Preprints.org

Review

Artificial Intelligence in the Non-Invasive Detection of Melanoma

Banu Ismail Mendi , Kivanc Kose , Lauren Fleshner , Richard Adam , Bijan Safai , Banu Farabi [*] ,
Mehmet Fatih Atak

Review

Artificial Intelligence in the Non-Invasive Detection of Melanoma

Banu İsmail Mendi [1,*], Kivanc Kose [2], Lauren Fleshner [3], Richard Adam [3], Bijan Safai [3,4,5], Banu Farabi [3,4,5] and Mehmet Fatih ATAK [3,4,5]

[1] Department of Dermatology, Niğde Ömer Halisdemir University, Niğde, Turkey
[2] Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York NY
[3] New York Medical College, Valhalla, NY
[4] Dermatology Department, NYC Health + Hospital/Metropolitan, New York, NY
[5] Dermatology Department, NYC Health + Hospital/Coney Island, Brooklyn, NY
[*] Correspondence: banu.mendi@saglik.gov.tr

Abstract: Skin cancer is one of the most prevalent cancers worldwide, with an increasing incidence. Skin cancer is typically classified as melanoma or non-melanoma skin cancer. Melanoma, despite being much less common than basal or squamous cell carcinomas, is the most deadly, with nearly 8,300 Americans expected to die from melanoma each year. Biopsies are currently the gold standard in diagnosing melanoma; however, they can be invasive and expensive and inaccessible to lower-income individuals. Currently, suspicious lesions are triaged with image-based technologies, such as dermatoscopy and confocal microscopy. While these techniques are useful, there is wide inter-user variability and minimal training for dermatology residents on how to properly use these devices. The use of artificial intelligence (AI)-based technologies in dermatology has emerged in recent years to assist in the diagnosis of melanoma that may be more accessible to all patients and more accurate than current methods of screening. This review explores the current status of the application of AI-based algorithms in the detection of melanoma, underscoring its potential to aid dermatologists in clinical practice. We will specifically focus on AI's application in clinical imaging, dermoscopic evaluation, algorithms that can distinguish melanoma from non-melanoma skin cancers, and in-vivo skin imaging devices.

Keywords: artificial intelligence; algorithms; melanoma; skin cancer; dermoscopy; noninvasive skin imaging; reflectance confocal microscopy; optical coherence tomography; diagnostic accuracy; skin cancer detection

1. Introduction

Skin cancer is the most commonly diagnosed cancer among fair-skinned populations with an increasing incidence worldwide [1]. Cancers of the skin are typically defined as either melanoma or non-melanoma. Melanoma, the most lethal among skin cancer subtypes, occurs due to uncontrolled proliferation of melanocytes [2]. The American Cancer Society reports that although melanoma cases constitute only 1% of total skin cancer cases, death rates from melanoma are much higher compared to other skin cancer subtypes [3].

An early diagnosis of skin cancer, especially melanoma, is highly effective in reducing mortality [4]. Currently, skin biopsies and histopathological evaluation are the gold standard in the diagnosis of skin cancer [5]. However, it is impractical to confirm all skin lesions with a biopsy for several reasons, including scar formation from excisions, time constraints in clinical practice, and financial burdens. As a result, several imaging technologies are utilized in determining the necessity of a biopsy [6,7]. One example is dermoscopy, an epiluminescence microscopy technique that utilizes a magnifying lens and a (non)polarized light source to capture subsurface morphologic features (including pigmentation) from epidermal and dermal layers of the skin. Dermatoscopy is also widely

used in the diagnosis of skin diseases, especially skin cancers. Furthermore, the use of high-resolution, non-invasive diagnostic devices such as confocal microscopes, which can acquire images of skin lesions at the cellular resolution, on par with histology, has also become widespread [8,9].

The use of such imaging technologies is successful in reducing unnecessary biopsies and increasing sensitivity; however, their success is highly correlated to the skill level of providers. For instance, while many residency programs incorporate imaging techniques into teaching, no such training exists in dermatology residency programs. Therefore, the clinicians' performance in utilizing these technologies for diagnostic assessment are variable and highly user-dependent.

Recently, there has been an attempt to streamline the diagnosis of skin cancer and provide more rapid diagnoses, such as in primary healthcare settings, with the utilization of AI. AI algorithms have been engineered to incorporate macroscopic, dermoscopic, and histopathological images to predict suspicious lesions that warrant further testing. Prior literature has demonstrated that AI algorithms can perform as well as or better than consultant dermatologists and can assist clinicians in the diagnosis of skin cancers [10,11].

In this review, we aim to discuss the current status of utilizing AI-based technology in diagnosing melanoma, their potential applications, and their drawbacks.

Artificial Intelligence: Fundamental Principles

Artificial intelligence is a vast range of technologies that enables machines to exhibit human-like intelligence and cognition. Machine Learning (ML) is a subfield of AI that can make predictions based on user input data [12]. ML presents an excellent opportunity for the automation of medical data analysis to impact clinical care, with its ability to learn and make predictions that can potentially support clinical decision-making processes.

Supervised models are currently the most prevalent form of ML utilized in dermatology. In this approach, each sample in the dataset is associated with a "label". During the training process, the model learns to estimate labels from the raw data of the samples, such as pixel values for images. The three primary tasks undertaken are classification, detection, and segmentation. In classification, each sample is associated with a single label, such as a dermoscopy image classified as melanoma. Detection involves identifying the presence or absence of a given structure within the sample, such as detecting atypical networks in a dermoscopy image. Segmentation goes a step further by identifying the existence, locating, and delineating the extent of the structure, exemplified by outlining the lesion area in a dermoscopic image.

Currently, the majority of the ML studies in dermatology involve applications of deep learning (DL) models (e.g., convolutional neural networks (CNN), transformers, or their variants/combinations) to classify images to improve the diagnosis of skin diseases [13]. In their most basic and widely used form, CNNs consist of multiple cascading non-linear modeling units called "layers". These layers filter the input data by filtering the redundant information, finding correlations, and summarizing critical information into a distilled representation called "features." These layers are typically followed by several classification layers, which map the extracted features to target diagnostic labels.

In recent ML models, CNNs are commonly replaced with a new architecture called Transformers. Transformer models are neural network architectures designed to capture long-range dependencies and contextual information in sequential data-like text. The main innovation is through the use of a novel mechanism, "attention." Attention allows the model to selectively focus on relevant data when generating the output. Transformers excel at tasks requiring capturing long-range dependencies and contextual information in the data. This capability of handling both local and contextual information has contributed to their widespread adoption. Transformer models are also adapted to visual tasks such as image classification and segmentation. In this setting, instead of processing sequential data like text, Vision Transformers take images input and process them similarly to words in transformer models. The image is divided into patches, which are subsequently encoded and processed through self-attention mechanisms to capture global relationships between patches. Vision Transformers have achieved state-of-the-art performance on various computer vision benchmarks, demonstrating their effectiveness in understanding and modeling visual data.

Evaluating Artificial Intelligence Algorithms

In the field of dermatological AI, evaluation metrics play a crucial role in assessing algorithm performance. The Area Under the Receiver Operating Characteristic curve (AUROC or AUC) stands as the predominant evaluation metric, quantifying an algorithm's discriminative ability between positive and negative cases. A perfect AUROC score of 1.00 indicates optimal discrimination, while 0.5 signifies discrimination by chance, equivalent to random guessing [14,15]. The ROC curve gives the user ability to assess the algorithm at different sensitivity and specificity operating points, enabling them to manage the decision thresholds for the diagnostic algorithms. Sensitivity measures the algorithm's ability to correctly identify true positive cases, while specificity evaluates its accuracy in identifying true negatives. Complementary metrics, including precision, and F1 score, also provide additional dimensions of performance assessment. Precision quantifies the accuracy of positive predictions, and the F1 score balances precision and recall. This comprehensive suite of metrics enables a nuanced evaluation of AI algorithms, offering insights into their capacity to accurately classify cases, the reliability of their predictions, and the inherent trade-offs between different performance aspects. Such thorough evaluation is essential for understanding an algorithm's potential clinical utility and limitations in dermatological applications. For segmentation tasks, DICE coefficient [16] or the Jacard index [17] are the most widely used evaluation metrics. These metrics quantify the overlap between two sets, ranging from 0 (no overlap) to 1 (perfect overlap). The Jaccard coefficient measures similarity by comparing set intersection to union, frequently used in text analysis and image segmentation. The Dice coefficient, while similar, weights commonalities more heavily than differences.

2. Artificial Intelligence in the Diagnosis of Melanoma

Utilization of Clinical Images

Melanoma has historically been screened through clinical examination using established visual assessment methods, most notably the ABCDE criteria. The ABCDE criteria focus on five key characteristics of a mole or lesion: Asymmetry (one half of the lesion does not match the other), Border irregularity (uneven or poorly defined edges), Color variation (multiple colors or shades), Diameter (usually larger than 6mm), and Evolving (any changes in size, shape, or color over time). These features help clinicians identify suspicious lesions that may require further investigation [18]. Although state-of-the-art diagnostic tools, including non-invasive imaging devices such as dermoscopes and confocal microscopes, have been developed to improve the accuracy of melanoma detection, visual assessment methods are still commonly used in patient skin self-exams and in primary care settings where non-invasive imaging devices are not available. Utilizing AI to enhance the accuracy of visual assessment methods for evaluating pigmented skin lesions with clinical images may contribute to an earlier diagnosis of melanoma.

Nasr-Esfahani et al. utilized a CNN that consisted of two convolutional layers followed by pooling layers and a fully connected layer with the goal of classifying images as benign or melanoma. Preprocessing techniques were also applied to reduce the illumination artifacts (from non-uniform light and/or reflections of incident light from skin) and noise effects (reducing the effects of normal skin's texture on classification process). The dataset consisted of 170 clinical images (70 melanoma and 100 benign nevus). Due to the small sample size, data augmentation techniques were employed such as cropping, scaling, and rotating to generate 6120 images, where 80% of images were used for training and 20% for testing. Their model achieved an accuracy of 81%, specificity of 80%, sensitivity of 81% (18), NPV of 86%, and a PPV of 86% [19].

Moreover, Yap et al. utilized CNN models (ResNet-50, with and without embedding networks) to extract features from both dermoscopic and clinical macroscopic images. They applied a late fusion technique (embedding networks) to combine features from both modalities and incorporated metadata such as age, gender, and body location to enhance classification performance. Their dataset included 2,917 skin lesion cases from five classes (naevus, melanoma, BCC, squamous cell carcinoma (SCC), and pigmented benign keratoses), with each case containing a dermoscopic image, a

macroscopic image, and patient metadata. Using macroscopic images with embedding networks, the AUC for melanoma detection was 0.791. This increased to 0.866 when both macroscopic images and dermoscopy were used; however, the AUC was 0.861 when patient metadata were integrated [20]. Additionally, Riazi Esfahani et al. utilized a CNN to analyze 793 dermatologic images—437 of malignant melanoma and 357 benign nevi. For melanoma detection, their model achieved an accuracy of 88.6%, with a specificity of 81.8% and sensitivity of 97.1%. However, the study's limitations were noted as variations in image quality and acquisition methods, which may affect the model's generalizability [21].

Dorj et al. employed a pre-trained CNN model, AlexNet with 11 layers (5 convolutional layers, 3 max pooling layers and 3 fully connected layers) to extract and classify features using an ECOC-SVM classifier. Their dataset consisted of 3,753 images (2985 for training and 758 for testing) representing four types of skin cancers: actinic keratoses, BCC, SCC, and melanoma (n=958, 768 training and 190 testing). For melanoma classification, the model achieved an average accuracy of 0.942, a specificity of 0.9074, and a sensitivity of 0.9783 [22]. Soenksen et al. assessed multiple deep convolutional neural networks (DCNNs) utilizing a dataset of 33,980 images, encompassing melanoma, SCC, BCC, and various benign lesions. 4063 images of suspicious pigmented lesions (SPLs) were included in this dataset of which 2906 images were melanoma. Data was also divided into 6 different classes: backgrounds, skin edges, bare skin sections, nonsuspicious pigmented lesions of low priority (NSPL), NSPL of medium priority and SPLs. Blob detection algorithm was initially done to accelerate analysis. Their baseline DCNN model had 3 convolutional neural networks and utilized 60% of the data as training, 20% as validation and 20% as testing. Furthermore, they also trained their DCNN on a 10x non-overlapping augmented dataset with class balancing (naug = 300,000). VGG16 ImageNet pre-trained network was applied as transfer learning to their DCNN as another model. Another transfer learning DCNN model based on ImageNet's Xception network was also generated to compare to VGG16's performance. The VGG16 transfer learning DCNN model demonstrated the highest performance, achieving an AUC of 0.935, The overall AUC across the 6 included classes (AUCmicro) for this model was .97 with a sensitivity of .903 and specificity of .899. This model was further applied to analyze wide-field images, using a "saliency-based" approach to detect "ugly duckling" lesions—those that are noticeably abnormal compared to other lesions on the same patient. The model exhibited a 96.3% agreement with the consensus of 10 dermatologists; however, this agreement dropped to 82.96% when examining a reduced number of neighboring lesions [23]. Pomponiu et al. employed a deep neural network (DNN) consisting of a CNN with 5 convolutional layers and 2 fully connected layers pre-trained on natural images. Additionally, a KNN classifier was applied to distinguish between benign nevi and melanoma lesions. The dataset consisted of 399 images of pigmented skin lesions (217 benign and 182 melanoma) from online dermatology image libraries (DermIS and DermQuest). Their model achieved an accuracy of 0.83, with a specificity of 0.95 and a sensitivity of 0.92 [24]. Han et al. utilized a DL algorithm (ResNet-152) to classify images of 12 skin diseases (BCC, SCC, intraepithelial carcinoma, actinic keratosis, seborrheic keratosis, melanoma, melanocytic nevus, lentigo, pyogenic granuloma, hemangioma, dermatofibroma, wart). The model was evaluated on multiple datasets, including the Asan and Edinburgh datasets. 19,9398 images from the Asan dataset, MED-NODE dataset, and atlas site images were used for training while 480 images from Asan and Edinburgh datasets were used for testing. For melanoma detection in the Asan dataset, the AUC, sensitivity, and specificity were 0.96, 0.91, and 0.904, respectively. For the Edinburgh dataset, these values were 0.88, 0.855, and 0.807, respectively. The model demonstrated strong diagnostic performance, comparable to that of dermatologists, with particularly good results on the Asan dataset. However, the slight performance drop on the Edinburgh dataset highlights the impact of demographic and ethnic differences, as well as variations in image contrast, on the algorithm's effectiveness [25].

Liu et al constructed a deep learning system (DLS) with Inception-v4 modules to process images and a shallow module to process metadata such as demographic information and medical history with the goal of identifying 26 of the most common skin cases in adults. Their model wasn't just used to give a single diagnosis, but also gave a list of top 3 differential diagnoses. Primary output was

classification from 26 skin conditions and 'other' while secondary output was classification from a full list of 419 skin conditions. Data came from teledermatology cases, and they performed a temporal split of their data where 80% of cases with metadata (64,837 images) were used for training of the DLS while 20% of their data with metadata (validation set A, 14,833 images) was used for validation. Validation set A was randomly subsampled to generate validation set B (3707 images) to compare the DLS performance to that of dermatologists. DLS performance with validation set A for top 1 diagnosis accuracy and sensitivity over 26 skin conditions was .71 and .58 respectively. These values increased to .93 for accuracy and .83 for sensitivity for top-3 diagnoses from the 26 skin conditions. These values were lower for both categories when looking at the full list of 413 skin conditions but still comparable. When using validation set B, DLS demonstrated a top 1 accuracy of .66 compared to .63 for that of dermatologists from the 26 skin conditions. Top 1 sensitivity for DLS with 26 skin conditions was .56 which was comparable to that of dermatologists at .51. Top 3 accuracy under the same conditions for DLS was substantially higher at .9 compared to .75 for that of dermatologists. Top 3 sensitivity for DLS with 26 skin conditions was .64 which was also substantially greater than that of dermatologists at .49. Top 1 and 3 accuracies on the 419 classifications were less than that on the 26 classification for both DLS and dermatologist but was still comparable between the two [26].

Sangers et al. conducted a prospective multicenter study to evaluate skin lesions using an app on iOS and Android devices, comparing the app's outcomes to histopathological diagnoses or clinical assessments made by dermatologists. They collected images of 785 skin lesions collected from 372 patients, with 418 classified as suspicious (premalignant or malignant) and 367 as benign. The app utilized CNN (version RD-174) to assess the risk of the photographed lesions, categorizing them as low or high risk. Overall app sensitivity and specificity were .869 and .704 respectively. For melanocytic lesions, the sensitivity and specificity were 0.819 and 0.733, respectively. One limitation of the study was that lesion photos were taken by trained researchers in outpatient settings rather than by patients, which may affect the app's external validity, as it is intended for general use in non-clinical environments. Additionally, the study employed two high-resolution smartphone models, raising concerns about the app's performance on devices with lower camera resolution or older hardware. Furthermore, over 80% of participants had Fitzpatrick skin types I or II, which may limit the study's applicability to individuals with darker skin tones. Lastly, the low number of melanoma cases (n=12) restricts conclusions about the app's capability to detect melanomas specifically. Despite these important limitations, the study introduces the concept of utilizing smartphone apps for self-skin examination, and self-assessment of skin cancer risk, which could be highly beneficial for early detection of skin cancers [27].

Polturu et al. employed an Automated Machine Learning model (AutoML) created using a no-code online service platform to analyze a dataset of 87 non-melanoma images and 119 melanoma images, all taken with a consumer-grade camera. The model attained an overall accuracy of 0.844, with a specificity of 0.857 and a sensitivity of 0.833 [28]. Algorithms used in the diagnosis of melanoma from clinical images are summarized in Table 1.

Table 1. Algorithms used in the diagnosis of melanoma from clinical images.

Publication	End-point	Dataset	Algorithm	Performance
Nasr-Esfahani et al. [19]	Classification (benign/melanoma)	170 clinical images that underwent data augmentation to generate 6120 images (80% training, 20% validation)	CNN with 2 convolutional layers each followed by pooling layers along with a fully connected layer	Acc: 81% Spe: 80% Sen: 81% NPV: 86% PPV: 86%
Yap et al. [20]	Classification of melanoma from 5 different types of lesions	2917 cases with each case containing patient metadata, macroscopic image and dermoscopic images with 5 classes (naevus, melanoma, BCC, SCC, and pigmented benign keratoses)	ResNet-50 with embedding networks	Macroscopic images alone AUC: .791 Macroscopic and dermoscopy AUC: .866 Macroscopic, dermoscopy and metadata AUC: .861
Riazi Esfahani et al. [21]	Classification (malignant melanoma/benign nevi)	793 images (437 malignant melanoma and 357 benign nevi)	CNN	Acc: 88.6% Spe: 88.6% Sen: 81.8%
Dorj et al.[22]	Classification of melanoma from 4 different skin cancers (actinic keratoses, BCC, SCC, melanoma)	3753 images (2985 training and 758 testing) including 958 melanoma	AlexNet with ECOC-SVM classifier	Acc: .942 Spe: .9074 Sen: .9783
Soenksen et al. [23]	Classification across 6 different classes as well as distinguishing SPLs	33,980 (including backgrounds, skin edges, bare skin sections, low priority NSPLs, medium priority NSPLs and SPLs) (60% training, 20% validation and 20% as testing)	DCNN with VGG16 Image Net pretrained network as transfer learning	Across all 6 classes AUC_{micro}: .97 Spe_{micro}: .903 Sen_{micro}:.899 For SPLs AUC:.935
Pomponiu et al. [24]	Classification (melanoma/benign nevi)	399 images (217 benign, 182 melanoma) from online image libraries	CNN with a KNN classifier	Acc: .83 Spe: .95 Sen: .92
Han et al. [25]	Melanoma detection from 12 different skin diseases	Training: 19,938 images from the Asan dataset [29], MED-NODE dataset [30], and atlas site images Testing: 480 images from Asan and Edinburgh datasets [31]	ResNet152	Asan AUC: .96 Spe: .904 Sen: .91 Edinburgh AUC: .88 Spe: .855 Sen: .807
Liu et al. [26]	Primary: classification among 26 different skin conditions Secondary: classification among a full set of 419 different skin conditions	Training: 64,837 images with metadata Validation set A: 14833 images with metadata Validation set B used to compare to dermatologists: 3707 images with metadata	DLS with Inception-v4 modules and shallow module	Validation set A for 26 image classification: Acc_{top1}:.71 Acc_{top3}: .93 Sen_{top1}:.58 Sen_{top3}: .83 Validation set B for 26 image classification:

				Acc$_{top1}$:.66 Acc$_{top3}$: .9 Sen$_{top1}$:.56 Sen$_{top3}$: .64 Dermatologists: Acc$_{top1}$:.63 Acc$_{top3}$: .75 Sen$_{top1}$:.51 Sen$_{top3}$: .49
Sangers et al. [27]	Classification (low/high risk)	785 images (418 suspicious, 367 benign)	RD-174	Overall app classification Sen: .869 Spe:.704 Classification for melanocytic lesions: Sen: .819 Spe: .733
Polturu et al. [28]	Classification (non-melanoma/melanoma)	206 images from DermIS [32] and Derm Quest [33] (87 nonmelanoma and 119 melanoma, 85% used for training and 15% used for testing)	AutoML was created using a no-code online service platform	Acc: .844 Sen: .833 Spe: .857

Convolutional Neural Network (CNN); Accuracy (Acc); Specificity (Spe); Sensitivity (Sen); Negative Predictive Value (NPV); Positive Predictive Value (PPV); Basal Cell Carcinoma (BCC); Squamous Cell Carcinoma (SCC); Area under the ROC curve (AUC); Error-Correcting Output Codes (ECOC); Support Vector Machine (SVM); Suspicious Pigmented Lesions (SPLs); Nonsuspicious Pigmented Lesions (NSLPs); Deep Convolutional Neural Network (DCNN); Visual Geometry Group (VGG); k nearest neighbor (KNN); Deep learning system (DLS); Automated Machine Learning (AutoML).

3. Dermoscopic Images

Dermoscopy is currently used as a non-invasive diagnostic measure of skin lesions. It is particularly useful in the differential diagnosis of skin tumors [34]. Recently, AI models and technologies have been applied to dermoscopic imaging, and successful results have been obtained in the differential diagnosis of skin tumors. We reviewed 37 studies that used dermoscopic images as a dataset (Tables 2 and 3). Of these 37 studies, 26 evaluated AI models in the diagnosis of melanoma, and 11 studies evaluated both melanoma and non-melanoma skin cancers.

3.1. Distinguishing Melanoma from Benign Lesions

Masood et al. classified clinical and dermoscopic photographs as benign/melanoma using ANN and compared the performances of three different ANN algorithms (Levenberg-Marquardt (LM), resilient backpropagation (RP), scaled conjugate gradient (SCG)). SCG gave the most successful results with 92.6% sensitivity and 91.4% specificity. LM achieved a specificity of 95.1% in benign lesions, but it was not as successful as SCG in melanoma [35]. In [36], a fusion ML model consisting of five individual top-ranked algorithms from the ISBI 2016 Challenge was applied for melanoma detection, and its performance was compared to dermatologists. Their model achieved an AUROC of 0.86 and was more accurate than dermatologists; however, applying ML classifications to dermatologist evaluations increased dermatologist sensitivity from 76.0% to 80.8% and specificity from 72.6% to 72.8% [37].

Recently, there has been a surge in studies on the discrimination of melanocytic lesions (benign/malignant) using CNNs on dermoscopic photographs. Chanki Yu et al. used a pre-trained CNN model (VCG-16) in diagnosing acral melanoma, compared to both general practitioners and dermatologists. They performed 2-fold cross-validation and split the dataset into a 50/50 train-test

split. The model achieved a similar AUROC value to experts and was significantly superior to the non-expert group [38]. Abbas et al. also designed a seven-layer deep CNN to discriminate between acral melanoma and benign nevus. They used 724 dermoscopic images from Chanki Yu et al's [38] dataset and 4,344 dermoscopic images generated by data augmentation techniques. The authors also applied transfer learning to the AlexNet and ResNet-18 and fine-tuned them by modifying their last layers. An AUC of 0.97, 0.96, and 0.91 was obtained with ResNet-18, AlexNet, and the proposed ConvNet, respectively [39]. Another study proposed a CNN model to distinguish combined nevi from melanoma. Moleanalyzer Pro, previously trained on more than 120,000 dermoscopic images, was used in the study, and 72 dermoscopic images (36 combined nevus and 36 melanoma) were evaluated. When compared to 11 dermatologists divided into 3 groups (beginner/qualified/expert), the model outperformed all of them, revealing 97.1% sensitivity and 78.8% specificity [40].

Although the success of dermatologists is compared with AI, few dermatologists were included. To address this drawback, Brinker et al. compared the performance of a CNN algorithm (Resnet) trained only on open-source dermoscopic images with 157 dermatologists, resulting in seven dermatologists being more accurate than CNNs [41]. Furthermore, Giulini et al. combined CNN and human expertise in the diagnosis of melanoma. In the study, 64 physicians (33 dermatologists, 11 dermatology residents, and 20 general practitioners) assessed 100 dermoscopic photographs of 50 melanomas and 50 benign nevi. After a duration of 4 months, the same photographs were reevaluated in a different order with CNN assistance by the physicians. In the session with CNN assistance, the mean sensitivity and specificity increased to 67.88% and 73.72% from 56.31% and 69.28%, respectively [42].

Hybrid models are also commonly studied in the literature; Mahbod et al. used 3 pre-trained CNN models (AlexNet, VGG16, and ResNet-18) for feature extraction, followed by an SVM-based classification step. The final classification result was obtained by averaging the output of the individual models. The resulting ensemble model was evaluated on 150 validation images and achieved an AUC of 90.69%, surpassing the performance of the individual CNN models (AlexNet, VGG16, and ResNet-18) [43]. Ningrum et al. constructed a hybrid model by integrating dermoscopic pictures and patient data to diagnose melanoma. They employed a model that utilized both CNN to analyze photos and ANN to analyze patient data in order to categorize patients as melanoma or nonmelanoma. The results were compared with CNNs from only analyzing images. The CNN+ANN model achieved an accuracy of 92.34%, surpassing the accuracy of the CNN model alone at 73.69% [44].

While AI demonstrates success in studies, its application and implementation in real-world scenarios are crucial. Hekler et al. assessed the efficacy of DL in categorizing lesions by employing multiple real-world lesion images, single lesion images, and modified lesion images. The model displayed markedly enhanced performance when utilizing multiple real-world images, particularly in uncertainty estimation and robustness [45]. Specifically, the utilization of AI in melanoma screening is poised to substantially alleviate the workload on clinicians. To showcase AI's potential as a melanoma screening tool, Crawford et al. explored the feasibility of employing AI to identify potential melanomas in self-referred patients concerned about the malignancy of their skin lesions. The AI successfully identified 11 of 17 malignant lesions, achieving an accuracy of 73.56%, exceeding the accuracy of 4 out of 5 dermatologists involved in the study [46].

The lack of transparency of AI techniques reduces their reliability for users. To address this issue, Chanda et al. developed an explainable AI (XAI) algorithm. In the task of predicting melanoma, the algorithm explains the basis of its prediction. The investigation revealed that the XAI increased clinicians' diagnostic confidence while also enhancing their trust in the assistance provided by XAI [47]. Correira et al. have introduced a method that utilizes an interpretable prototypical-part model that integrates binary masks, automatically generated by a segmentation network and user-refined prototypes. This model is designed to incorporate non-expert feedback, ensuring that the learned prototypes specifically relate to important areas within the skin lesion while excluding irrelevant factors beyond its boundaries. By following these two distinct information pathways, the proposed

approach demonstrates superior diagnostic performance when compared to non-interpretable models [48].

Table 2. Algorithms that distinguish between melanoma and benign lesions through dermoscopic images.

Publication	End-point	Dataset	Algorithm	Performance
Masood et al. [35]	Classification (benign/melanoma)	135 images (Clinical + dermoscopic) 107 for training, 14 for validation 14 for testing	Compared 3 ANN algorithms (RP, LM, SCG)	SCG: Acc: 91,9% Sen: 92.6% Spe: 91.4% LM: Acc: 91,1% Sen: 85.2% Spe: 95.1% RP: Acc: 88,1% Sen: 77.8% Spe: 95.1%
Aswin et al. [49]	Classification (Cancerous/Non-cancerous)	30 dermoscopic images for training 50 dermoscopic images for testing	Hybrid Genetic Algorithm + ANN	Acc: 88%
Xie et al. [50]	Classification (MM/BN)	Dermoscopic images Xanthous race:240 images (80 MM, 160 BN) Caucasian race: 360 images (120 MM, 240 BN)	Proposed: meta-ensemble model of multiple neural network ensembles Ensemble 1: single-hidden-layer BP nets with same structures Ensemble 2: single-hidden-layer BP nets and fuzzy nets Ensemble 3: double-hidden-layer BP nets with different structures	Xanthous race: Sen: 95% Spe: 93.75% Acc: 94.17% Caucasian race: Sen: 83.33% Spe: 95% Acc: 91.11%
Marchetti et al. [36]	Classification (MM/BN)	ISBI 2016 challenge dataset [51], MM:248 images BN:1031 images Train set:900 images Test set:379 images Reader study:100 images (50 MM, 50 BN)	Five methods (unlearned and machine learning) were used to combine individual automated predictions into "fusion" algorithms	Top Fusion Algorithm: Greedy Fusion: Sen: 58% Spe: 92% AUC: 86% Dermatologists: Sen: 82% Spe: 59% AUC: 71%
Marchetti et al. [37]	Classification (MM/BN/SK) and (biopsy/observation)	ISIC Archive [52]: 2,750 dermoscopy images (521 (19%) MM, 1,843 (67%) BN, and 386 (14%) SK) Training set: 2,000 images Validation: 150 images Test set: 600 images	ISBI 2017 Challenge top ranked algorithm	Algorithm: Sen: 76% Spe: 85% AUC: 0.87 Dermatologists: Sen: 76.0% Spe: 72.6% AUC: 0.74
Cueva et al. [53]	Classification (Cancerous/Non-cancerous)	PH² database [54] Training set: 30 images (10 MM, 10 common mole, 10 no-common mole) Test set: 201 images (80 common mole, 80 no-common mole, 41 MM)	ANN with backpropagation algorithm	After an analysis of 201 images in the algorithm developed a performance of 97.51% was obtained

Navarro et al. [55]	Segmentation and registration to evaluate lesion change	ISIC archive [52]: Training set: 2000 dermoscopic images Validation: 150 dermoscopic images Test set: 600 dermoscopic images	Segmentation: LF-SLIC Registration:SP-SIFT	Acc: 0.96 for segmentation
Yu C. et al. [38]	Classification (melanoma/non-melanoma)	725 images (AM: 350 images, BN: 374 images) Group A: 175 images AM, 187 images BN Group B: 175 images AM, 187 images BN Training set: Group A images for training Group B Group B images for training Group A Test set: Group A images for Group A Group B images for Group B	CNN (VCG-16)	Group A: <u>CNN:</u> Sen: 92.57 Spe: 75.39 Acc: 83.51 <u>Expert:</u> Sen: 94.88 Spe: 68.72 Acc: 81.08 <u>Non-expert:</u> Sen: 41.71 Spe: 91.28 Acc: 67.84 Group B: <u>CNN:</u> Sen: 92.57 Spe: 68.16 Acc: 80.23 <u>Expert:</u> Sen: 98.29 Spe: 65.36 Acc: 81.64 <u>Non-expert:</u> Sen: 48.00 Spe: 77.10 Acc: 62.71
Abbas et al. [39]	Classification (benign nevus/acral melanoma)	724 images from Yonsei University [38] (350 acral melanoma, 374 benign nevi) 4344 images with data augmentation (2100 acral melanoma, 2244 benign nevi)	Compared three CNN algorithms (Seven-layered deep CNN, ResNet-18, AlexNet)	<u>ResNet-18</u> Acc: 0.97 AUC: 0.97 <u>AlexNet:</u> Acc: 0.96 AUC: 0.96 <u>Proposed ConvNet</u> Acc: 0.91 AUC: 0.91
Fink et al. [40]	Classification (Benign/Malignant)	Training set: >120.000 dermoscopic images and labels Test set: 72 images (36 combined naevi, 36 melanomas)	CNN (Moleanalyzer-Pro) based on a GoogleNet Inception_v4 architecture	<u>CNN:</u> Sen: 97.1% Spe: 78.8% <u>Dermatologists:</u> Sen: 90.6% Spe: 71.0 %
Phillips et al. [56]	Classification (MM/dysplastic nevi/other)	Pretrained algorithm Training set (in study): 289 images (36 melanoma lesions; 67 nonmelanoma lesions, 186 control lesions) Test set:1550 images	SkinAnalytics (CNN)	<u>The algorithm:</u> <u>İphone 6s image:</u> AUC: 95,8% Spe: 78,1% <u>Galaxy S6 image:</u> AUC: 93,8% Spe: 75,6% <u>DSLR image:</u> AUC: 91,8% Spe: 45,5% <u>Specialists:</u> AUC: 77,8% Spe: 69,9%

Martin-Gonzalez et al. [57]	Classification (benign/ malignant skin lesion)	Pretrained with 37,688 images from ISIC Archive [52] 2019 and 2020 Training set: 339 images (143 MM, 196 BN) Test set:232 images (55 MM, 177 BN)	QuantusSKIN (CNN)	AUC: 0,813 Sen: 0,691 Spe: 0,802 Acc: 0,776
Brinker et al.[41]	Classification (Melanoma/Nevi)	Training set: 12,378 dermoscopic images from ISIC dataset [52] Test set:100 dermoscopic images (20 MM, 80 Nevi)	ResNet-50 (CNN)	<u>Algorithm:</u> Sen: 74.1% Spe: 86.5% <u>Dermatologists:</u> Sen: 74.1% Spe: 60%
Giulini et al.[42]	Classification (Melanoma/Nevi)	Over 28,000 dermoscopic images CNN test set: 2489 images (344 melanomas, 2155 nevi) Physician test set: 100 images (50 MM, 50 nevi)	Session 1: Physicians without CNN Session 2: Physicians with CNN	<u>Physicians without CNN</u> Sen: 56.31% Spe: 69.28% <u>Physicians with CNN</u> Sen: 67.88% Spe: 73.72%
Ding et al.[58]	Classification (Binary:melanoma/non-melanoma and multiclass: benign nevi, seborrheic keratosis or melanoma)	ISIC Dataset [52] Training set: 2000 images (374 MM, 254 SK, 1,372 BN) Validation set: 150 images (30 MM, 42 SK, 78 BN) Test set: 600 images (117 MM, 90 SK, 393 BN)	Segmentation: U-Net Classification: Five CNNs (Inception-v3, ResNet-50, Densenet169, Inception-ResNet-v2 and Xception) with SE-block and the neural network for ensemble learning consisting of two local connected layers and a softmax layer	<u>Binary:</u> <u>Inception-v3</u> Acc: 0.885 AUC: 0.883 <u>ResNet-50</u> Acc: 0.88 AUC: 0.882 <u>Densenet169</u> Acc: 0.893 AUC: 0.882 <u>Inception-ResNet-v2</u> Acc: 0.89 AUC: 0.894 <u>Xception</u> Acc: 0.891 AUC: 0.896 <u>Ensemble</u> Acc:0.909 AUC: 0.911 <u>Multiclass:</u> <u>Inception-v3</u> Acc: 0.792 AUC: 0.883 <u>ResNet-50</u> Acc: 0.762 AUC: 0.864 <u>Densenet169</u> Acc: 0.800 AUC: 0.881 <u>Inception-ResNet-v2</u> Acc: 0.800 AUC: 0.873 <u>Xception</u> Acc: 0.810 AUC: 0.896 <u>Ensemble</u> Acc: 0.851 AUC: 0.913

Yu L. Et al.[59]	Segmentation and Classification (Benign/Malignant)	ISIC dataset [52] Training set: 900 images Test set: 350 images	FCRN for skin lesion segmentation and very deep residual network for classification	Segmentation: Sen: 0,911 Spe: 0,957 Acc: 0,949 Classification with segmentation: Sen: 0,547 Spe: 0,931 Acc: 0,855
Bisla et al. [60]	Classification (Nevus, SK, MM)	Training set: ISIC dataset [52]: 803 MM, 2107 nevus, 288 SK PH² dataset [54]: 40 MM, 80 Nevus Edinburgh dataset [31]: 76 MM, 331 nevus, 257 SK Test set: ISIC data sets 600 images (117 MM, 90 SK, and 393 nevus)	Segmentation:Modified U-Net (CNN) Augmentation: de-coupled DCGANs Classification:ResNet-50	AUC: 0,915 Acc: 81,6%
Mahbod et al. [43]	Classification (MM/All, SK/All)	ISIC dataset [52] Training: 2037 dermoscopic images (411 MM, 254 SK, 1372 BN)	Feature Extraction: Pretrained CNNs (AlexNet, ResNet-18 and VGG16) Classification: SVM	AUC: 90,69
Bassel et al. [61]	Classification (Benign/Malignant)	ISIC dataset [52]: 1800 images of benign type and 1497 pictures of malignant cancer Training set: 70% of images (1440 benign, 1197 malignant) Test set: 30% of images (360 benign, 300 malignant)	Model 1:Feature Extraction: ResNet50 Model 2:Feature Extraction: VCG-16 Model 3:Feature Extraction: Xception Classification: Stacked CV model (SVM+NN+RF+KNN)	ResNet Model: Acc: 81,6% AUC: 0,818 VCG-16 Model: Acc: 86,5 % AUC: 0,843 Xception Model: Acc: 90,9 AUC: 0,917
Ningrum et al. [44]	Classification (Melanom/benign)	ISIC dataset [52] 900 images Training set: 720 images Validation set: 180 images Test set: 300 (93 malignant, 207 nonmalignant)	Classification: CNN model for images + ANN model for patient metadata .	CNN Acc: 73.69 AUC: 82.4 CNN+ANN Acc: 92.34 AUC: 97.1
Nambisan et al.[62]	Segmentation and classification (Melanoma/Benign)	ISIC dataset [52] Segmentation task: 487 MM images Classification task: 1000 images (500 MM, and 500 benign (100 images per class from the Actinic keratosis, Melanocytic nevus, Benign keratosis, Dermatofibroma, and Vascular lesion)	Segmentation (Classification dataset+Segmentation dataset (Irregular networks)) U-Net/U-Net++/MA-Net/PA-Net Handcrafted Feature Extraction Classification: Level 0 (without segmentation): DL classification model Level 1 (With segmentation and with level 0 model's results): Conventional classification model	Conventional Ensemble Acc: 0.793 DL Ensemble Acc: 0.838 EfficientNet-B0 + Conventional Ensemble Acc: 0.862

Collenne et al. [63]	Classification (Melanoma/Nevi)	ISIC dataset [52] (6371 nevi and 1301 melanoma) Training set 70% of images: Validation set: 10% of images Test set: 20% of images	Segmentation: U-Net Classification ANN(for asymmetry features + CNN (EfficientNet)	<u>Handcrafted Model with asymmetry features (ANN):</u> Acc: 79% AUC: 0.87 Sen: 90% Spe: 67% <u>ANN+CNN:</u> Sen: 0.92 Spe: 0.82 Acc: 0.87 AUC: 0.942
Hekler et al. [45]	Classification (Melanoma/Nevi)	HAM10000 [64] and BCN20000 [65] Datasets 29,562 images (7,794 melanoma and 21,768 nevi) %80 training, %20 validation Test set: SCP2 dataset, 293 melanoma and 363 melanocytic nevi from 617 patients	ConvNeXT architecture 1. Classification using single image 2. Classification using multiple real-world images 3. Classification using multiple artificially modified images	<u>Single image approach:</u> Acc: 0.905 ECE: 0.131 <u>Multiview real-world approach:</u> Acc: 0.930 ECE: 0.072 <u>Multiview artificial approach:</u> Acc:0.929 ECE: 0.086
Crawford et al.[46]	Classification (Excision/no excision)	Self-referred patients	MoleAnalyzer Pro	<u>AI</u> Sen: 64.7% Spe: 75.76% PPV: 40.0% NPV: 89.6% Acc: 73.56%

Artificial Neural Network (ANN); Levenberg-Marquardt (LM); Resilient Back propagation (RP); Scaled Conjugate Gradient (SCG); Accuracy (Acc); Sensitivity (Sen); Specificity (Spe); Malign Melanoma (MM); Benign Nevi (BN); Back propagation (BP); International Symposium on Biomedical Imaging (ISBI) challenge 2016; Area under the ROC curve (AUC); International Skin Imaging Collaboration (ISIC); Seborrheic Keratosis (SK); Local Features-Simple Linear Iterative Clustering (LF-SLIC); Scale Invariant Feature Transform (SIFT); Acral Melanoma (AM); Convolutional Neural Network (CNN); Visual Geometry Group (VGG), Squeeze-and-Excitation block (SE-Block); Fully Convolutional Residual Network (FCRN); Deep Convolutional Generative Adversarial Network (DCGAN); Neural network (NN); Random forest (RF); Human Against Machine with 10000 training images (HAM10000); Expected calibration error (ECE); Artificial Intelligence (AI); Negative predictive value (NPV); Positive predictive value (PPV).

3.2. Distinguishing Melanoma from Other Skin Cancers

Esteva et al. used a pre-trained GoogLeNet Inception v3 architecture and performed transfer learning on 127,463 clinical images, including 3,374 dermoscopy images containing 2032 diseases. After the CNN model was trained, comparisons were made on 135 epidermal (65 malignant, 70 benign), 130 melanocytic (33 malignant, 97 benign), and 111 melanocytic-dermoscopic (71 malignant, 40 benign) images by 21 board-certified dermatologists. CNN performed on par with dermatologists on all 3 criteria. The AUC from clinical photographs was 0.94 for melanoma and 0.91 for melanoma from dermoscopic photographs [66]. Rezvantalab et al. also used pre-trained models (DenseNet 201, ResNet 152, Inception v3, InceptionResNet v2) in the classification of 8 diagnostic categories (melanoma, melanocytic nevus, BCC, benign keratosis, actinic keratosis, intraepithelial carcinoma, dermatofibroma, vascular lesions, and atypical nevus). All of the models performed better than dermatologists in detecting melanoma and BCC. The most successful model was ResNet 152 with 94.4% AUC in melanoma [67]. Maron et al. included more dermatologists in their study and found that CNNs outperformed dermatologists on both endpoints except BCC [68]. In another study, Tschandl et al. evaluated the success of CNN in nonpigmented cancers, the most common skin cancer manifestation. They trained the model with dermoscopic and clinical images and compared it to 95

human raters. The evaluators were divided into 3 groups: beginner, intermediate, and expert, according to their dermoscopy experience. The model's AUC was higher than the human rating; however, it was less accurate than experts [69]. Tschandl et al. then evaluated the success of ML in benign and malignant pigmented skin lesions. They compared the top 3 algorithms of the ISIC 2018 challenge with human readers and experts and ultimately outperformed both groups [70]. However, these studies only include images of the lesions and do not include clinical information, which has a very important impact on diagnosis. Therefore, a two-level comparison study including textual information was conducted by Haenssle et al. In level I, only dermoscopic images were used, while in level II, clinical and dermoscopic images and textual information were used. At level I, CNN achieved a higher accuracy than dermatologists, but at level II, dermatologists achieved a higher accuracy rate [71].

Although dermatologists and AI are seen as competitors in studies, superior results are often recorded with the combination of classifiers. Hekler et al. investigated the potential benefit of combining humans and AI in skin cancer classification. The primary endpoint was the correct classification of images into five designated categories, while the secondary endpoint was the classification of lesions as benign or malignant. Ultimately, the combination of humans and machines achieved 82.95% accuracy; this was 1.36% higher than the best of the two individual classifiers [72].

Table 3. Algorithms that distinguish melanoma from other skin cancers through dermoscopic images.

Publication	End-point	Dataset	Algorithm	Performance
Esteva et al. [66]	Classification Binary: Keratinocyte carcinoma/SK; melanoma/nevi 3 way: Benign/Malign/Non-neoplastic 9 way: Cutaneous lymphoma and lymphoid infiltrates/ Benign dermal tumors, cysts, sinuses/ Malignant dermal tumor/ Benign epidermal tumors, hamartomas, milia, and growths/ Malignant and premalignant epidermal tumors/ Genodermatoses and supernumerary growths/ Inflammatory conditions/ Benign melanocytic lesions/ Malignant Melanoma	ISIC [52] and Edinburgh dataset [31] and the Stanford Hospital: 129,450 clinical images, including 3,374 dermoscopic images of 757 disease classes Training set: 127,463 images Test set:1,942 images	Google Inception v3 (CNN)	Binary classification (Algorithm AUC) Carcinoma: 0,96 Melanoma: 0,0,94 Melanoma (Dermoscopic images): 0,91 3 way classification: Dermatologist 1 Acc: 65.6% Dermatologist 2 Acc: 66.0% CNN Acc: 69.4 ± 0.8% CNN partitioning algorithm Acc: 72.1 ± 0.9% 9 way classification: Dermatologist 1 Acc: 53.3% Dermatologist 2 Acc: 55.0% CNN Acc: 48.9 ± 1.9% CNN partitioning algorithm Acc: 55.4 ± 1.7%

Rezvantalab et al. [67]	Classification (MM/Melanocytic Nevi/BCC/AKIEC/Benign keratosis/DF/Vascular lesion)	HAM10000 dataset [64] :10015 dermatoscopic images (1113 MM, 6705 nevi, 514 BCC, 327 AK and intraepithelial carcinoma (AKIEC), 1099 benign keratosis, 115 DF, 142 vascular lesions) PH² set (55): 80 nevi, 40 MM Training set: 70 % Validation set: 15% Test set: 15%	Compared CNNs for classification: Inception v3/InceptionResNet v2/ResNet 152/DenseNet 201	<u>AUC (Melanoma)</u> Dermatologist: 82,26 DenseNet 201: 93,80 ResNet 152: 94,40 Inception v3: 93,40 InceptionResNet v2: 93,20 <u>AUC (BCC)</u> Dermatologist: 88,82 DenseNet 201: 99,30 ResNet 152: 99,10 Inception v3: 98,60 InceptionResNet v2: 98,60
Maron et al. [68]	Classification Two way:Benign/Malignant Five way:AKIEC/BCC/MM/Nevi/BKL (benign keratosis, including seborrhoeic keratosis, solar lentigo and lichen planus like keratosis)	Training set: 11,444 images (ISIC Archive[52] and HAM10000 dataset [64]) Test set: 300 test images (60 for each of the five disease classes) (HAM10000 dataset)	CNN (ResNet50)	<u>Two way classification:</u> CNN AUC: 0,928 CNN Spe: 91,3% Dermatologist Spe: 59,8% <u>Five way classification:</u> CNN AUC: 0,960 CNN Spe: 89,2% Dermatologist Spe: 98,8%
Tschandl et al [69]	Classification (Benign/Malignant)	Training set:7895 dermoscopic and 5829 close-up images Test set: 2,072 dermoscopic and close-up images	Combined convolutional neural network (cCNN) (InceptionResNetV2, InceptionV3, Xception, ResNet50)	<u>cCNN:</u> AUC: 0,695 Sen: 80,5% Spe: 53,5% <u>Human Raters:</u> AUC: 0,742 Sen: 77,6% Spe: 51,3%
Tschandl et al. [70]	Classification (7 way classification: intraepithelial carcinoma including AK and Bowen's disease; BCC; benign keratinocytic lesions including solar lentigo, SK, and LPLK; dermatofibroma; melanoma; melanocytic nevi; and vascular lesions)	Training set: 10,015 dermoscopic images Test set: 1,195 images	Top 3 algorithms of the ISIC 2018 challenge[73]	<u>Algorithms (mean):</u> Sen: 81,9% Spe: 96,2% <u>Human readers (mean):</u> Sen: 67,8% Spe: 94,0%

Haenssle et al.[71]	Classification (Benign/Malignant) Management decision (treatment/ excision, no action, follow-up examination)	Pretrained CNN Test set: 100 images including pigmented/ non-pigmented and melanocytic/non-melanocytic skin lesions	Inception v4/ Moleanalyzer Pro (CNN)	CNN Management Decision: Sen: 95.0% Spe: 76.7% Acc: 84.0% AUC: 0.918 CNN Diagnosis (Benign/Malignant) Sen: 95.0% Spe: 76.7% Acc: 84.0% Level 1 Management Decision: Dermatologist: Sen: 89.0% Spe: 80.7% Acc: 84.0 % Level 1 Diagnosis (Benign/Malignant) Dermatologist: Sen: 83.8% Spe: 77.6% Acc: 80.1% Level 2 Management Decision: Dermatologist: Sen: 94.1% Spe: 80.4% Acc: 85.9% Level 2 Diagnosis (Benign/Malignant) Dermatologist: Sen: 90.6% Spe: 82.4% Acc: 85.7%
Hekler et al. [72]	Primary end point: Classification to 5 categories (MM/nevus/BCC/AK,Bowen's disease or squamous cell carcinoma/seborrhoeic keratosis, lentigo solaris or lichen ruber planus) Secondary end-point: Binary classification (Benign/malignant)	Training set: 12336 dermoscopic images (585 images of AK,Bowen,SCC, 910 images of BCC, 3101 images of seborrhoeic keratosis,lentigo Solaris,lichen ruber planus, 4219 images of nevi,3521 images of MM)	CNN (ResNet50)	Multiclass classification: Physician Acc: 42.94% CNN Acc: 81.59% Physician+CNN Acc: 82.95% Binary classification: Physician Sen: 66% Spe: 62% CNN Sen: 86.1% Spe: 89.2% Physician+CNN Sen: 89% Spe: 84%

Xinrong Lu et al. [74]	Classification (normal, carcinoma, and melanoma)	HAM10000 dataset [64] Training set: 8012 images (%80) Test set: 2003 images (%20)	Proposed Xception (The ReLU activation function of the model was replaced with the swish activation function) compared with VGG16, InceptionV3, AlexNet and Xception	VGG16: Acc: 48.99 Sen: 53.7 InceptionV3 Acc: 52.99 Sen: 53.99 AlexNet Acc: 75.99 Sen: 76.99 Xception Acc: 92.90 Sen: 91.99 Proposed Xception Acc: 100 Sen: 94.05
Mengistu et al. [75]	Classification (BCC, SCC, MM)	235 images (162 images for training and 73 images for testing)	Combined SOM and RBFNN and compared them with KNN, ANN, and naïve-Bayes	Proposed model Acc: 93,15% KNN Acc:71,23% ANN Acc: 63,01% Naïve-Bayes Acc: 56,16%
Rashid et al. [76]	Classification (MM/Melanocytic Nevus/BCC/AKIEC/Benign Keratosis/DF/Vascular Lesion)	ISIC dataset [52] Training set: 8000 images Test set: 2000 images	GAN compared with CNN (DenseNet and ResNet-50)	GAN Acc: 0,861 DenseNet Acc: 0.815 ResNet-50 Acc: 0.792
Alwakid et al. [77]	Classification (MM/BN/BCC/Vascular lesion/Benign keratosis/Actinic Carcinoma/DF)	HAM10000 dataset [64] 10015 dermatoscopic images Training set: 8029 images Validation set: 993 images Test set: 993 images	Inception-V3, InceptionResnet-V2	Inception-V3 Acc: 0.897 Spe: 0.89 Sen: 0.90 InceptionResnet-V2 Acc: 0.913 Spe: 0.90 Sen: 0.91

Seborrheic Keratosis (SK); International Skin Imaging Collaboration (ISIC); Convolutional Neural Network (CNN); Area under the ROC curve (AUC); Accuracy (Acc); Malign melanoma (MM); Basal Cell Carcinoma (BCC); Squamous Cell Carcinoma (SCC); Actinic keratosis and intraepithelial carcinoma (AKIEC); Dermatofibroma (DF); Actinic keratosis (AK); Human Against Machine with 10000 training images (HAM10000); Benign keratosis, including seborrhoeic keratosis, solar lentigo and lichen planus-like keratosis (BKL); Sensitivity (Sen); Specificity (Spe); Combined convolutional neural network (cCNN); Lichen planus-like keratosis (LPLK); Self-organizing map (SOM); Radial basis function (RBF); Neural network (NN); K-Nearest Neighbors (KNN); Artificial Neural Network (ANN); Generative Adversarial Network (GAN).

The low quality of dermoscopic images and artifacts in the data sets can reduce the success of the models, causing false positives and false negatives or causing overfitting problems. To investigate how much image quality affects the models' success, Maier et al. measured the accuracy and precision of three AI models (AlexNet, ResNet50, and MobileNetv2) in classifying images with different degrees of blurriness and brightness. They report that as blurriness and disparity change, misclassifications occur, model accuracy and improvement decrease, and different models behave differently [78]. Winkler et al. investigated the effect of surgical skin markings, one of the frequently encountered artifacts, on the diagnostic performance of AI models. In unmarked lesions, CNN reached 95.7% sensitivity, 84.1% specificity, and 0.969 AUC, while in marked lesions, specificity decreased to 45.8% and AUC decreased to 0.922 [79]. In another study, the same group investigated the negative impact of the dark corner artifact caused by the tubular lenses of the dermoscope. The small and moderate dark edge artifact did not change the specificity and AUC value, but the large

size reduced the specificity and the AUC remained constant [80]. Artifacts such as hair, gel bubbles, and rulers are also frequently encountered in dermoscopic images. Images may need to be preprocessed to avoid reducing the success of the model. However, the image may become corrupted during this process, and this process itself may cause errors [81]. With improvements in segmentation, the impact of these problems on model success is decreasing [82].

4. In Vivo Skin Imaging Devices

4.1. RCM

In recent decades, non-invasive optical imaging methods have been developed to enhance specificity and enable earlier detection of skin cancers. Confocal microscopy (CM) is among these innovative techniques. There are two primary types of CM: reflectance confocal microscopy (RCM) and ex vivo confocal microscopy (EVCM). RCM, in particular, allows for the in vivo imaging of skin lesions with a "quasi-histologic" resolution, eliminating the need for a biopsy. This imaging technique depends solely on the inherent reflectance contrast of various skin tissue components, without the need for external contrast agents or dyes. As a result, RCM images are presented in grayscale and captured in an en face orientation, in contrast to the "vertical" (i.e., perpendicular to the skin surface) sections commonly used in pathology. RCM has shown to enhance the specificity and sensitivity in diagnosing melanoma, reduce the number of unnecessary biopsies, and assist in margin assessment and surveillance of melanoma. However, RCM also has certain limitations, such as the production of gray-scale raw images, susceptibility to technical artifacts, and a reliance on the expertise of the reader for accurate interpretation [83]. AI can help to overcome these limitations. Algorithms used in melanoma diagnosis with RCM are summarized in Table 4.

Due to the nature of the technique, RCM is susceptible to various artifacts that may impact image quality and diagnostic accuracy. These include faulty reflectance caused by corneal layer reflection or foreign objects such as air or oil bubbles, artifacts from the convexity of nodular lesions or skin creases, and motion artifacts like shifting and misalignment of RCM mosaics due to subtle movements by the patient or technician. Different AI techniques can be employed to detect and eliminate these artifacts. Kose et al. showed that an automated semantic segmentation method called Multiscale Encoder-Decoder Network (MED-Net) could automatically detect artifacts in RCM images of melanocytic lesions with 83% sensitivity and 92% specificity [84].

Another pitfall of RCM is the significant training required for accurate image interpretation, which is essential for achieving high diagnostic accuracy. Gerger et al. developed an automated diagnostic image analysis system using Classification and Regression Trees (CART) to differentiate melanoma from benign nevi in RCM images. The system correctly classified 97.31% of the images in the learning set and 81.03% in the test set [85]. Koller et al. employed a similar machine learning algorithm using Classification and Regression Trees (CART) analysis software to distinguish between benign melanocytic nevi and melanoma in RCM images [86]. The algorithm successfully classified 93.60% of the melanoma images and 90.40% of the nevi images within the learning set. However, its success did not extend to an independent test set, indicating limitations in its generalizability. Wodzinski et al. used a CNN based on ResNet architecture, which achieved an 87% accuracy rate in identifying common skin neoplasms, such as melanoma, BCC, and nevi, using in vivo RCM images. This performance slightly surpassed the diagnostic accuracy of human experts [87]. Kose et al. developed an automated semantic segmentation method known as the Multiscale Encoder-Decoder Network (MED-Net). Their findings demonstrated that MED-Net with "deep supervision" method achieved a pixel-wise mean sensitivity of 70 ± 11% and a specificity of 95 ± 2% for detecting various patterns of melanocytic lesions at the dermal/epidermal junction (DEJ) in in vivo RCM images. Furthermore, MED-Net accurately identified the location and extent of these patterns, achieving a Dice coefficient of 0.71 ± 0.09 [88]. Similarly, D'Alonzo et al. implemented a weakly supervised semantic segmentation model based on EfficientNet, a deep neural network (DNN), to analyze RCM mosaics of pigmented lesions at the dermal-epidermal junction (DEJ). This model was designed to distinguish between non-worrisome ("benign") areas and those suggestive of melanoma ("aspecific").

The trained model achieved an average area under the ROC curve of 0.969 and a Dice coefficient of 0.778, demonstrating the potential for spatial localization of aspecific regions in RCM images, thereby enhancing the interpretability of diagnostic decisions for clinicians [89]. Finally, in Mandal et al., aiming to distinguish Lentigo maligna (LM) from Atypical intraepidermal melanocytic proliferation (AIMP). Authors developed a method that first merges an RCM stack into a single image via local-z projection [90] and then processes the resulting image using DenseNet169, a CNN classifier. Trained and tested over a dataset of 517 RCM stacks (389 LM and 148 AIMP) collected from 110 patients (split into ~80% training vs ~20% testing).   The model achieved an accuracy of 0.80 [91].

Table 4. Algorithms used in melanoma diagnosis with RCM.

Publication	End-Point	Dataset	Algorithm	Performance
Kose et al. [84]	Segmentation; detection of artifacts	117 RCM mosaics	MED-Net; an automated semantic segmentation method	Sensitivity:82%, Specificity: 93%
Gerger et al. [85]	Classification; benign nevi vs melanoma	408 benign nevi and 449 melanoma images	CART (Classification and Regression Trees)	Learning set: 97.31% of images correctly classified Training set: 81.03% of images correctly classified
Koller et al. [86]	Classification; benign nevi vs melanoma	4669 melanoma and 11 600 benign nevi RCM images	CART (Classification and Regression Trees)	Learning set: 93.60% of the melanoma and 90.40% of the nevi images correctly classified
Wodzinski et al. [87]	Classification; benign nevi vs melanoma vs BCC	429 RCM mosaics	a CNN based on ResNet architecture	F1 score for melanoma in test set: 0.84 ± 0.03
Kose et al. [88]	Segmentation; six distinct patterns (aspecific, non-lesion, artifact, ring, nested, meshwork)	117 RCM mosaics	an automated semantic segmentation method, MED-Net	Pixel-wise mean sensitivity: 70 ± 11% Pixel-wise mean specificity: 95 ± 2%, respectively, with 0.71 ± 0.09 Dice coefficient over six classes.
D'Alonzo et al. [89]	Segmentation; "benign" and "aspecific (nonspecific)" regions	157 RCM mosaics	Efficientnet, a deep neural network (DNN)	AUC of 0.969, and Dice coefficient of 0.778
Mandal et al. [91]	Classification; Atypical intraepidermal melanocytic proliferation (AIMP) vs Lentigo Maligna (LM)	517 RCM stacks (389 LM and 148 AIMP) from 110 patients	DenseNet169, a CNN classifier.	Accuracy: 0.80 F1 score for LM: 0.87

Reflectance confocal microscopy (RCM); CART (Classification and Regression Trees); Basal Cell Carcinoma (BCC); Convolutional Neural Network (CNN); Deep neural network (DNN); Atypical intraepidermal melanocytic proliferation (AIMP); Lentigo Maligna (LM).

4.2. Optical Coherence Tomography (OCT) and OCT-like Devices

Optical Coherence Tomography (OCT) is a noninvasive imaging method that captures the echo delays and intensity of reflected infrared or near-infrared light [92]. It enables real-time visualization of the skin, with the ability to penetrate depths of 1–2 mm and deliver a resolution ranging from 3 to 15 μm [93]. Building on the principles of OCT technology, several new devices, such as full-field OCT (FF-OCT), vibrational optical coherence tomography (VOCT), and combination devices like an OCT module with near-infrared Raman spectroscopy, have been developed to enhance accuracy and image quality.

AI has been specifically utilized to assist with image interpretation at various levels in these devices. The delineation of the dermal-epidermal junction (DEJ) is also crucial for diagnosing melanoma in OCT images. Chou et al. utilized a multi-directional CNN to successfully predict the DEJ in FF-OCT images [94]. Silver et al. demonstrated that a machine learning model based on logistic regression achieved a specificity of 77.8% and a sensitivity of 83.3% in distinguishing melanoma from normal skin in VOCT images [95]. Lee et al. trained a SVM using OCT images of melanoma and

20

benign nevi, and subsequently applied this machine learning model to successfully identify pigmented non-malignant lesions in a patient with phacomatosis pigmentokeratotica [96]. You et al. developed an integrated OCT-Raman spectroscopy device and utilized several machine learning models to differentiate between various skin cancer cell types (BCC, SCC, and melanoma) and normal cells in experimentally cultivated cell line models. By applying the decision tree algorithm to OCT features, an accuracy of 85.9% was obtained in distinguishing between cancerous and healthy cells. Moreover, impressively, the discrimination accuracy between melanoma and keratinocytic tumors using all Raman spectra reached 98.9% with the KNN algorithm and 91.6% with the decision tree (TREE) algorithm [97].

5. Conclusions

As dermatologists inherently rely on visual assessments, it remains a field well-suited for the integration of artificial intelligence. The increasing incorporation of AI into dermatological practice holds promise in mitigating clinician workload by optimizing the identification and prioritization of benign lesions at primary care settings, thereby reducing unnecessary invasive testing and ultimately diminishing morbidity and mortality rates. However, despite these advancements, achieving a 100% accurate diagnosis with AI remains an elusive goal, occasionally leading to overdiagnosis and overtreatment of malignant lesions, particularly premalignant lesions that may not progress to cancer.

Another significant challenge arises from the generalizability of AI models to diverse patient populations not adequately represented in training datasets. This issue can manifest in various ways, including differences in image acquisition devices, methods, and demographic characteristics between datasets. Furthermore, the existing literature highlights the need for increased diversity in dermatological imaging datasets, as they predominantly feature individuals with lighter skin tones. This underrepresentation underscores the importance of developing and utilizing AI models that incorporate diverse patient populations to minimize performance disparities.

To address the challenges of generalizability and diversity in AI-driven skin cancer diagnosis, large-scale, inclusive datasets are being developed as standardized benchmarks for evaluating AI performance. The International Skin Imaging Collaboration (ISIC) has played a major role in this effort by aggregating over 1.1 million images from leading cancer research institutions on five continents. Through its regular machine learning challenges, ISIC fosters collaboration between the AI and medical communities, enabling researchers to compare their approaches against standardized datasets and accelerate the development of more accurate and effective AI models for skin cancer diagnosis. Furthermore, ISIC leads the efforts within Digital Imaging and Communications in Medicine (DICOM) for dermatology imaging modalities, driving the standardization of image acquisition protocols and metadata collection practices to minimize discrepancies between datasets. This collaborative approach enables the creation of a more robust and representative dataset, which is essential for developing AI models that can accurately diagnose skin cancers in diverse patient populations. By promoting interoperability and consistency across datasets, ISIC facilitates the development of more reliable and generalizable AI solutions for skin cancer diagnosis.

Despite these advancements, it is essential to acknowledge that dermatological expertise and clinical correlation remain indispensable for ensuring precision in diagnostic evaluations and treatment strategies, particularly for complex cases requiring contextual insights.

References

1. Narayanan, D.L.; Saladi, R.N.; Fox, J.L. Ultraviolet radiation and skin cancer. *International journal of dermatology* **2010**, *49*, 978-986.
2. Society, A.C. What Is Melanoma Skin Cancer? **2023**.
3. Society, A.C. Key Statistics for Melanoma Skin Cancer. **2023**.
4. Rastrelli, M.; Tropea, S.; Rossi, C.R.; Alaibac, M. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo* **2014**, *28*, 1005-1011.
5. Jones, S.; Henry, V.; Strong, E.; Sheriff, S.A.; Wanat, K.; Kasprzak, J.; Clark, M.; Shukla, M.; Zenga, J.; Stadler, M.; et al. Clinical Impact and Accuracy of Shave Biopsy for Initial Diagnosis of Cutaneous Melanoma. *J Surg Res* **2023**, *286*, 35-40, doi:10.1016/j.jss.2022.12.042.
6. Alam, M.; Lee, A.; Ibrahimi, O.A.; Kim, N.; Bordeaux, J.; Chen, K.; Dinehart, S.; Goldberg, D.J.; Hanke, C.W.; Hruza, G.J.; et al. A multistep approach to improving biopsy site identification in dermatology: physician, staff, and patient roles based on a Delphi consensus. *JAMA Dermatol* **2014**, *150*, 550-558, doi:10.1001/jamadermatol.2013.9804.
7. St John, J.; Walker, J.; Goldberg, D.; Maloney, M.E. Avoiding Medical Errors in Cutaneous Site Identification: A Best Practices Review. *Dermatol Surg* **2016**, *42*, 477-484, doi:10.1097/dss.0000000000000683.
8. Dubois, A.; Levecq, O.; Azimani, H.; Siret, D.; Barut, A.; Suppa, M.; Del Marmol, V.; Malvehy, J.; Cinotti, E.; Rubegni, P.; et al. Line-field confocal optical coherence tomography for high-resolution noninvasive imaging of skin tumors. *J Biomed Opt* **2018**, *23*, 1-9, doi:10.1117/1.Jbo.23.10.106007.
9. Cinotti, E.; Couzan, C.; Perrot, J.L.; Habougit, C.; Labeille, B.; Cambazard, F.; Moscarella, E.; Kyrgidis, A.; Argenziano, G.; Pellacani, G.; et al. In vivo confocal microscopic substrate of grey colour in melanosis. *J Eur Acad Dermatol Venereol* **2015**, *29*, 2458-2462, doi:10.1111/jdv.13394.
10. Jones, O.T.; Matin, R.N.; van der Schaar, M.; Prathivadi Bhayankaram, K.; Ranmuthu, C.K.I.; Islam, M.S.; Behiyat, D.; Boscott, R.; Calanzani, N.; Emery, J.; et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. *Lancet Digit Health* **2022**, *4*, e466-e476, doi:10.1016/s2589-7500(22)00023-1.
11. Chu, Y.S.; An, H.G.; Oh, B.H.; Yang, S. Artificial Intelligence in Cutaneous Oncology. *Frontiers in Medicine* **2020**, *7*, doi:10.3389/fmed.2020.00318.
12. Hogarty, D.T.; Su, J.C.; Phan, K.; Attia, M.; Hossny, M.; Nahavandi, S.; Lenane, P.; Moloney, F.J.; Yazdabadi, A. Artificial Intelligence in Dermatology-Where We Are and the Way to the Future: A Review. *Am J Clin Dermatol* **2020**, *21*, 41-47, doi:10.1007/s40257-019-00462-6.
13. Patel, S.; Wang, J.V.; Motaparthi, K.; Lee, J.B. Artificial intelligence in dermatology for the clinician. *Clin Dermatol* **2021**, *39*, 667-672, doi:10.1016/j.clindermatol.2021.03.012.
14. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **1997**, *30*, 1145-1159, doi:https://doi.org/10.1016/S0031-3203(96)00142-2.
15. Nahm, F.S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol* **2022**, *75*, 25-36, doi:10.4097/kja.21209.
16. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297-302.
17. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New phytologist* **1912**, *11*, 37-50.
18. Duarte, A.F.; Sousa-Pinto, B.; Azevedo, L.F.; Barros, A.M.; Puig, S.; Malvehy, J.; Haneke, E.; Correia, O. Clinical ABCDE rule for early melanoma detection. *European Journal of Dermatology* **2021**, *31*, 771-778.
19. Nasr-Esfahani, E.; Samavi, S.; Karimi, N.; Soroushmehr, S.M.R.; Jafari, M.H.; Ward, K.; Najarian, K. Melanoma detection by analysis of clinical images using convolutional neural network. In Proceedings of the 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2016; pp. 1373-1376.
20. Yap, J.; Yolland, W.; Tschandl, P. Multimodal skin lesion classification using deep learning. *Experimental dermatology* **2018**, *27*, 1261-1267.
21. Esfahani, P.R.; Mazboudi, P.; Reddy, A.J.; Farasat, V.P.; Guirgus, M.E.; Tak, N.; Min, M.; Arakji, G.H.; Patel, R. Leveraging machine learning for accurate detection and diagnosis of melanoma and nevi: an interdisciplinary study in dermatology. *Cureus* **2023**, *15*.
22. Dorj, U.-O.; Lee, K.-K.; Choi, J.-Y.; Lee, M. The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications* **2018**, *77*, 9909-9924.
23. Soenksen, L.R.; Kassis, T.; Conover, S.T.; Marti-Fuster, B.; Birkenfeld, J.S.; Tucker-Schwartz, J.; Naseem, A.; Stavert, R.R.; Kim, C.C.; Senna, M.M. Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Science Translational Medicine* **2021**, *13*, eabb3652.
24. Pomponiu, V.; Nejati, H.; Cheung, N.-M. Deepmole: Deep neural networks for skin mole lesion classification. In Proceedings of the 2016 IEEE international conference on image processing (ICIP), 2016; pp. 2623-2627.
25. Han, S.S.; Kim, M.S.; Lim, W.; Park, G.H.; Park, I.; Chang, S.E. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology* **2018**, *138*, 1529-1538.

26. Liu, Y.; Jain, A.; Eng, C.; Way, D.H.; Lee, K.; Bui, P.; Kanada, K.; de Oliveira Marinho, G.; Gallegos, J.; Gabriele, S. A deep learning system for differential diagnosis of skin diseases. *Nature medicine* **2020**, *26*, 900-908.

27. Sangers, T.; Reeder, S.; van der Vet, S.; Jhingoer, S.; Mooyaart, A.; Siegel, D.M.; Nijsten, T.; Wakkee, M. Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: a prospective multicenter diagnostic accuracy study. *Dermatology* **2022**, *238*, 649-656.

28. Potluru, A.; Arora, A.; Arora, A.; Joiya, S.A. Automated Machine Learning (AutoML) for the Diagnosis of Melanoma Skin Lesions From Consumer-Grade Camera Photos. *Cureus* **2024**, *16*.

29. Asan and Hallym Dataset (Thumbnails). **2017**, doi:10.6084/m9.figshare.5406136.v1.

30. Giotis, I.; Molders, N.; Land, S.; Biehl, M.; Jonkman, M.F.; Petkov, N. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert systems with applications* **2015**, *42*, 6578-6585.

31. Ballerini, L.; Fisher, R.B.; Aldridge, B.; Rees, J. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. *Color medical image analysis* **2013**, 63-86.

32. DermIS.

33. Boer, A.; Nischal, K. www. derm101. com: A growing online resource for learning dermatology and dermatopathology. *Indian Journal of Dermatology, Venereology and Leprology* **2007**, *73*, 138.

34. Kato, J.; Horimoto, K.; Sato, S.; Minowa, T.; Uhara, H. Dermoscopy of melanoma and non-melanoma skin cancers. *Frontiers in medicine* **2019**, *6*, 180.

35. Masood, A.; Al-Jumaily, A.A.; Adnan, T. Development of automated diagnostic system for skin cancer: Performance analysis of neural network learning algorithms for classification. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2014: 24th International Conference on Artificial Neural Networks, Hamburg, Germany, September 15-19, 2014. Proceedings 24, 2014; pp. 837-844.

36. Marchetti, M.A.; Codella, N.C.; Dusza, S.W.; Gutman, D.A.; Helba, B.; Kalloo, A.; Mishra, N.; Carrera, C.; Celebi, M.E.; DeFazio, J.L. Results of the 2016 international skin imaging collaboration isbi challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *Journal of the American Academy of Dermatology* **2018**, *78*, 270.

37. Marchetti, M.A.; Liopyris, K.; Dusza, S.W.; Codella, N.C.; Gutman, D.A.; Helba, B.; Kalloo, A.; Halpern, A.C.; Soyer, H.P.; Curiel-Lewandrowski, C. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. *Journal of the American Academy of Dermatology* **2020**, *82*, 622-627.

38. Yu, C.; Yang, S.; Kim, W.; Jung, J.; Chung, K.-Y.; Lee, S.W.; Oh, B. Acral melanoma detection using a convolutional neural network for dermoscopy images. *PloS one* **2018**, *13*, e0193321.

39. Abbas, Q.; Ramzan, F.; Ghani, M.U. Acral melanoma detection using dermoscopic images and convolutional neural networks. *Vis Comput Ind Biomed Art* **2021**, *4*, 25, doi:10.1186/s42492-021-00091-z.

40. Fink, C.; Blum, A.; Buhl, T.; Mitteldorf, C.; Hofmann-Wellenhof, R.; Deinlein, T.; Stolz, W.; Trennheuser, L.; Cussigh, C.; Deltgen, D. Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *Journal of the European Academy of Dermatology and Venereology* **2020**, *34*, 1355-1361.

41. Brinker, T.J.; Hekler, A.; Enk, A.H.; Klode, J.; Hauschild, A.; Berking, C.; Schilling, B.; Haferkamp, S.; Schadendorf, D.; Holland-Letz, T. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer* **2019**, *113*, 47-54.

42. Giulini, M.; Goldust, M.; Grabbe, S.; Ludwigs, C.; Seliger, D.; Karagaiah, P.; Schepler, H.; Butsch, F.; Weidenthaler-Barth, B.; Rietz, S. Combining artificial intelligence and human expertise for more accurate dermoscopic melanoma diagnosis: A 2-session retrospective reader study. *Journal of the American Academy of Dermatology* **2024**, *90*, 1266-1268.

43. Mahbod, A.; Schaefer, G.; Wang, C.; Ecker, R.; Ellinge, I. Skin lesion classification using hybrid deep neural networks. In Proceedings of the ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2019; pp. 1229-1233.

44. Ningrum, D.N.A.; Yuan, S.-P.; Kung, W.-M.; Wu, C.-C.; Tzeng, I.-S.; Huang, C.-Y.; Li, J.Y.-C.; Wang, Y.-C. Deep learning classifier with patient's metadata of dermoscopic images in malignant melanoma detection. *Journal of multidisciplinary healthcare* **2021**, 877-885.

45. Hekler, A.; Maron, R.C.; Haggenmüller, S.; Schmitt, M.; Wies, C.; Utikal, J.S.; Meier, F.; Hobelsberger, S.; Gellrich, F.F.; Sergon, M. Using multiple real-world dermoscopic photographs of one lesion improves melanoma classification via deep learning. *Journal of the American Academy of Dermatology* **2024**, *90*, 1028-1031.

46. Crawford, M.E.; Kamali, K.; Dorey, R.A.; MacIntyre, O.C.; Cleminson, K.; MacGillivary, M.L.; Green, P.J.; Langley, R.G.; Purdy, K.S.; DeCoste, R.C. Using artificial intelligence as a melanoma screening tool in self-referred patients. *Journal of Cutaneous Medicine and Surgery* **2024**, *28*, 37-43.

47. Chanda, T.; Hauser, K.; Hobelsberger, S.; Bucher, T.-C.; Garcia, C.N.; Wies, C.; Kittler, H.; Tschandl, P.; Navarrete-Dechent, C.; Podlipnik, S. Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. *Nature Communications* **2024**, *15*, 524.
48. Correia, M.; Bissoto, A.; Santiago, C.; Barata, C. XAI for Skin Cancer Detection with Prototypes and Non-Expert Supervision. *arXiv preprint arXiv:2402.01410* **2024**.
49. Aswin, R.; Jaleel, J.A.; Salim, S. Hybrid genetic algorithm—Artificial neural network classifier for skin cancer detection. In Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014; pp. 1304-1309.
50. Xie, F.; Fan, H.; Li, Y.; Jiang, Z.; Meng, R.; Bovik, A. Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model. *IEEE Trans Med Imaging* **2017**, *36*, 849-858, doi:10.1109/tmi.2016.2633551.
51. Gutman, D.; Codella, N.C.; Celebi, E.; Helba, B.; Marchetti, M.; Mishra, N.; Halpern, A. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1605.01397* **2016**.
52. ISIC Archive.
53. Cueva, W.F.; Muñoz, F.; Vásquez, G.; Delgado, G. Detection of skin cancer" Melanoma" through computer vision. In Proceedings of the 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), 2017; pp. 1-4.
54. Mendonca, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.; Rozeira, J. PH² - a dermoscopic image database for research and benchmarking. *Annu Int Conf IEEE Eng Med Biol Soc* **2013**, *2013*, 5437-5440, doi:10.1109/embc.2013.6610779.
55. Navarro, F.; Escudero-Vinolo, M.; Bescós, J. Accurate segmentation and registration of skin lesion images to evaluate lesion change. *IEEE journal of biomedical and health informatics* **2018**, *23*, 501-508.
56. Phillips, M.; Marsden, H.; Jaffe, W.; Matin, R.N.; Wali, G.N.; Greenhalgh, J.; McGrath, E.; James, R.; Ladoyanni, E.; Bewley, A. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA network open* **2019**, *2*, e1913436-e1913436.
57. Martin-Gonzalez, M.; Azcarraga, C.; Martin-Gil, A.; Carpena-Torres, C.; Jaen, P. Efficacy of a deep learning convolutional neural network system for melanoma diagnosis in a hospital population. *International Journal of Environmental Research and Public Health* **2022**, *19*, 3892.
58. Ding, J.; Song, J.; Li, J.; Tang, J.; Guo, F. Two-stage deep neural network via ensemble learning for melanoma classification. *Frontiers in Bioengineering and Biotechnology* **2022**, *9*, 758495.
59. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.-A. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging* **2016**, *36*, 994-1004.
60. Bisla, D.; Choromanska, A.; Berman, R.S.; Stein, J.A.; Polsky, D. Towards automated melanoma detection with deep learning: Data purification and augmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2019; pp. 0-0.
61. Bassel, A.; Abdulkareem, A.B.; Alyasseri, Z.A.A.; Sani, N.S.; Mohammed, H.J. Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics* **2022**, *12*, 2472.
62. Nambisan, A.K.; Maurya, A.; Lama, N.; Phan, T.; Patel, G.; Miller, K.; Lama, B.; Hagerty, J.; Stanley, R.; Stoecker, W.V. Improving Automatic Melanoma Diagnosis Using Deep Learning-Based Segmentation of Irregular Networks. *Cancers (Basel)* **2023**, *15*, doi:10.3390/cancers15041259.
63. Collenne, J.; Monnier, J.; Iguernaissi, R.; Nawaf, M.; Richard, M.A.; Grob, J.J.; Gaudy-Marqueste, C.; Dubuisson, S.; Merad, D. Fusion between an Algorithm Based on the Characterization of Melanocytic Lesions' Asymmetry with an Ensemble of Convolutional Neural Networks for Melanoma Detection. *J Invest Dermatol* **2024**, *144*, 1600-1607.e1602, doi:10.1016/j.jid.2023.09.289.
64. Tschandl, P. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. **2018**, doi:doi:10.7910/DVN/DBW86T.
65. Hernández-Pérez, C.; Combalia, M.; Podlipnik, S.; Codella, N.C.F.; Rotemberg, V.; Halpern, A.C.; Reiter, O.; Carrera, C.; Barreiro, A.; Helba, B.; et al. BCN20000: Dermoscopic Lesions in the Wild. *Scientific Data* **2024**, *11*, 641, doi:10.1038/s41597-024-03387-w.
66. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115-118, doi:10.1038/nature21056.
67. Rezvantalab, A.; Safigholi, H.; Karimijeshni, S. Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. *arXiv preprint arXiv:1810.10348* **2018**.
68. Maron, R.C.; Weichenthal, M.; Utikal, J.S.; Hekler, A.; Berking, C.; Hauschild, A.; Enk, A.H.; Haferkamp, S.; Klode, J.; Schadendorf, D.; et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer* **2019**, *119*, 57-65, doi:10.1016/j.ejca.2019.06.013.

69. Tschandl, P.; Rosendahl, C.; Akay, B.N.; Argenziano, G.; Blum, A.; Braun, R.P.; Cabo, H.; Gourhant, J.Y.; Kreusch, J.; Lallas, A.; et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol* **2019**, *155*, 58-65, doi:10.1001/jamadermatol.2018.4378.

70. Tschandl, P.; Codella, N.; Akay, B.N.; Argenziano, G.; Braun, R.P.; Cabo, H.; Gutman, D.; Halpern, A.; Helba, B.; Hofmann-Wellenhof, R.; et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* **2019**, *20*, 938-947, doi:10.1016/s1470-2045(19)30333-x.

71. Haenssle, H.A.; Fink, C.; Toberer, F.; Winkler, J.; Stolz, W.; Deinlein, T.; Hofmann-Wellenhof, R.; Lallas, A.; Emmert, S.; Buhl, T.; et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol* **2020**, *31*, 137-143, doi:10.1016/j.annonc.2019.10.013.

72. Hekler, A.; Utikal, J.S.; Enk, A.H.; Hauschild, A.; Weichenthal, M.; Maron, R.C.; Berking, C.; Haferkamp, S.; Klode, J.; Schadendorf, D.; et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer* **2019**, *120*, 114-121, doi:10.1016/j.ejca.2019.07.019.

73. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368* **2019**.

74. Lu, X.; Firoozeh Abolhasani Zadeh, Y.A. Deep Learning-Based Classification for Melanoma Detection Using XceptionNet. *J Healthc Eng* **2022**, *2022*, 2196096, doi:10.1155/2022/2196096.

75. Mengistu, A.D.; Alemayehu, D.M. Computer vision for skin cancer diagnosis and recognition using RBF and SOM. *International Journal of Image Processing (IJIP)* **2015**, *9*, 311-319.

76. Rashid, H.; Tanveer, M.A.; Khan, H.A. Skin lesion classification using GAN based data augmentation. In Proceedings of the 2019 41St annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2019; pp. 916-919.

77. Alwakid, G.; Gouda, W.; Humayun, M.; Jhanjhi, N.Z. Diagnosing Melanomas in Dermoscopy Images Using Deep Learning. *Diagnostics (Basel)* **2023**, *13*, doi:10.3390/diagnostics13101815.

78. Maier, K.; Zaniolo, L.; Marques, O. Image quality issues in teledermatology: A comparative analysis of artificial intelligence solutions. *J Am Acad Dermatol* **2022**, *87*, 240-242, doi:10.1016/j.jaad.2021.07.073.

79. Winkler, J.K.; Fink, C.; Toberer, F.; Enk, A.; Deinlein, T.; Hofmann-Wellenhof, R.; Thomas, L.; Lallas, A.; Blum, A.; Stolz, W.; et al. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. *JAMA Dermatol* **2019**, *155*, 1135-1141, doi:10.1001/jamadermatol.2019.1735.

80. Sies, K.; Winkler, J.K.; Fink, C.; Bardehle, F.; Toberer, F.; Kommoss, F.K.F.; Buhl, T.; Enk, A.; Rosenberger, A.; Haenssle, H.A. Dark corner artefact and diagnostic performance of a market-approved neural network for skin cancer classification. *J Dtsch Dermatol Ges* **2021**, *19*, 842-850, doi:10.1111/ddg.14384.

81. Sultana, N.N.; Puhan, N.B. Recent deep learning methods for melanoma detection: a review. In Proceedings of the Mathematics and Computing: 4th International Conference, ICMC 2018, Varanasi, India, January 9-11, 2018, Revised Selected Papers 4, 2018; pp. 118-132.

82. Jafari, M.H.; Karimi, N.; Nasr-Esfahani, E.; Samavi, S.; Soroushmehr, S.M.R.; Ward, K.; Najarian, K. Skin lesion segmentation in clinical images using deep learning. In Proceedings of the 2016 23rd International conference on pattern recognition (ICPR), 2016; pp. 337-342.

83. Atak, M.F.; Farabi, B.; Navarrete-Dechent, C.; Rubinstein, G.; Rajadhyaksha, M.; Jain, M. Confocal microscopy for diagnosis and management of cutaneous malignancies: clinical impacts and innovation. *Diagnostics* **2023**, *13*, 854.

84. 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. *Journal of Investigative Dermatology* **2020**, *140*, 1214-1222.

85. Gerger, A.; Wiltgen, M.; Langsenlehner, U.; Richtig, E.; Horn, M.; Weger, W.; Ahlgrimm-Siess, V.; Hofmann-Wellenhof, R.; Samonigg, H.; Smolle, J. Diagnostic image analysis of malignant melanoma in in vivo confocal laser-scanning microscopy: a preliminary study. *Skin Research and Technology* **2008**, *14*, 359-363.

86. 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. *Journal of the European Academy of Dermatology and Venereology* **2011**, *25*, 554-558.

87. Wodzinski, M.; Skalski, A.; Witkowski, A.; Pellacani, G.; Ludzik, J. Convolutional neural network approach to classify skin lesions using reflectance confocal microscopy. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019; pp. 4754-4757.

88. Kose, K.; Bozkurt, A.; Alessi-Fox, C.; Gill, M.; Longo, C.; Pellacani, G.; Dy, J.G.; Brooks, D.H.; Rajadhyaksha, M. Segmentation of cellular patterns in confocal images of melanocytic lesions in vivo via a multiscale encoder-decoder network (MED-Net). *Medical image analysis* **2021**, *67*, 101841.

89.	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. *Scientific Reports* **2021**, *11*, 3679.

90.	Herbert, S.; Valon, L.; Mancini, L.; Dray, N.; Caldarelli, P.; Gros, J.; Esposito, E.; Shorte, S.L.; Bally-Cuif, L.; Aulner, N. LocalZProjector and DeProj: a toolbox for local 2D projection and accurate morphometrics of large 3D microscopy images. *BMC biology* **2021**, *19*, 1-13.

91.	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.

92.	Gambichler, T.; Jaedicke, V.; Terras, S. Optical coherence tomography in dermatology: technical and clinical aspects. *Archives of dermatological research* **2011**, *303*, 457-473.

93.	Sattler, E.; Kästle, R.; Welzel, J. Optical coherence tomography in dermatology. *Journal of biomedical optics* **2013**, *18*, 061224-061224.

94.	Chou, H.-Y.; Huang, S.-L.; Tjiu, J.-W.; Chen, H.H. Dermal epidermal junction detection for full-field optical coherence tomography data of human skin by deep learning. *Computerized Medical Imaging and Graphics* **2021**, *87*, 101833.

95.	Silver, F.H.; Mesica, A.; Gonzalez-Mercedes, M.; Deshmukh, T. Identification of Cancerous Skin Lesions Using Vibrational Optical Coherence Tomography (VOCT): Use of VOCT in Conjunction with Machine Learning to Diagnose Skin Cancer Remotely Using Telemedicine. *Cancers* **2022**, *15*, 156.

96.	Lee, J.; Beirami, M.J.; Ebrahimpour, R.; Puyana, C.; Tsoukas, M.; Avanaki, K. Optical coherence tomography confirms non-malignant pigmented lesions in phacomatosis pigmentokeratotica using a support vector machine learning algorithm. *Skin Research and Technology* **2023**, *29*, e13377.

97.	You, C.; Yi, J.-Y.; Hsu, T.-W.; Huang, S.-L. Integration of cellular-resolution optical coherence tomography and Raman spectroscopy for discrimination of skin cancer cells with machine learning. *Journal of Biomedical Optics* **2023**, *28*, 096005-096005.