *30*, 799–814. [CrossRef]
29. Agozzino, M.; Moscarella, E.; Babino, G.; Caccavale, S.; Piccolo, V.; Argenziano, G. The Use of in Vivo Reflectance Confocal Microscopy for the Diagnosis of Melanoma. *Expert Rev. Anticancer Ther.* **2019**, *19*, 413–421. [CrossRef]
30. Jemal, A.; Siegel, R.; Ward, E.; Hao, Y.; Xu, J.; Murray, T.; Thun, M.J. Cancer Statistics, 2008. *Cancer J. Clin.* **2008**, *58*, 71–96. [CrossRef] [PubMed]
31. Argenziano, G.; Soyer, H.P.; Chimenti, S.; Talamini, R.; Corona, R.; Sera, F.; Binder, M.; Cerroni, L.; De Rosa, G.; Ferrara, G.; et al. Dermoscopy of Pigmented Skin Lesions: Results of a Consensus Meeting via the Internet. *J. Am. Acad. Dermatol.* **2003**, *48*, 679–693. [CrossRef]
32. Soyer, H.P.; Kenet, R.O.; Wolf, I.H.; Kenet, B.J.; Cerroni, L. Clinicopathological Correlation of Pigmented Skin Lesions Using Dermoscopy. *Eur. J. Dermatol.* **2000**, *10*, 22–28. [PubMed]
33. Langley, R.G.B.; Rajadhyaksha, M.; Dwyer, P.J.; Sober, A.J.; Flotte, T.J.; Anderson, R.R. Confocal Scanning Laser Microscopy of Benign and Malignant Melanocytic Skin Lesions in vivo. *J. Am. Acad. Dermatol.* **2001**, *45*, 365–376. [CrossRef]
34. Pellacani, G.; Cesinaro, A.M.; Seidenari, S. In Vivo Assessment of Melanocytic Nests in Nevi and Melanomas by Reflectance Confocal Microscopy. *Mod. Pathol.* **2005**, *18*, 469–474. [CrossRef]
35. Guitera, P.; Menzies, S.W.; Longo, C.; Cesinaro, A.M.; Scolyer, R.A.; Pellacani, G. In Vivo Confocal Microscopy for Diagnosis of Melanoma and Basal Cell Carcinoma Using a Two-Step Method: Analysis of 710 Consecutive Clinically Equivocal Cases. *J. Invest. Dermatol.* **2012**, *132*, 2386–2394. [CrossRef] [PubMed]
36. Alarcon, I.; Carrera, C.; Palou, J.; Alos, L.; Malvehy, J.; Puig, S. Impact of in Vivo Reflectance Confocal Microscopy on the Number Needed to Treat Melanoma in Doubtful Lesions. *Br. J. Dermatol.* **2014**, *170*, 802–808. [CrossRef] [PubMed]
37. Pellacani, G.; Vinceti, M.; Bassoli, S.; Braun, R.; Gonzalez, S.; Guitera, P.; Longo, C.; Marghoob, A.A.; Menzies, S.W.; Puig, S.; et al. Reflectance Confocal Microscopy and Features of Melanocytic Lesions. *Arch. Dermatol.* **2009**, *145*, 1137–1143. [CrossRef]
38. Ahlgrimm-Siess, V.; Laimer, M.; Rabinovitz, H.S.; Oliviero, M.; Hofmann-Wellenhof, R.; Marghoob, A.A.; Scope, A. Confocal Microscopy in Skin Cancer. *Curr. Dermatol. Rep.* **2018**, *7*, 105–118. [CrossRef] [PubMed]
39. Kurugol, S.; Dy, J.G.; Rajadhyaksha, M.; Gossage, K.W.; Weissmann, J.; Brooks, D.H. Semi-Automated Algorithm for Localization of Dermal/Epidermal Junction in Reflectance Confocal Microscopy Images of Human Skin. In *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVIII, Proceedings of the SPIE BIOS, San Francisco, CA, USA, 22–27 January 2011*; Conchello, J.-A., Cogswell, C.J., Wilson, T., Brown, T.G., Eds.; The International Society for Optics and Photonics (SPIE): Bellingham, WA, USA, 2011; p. 79041A.
40. Kurugol, S.; Dy, J.G.; Brooks, D.H.; Rajadhyaksha, M. Pilot Study of Semiautomated Localization of the Dermal/Epidermal Junction in Reflectance Confocal Microscopy Images of Skin. *J. Biomed. Opt.* **2011**, *16*, 036005. [CrossRef]
41. Vural, V.; Fung, G.; Krishnapuram, B.; Dy, J.G.; Rao, B. Using Local Dependencies Within Batches to Improve Large Margin Classifiers. *J. Mach. Learn. Res.* **2009**, *10*, 183–206.
42. Kurugol, S.; Rajadhyaksha, M.; Dy, J.G.; Brooks, D.H. Validation Study of Automated Dermal/Epidermal Junction Localization Algorithm in Reflectance Confocal Microscopy Images of Skin. In *Photonic Therapeutics and Diagnostics VIII, Proceedings of the SPIE BIOS, San Francisco, CA, USA, 21–26 January 2012*; SPIE: Bellingham, WA, USA, 2012; p. 820702.

43. Kurugol, S.; Kose, K.; Park, B.; Dy, J.G.; Brooks, D.H.; Rajadhyaksha, M. Automated Delineation of Dermal–Epidermal Junction in Reflectance Confocal Microscopy Image Stacks of Human Skin. *J. Invest. Dermatol.* **2015**, *135*, 710–717. [CrossRef]
44. Avenel, C.; Kulikova, M.S. Marked Point Processes with Simple and Complex Shape Objects for Cell Nuclei Extraction from Breast Cancer H&E Images. In Proceedings of the Medical Imaging 2013: Digital Pathology, Lake Buena Vista (Orlando Area), FL, USA, 9–14 February 2013; p. 86760Z.
45. Somoza, E.; Cula, G.O.; Correa, C.; Hirsch, J.B. Automatic Localization of Skin Layers in Reflectance Confocal Microscopy. In *Image Analysis and Recognition, Proceedings of the 11th International Conference, ICIAR 2014, Vilamoura, Portugal, 22–24 October 2014*; Proceedings, Part II; Springer: Cham, Switzerland, 2014; pp. 141–150.
46. Nister, D.; Stewenius, H. Scalable Recognition with a Vocabulary Tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006*; IEEE: Piscataway, NJ, USA, 2006; Volume 2, pp. 2161–2168.
47. Fulkerson, B.; Vedaldi, A.; Soatto, S. Localizing Objects with Smart Dictionaries. In *Computer Vision—ECCV 2008, Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008*; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2008; pp. 179–192.
48. Dhillon, I.S.; Modha, D.S. Concept Decompositions for Large Sparse Text Data Using Clustering. *Mach. Learn.* **2001**, *42*, 143–175. [CrossRef]
49. Kaur, P.; Dana, K.J.; Cula, G.O.; Mack, M.C. Hybrid Deep Learning for Reflectance Confocal Microscopy Skin Images. In *2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016*; IEEE: Piscataway, NJ, USA, 2016; pp. 1466–1471.
50. Robic, J.; Perret, B.; Nkengne, A.; Couprie, M.; Talbot, H. Three-Dimensional Conditional Random Field for the Dermal–Epidermal Junction Segmentation. *J. Med. Imaging* **2019**, *6*, 024003. [CrossRef]
51. Lafferty, J.; McCallum, A.; Pereira, F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *ICML'01: Proceedings of the Eighteenth International Conference on Machine Learning, Williamstown, MA, USA, 28 June–1 July 2001*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2001; pp. 282–289.
52. Bozkurt, A.; Gale, T.; Kose, K.; Alessi-Fox, C.; Brooks, D.H.; Rajadhyaksha, M.; Dy, J. Delineation of Skin Strata in Reflectance Confocal Microscopy Images with Recurrent Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HA, USA, 21–26 July 2017*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 777–785.
53. Bozkurt, A.; Kose, K.; Alessi-Fox, C.; Gill, M.; Dy, J.; Brooks, D.; Rajadhyaksha, M. A Multiresolution Convolutional Neural Network with Partial Label Training for Annotating Reflectance Confocal Microscopy Images of Skin. In *21st International Conference Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Granada, Spain, 16–20 September 2018*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11071, pp. 292–299.
54. Gareau, D.; Hennessy, R.; Wan, E.; Pellacani, G.; Jacques, S.L. Automated Detection of Malignant Features in Confocal Microscopy on Superficial Spreading Melanoma versus Nevi. *J. Biomed. Opt.* **2010**, *15*, 061713. [CrossRef]
55. Kose, K.; Alessi-Fox, C.; Gill, M.; Dy, J.G.; Brooks, D.H.; Rajadhyaksha, M. A Machine Learning Method for Identifying Morphological Patterns in Reflectance Confocal Microscopy Mosaics of Melanocytic Skin Lesions In-Vivo. In *Photonic Therapeutics and Diagnostics XII, SPIE BIOS, San Francisco, CA, USA, 13–18 February 2016*; Choi, B., Kollias, N., Zeng, H., Kang, H.W., Wong, B.J.F., Ilgner, J.F., Tearney, G.J., Gregory, K.W., Marcu, L., Skala, M.C., et al., Eds.; The International Society for Optics and Photonics (SPIE): San Francisco, CA, USA, 2016; p. 968908.
56. Kose, K.; Bozkurt, A.; Ariafar, S.; Alessi-Fox, C.A.; Gill, M.; Dy, J.G.; Brooks, D.H.; Rajadhyaksha, M. Deep Learning Based Classification of Morphological Patterns in RCM to Guide Noninvasive Diagnosis of Melanocytic Lesions (Conference Presentation). In *Photonics in Dermatology and Plastic Surgery, San Francisco, CA, USA, 28–29 January 2017*; Choi, B., Zeng, H., Kollias, N., Eds.; The International Society for Optics and Photonics (SPIE): San Francisco, CA, USA, 2017; p. 11.
57. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *18th International Conference Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
58. D'Alonzo, M.; Bozkurt, A.; Alessi-Fox, C.; Gill, M.; Brooks, D.H.; Rajadhyaksha, M.; Kose, K.; Dy, J.G. Semantic Segmentation of Reflectance Confocal Microscopy Mosaics of Pigmented Lesions Using Weak Labels. *Sci. Rep.* **2021**, *11*, 3679. [CrossRef]
59. Mandal, A.; Priyam, S.; Chan, H.H.; Gouveia, B.M.; Guitera, P.; Song, Y.; Baker, M.A.B.; Vafaee, F. Computer-Aided Diagnosis of Melanoma Subtypes Using Reflectance Confocal Images. *Cancers* **2023**, *15*, 1428. [CrossRef]
60. Wiltgen, M.; Gerger, A.; Wagner, C.; Smolle, J. Automatic Identification of Diagnostic Significant Regions in Confocal Laser Scanning Microscopy of Melanocytic Skin Tumors. *Methods Inf. Med.* **2008**, *47*, 14–25. [CrossRef]
61. Wiltgen, M.; Gerger, A.; Smolle, J. Tissue Counter Analysis of Benign Common Nevi and Malignant Melanoma. *Int. J. Med. Inform.* **2003**, *69*, 17–28. [CrossRef] [PubMed]
62. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017; ISBN 9781315139470.

63. Wiltgen, M.; Bloice, M.; Koller, S.; Hoffmann-Wellenhof, R.; Smolle, J.; Gerger, A. Computer-Aided Diagnosis of Melanocytic Skin Tumors by Use of Confocal Laser Scanning Microscopy Images. *Anal. Quant. Cytol. Histol.* **2011**, *33*, 85–100. [PubMed]
64. Wiltgen, M.; Bloice, M. Automatic Interpretation of Melanocytic Images in Confocal Laser Scanning Microscopy. In *Microscopy and Analysis*; InTech: London, UK, 2016.
65. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
66. Wodzinski, M.; Skalski, A.; Witkowski, A.; Pellacani, G.; Ludzik, J. Convolutional Neural Network Approach to Classify Skin Lesions Using Reflectance Confocal Microscopy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2019; pp. 4754–4757.
67. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]