SURVEY

Recent Advances in Diagnosis of Skin Lesions Using Dermoscopic Images Based on Deep Learning

YALI NIE [1], (Member, IEEE), **PAOLO SOMMELLA** [2], (Member, IEEE),
MARCO CARRATÙ [2], (Member, IEEE), **MATTEO FERRO** [2], (Member, IEEE),
MATTIAS O'NILS [1], AND **JAN LUNDGREN** [1]

[1]Department of Electronics Design, Mid Sweden University, 85170 Sundsvall, Sweden
[2]Department of Industrial Engineering, University of Salerno, 84084 Fisciano, Italy

Corresponding authors: Yali Nie (yali.nie@miun.se) and Jan Lundgren (jan.lundgren@miun.se)

ABSTRACT Skin cancer is one of the most threatening cancers, which spreads to the other parts of the body if not caught and treated early. During the last few years, the integration of deep learning into skin cancer has been a milestone in health care, and dermoscopic images are right at the center of this revolution. This review study focuses on the state-of-the-art automatic diagnosis of skin cancer from dermoscopic images based on deep learning. This work thoroughly explores the existing deep learning and its application in diagnosing dermoscopic images. This study aims to present and summarize the latest methodology in melanoma classification and the techniques to improve this. We discuss advancements in deep learning-based solutions to diagnose skin cancer, along with some challenges and future opportunities to strengthen these automatic systems to support dermatologists and enhance their ability to diagnose skin cancer.

INDEX TERMS Skin cancer, dermoscopy images, deep learning, classification, literature review.

## I. INTRODUCTION

A. BACKGROUND

Melanoma of the skin is the 19th most commonly occurring cancer in men and women [1]. Skin cancer, and melanoma specifically, is a complex disease. One type of malignant melanoma accounts for about 1 % of all skin cancers, but the vast majority of skin cancer deaths. The most affected regions are Europe, North America, and Oceania [2]. Figure 1 presents a heat map of estimated national, age-standardized melanoma incidence rates in 185 countries in 2020. The countries with the 20 highest rates of skin melanoma in 2020 are given in Figure 2 [2]. Invasive melanoma incidence has been increasing rapidly since the mid-1970s. From 2008 to 2017, the rate increased by about 2 % per year [3]. According to the American Cancer Organization, 106,110 new cases of melanoma of the skin were diagnosed in the U.S. in 2021, while in the same year, 7,180 people died from the disease [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar .
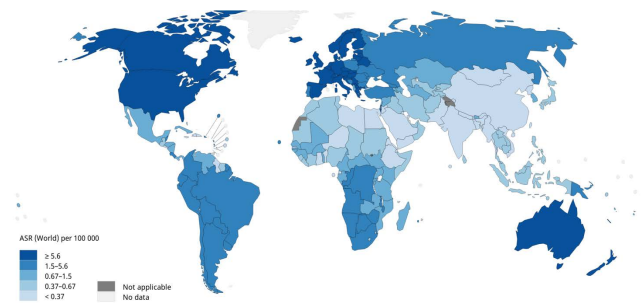
FIGURE 1. Global heat map showing estimated age-standardized incidence rates, in 2020, of melanoma of the skin in all sexes, all ages. The map shows melanoma incidence in all parts of the world, except Greenland in the Arctic Circle. The regions most affected by skin melanoma globally are Europe, the United States, Canada, and Australia [2].

Although the 5-year survival for melanoma of the skin is high, at 93%, early detection of the disease is critically important to reduce melanoma-related mortality [4].
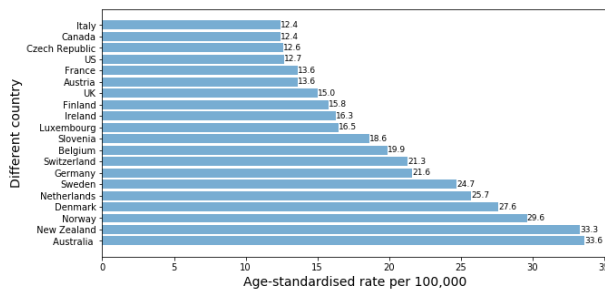
FIGURE 2. The age-standardized rate of skin melanoma per 100,000 in the 20 countries with the highest rates, 2020.

Dermatologists use the two most popular non-invasive techniques, macroscopic (clinical) and dermoscopic, to acquire color images of skin lesions. Dermoscopy is a microscopy-based tool to improve non-invasive diagnostic discrimination of skin lesions based on color and structure analysis [5]. This paper focuses on dermoscopy images. Because dermoscopic structures have direct histopathologic correlates, dermoscopic images help the dermatologist select management and treatment options for particular types of skin cancers [6]. In addition, dermoscopy can be useful for helpful in detecting thinner and smaller cancers and gaining more precision. Pattern analysis, the dermoscopic interpretation method preferred by pigmented lesion specialists, requires assessing numerous lesion patterns simultaneously depending on the location of the body [7]. Some traditional dermoscopic algorithms have been further developed to focus on the most common features of melanoma to aid practitioners with the interpretation of dermoscopy findings: the 7-point checklist (1998), the Menzies method (1996), the asymmetry, border, color, and differential structures (ABCD) rule (1994), the triage amalgamated dermoscopic algorithm (TADA) method (2016), and the color, architecture, symmetry, and homogeneity (CASH) (2006) algorithm [5]. However, the skin melanoma recognition accuracy is not ideal because of the similarity between different skin melanoma and the limited number of dermatologists with professional knowledge. The identification of skin melanoma has become a serious scientific challenge.

More recently, with the rapid development of artificial intelligence (AI) technology, deep learning (DL) has quickly been applied in diagnosis of skin lesions diagnosis. As a result, the medical image processing of skin disease has become an essential component and has received significant attention in the cross-field of image processing, machine science, and intelligent medicine. As a result, many experts and scholars have been engaged in the image recognition of skin disease.

Other survey papers in the field focus either on mature technologies using deep neural networks [8], or they focus on more traditional machine learning [9]. This survey paper instead summarizes in part the improvement of classification results but also innovative technologies for enhancing the CNN frameworks commonly used in skin disease

classification and proposes some directions for current research status and future research.

B. CHALLENGES

The so-called skin lesion classification is that there is a fixed set of classification labels. For each input image, a classification label is found from the classification label set, and classification label is assigned to the input image. Although the classification task seems simple, this is one of the core problems in the field of computer vision. Many seemingly different problems in the field of computer vision (such as object detection and segmentation) can be attributed to image classification problems. The difficulties and challenges of skin disease classification and detection are summarized in three levels in this article: the instance level, the category level, and the semantic level, as outlined below.

1) INSTANCE LEVEL

For a single instance of skin cancer, the size change caused by the difference in the image acquisition process, the lighting conditions, and the shooting angle of view, as well as the distance, the non-rigid body deformation of the object itself, and the partial occlusion of other objects, usually make the apparent characteristics of the object instance.

2) CATEGORY LEVEL

Difficulties and challenges usually come from two directions. Firstly, there is a large intra-class difference when the apparent characteristics of objects belonging to the same class are quite different. The reasons are the changes in the various instance levels mentioned above. Secondly, the difference between different instances in the class has to do with interference from the background: In the actual scene, the object might not appear against a spotless background - in fact, often the background may be very complicated and interfere with the object of interest. This greatly dramatically increases the difficulty of identifying the skin lesion.

3) SEMANTIC LEVEL

Difficulties and challenges are related to the visual semantics of images. Difficulties at this level are often very tough to deal with. Especially for the current level of computer vision theory, a typical problem is what is called "multiple stability". Having the same image but different interpretations are related not only to the physical conditions such as the person's viewing angle and focus, but also to the personality and experience of the person, and this is precisely the part that the visual recognition system finds difficult to handle.

It is a significant challenge for researchers aiming for an accurate diagnosis to tackle these kinds of distortion for precise diagnoses such as: skin hairs, gel bubbles, dark corners, ruler markings, color charts, ink marks, low contrast, incomplete photos and other distortions, as shown in Figure 3.

(a) Skin hairs

(b) Gel bubbles

(c) Dark corners

(d) Ruler markings

(e) Color charts

(f) Marker ink

(g) Distortion

(h) Low contrast

(i) Incomplete

FIGURE 3. The challenges of reaching a diagnosis based on dermoscopic images (the above images were selected from the International Skin Imaging Collaboration (ISIC) archive).
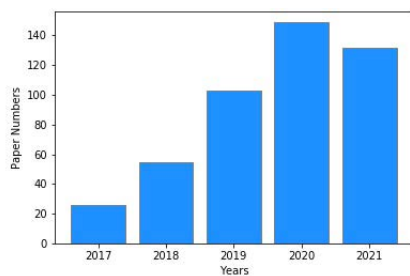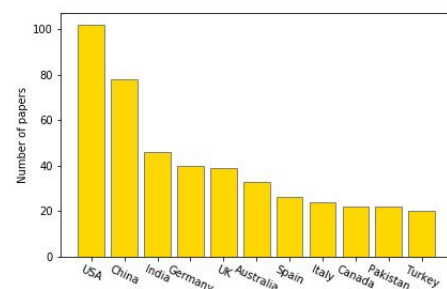

FIGURE 4. Distribution of selected papers, by year of publication.

C. RESEARCH METHOD

This review is mainly based on a literature search on AI and DL in dermatology, performed in Web of Science databases of artificial intelligence and DL in dermatology. The investigation was conducted in November 2021. Most articles from the last 5 years (2017 - 2021) were included to focus on emerging methods. The following primary keywords were used: "deep learning", and "melanoma." Our literature search yielded a total of 441 articles, including 279 journal articles, 19 reviews, 15 meeting abstracts, ten early access articles, and 118 conference papers. Our search showed that research on this aspect of skin diseases is rapidly increasing, as shown

in Figure 4. We have ranked the countries according to the number of articles: see Figure 5 for the eleven countries with the most significant number of articles.

FIGURE 5. Distribution of selected papers on articifial intelligence (AI) and deep learning (DL) in dermatology, by country, 2017–2021.

This study investigates the research status regarding the topic, and diagnosis of a skin lesion in recent years, and summarizes the datasets used by researchers, as well as analyses of image preprocessing, data augmentation, DL models, and framework performance indicators. We aim to provide a reference for DL methods for dermatologists. In addition, the

aim is to enable researchers to quickly and accurately retrieve the literature related to dermatological image recognition. The study's foundation is the rapidly developing AI-based diagnosis technology in the increasing medical AI field.

This study paper is organized as follows. Section I introduces the background, challenges and our research methods of skin lesion. Section II discusses DL and its application in dermoscopic images, while Section III provides some essential techniques utilized to improve melanoma classification in the literature. An overview of classification performance and a discussion are presented in the Sections IV, and V. Section VI concludes the paper.

II. DEEP LEARNING AND ITS APPLICATION IN DERMOSCOPIC IMAGES RECOGNITION

In the following, the basic technical components (frameworks, datasets, and metrics) typically adopted for developing and testing automatic classification systems based on DL are detailed, together with the most current strategies proposed for improving performance in diagnosis of skin cancer.

A. FRAMEWORKS AND BACKBONES
1) DEEP LEARNING FRAMEWORKS

Deep learning frameworks include interfaces, libraries, and tools that allow programmers to develop deep and machine learning models more efficiently than is the case with coding them from scratch. In addition, they provide concise ways for defining models using prebuilt and optimized functions. In addition to speeding up the process of creating machine or DL algorithms, the frameworks offer accurate and research-backed ways to do it, making the end product far more accurate than would be achieved if the entirety of the model was built from scratch. More than two dozen DL libraries developed by tech giants, tech foundations, and academic institutions are available to the public. While each framework has its advantage in a particular subdiscipline of DL, many of them are not currently being maintained by their designers. Therefore, we can talk about only a handful of active and reliable DL frameworks. In this paper, we will discuss three DL frameworks: TensorFlow (TF) [10], Keras [11], and PyTorch [12], which are the most important DL frameworks today (2021). The three are shown detailed in Table 1. The Table also includes some other DL frameworks that have been mentioned in the literature in recent years, namely MatConvNet [13], Caffe [14], and Theano [15].

Excelling in TF with Keras application programming interface (API) is the soundest option. TensorFlow is an open-source machine learning platform focusing on neural networks, which was developed by the Google Brain team. The main reason for choosing TF over other DL frameworks is its popularity. TensorFlow is mighty and easy to use and has excellent community support.

Keras was designed by Google to enable fast experimentation with neural networks. It is very user-friendly, modular, and extensible. Keras also has the advantage of being simple, flexible, and powerful. Because of these features, Keras is viewed by newcomers as the go-to DL framework. Since PyTorch was developed by Facebook and offers an easy-to-use interface, its popularity has gained momentum, particularly in academia. PyTorch is the main competitor of TF.

MatConvNet is a toolkit based on CNN for Matlab, supporting both CPU and GPU. In fact, this toolkit not only supports CNN, but also supports some other networks such as RNN, LSTM, etc. Caffe is an early DL framework made with expression, speed, and modularity. It is ideal for feedforward neural networks and image processing tasks. Theano is based on python whose development started in 2007. This library is good at dealing with multidimensional arrays. With the strong rise of Tensorflow, Keras and Pytorch, MatConvNet, Caffe, Theano are declining day by day, and fewer and fewer researchers use them.

2) CONVOLUTIONAL NEURAL NETWORKS BACKBONES FOR IMAGE CLASSIFICATION

A convolutional neural network (CNN), also known as "ConvNet", is a specific type of feed-forward neural network with a stack of convolutional layers, each followed by pooling layers in order to extract features from the input data and produce a set of high level feature maps at each level of convolution. The feature maps information is summarized using pooling layers in order to reduce the number of parameters and uses a fully connected layer to produce the final classification [16].

The CNN structure evolution summarized in this article started with the neurocognitive machine model. At the same time, the convolutional structure has appeared. The LeNet [17] CNN structure became available in 1998. However, the CNN's edge began to be overshadowed by hand-designed features such as support vector machine (SVM). With the introduction of rectified linear unit (ReLU) and Dropout, as well as the historic opportunities brought by graphics processing units (GPUs) and big data, CNN ushered in a landmark breakthrough in 2012 - AlexNet [16]. Figure 6 presents the evolution of the CNN structure.

Today, researchers rarely build models from start to finish. Common features of classic models have been encapsulated in DL frameworks (such as TF or PyTorch). Researchers only make some modifications on this basis. All the literature collected in this study is based on the CNN model. Compared with traditional machine learning, the CNN model has excellent feature representation (automatically learned from raw data). Currently, the primary method of skin disease image recognition is to use a CNN in DL, and then to use pooling for image recognition. The research work collected in this study adopted famous CNN architecture, such as AlexNet [16], VGG (short for "Visual Geometry Group") [18], Inception [19], ResNet (short for "residual neural network") [20], DensenNet [21], EfficientNet [22], and so on. Figure 7 plots the state-of-art models' performances in dataset ImageNet [23] from 2011 to 2021. Some researchers [24], [25], [26], [27], [28], [29], [30] have
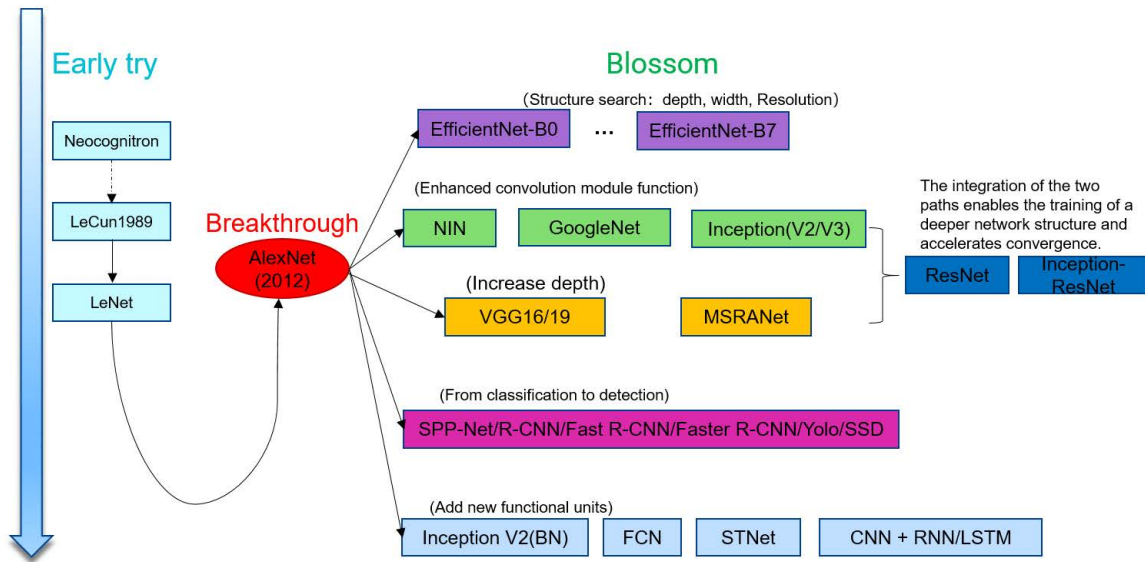
FIGURE 6. The historical evolution of CNN structure has changed from an early attempt to a historic breakthrough, and then to the current prosperity.
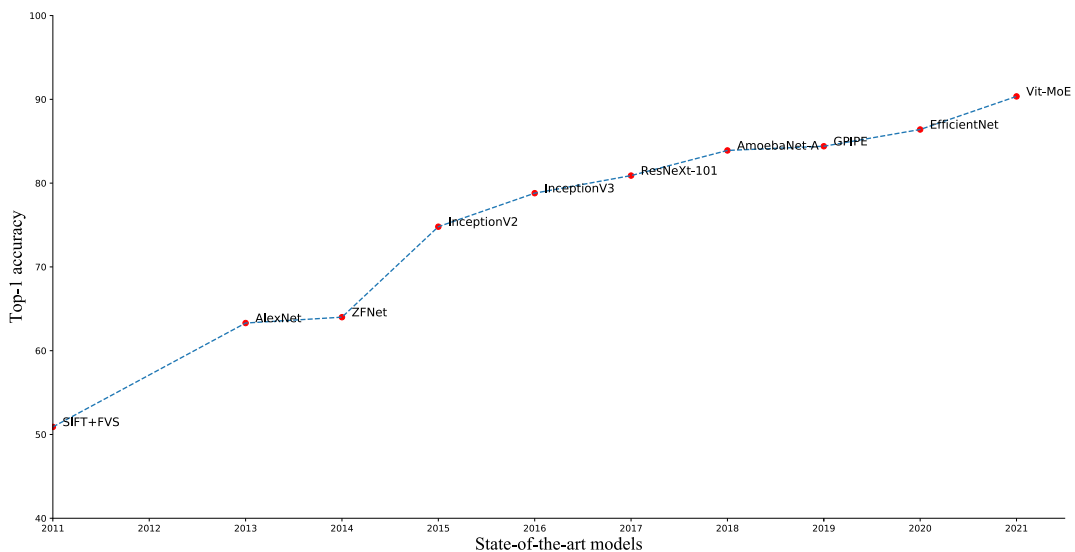
FIGURE 7. The state-of-the-art model in each year from 2011 to 2021. The horizontal axis represents the top-1 accuracy in ImageNet.

preferred to use multiple models to conduct experiments because they allow the opportunity to compare the performance of different models.

B. STANDARD SKIN LESION DERMOSCOPIC IMAGES DATASETS

There are many datasets available for skin lesion classification. Some are publicly available and some are licensed. Deep learning requires a large amount of data to extract features during training. However, large-scale image data of skin lesion are challenging to obtain because images of skin lesions involve patients' privacy; also, there are various skin diseases, and some are rare diseases. Skin lesion

images need to be labeled by experts with appropriate medical knowledge due to the similarity of lesion manifestations between various skin diseases. Currently, the acquisition of skin disease datasets is mainly divided into self-collected and public datasets. Self-collected datasets are usually not publicly available. Most published dermatological datasets are image data obtained by using dermoscopic imaging and collected from dermatological image databases. Universities, in collaboration with renowned hospitals, also collect some datasets.

Regarding public datasets for studying melanoma, the most extensive collection of datasets can be found in the International Skin Imaging Collaboration (ISIC) repository, which

TABLE 1. The most important deep learning (DL) frameworks that were used in study papers and their features.

Framework	Year	Features	References
TensorFlow	2015	Developed by Google; The two programming languages with stable and official TensorFlow APIs are Python and C; Specifically optimized for the training and inference of neural networks; Supports these large numerical computations.	[31]–[44]
Keras	2015	Acquired by Google; Very user-friendly, modular, and extensible; Acts as an interface for the TensorFlow library; Is viewed by newcomers as the go-to DL library.	[26], [31], [32], [42]–[52]
PyTorch	2016	Developed and maintained by Facebook; Offers an easy-to-use interface; Tensor computing with strong acceleration via GPU and deep neural networks built on top of a tape-based automatic differentiation system; Includes the Optim and Neural network (nn) module	[44], [52]–[62]
MatConvNet	2014	An implementation of CNNs for MATLAB; Toolbox is designed with an emphasis on simplicity and flexibility; Exposes the building blocks of CNNs as easy-to-use MATLAB functions, providing routines for computing linear convolutions with filter banks, feature pooling, and many more features [13].	[30], [44], [50], [52], [63]–[76], [76]–[84]
Caffe	2013	A DL framework characterized by its speed, scalability, and modularity.	[84]–[90]
Theano	2009	A Python library that allows to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently [15].	[51], [91], [92]

comprises images labeled by expert dermatologists. Human Against Machine with 10000 training images (HAM10000), Memorial Sloan-Kettering (MSK) and UDA [108] datasets, for example, are held in this repository. Furthermore, this repository provides the different datasets presented in the annual ISIC challenges, commonly used as benchmarks by the researchers. In 2016 [108], the ISIC hosted the International Symposium on Biomedical Imaging (ISBI), and named its 2016 dataset after the ISBI. The ISIC have released five challenging datasets so far: ISBI 2016 [110], ISIC 2017(also known as ''ISBI 2017''), ISIC 2018 [103], ISIC 2019 [95] and ISIC 2020 [93]. The first challenge, ISBI 2016 consisted of two classes with 1,279 images. In the second challenge, ISIC 2017, the number of images and classes increased to 2,000 images while the number of classes increased to three. Thereafter, ISIC 2018 contained 12,500 images, divided into seven classes of skin lesions. The next challenge, ISIC 2019, contained 25,331 images divided into eight classes. The most

recent challenging dataset, ISIC 2020, contains 33,126 different images gathered from more than 2,000 patients at multiple medical centers on three continents, including the Melanoma Institute Australia, the Sydney Melanoma Diagnostic Centre, and the Medical University of Vienna. Each image's metadata included the patient's approximate age at the time of image capture, gender, general anatomic location of the lesion, patient identification number (patient ID), benign/malignant type, and the precise diagnosis (if available). There are 9 subcategories of ISIC 2020. It is indeed an extremely unbalanced database. Moreover, the data can be downloaded in two different formats, Joint Photographic Experts Group (JPEG) or TFRecord. The ISIC Archive contains over 150,000 total images, of which approximately 70,000 have been made public [114] (as of November 12th 2021).

The HAM10000 collected over a period of 20 years from the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff

TABLE 2. The most popular skin lesion datasets.

Name	Number of images	Type	Disease classes	P/N	Reference
ISIC 2020 [93]	33,126	D	9	P	[27], [94]
ISIC 2019 [95]	25,331	D	9	P	[56], [69], [70], [96]–[102]
ISIC 2018 [103]	12,500	D	7	P	[24], [29], [48], [58], [59], [104]–[107]
ISIC 2017 [108]	∼2,000	D	3	P	[46], [85], [104], [109]
ISBI 2016 [110]	1,279	D	2	P	[28], [76], [77], [90], [101], [111]–[113]
ISIC Archive(2018) [114]	23,665	D	7	P	[44], [49], [63], [115]–[119]
HAM 10000 [120]	10,015	D	7	P	[87], [121]–[130]
PH2 [131]	200	D	2	P	[26], [72]–[74], [79], [81], [101]
Atlas [132]	2,022	D & C	2	P	[43], [104]
Dermofit [133]	1,300	D	10	N	[41], [68], [104], [134]
Dermnet NZ [135]	23,000	D & C & H	23	P	[115], [136], [137]
MED-NODE [138]	170	D	2	P	[89], [139]–[141]

1 D: dermoscopic images; C: clinical images; H: histological images
2 P/N: public available or not
3 ISIC: International Skin Imaging Collaboration
4 HAM10000: Human Against Machine with 10000 training images
5 PH2: A dermoscopic image database
6 Dermnet NZ: DermNet New Zealand
7 Atlas: Interactive atlas of dermoscopy

Rosendahl in Queensland, Australia. It consists of 10,015 dermoscopic dermatoscopic images which are released as a training set for academic machine learning purposes and are publicly available through the ISIC archive [120]. A dermoscopic image database (PH2) dataset was built up through a joint research collaboration between the Universidade do Porto, Tecnico Lisboa, and the Dermatology Service of Hospital Pedro Hispano in Matosinhos, Portugal [131]. It has overall 200 melanocytic lesion images.

The interactive atlas of dermoscopy [132] (Atlas) dataset has 1,011 dermoscopic images (252 melanoma and 759 nevi cases), with 7-point checklist criteria. There are also 1,011 clinical color images corresponding to dermoscopic images. The Dermofit Image Library [133] consists of 1,300 high-resolution images with ten classes of skin lesions; use is subject to a licensing agreement, with a one-off license fee of 75 (an academic license is available). DermNet New Zealand (Dermnet NZ) [135] has one of the largest and most diverse collections of clinical, dermoscopic, and histological images of various skin diseases. These images can be used for academic research purposes. Additional high-resolution images are available for purchase. The MED-NODE dataset, created by the Department of Dermatology of the University Medical Center Groningen (UMCG) in the Netherlands, was initially used to train the MED-NODE computer-assisted melanoma detection system [138]. There are 170 non-dermoscopic images in this dataset, 70 of which are melanoma and 100 which are nevi in this dataset.

A summary of the abovementioned skin lesion datasets, including the total number of images, total number of disease classes, whether the dataset is publicly available (and free to use), and the papers using different datasets, are presented in Table 2.

C. METRICS

Standard metrics are needed to assess the performance of different models. Melanoma diagnosis models are assessed according to a variety of metrics based on the number of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) from a DL prediction. These metrics include accuracy(ACC), precision (PREC), sensitivity (SE) and specificity (SP). The ACC metric measures how close the predicted value is to the actual data values. The PREC metric tests the ability of the classifier to reject irrelevant samples. Sensitivity and Specificity are important metrics used in medical diagnosis. The higher the value, the lower the probability of a missed diagnosis. The Sensitivity metric measures the proportion of the correctly detected, relevant samples, which is also known as recall or the "true positive rate (TPR)". Specificity is also called the "true negative rate (TNR)", and the higher the value is, the higher the probability of diagnosis. SP describes the ability of the classifier to detect the TNR.

The F-score is a trade-off between PREC and recall also known as the "F-measure". The formula is expressed as:

$$F_\beta = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall} \quad (1)$$

where β is used to reconcile the importance of PREC and recall. When $\beta = 1$, they are equally important and this is

called "F1-score". The F1-score (or "dice coefficient (DC)") can be obtained by the weighted average of SE (recall) and PREC, where the relative contribution of both recall and PREC to the F1-score is equal. The Matthews correlation coefficient (MCC) is a correlation coefficient that yields a value between -1 and $+1$ for actual and estimated binary classifications. A coefficient of $+1$ shows ideal prediction, 0 shows random prediction, and -1 indicates complete disagreement between predictions and the ground truth. It is generally considered that this indicator is a relatively balanced indicator, and it can be applied even when the sample content of the two categories differs significantly.

The receiver operating characteristic (ROC) curve is plotted with a TP fraction (SE) versus FP fraction (1-SP) by varying the threshold on the probability map. The Area Under the Receiver Operating Characteristics (AUC or AUROC) measures the area under the ROC curve. The term AUC curve refers to the probability that the classifier outputs positive and negative samples, and the likelihood that the classifier outputs a positive sample is greater than of it outputting a negative sample. It represents the complete two-dimensional area within the entire ROC curve from origin (0,0) to point (1,1). The AUC is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve.

ROC curves make it easy to identify the best threshold when making a decision. AUC helps to decide which model is better. Furthermore, AUC is not affected by the class imbalance problem, and different sample ratios will not affect the evaluation results of AUC.

In the AUC calculation formula, the predicted probability is sorted from high to low, and then a rank value is set for each probability value. The rank represents the number of samples that the predicted probability exceeds. To find that the predicted probability value of the positive sample in the combination is greater than that of the negative sample, if the score value of all the positive samples is greater than that of the negative sample, then the first and any combination of the predicted probability value must be larger. Its rank value is n, but M-1 in n-1 is a combination of positive samples and positive samples, which is not within the statistical scope, so it must be subtracted, and so on. Finally, divide by M × N.

These are the most popular measurements typically used for classification evaluation. The specific performance indicators are presented in Table 3.

In addition, for multi-class problems, micro-average and macro-average are used. (1) To calculate the micro-average, the total precision and recall of all categories are calculated and then combined. The calculated average value is the micro-average score. A usage scenario might be that the number of each category is considered in the calculation formula, so it is suitable for data distribution in an unbalanced situation. At the same time, because of the amount of data taken into account, when the data is extremely unbalanced, a larger number of classes will greatly affect the value of average. (2) For the macro-average, the calculation method

is as follows: For all the categories, average the precision and recall, and then calculate the average value as macro-average. A usage scenario might be the following: The amount of data is not considered, so each category will be treated equally (because the precision and recall of each category are between 0 and 1), and will be relatively highly affected by PREC and high recall classes.

Generally speaking, a macro-average will compute the metric independently for each class and then take the average (hence treating all classes equally), whereas a micro-average will aggregate the contributions of all classes to compute the average metric. In a multi-class classification setup, micro-average is preferable if you suspect there might be class imbalance.

Top-N accuracy is another metric, which indicates the capability of a classifier to predict correct class in first N attempts. This metric gives a deeper insight into the classifier's learning and discriminating ability.

A much better way to evaluate the performance of a classifier is to look at the confusion matrix. The general idea is to count the number of times instances of class A are classified as class B. The number of correct and incorrect predictions are summarized with count values and broken down by each class [142].

D. DERMOSCOPIC APPLICATION OF DEEP LEARNING

Because of the similarity in color, texture, edge contour, and other features between different skin lesions, and the difference in pathological tissues between different patients, it is a big challenge to classify skin cancer. Deep convolutional neural networks have been used for general and highly variable tasks across many studies [117], [139], [140], [143], [144], [145], [146], [147], [148], [149], [150].

They can be used to classify skin lesions in two fundamentally different ways.

In the first, a CNN pretrained on another large dataset, such as ImageNet, can be applied as a feature extractor. In this case, classification is performed by another classifier, such as the k-nearest neighbors (kNN) algorithm, SVM, or artificial neural networks (ANNs). In the second way, a CNN can directly learn the relationship between the raw pixel data and the class labels through end-to-end learning. In contrast to the classic workflow typically applied in machine learning, feature extraction becomes an integral part of classification and is no longer considered a separate, independent processing step. If the CNN is trained with end-to-end learning, the research can be divided into two different approaches: learning the model from scratch, and transfer learning.

The landmark publication by Esteva *et al.* [41] belongs to the latter approach and is further discussed below. The proposed CNN model adopts the GoogLeNet Inception v3 model pre-trained with the extensive image database ImageNet and then fine-tuned to classify skin lesions using transfer learning involving more than 120,000 clinical images. The model achieved a value equal to 0.94 for the AUC of the corresponding ROC curves for skin lesions classified

TABLE 3. Evaluation metrics measuring the performance of a predictive model, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values, and their correspondence to the model's accuracy (ACC), precision (PREC), sensitivity (SE), F1-score, Matthews correlation coefficient (MCC), the number of positive samples (M), the number of negative samples(N)and so on.

Metrics	Formula	Explanation	Reference
Accuracy (ACC)	$ACC = \frac{TP+TN}{TP+FP+FN+TN}$	The number of correct predictions divided by the total number of predictions. Ratio of true detected cases to all cases.	[27], [34], [35], [47], [49], [56]–[58], [63]–[66], [122], [125], [151]–[155]
Precision (positive predictive value(PPV))	$PREC(PPV) = \frac{TP}{TP+FP}$	Fraction of relevant instances among the retrieved instances. This is also equivalent to the PPV.	[27], [34], [47], [56], [58], [63]–[65], [122], [125], [154]
Sensitivity (true positive rate(TPR)) Recall	$SE(TPR) = \frac{TP}{TP+FN}$	The ability of the test to correctly identify the diseased state.	[27], [33]–[35], [47], [49], [56]–[59], [63], [65], [66], [122], [125], [151]–[156]
Specificity (true negative rate(TNR))	$SP(TNR) = \frac{TN}{FP+TN}$	This ability of the test to correctly diagnose the benign cases.	[27], [33], [35], [47], [49], [56]–[59], [64]–[66], [151]–[153], [155], [156]
Negative predictive value(NPV)	$NPV = \frac{TN}{TN+FN}$	The proportion of negative examples wrongly categorized as positive.	[65], [74], [141]
F1-Score	$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	This is also called the "F-Measure". The F1-score conveys the balance between the precision and the recall.	[27], [49], [59], [63], [64], [122], [125], [154]
MCC	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	Matthews correlation coefficient is specially designed to analyze the predictive performance of unbalanced data.	[26], [38], [51], [74], [141], [154]
AUC	$AUC = \frac{\sum_{i \in positiveClass} rank_i - \frac{M(1+M)}{2}}{M \times N}$	Area under the receiver operating characteristic (ROC) curve (AU-ROC). This is a probability curve that plots the TPR against FPR at various threshold values.	[27], [33], [35], [49], [56], [58], [66], [151], [154]–[156]

exclusively with dermoscopic images. The very similar approach presented in Haenssle *et al.* [140] (where the modified version of the GoogleNet Inception CNN architecture was additionally trained with more than 100,000 digital images) showed significantly lower diagnostic accuracy (0.86, achieved as AUC for the classification task of melanomas versus benign nevi). In that study, the diagnostic performance of CNN model was compared to that of a group of dermatologists based on a collection of 100 dermoscopic images representing the spectrum of melanocytic lesions typically encountered in daily clinical routine [140].

Regarding the former approach (i.e., learning the model from scratch), the most recent works and meta-analyses carried out by experts in both computer science and dermatology highlight the exploitation of the CNN. Feature extraction can lead to satisfying diagnostic performance (similar to the performance of physicians with long clinical experience) also when DL is applied to small proprietary datasets (typically including < 2,000 dermoscopic images and the corresponding expert annotations and biopsy results) that are often available from the involved clinical institution.

However, in Brinker *et al.* [117], a CNN trained with open-source images was exclusively capable of outperforming dermatologists of all levels hierarchical categories of experience (from junior to chief physicians) in dermoscopic melanoma image classification. The CNN had a more minor variance of results indicating a higher computer vision robustness than human assessment for dermatologic image classification tasks [139]. Maron *et al.* [145] showed that the automated binary classification of dermoscopic melanoma and nevus images can be extended to a multi-class classification problem, thus better reflecting clinical differential diagnoses, while still outperforming dermatologists at a significant level.

III. TECHNIQUES TO IMPROVE CONVOLUTIONAL NEURAL NETWORKS FOR MELANOMA DIAGNOSIS

A. THE BASIC PROCESS OF SKIN CANCER CLASSIFICATION

The skin cancer image classification method based on DL can learn hierarchical feature descriptions in a supervised or unsupervised manner, thus replacing the manual design or selection of image features. The CNN DL model has in recent years achieved impressive results in the image field. Convolutional neural networks directly use image pixel information as input, retaining all the information of the input image to a great extent, through convolution. The operation performs feature extraction and high level abstraction, and the model output is the direct result of image recognition. This direct end-to-end, ''input–output'' learning method has achieved outstanding results and is widely used.

Figure 8 illustrates the flow of melanoma classification which includes: **Data preparation** (the preprocessing techniques also include methods such as contrast enhancement and intensity adjustment, space correction, binarization,

morphological operations, gray-scaling, and noise reduction. At this stage, noise and other artifacts are removed from images. Fekri-Ershad *et al.* [157] applied a color based image retrieval method to perform melanoma detection); **model structure** (which involves defining data input and dimensions, as well as network core modules, classifiers, and loss function and network output); **training the model** (which involves choosing backbone, defining parameters, and constructing and performing training); and **testing and applying the model**. We can also roughly divide the process into four parts: Input, network, training, and output. When we try to improve the effect of model training, we can optimize these four aspects. The traditional melanoma image classification method consists of multiple stages, and the framework is more complicated. The end-to-end CNN model structure can be put in place in one step, and the classification accuracy is greatly improved.

In the past few years, there has been an increasing tendency, not only to develop and use different modern CNN backbones to solve complex real-world problems, but also to apply advanced techniques for achieving better training of these models. Examples include using generative adversarial network (GAN) models, and focusing on focal loss [28], [36], [52], [158], [159], transfer learning techniques, data augmentation methods, and the development of ensembles of CNNs.

This study summarizes several basic guidelines regarding factors that influence model performance, as described by Ng [160]: (1) The expressive ability of the model (depth and width); (2) the learning rate; (3) the optimizer; (4) the learning rate adjustment strategy. In DL, model overfitting often occurs, and methods to reduce the impact of model overfitting usually include data augmentation (data enhancement can increase the data size) and regularization.

B. TRANSFER LEARNING

Transfer learning is a new task that improves learning by transferring knowledge from related tasks that have been learned. For example, there are three tasks: task A, B, and C. They use the same network structure. For a deep neural network, the weights of the CNN layers in the front layer are very close. Here the process of extracting an object features in a CNN model, the first three layers may first extract vertical edges, and then extract horizontal Edge, then extract the round area. So the previous CNN weights do not need to be trained. In order to avoid similar repeating tasks, task C can then use the training results of task A or B to continue training, which can reduce the number of parameters and training time.

Migration ability is the criterion we need to consider when deciding which task model to use. The larger the amount of data in the original model, the stronger the migration capability; and the more similar the problem scenarios of the original model and the new problem, the stronger the migration ability. The stronger the migration ability, the lower the number of layers that need to be frozen, and vice versa.
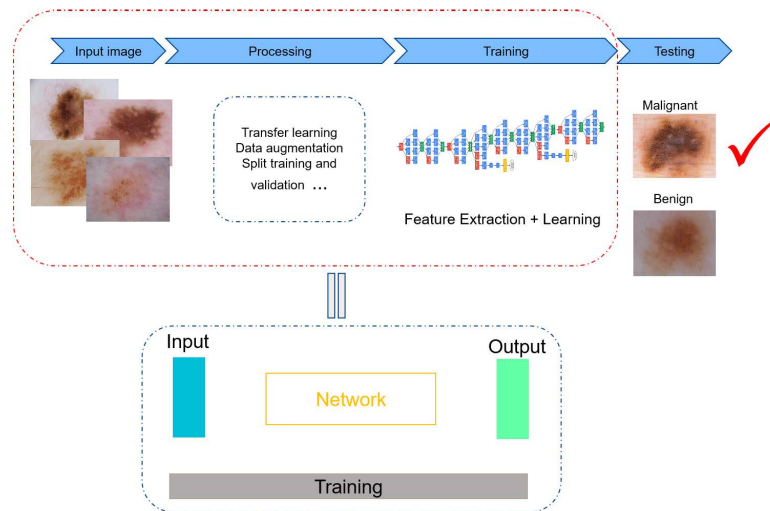
FIGURE 8. Flow chart of melanoma diagnosis based on a general convolutional neural networks (CNN) model in a general way. Image processing is divided into image acquisition, image prepossessing, and dataset division. Image prepossessing includes image size adjustment, normalization, and noise removal. Melanoma image recognition mainly includes image feature extraction and classification models to classify the extracted features and output the results.

For example, task A is trained with more pictures, but task B is a closer training task, so the selection will be contradictory.

Today, with DL being popular, the training of neural networks is becoming more and more time-consuming. The main reason that needs transfer learning is because malignant and benign lesions have high similarity, so it takes a long time to identify and classify them. Moreover, transfer learning is more efficient in classifying between similar lesions, making it a first choice [161]. These papers used transfer learning in the literature we surveyed [25], [26], [28], [30], [33], [34], [35], [36], [37], [38], [39], [41], [42], [46], [52], [58], [61], [62], [64], [66], [67], [68], [70], [71], [72], [73], [75], [76], [76], [77], [85], [86], [87], [92], [102], [112], [113], [122], [124], [126], [127], [129], [141], [151], [152], [158], [159], [162], [163], [164], [165], [166], [167], [168], [169], [170], [171], [172], [173], [174], [175], [176], [177], [178], [179]. Transfer learning can transfer the parameters of the trained model (pre-training model) to the new model to help the new model training. Here are three benefits of transfer learning: firstly, before fine-tuning, the initial performance of the model is higher; secondly, during the training process, the rate of model improvement is faster; thirdly, after the training, the obtained model converges better. Therefore, it is becoming more and more common to use trained neural networks for other tasks such as transfer learning [32].

By using pre-trained models which have been previously trained on large datasets, we can directly use the weights and architecture obtained and apply the learning to our problem statement. This is known as transfer learning. We "transfer the learning" of the pre-trained model to our specific problem statement. You should be very careful while choosing what pre-trained model you should use in your case. If the

problem statement we have at hand is very different from the one on which the pre-trained model was trained – the prediction we would get could be very wildly inaccurate. For example, a model previously trained for speech recognition would most likely be very inaccurate if we try to use it to identify objects. Imagenet data set has been widely used to build various architectures since it is large enough (1.2M images) [23] to create a generalized model. These pre-trained networks demonstrate a strong solid ability to generalize to images outside the ImageNet dataset via transfer learning. There are three ways to fine-tune the model: (1) use a pre-trained model as feature extraction and remove the out layer; (2) use the architecture of the model while we initialize all the weights randomly and train the model according to our dataset again; (3) train some layers while freezing others. AlexNet, SqueezeNet, MobileNet, Google Inception, ResNet, Xception, VGGNet, DenseNet are examples of commonly used pre-trained CNNs [25].

C. DATA AUGMENTATION

Deep learning models show remarkable results in automated skin lesion analysis. However, these models require considerable amounts of data, while the availability of annotated skin lesion images is often limited. Data augmentation is a way to expand the training dataset by transforming input images without having to collect new datasets for model training, thus avoiding the overfitting issue that might occur during the training process when a small amount of training data is used. These papers use data augmentation for performance enhancement: [25], [26], [34], [35], [36], [40], [41], [46], [46], [47], [49], [50], [52], [56], [58], [60], [60], [61], [64], [67], [68], [85], [88], [90], [91], [98], [105], [107],

[109], [115], [116], [122], [124], [125], [126], [127], [130], [151], [152], [154], [155], [158], [159], [162], [163], [164], [165], [166], [168], [169], [170], [180], [181], [182], [183], [184], [185], [186]. The literature includes several works on data augmentation. Perez *et al.* [187], describe the impact of 13 data augmentation scenarios for melanoma classification trained on three different CNNs, such as contrast, flips, random crops, scaling. Kato *et al.* [188] used data augmentation to demonstrate how the system improves diagnostic performance by executing vertical or horizontal inversion (or both) to the original single-wavelength images, thus increasing the training dataset fourfold. Zhao *et al.* [56] applied flip vertical and flip horizontal resizing and rotation on ISIC2019 to perform skin lesion image classification. In the following, we summarize several commonly used data augmentation strategies:

1) GEOMETRICAL TRANSFORMATION
Geometrical transformation methods include random reflection, rotation, translation, shearing, minimizing, zooming, and scaling [25], [34], [36], [164], [172], [189].

2) COLOUR JITTER
Common color jitter methods are adjustments of brightness, contrast, saturation, and HVS (hue, value, and saturation). They change the ratio between each color channel, or values of the multiplication factor or different magnitudes. Oukil *et al.* [190] applied color features in dermoscopic images and achieved good results.

3) NOISE ADDITION
Noise addition consists of addition of a random value drawn from different noise distributions while preserving the important features of the images. Gaussian noise, Poisson noise, and Salt & Pepper noise are common types. When the neural network is trying to learn high-frequency features that may be useless, adding a moderate amount of noise can avoid overfitting. Noise addition is usually used with GAN algorithms. The use of informative noise allows the GAN to avoid mode collapse and creates faster convergence [191].

4) MULTISAMPLE TECHNIQUE
Synthetic Minority Over-sampling Technique (SMOTE) [192], based on interpolation method, can synthesize new samples for small sample classes. It is used to deal with the sample imbalance problem by artificially synthesizing new samples, thereby improving the performance of the classifier. Sample pairing [193]is another way to enhance the training data. In this technique, two images are randomly selected from the training set and processed by basic data enhancement operations (such as random flip); thereafter, the pixels are superimposed to create a new sample in the form of averaging, and the label is one of the original sample labels. The third technique is mixup [194]. Lee and Chin [195] applied vertical half mixing, horizontal half

mixing, diagonal–quadrant mixing, four-quadrant mixing, four-column mixing, and region of interest (ROI) mixing to augment data. All these techniques aim to augment the discrete sample points to fit the true sample distribution.

5) GENERATIVE ADVERSARIAL NETWORKS
Generative adversarial networks (GANs) [196] provide a path for sophisticated domain-specific data augmentation and a solution to problems that require a generative solution. They are based on a game theoretic scenario in which the generator network must compete against an adversary. The generator network directly produces samples. During the past few years, GANs develop rapidly. These [56], [62], [109], [124], [158], [169] applied GANs algorithm to skin lesion classification.

Abdelhalim *et al.* [124] used GANs to generate fine-grained 256×256 skin lesion images for CNN-based melanoma detection, which led to significant improvements with sensitivity increased by 5.6 % over non-augmented counterparts. Zhao *et al.* [56] proposed a skin lesion image classification approach based on a skin lesion augmentation according to style-based GAN and DenseNet201. This method generated high quality skin lesion images and performed well on the ISIC 2019 dataset(its balanced multiclass accuracy achieved 93.64%). Qin *et al.* [169] also applied style-based GANs data augmentation technology to improve the skin lesion classification performance. While a cycle consistent adversarial networks (cycle-GAN) for skin lesion image synthesizing was adopted by Gu *et al.* [62]. Pollastri *et al.* [109] proved that a Laplacian Generative Adversarial Network (LAPGAN) can be employed to obtain an accuracy boost equivalent to 138% more real annotated images when the dataset is over 500 images.

6) AUTOAUGMENT
The basic idea of Autoaugment [197] is to use reinforcement learning to find the best image transformation strategy from the data itself, and learn different augmentation methods for different tasks.

The latter two methods are often used for unsupervised data augmentation.

D. ENSEMBLE LEARNING
The classification of skin lesions has in recent years relied on the ensemble method to achieve highly accurate performance [29], [30], [31], [32], [38], [69], [72], [76], [80], [81], [82], [85], [87], [96], [105], [105], [111], [121], [129], [177], [182], [198], [199], [200], [201], [202], [203]. Generally, current researchers applying ensemble methods follow a similar workflow. First, several multiclass CNNs that are trained for a specific task, and then their outputs are merged using an aggregation approach. An overview of related works applying ensemble methods is provided in Table 4. The most used aggregation methods are:

1) WEIGHTED MAJORITY VOTING STRATEGY

Weighted majority vote strategy is used in popular ensemble learning algorithms, which tends to select among high probability values of the class that has received the highest number of votes [204].

2) MODEL AVERAGING STRATEGY

The ensemble prediction is calculated as the average of the member predictions [205]. There is a requirement that all ensemble members have skill as compared to random chance, although some models are known to perform much better, or much worse, than other models.

3) WEIGHTED AVERAGE STRATEGY

The weighted ensemble is an extension of a model averaging ensemble where the contribution of each member to the final prediction is weighted by the performance of the model [206]. The model weights are small positive values and the sum of all weights equals 1, allowing the weights to indicate the percentage of trust or expected performance from each model.

4) DECISION DIRECTED ACYCLIC GRAPH STRATEGY

The decision directed acyclic graph (DDAG) is a graph whose edges have an orientation and no cycles. The DDAG ensemble method is a decision tree that combines a set of binary classifiers into a multiclass classifier [105].

5) GEOMETRIC AVERAGING STRATEGY

The geometric averaging method (also called ''geometric mean method'') aims to find diverse networks with relatively small steps in the weight space, without leaving a region that corresponds to low test error [207].

IV. OVERVIEW OF CLASSIFICATION PERFORMANCE

The publication by Esteva *et al.* [41] was important because, although not strictly focused on dermoscopic images, it clearly showed the potential of DL techniques when applied to the domain of cutaneous oncology. In the years following their study, great research efforts were invested in introducing new DL solutions to solve the problems arising from the application to dermoscopy, first of all represented by the availability of small datasets (when compared to clinical image sets). Very important were the ISIC challenges which provided the opportunity to compare original proposals from many international research groups. For example, the new ResNet models [24] were introduced and emerged as a valid technique that was able to guarantee better results (with respect to the performance exhibited by traditional models such as AlexNet, GoogleNet, and VGG models) for both skin lesion segmentation and the melanoma classification problems. Table 5 presents the performance of the top five research groups on ISIC challenges of 2016–2019.

Better results are also reported in a comparative study of DL architecture on melanoma detection using dermoscopic

images [208]. Preprocessing methods such as illumination correction, contrast enhancement, and artefact removal are suggested to improve image quality and obtain a better generalization ability. Due to the imbalanced class distributions of skin lesions, various augmentation approaches are adopted in these methods. Various standard evaluation metrics, such as SP, SE, ACC, and F-measure, are employed to evaluate the obtained results. Finally, experiments show that ResNet50 outperforms its counterparts AlexNet, Xception, VGGNet16, and VGGNet19 architecture, with a classification ACC as high as 92.08% and an F-score equal to 92.74%.

A very interesting meta-analysis including more than 200 studies on the research emanating from the field of computer science is reported by Dick *et al.* [208]. Combining all the results for automated systems gave a melanoma SE of 0.74 (95% CI 0.66–0.80) and an SP of 0.84 (95% CI 0.79–0.88). Although the SE was lower in studies that used independent test sets than in those that did not, the SP was similar. Moreover, in comparison with dermatologists' diagnoses, computer-aided diagnoses showed similar SEs and a 10 percentage point lower SP, but the difference was not statistically significant. As main conclusion of the meta-analysis, the ACC of computer-aided diagnosis for melanoma detection may be considered comparable to that of experts; nevertheless, the real-world applicability of these systems is as yet unknown and potentially limited owing to overfitting and the risk of bias of the available studies.

Responses to the main doubts arising from this type of analysis may be found in studies carried out mainly by physicians and focused on the well-recognized DL CNN models. Among them, interesting results are reported by Brinker *et al.* [150] who compared AI algorithms to classifications made by 157 German dermatologists. Haenssle *et al.* [149] report results where, under less artificial conditions and in a broader spectrum of diagnoses, the CNN and most dermatologists performed on the same level; they [140] also compared the diagnostic performance of a CNN with that of a large international group of 58 dermatologists from 17 countries, including 30 experts with more than 5 years of dermoscopic experience. Their data clearly show that a CNN algorithm may be a suitable tool to aid physicians in melanoma detection, irrespective of their level of experience and training. An adequately trained DL CNN can provide a highly accurate diagnostic classification of dermoscopic images of melanocytic origin. Therefore, physicians of all levels of training and experience may benefit from assistance in the form of a CNN image classification. In a study by Brinker *et al.* [117], a CNN trained with open-source images was exclusively capable of outperforming dermatologists of all levels of experience in dermoscopic melanoma image classification. The CNN had lower variance of results, indicating a higher robustness of computer vision, compared to human assessment, for dermatologic image classification tasks [139]. Maron *et al.* [145] showed that the automated binary classification of dermoscopic melanoma and nevus images can be extended to a multiclass

TABLE 4. Overview of the related studies using ensemble methods with convolutional neural networks (CNNs) for skin disease diagnosis.

Paper	CNN models	Aggregation methods	Year
[198]	10 CNN models	Geometric averaging	2017
[111]	10 U-Nets	Model averaging	2017
[82]	GoogleNet, AlexNet, ResNet50, and VGGNet16	Weighted majority voting	2018
[199]	ResNet and Inception	Model averaging	2018
[200]	13 CNN models	Model averaging	2019
[201]	AlexNet, VGG16, and ResNet18	Model averaging	2019
[31]	VGG16, VGG19, ResNet50, InceptionV3, Xception, and DenseNet121	Model averaging	2019
[32]	EfficientNetB0, EfficientNetB1 and SeReNeXt-50	Model averaging	2020
[96]	EN B0-B6, ResNeXt, and SENet154	Model averaging	2020
[105]	VGG16, VGG19 and ResNet50	DDAG	2021
[202]	InceptionV4, ResNet, DenseNet121, and DenseNet145	Weighted average	2021
[203]	EfficientNetB0, Xception, and DenseNet121	Model averaging	2021

¹ CNN: convolutional neural network
² DDAG: the decision directed acyclic graph

classification problem, thus better reflecting clinical differential diagnoses, while still outperforming dermatologists at a significant level.

The promising results in a clinical setting have further led to testing the combination of human and AI. Regarding the multiclass task, the combination of "man and machine" reported by Hekler *et al.* [147] achieved an ACC of 82.95%. This was 1.36% higher than the best of the two individual classifiers (e.g., 81.59% achieved by the CNN). Owing to the class imbalance in the binary problem, SE, but not ACC, was examined and demonstrated to be superior (89%) to the best individual classifier (CNN, with 86.1%). The SP in the combined classifier decreased from 89.2% to 84%. However, at an equal SE of 89%, the CNN achieved a SP of only 81.5%. Therefore, the findings clearly indicate that the combination of human and AI classification achieves superior results over the independent results of either of these classifiers.

V. DISCUSSION

Most experiments are conducted on a GPU to speed up the training and deployment process. We have mentioned that, to enhance the quality of images, some employ different preprocessing steps. Data augmentation, transfer learning, and ensemble techniques all address the class ACC problem. In this section, we will discuss some salient aspects of melanoma classification and the outlook for the future.

A. THE HAIR REMOVAL

Hair should preferably be removed in dermoscopy applications because it causes undesired effects such as occlusions in lesion areas. Kim and Hong [27] used a CycleGAN to remove hair in melanoma classification. Their results in ISIC 2020 verify that applying the proposed hair elimination algorithm significantly enhances the performance of the melanoma classification, outperforming the benchmarks. Zhao *et al.* [56] applied inpainting algorithms to replace the

pixel values and used a black top-hat filter with a grayscale image. Attia *et al.* [79] performed a survey on hair detection and also conducted experiments with hybrid CNNs. Since DL uses a set of cascaded, sequential layers that operate on the input data, each layer performs a non-linear processing operation to extract a hierarchical representation (achieved by extraction of feature maps) of the input pixels based on the neighborhood. As the activation maps have higher values at the "hair" or "ruler marking" pixels, this achieves the purpose of detecting hair. After removal of the hair, the skin lesion becomes clearer; removing hair can help the classification model to better identify the lesion location in the skin lesion image and improve the ACC of classification results [56].

B. DATA BALANCE

Imbalanced classification is the problem of classification when there is an unequal distribution of classes in the training dataset. The imbalance in the class distribution may vary, but a severe imbalance is more challenging to model and may require specialized techniques. Zhao *et al.* [56] propose a skin lesion augmentation style-based GAN to address insufficient data samples, unbalanced data, and missing labels data. They also introduced the use of A-SoftMax and focal loss to solve the imbalance problems of ISIC 2019. Vasconcelos and Vasconcelos [112] used data augmentation to deal with small and unbalanced ISBI 2016 datasets. Pham *et al.* [126] used a combination of balanced mini-batch logic and real-time image augmentation, which is effective in training the networks with imbalanced skin datasets. Dong *et al.* [210] addressed the class imbalance in large-scale image classification with a novel loss function and hard sample mining. Johnson and Khoshgoftaar [211] have made a summary of DL class imbalance methods and hybrid methods, detailing methods that can be classified as data level-based, and as algorithm level-based. To alleviate the data imbalance problem,

TABLE 5. The top five dermatological classifications and their performance in the annual International Skin Imaging Collaboration (ISIC) challenges from 2016 to 2019 [209].

Dataset	Approach name	AUC	Average PREC	ACC	SE	SP	F1-score	PPV	NPV
ISIC 2016	CUMED	0.804	0.640	0.855	0.507	0.941	0.580	0.679	0.885
	GTDL	0.802	0.622	0.813	0.573	0.872	0.548	0.524	0.892
	BF_TB	0.826	0.601	0.834	0.320	0.961	0.432	0.667	0.851
	ThrunLab-Mjolnir	0.796	0.567	0.786	0.667	0.816	0.552	0.472	0.908
	Jordan Yap	0.775	0.563	0.844	0.240	0.993	0.379	0.900	0.841
ISIC 2017	ResNet ensemble with normalized image	0.911	0.750	0.816	0.856	0.812	0.612	0.488	0.962
	FCN + modified ResNet-50	0.910	0.748	0.849	0.140	0.998	0.242	0.932	0.847
	VGG + U-shape	0.908	0.754	0.883	0.451	0.970	0.564	0.796	0.897
	EResNet	0.896	0.733	0.888	0.508	0.970	0.612	0.775	0.902
	Multi-task deep learning model	0.886	0.667	0.873	0.568	0.940	0.608	0.659	0.909
ISIC 2018	Ensembling CNNs + 5-fold	0.983	0.917	0.958	0.833	0.986	0.823	0.826	0.952
	Large ensemble with heavy multi-cropping and loss weighting	0.987	0.931	0.972	0.809	0.984	0.841	0.888	0.972
	Ensemble Of SENET and PNANET with Data augmentation	0.978	0.891	0.968	0.804	0.980	0.830	0.861	0.970
	Densenet	0.980	0.892	0.969	0.789	0.976	0.828	0.875	0.975
	Approach 3: average of approach 1 and 2	0.960	0.833	0.939	0.758	0.964	0.750	0.763	0.949
ISIC 2019	Ensemble of Multi-Res EfficientNets + SEN154 2	0.923	0.569	0.926	0.507	0.977	0.515	0.597	0.940
	Ensemble of EfficienetB3-B4-Seresnext101	0.780	0.364	0.917	0.607	0.952	0.532	0.507	0.952
	Ensemble	0.886	0.560	0.924	0.540	0.963	0.520	0.584	0.950
	13 models + hierarchical approach	0.892	0.550	0.919	0.507	0.965	0.502	0.560	0.943
	Densenet-161 with heavy use of random crops	0.870	0.489	0.910	0.473	0.967	0.432	0.450	0.933

[1] CNN = convolutional neural network; ACC = accuracy; AUC = area under the curve; NPV = negative predictive value; PREC = precision; PPV = positive predictive value; SE = sensitivity; SP = specificity
[2] So far, ISIC 2020 is still in competition, with 3308 teams and $30,000 prize money(Dec. 3rd, 2021)

Sayed *et al.* [25] used a random oversampling method followed by data augmentation. When the dataset is imbalanced, using only ACC for evaluation of results is not enough; a confusion matrix, and PREC, recall and the F1-score also need to be applied. Imbalanced data is one of the potential problems in the field of DL for skin cancer classification. This problem can be approached by properly analyzing the melanoma data.

C. THE IMPACT OF SEGMENTATION
Recent advances in CNN architectural models with the ability of semantic segmentation have been utilized by academics to segment skin lesion images [152]. Skin lesion segmentation plays a vital role in the proper classification of skin cancer using computer-based models. Khouloud *et al.* [212] used W-net and inception residual network and report best performance with an ACC of 97.39% and a dice coefficient of 93% for the segmentation process on the ISIC 2018 dataset. A fully convolutional neural network (FCN) [213], and U-Net [214], SegNet [215], and DeepLab [216] are the classic semantic segmentation networks in DL. More and more researchers [28], [36], [46], [48], [60], [67], [68], [91], [109], [123], [164] have worked on skin lesion segmentation in recent studies.

D. METADATA INFORMATION
Metadata consists of data on lesion location, lesion size, lesion anatomy site, and history of psoriasis, along with the age and gender of the patient. Pacheco and Krohling [217] present an improvement of approximately 7% in balanced ACC when applying metadata information on DL models. Ningrum *et al.* [49] used a CNN model, with only image input information, yielding an AUROC of 82.40%. For comparison, using of the CNN + ANN model with a combination of image and patient metadata yields an AUROC of 97.10% [49]. Liu *et al.* [34] report that in their study, the type of self-reported skin problem (e.g., acne, hair loss, or rash) and history of psoriasis had the greatest impact on ACC. Overall, the impact of metadata on the performance of models is significant and shows the importance of including these features in automated skin cancer detection.

E. CLINICAL AND HISTOPATHOLOGICAL IMAGES
In the clinical setting, diagnosis of skin cancer is conducted by inspecting the skin lesion with or without dermoscopy, followed by confirmatory biopsy and pathological examination. Dermoscopic images alone cannot provide all the skin lesion information. Pacheco & Krohling [217] have demonstrated the importance of clinical features in skin cancer detection and confirm the hypothesis that patient clinical information is important for skin lesion classification. The Atlas dataset provides clinical and dermoscopy images. Wang *et al.* [163] propose using two stream CNN processing based on this dataset in clinical and dermoscopy images. Still, these dermoscopic and clinical skin lesion datasets do not have corresponding pathological classification labels to develop a complete diagnosis pipeline for the computer-aided systems in current

publicly available skin lesion datasets. Moreover, most of the classification labels for dermoscopic skin lesion images are determined by pathological examination. Hekler *et al.* [218] illustrate the potential of DL to assist human assessment for a histopathologic melanoma diagnosis.

F. SMARTPHONE APPLICATIONS
Smartphone applications (apps) provide users with an instant assessment of skin cancer risk and offer the potential for earlier detection and treatment, which could improve the survival of patients. Against the background of the high burden of skin cancer in the world and limited access to dermatological care, particularly in remote areas, AI diagnostic tools provide the possibility to improve triage and reduce the time to excision for correctly diagnosed melanomas. If the mobile device is used properly, this could also reduce morbidity resulting from unnecessary biopsies. In a review paper [219], Freeman *et al.* show currently available apps, such as skinScan, SkinVision, and TeleSkin. There is no skin cancer risk stratification smartphone app that has received U.S. Food and Drug Administration (FDA) approval to date [219]. A combined reference standard comprising histology and clinical follow-up of benign lesions would provide more reliable and generalizable results. Smartphone algorithm-based apps for skin cancer all include disclaimers that the results should only be used as a guide and cannot replace health care advice [219].

G. LIGHT AND SOUND INFORMATION FOR SKIN LESION DIAGNOSIS
In recent years, some researches have emerged that use wavelength or polarization of light and combine sound information with skin lesion image information. In the field of biomedical imaging and diagnostics, polarization speckle is a growing fast. Wang *et al.* [162] used DL to extract skin lesion information from polarization speckle, and improved the performance in classifying benign and malignant skin lesions by 20%. Pölönen *et al.* [220] showed that use of the spectral and spatial domain will increase classification performance of CNNs. Dascalu *et al.* [221] acquired dermoscopy images by skin magnifier with polarized light with DL algorithm and sonified in the first phase; in the second phase, they did further analysis with a different DL. Whether it is spectral information or sound information, it has opened up a new way of thinking for skin lesion diagnosis. However, the existing public datasets hardly provide skin lesion data with light or sound information. So this is a challenge for most researchers.

H. FUTURE PROSPECTS
Deep learning shows great potential in the image-based diagnosis of skin cancer. However, there is still a significant discrepancy between expectations and true relevance of DL in current dermatological practice based on dermoscopy. In numerous studies we have cited, e.g. [27], [33], [56], [63], [105], [116], [123], [151], [163], [164], [202]. In this study, computer algorithms were able to detect pigmented

and non-pigmented neoplasms of the skin with high precision, comparable to dermatologists. The combination of the physician's assessment and AI has shown the best results.

Computer-based diagnostic systems are widely accepted among patients and physicians. Nevertheless, they are still not applicable in daily practice, as they have been tested chiefly in experimental environments. Some cases involving less artificial conditions and a broader spectrum of diagnoses have been reported where the CNN and most dermatologists performed on the same level [149]; however, many digital diagnostic criteria that help AI to classify skin lesions remain unclear. This lack of transparency still needs to be addressed. On the other hand, dermatologists are trained to integrate information from a range of sources, rendering comparative studies that are solely based on one single case image inadequate. Therefore, further and different clinical studies on the use of AI-based assistance systems are needed to prove the applicability of AI in daily dermatologic practice. Indeed, the different CNN-based approaches proposed in the literature should be revised and compared, not only regarding ACC, but also considering real possibilities to: create algorithms representing diverse patient populations; ensure that algorithm output is ultimately interpretable; prospectively validate algorithm performance; preserve human–patient interaction where necessary, and demonstrate validity in the eyes of regulatory bodies. In other words, future research on the development of DL techniques for dermoscopic application should take better account of the main deontological, legislative, and economic requirements involved in the complex clinical process of the skin cancer diagnosis.

Indeed, the dermoscope to catch ELM images is classified as a Medical Device whose commercialization and adoption in European market should be in accordance with the European Union Medical Device Regulation (EU MDR) [222]. In detail, the dermoscope may be purchased only by medical personnel and/or adopted by institution and enterprise for research purpose. Thus, the adoption of dermatoscope by the patients jointly with smartphone applications for instant assessment of the skin cancer risk is actually and will continue to be allowed only in research projects and setting, but not as a routine clinical scenario.

Moreover, according to EU MDR, the application software itself for processing the dermoscopic images and evaluating the skin cancer risk should be classified as Class II or III Medical Device (because of the important consequences and danger on the patient health status in the case of erroneous diagnostic indications provided). Thus, the whole life-cycle of the DL-based software system for dermoscopic analysis and classification (also including the development, the clinical validation and post-market surveillance) should address the stringent requirements of extensive normative (such as the Standard EN 62304:2006 about the software design, the ISO Standard 14971:2019 for Medical Device Risk management, the ISO Standard 13485:2016 about the Medical Device Quality Management systems, the MDCG 2019-9 Guide

for safety and clinical performance concerning the clinical investigation) and evaluated by the Certification Body during the CE mark certification process. The normative framework seems to limit the actual possibility by Small Medium Enterprises to introduce smartphone applications, whereas the corresponding market may be more easily approached by large companies already qualified as Medical Device Manufacturers for other SW systems and/or equipment.

On the basis of the legislative framework, according to the authors'opinion, the future research efforts should be better focused on the adoption of the DL-based software system only by dermatologist, thus matching also the following deontological features involved with the diagnosis of skin cancers:

i. promotion of periodic visiting by the specialist whose attention may be captured by skin lesions that do not appear as suspicious for the unexpert patient and will not be ever examined through smartphone application;

ii. improvements of psychological behavior against the pathology by the patient affected by melanoma that may be addressed on the correct diagnostic and successfully therapeutic pattern rather than be abruptly informed by an app on the high oncological risk of the self-examined lesion.

According to the presented perspective, the main research topic should be the development of DL-based systems able to improve the diagnostic expertise of the dermatologist (not only to provide support and second opinion for the examination of the single suspicious nevus). For the user the software system should not appear as a black-box; rather, the classification results should be easily related to well-knowledge diagnostic methods (such as ABCDE rules, 7-Point Check List, and Menzie'score). As an example, the approach of the Semantic Segmentation [223] based on DL (already successfully experimented in other applications such as the real-time segmentation of road traffic video for the autonomous driving) could be investigated to provide an automatic system able to recognize the atypical features within the dermoscopic images of suspicious lesions. Moreover, the metrics themselves adopted to analyze the performance of the proposed software systems should be revised for better show the efficacy in the clinical setting end the new intended aims. In detail, the differentiation among suspicious lesions to be excided and other types of classified nevi should be emphasized when the ROC curve is analyzed for the optimal tuning of DL-software systems.Finally, the economic impact supported by the clinical organizations in terms of the savings for the number of excisions as well as the costs associated with the erroneous diagnosis should be taken into account during the performance evaluation of the developed or systems.

VI. CONCLUSION

New techniques in machine learning, also known as "deep learning (DL)," were introduced around 2010. However, if we consider the full future potential of automating repetitive tasks; optimizing time-consuming tasks; augmenting

TABLE 6. Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[57]	2021	China	C	Private	–	yes	EfficientNet-b4	–	–	ACC, SE, SP	Pytorch	–	Frontiers in Medicine	5.091	Comparison Between CNN Model and Dermatologists
[56]	2021	China	C	ISIC2019	REMOVING HAIR:horizontal and vertical flipping, resizing, and rotation	–	GANs and DenseNet201	combining focal loss with weighted cross-entropy and A-SoftMax loss	Adam	AUC, ACC, APSE, SP, PPV, NPV, BMA, DC	Pytorch	GTX2080Ti	IEEE Access	3.367	GAN/Imbalance sample
[162]	2021	Canada	C	Pravite	right-left flipping, top-bottom flipping and random rotation	yes	ResNet-101, SVM, KNN, RF	cross-entropy	Adam	Acc, Sen, Spe	Tensorflow	GTX1080Ti	OPTICS AND LASER TECHNOLOGY	3.867	Use polarization speckle
[163]	2021	Canada	C	Atlas	flipping, rotating, and zooming in	yes	ResNet-50	cross-entropy	Adam	Acc, Sen, Spe, ROC	Keras	GTX1080Ti	COMPUTERS IN BIOLOGY AND MEDICINE	4.589	7-point checklist
[164]	2021	Singapore	C and S	ISIB2016 and ISIB2017	random flipping and cropping, random scaling	yes	ResNet-101	cross-entropy	SGD	Acc, Sen, Spe, AUC	–	–	PATTERN RECOGNITION	7.74	effective diagnosis guided feature fusion,recursive mutual learning
[25]	2021	Egypt	C	ISIC 2020	translations, rotations, cropping, flipping, scaling, adding noise	yes	VGG19, GoogleNet, and ResNet50, SqueezeNet	cross-entropy	Adam	Acc, Spe, Sen, F-score, AUC	MATLAB 2020	–	COMPUTERS IN BIOLOGY AND MEDICINE	4.589	hybrid CNN, random over-sampling (ROS) to solve imbalance
[180]	2021	Pakistan	C	ISIB2016 and ISIC2017	rotated, and flipped horizontally	–	Faster R-CNN and SVM	–	–	Sen, Spe, Acc, AUC	Tensorflow	Nvidia GTX1070	MULTIMEDIA TOOLS AND APPLICATIONS	2.757	Faster R-CNN
[105]	2021	France	C	ISIC 2018	horizontal flipping, vertical flipping, rotation, width, and height shift	–	ResNet50, VGG16, VGG19, DDAG Theory	cross-entropy	Adam	BACC, Sen, AUC	MATLAB 2020	Nvidia GeForce Rtx 2080	SENSORS	3.576	Ensemble Method
[202]	2021	China	C	ISIC2017 and ISIC2018	–	–	Inception-V4, ResNet and DenseNet, R-CNN mask	cross-entropy	–	Acc, Sen, Spe, F1-score, Pre, AUC	–	–	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE	5.428	Mask R-CNN, K-fold cross-validation
[165]	2021	Tunisia	C	ISIC2018	random flips and rotations	Yes	ResNet50	–	SGD, Adam, RMSProp	Pre, Recall, F1-score, mAP	–	NVIDIA Tesla K80	COMPUTERS IN BIOLOGY AND MEDICINE	4.589	Query-driven distance, transfer learning
[citeattallah2020...]	2021	Egypt	C	HAM10000	shearing, rotation, and top and bottom hat filtering	–	ResNet-50, DenseNet-201, and DarkNet-53, DWT, LBP	–	–	Acc, Sen, Spe, Pre, F1-score	–	–	CONTRAST MEDIA & MOLECULAR IMAGING	3.161	High-Level Radiomics Features and Low-Level Features
[166]	2021	UK	C	ISIC 2017	vertical and horizontal flipping, random translation and scaling	Yes	19 CNN models	cross-entropy	SGDM	accuracy, classification error, precision, sensitivity, specificity, F1-score, false-positive rate, false-negative rate, MCC	–	–	CANCERS	6.639	19 CNN models
[212]	2021	Algeria	S and C	ISIC 2016, ISIC 2017 and ISIC 2018, PH2	–	–	W-net, Inception Resnet	–	–	accuracy, sensitivity, specificity, dice coefficient and precision	Keras	Nvidia GeForce 840 M	APPLIED INTELLIGENCE	5.086	W-Net
[167]	2021	Pakistan	C	HAM10000, ISIB2018, and ISIC2019	–	Yes	DenseNet201, MobileNetV2, two-stream	–	a skewness-controlled moth-flame(ScMFO)	Sen, Prec, F1-score, Acc, time	MATLAB 2020b	–	INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS	8.709	two-stream, 10-fold cross-validation
[63]	2021	India	C	ISIC Archive	–	–	Cascaded CNN and SOM classifiers	–	–	precision, accuracy, recall and F1 score	MATLAB	–	Journal of Ambient Intelligence and Humanized Computing	7.104	–
[33]	2021	Italy	C	iDScore dataset	vertical and horizontal, flip and rotation	yes	ResNet50	–	–	SE, SP, ROC, AUC	Tensorflow	–	Journal of Dermatological Science	4.563	–
[151]	2021	US	C	33,980	rotation, scaling, and translations	yes	VGG16 and Xception	cross-entropy	Adam	SE, SP, ACC, AUC	Tensorflow and Keras	2 NVIDIA Tesla K80 GPUs	Science Translational Medicine	17.956	Using a multilesion saliency ranking or score
[152]	2021	Saudi Arabia	C	ISIC	–	No	Inception v4, ResNet-50, VGG 19	–	–	SE, SP, ACC	–	–	Journal of ambient intelligence and humanized computing	7.104	Preprocessing; GrabCut;neuro-fuzzy
[122]	2021	Turkey	C	HAM10000	rotation, shifting, reflecting,scaling and flipping	No	CNN	–	Adam	ACC, precision, recall, F1 score	–	–	Neural Computing and Applications	Springer Journal	Web application
[46]	2021	China	S	ISIC2017	flipped and rotated	no	DensNet121; two attention modules	the sum of dice, loss and binary cross entropy loss	Adam	Dice, JI, ACC, SE	Keras	NVIDIA Tesla V100	Journal of Ambient Intelligence and Humanized Computing	7.104	Spatial attention module
[47]	2021	Finland	C	Private(Finland)	mirrored and flipped, horizontally and vertically	–	CNN	–	–	SE, SP, PPV, ACC	Keras	2 Nvidia Tesla V100-SXM2	Acta dermato-venereologica	4.437	hyperspectral imaging:using the majority of the pixels to predict the class of the whole lesion

TABLE 6. (Continued.) Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[48]	2021	China	S	ISIC2018	rotation, flipping, translation, color shifting, and scaling operations	yes	DenseNet; ResNet; U-Net	cross-entropy loss	Adam	DI, JI, ACC, SE, SP	Keras	NVIDIA TESLA K40	Cognitive Computation	5.418	atrous spatial pyramid pooling
[26]	2021	Spain	C	PH2, MED-NODE, HAM10000, ISBI2016, ISIC2017, MSK-2, MSK-3 and MSK-4 and UDA-2	flips, rotations, zooms, and shifts	yes	DenseNet201, InceptionV3, NasNetMobile,DenseNet121, DenseNet169, InceptionResNetV2, InceptionV4, NasNetMobile, ResNet50, Xception, VGG16, and VGG19	–	SGD, ADAM and RMSprop	MCC	Keras	2 NVIDIA Geforce GTX 1080-Ti	Medical Image Analysis	3.11	–
[49]	2021	Taiwan	C	ISIC Archive	cropped,vertical and horizontal flips, random rotation, zoom	–	CNN and a combination of CNN with ANN	–	–	ACC, SE, SP,F1 score, AUC	Keras	–	Journal of Multidisciplinary Healthcare	2.404	Patient's Metadata
[123]	2021	Germany	S	HAM10000 and ISIC2018, ISIC2017, PH2,PROP	cropped	–	U-Net; ResNet50	–	–	JI, AUROC, ACC,SE, SP	–	–	Journal of Medical Internet Research	5.428	–
[116]	2021	Germany	C	ISIC archive, HAM10000, PH2, SKIN1.2, BCN20000, and PROP	flip, rotation, zoom	–	ResNet50, DenseNet121, VGG16	–	–	AUROC	torch	NVIDIA GeForce RTX 2080 Ti	European Journal of Cancer	9.162	–
[153]	2021	India	C	ISIC2017, PH2	–	–	DWT+GA and DLNN	Cross-entropy	SGD	ACC, SE, SP	–	–	Neural Computing and Applications	Springer Journal	–
[27]	2021	South Korea	C	ISIC 2020	–	–	CycleGAN; U-net+L2;ResNet-18; ResNet-152; ResNeXt-101; EfficientNet-B3; EfficientNet-B4	cross-entropy	–	accuracy, AUC, F1 score, precision, recall and specificity	–	–	IEEE Access	3.367	hair removal
[64]	2021	Pakistan	S and C	ISBI2016 and ISIC2017, HAM10000	yes	yes	VGG16; AlexNet; ResNet50	–	–	sensitivity (Sen), precision (Pre), accuracy (Acc), F1-Score (F1-S); and mean over coefficient (MOC)	MATLAB	–	Pattern Recognition Letters	3.756	–
[154]	2021	China	C	ISIC2017; ISIC2018; ISIC2019	flip, translation, rotation	–	CSLNet	cross entropy	Adamax	precision, sensitivity, specificity, accuracy, F1 score, Jaccard similarity coefficient (JSC), geometric mean (G-mean), Matthews correlation coefficient (MCC), Cohen's kappa score (CKS), AUROC, precision-recall curve (PR-AUC), and time.	Keras	NVIDIA GeForce GTX 1080	Computerized Medical Imaging and Graphics	4.79	–
[102]	2021	Spain	C	ISIC2019	shift, rotation, and reflection	yes	ResNet-50	cross-entropy	SGD	ROC, AUC	Keras	Titan Xp	Frontiers in Medicine	5.091	–
[224]	2021	China	C	ISIC2017; HAM10000	rotating, flipping, and crop	yes	ResNet-50	cross-entropy	–	accuracy, sensitivity, specificity, and AUC	–	–	IEEE Access	3.367	TRANSFER LEARNING
[124]	2021	Egypt	C	HAM10000	geometric, intensity geometric transformations	yes	GAN, ResNet-18	–	SGD	ACC, Macro Precision (MP), Sensitivity (Macro Recall (MR), Macro F1-score(MF), and the area under the ROC curve (AUC).	–	–	Expert Systems With Applications	6.954	Self-attention, GAN, 3-fold cross-validation
[158]	2020	Canada	C	ISIC2016	horizontal and flip, vertical Gaussian noise, brightness and zoom, horizontal and vertical shift, sampling noise once per pixel, color space conversion, and rotation.	yes	VGG-GAP; CycleGAN	focal loss	Adadelta	SE, ROC	Keras	4x NVIDIA P100 Pascal	Physics in Medicine & Biology	3.609	reach an expert dermatologist level; GAN
[34]	2020	US	C	private	random flipping, rotating, cropping and color perturbation	yes	Inception-v4	softmax cross-entropy loss	SGD	accuracy, sensitivity, average overlap (AO), PPV, top-k accuracy	TensorFlow	–	Nature medicine	53.44	Meta data
[58]	2020	China	C	ISIC2018	–	yes	SVM, SNN,Resnet152	–	–	ACC, SE, SP, PRE,AUC	Pytorch	RTX 2080TI	IEEE Access	3.367	SNN; four-fold cross validation

TABLE 6. (Continued.) Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[59]	2020	China	C	ISIC2018	–	–	SNN,R-STDP, CNN	DQNimb; focal loss; mean squared false error (MSFE); cross-entropy	–	SE, SP, G-mean, F-score	Pytorch	–	IEEE Access	3.367	SNN
[65]	2020	China	C	Dermquest and DermIS	–	–	CNN	–	WOA	SP, SE, ACC, NPV,PPV	Matlab	NVIDIA GeForce GTX Titan X	Artificial intelligence in medicine	5.326	–
[66]	2020	China	C	ISIC2016, ISIC2017	Color-based normalization and augmentation; Geometrical transformation(rotation, flipping, translation, and cropping)	yes	VGGNet, ResNet, Inception v3,SVM, SCDA	–	–	mean average precision (mAP), accuracy (Acc), area under receive operation curve (AUC), sensitivity (Sen) and specificity (Spec)	MatConvNet	NVIDIA TI-TAN XP	Applied Soft Computing	6.725	–
[155]	2020	Australia	C	Private	flipping and cropping	–	ResNet-50, Spatial Transformer Network	SegLoss	–	SE, SP, AUC, ACC	–	2 NVIDIA GTX 1080 GPUs	IEEE transactions on medical imaging	10.048	Paired images
[67]	2020	Pakistan	S and C	ISIC2016, ISIC2017, ISIC2018,PH2,HAM10000	–	yes	ResNet 101, DenseNet 201	–	–	ACC, time, SE,FNR	MATLAB	16-GB graphics card	Diagnostics	3.706	10-fold cross-validation
[156]	2020	Germany	–	Private	–	–	CNN	–	–	sensitivities, specificities, accuracy, and ROC	–	–	European Journal of Cancer	9.162	DATA PREPROCESSING
[28]	2020	China	S and C	ISBI 2016	rotation, mirroring, center cropping, brightness change, random occlusion	yes	MobileNet; DenseNet-121; U-Net, SqueezNet; ShuffleNet	focal loss, cross-entropy; BCE Dice Loss,	Adam	accuracy (AC), jaccard index (JA), dice coef cient (DC), speci city (SP), and sensitivity (SE),AUC	Keras	NVIDIA Tesla K80 GPU (12G)	IEEE Access	3.367	–
[35]	2020	China	C	Private	–	yes	Inception v3	Cross entropy mean	SGD	accuracy, specificity, sensitivity, and area under the receiver operating characteristic (ROC) curve (AUC); Kappa coefficients	Tensorflow	–	Chinese Medical Journal	2.628	–
[60]	2021	Hong Kong	S	ISIC2017 and ISIC2018	vertical and horizontal flipping, rotations, random cropping, scaling, sharpening, adjustments of color distributions, and adding noises (Gaussian, salt or pepper noises).	–	adaptive dual attention module: dual encoder	dice loss and bce loss	–	JSI, DSC, SE, SP, ACC, TJI	Pytorch	Nvidia GeForce GTX 1080Ti	IEEE transactions on medical imaging	11.048	multi-scale resolution fusion: use new network
[112]	2020	Brazil	C	ISBI 2016	Geometric, Color transformations, Data augmentation by image processing techniques; Data augmentation based on specialists knowledge	yes	GoogLeNet	–	–	average precision, accuracy, sensitivity and specificity	–	Titan X	Pattern Recognition Letters	3.756	Transfer learning; Imbalanced dataset; Data augmentation; 5-fold cross validation
[125]	2020	Spain	C	HAM10000	Horizontal and vertical flipping, rotations.	–	DenseNet201; GoogLeNet; Inception-ResNetV2; InceptionV3; MobileNetV2	–	SGD	Accuracy, Precision, Recall and F-measure	–	2 Titan X	Neural Processing Letters	2.908	–
[182]	2020	China	C	ISIC 2016 and ISIC 2017	vertical and horizontal flips	–	GP-CNN	–	SGD	Accuracy (Acc), sensitivity (Sen), specificity (Spe), and area under the receiver operating characteristic curve (Auc)	Keras	RTX 2080Ti	IEEE journal of biomedical and health informatics	5.772	Fusion Strategy; a) Averaging predictions. (b) SVM stacking.(c) Weighted ensemble of predictions
[68]	2020	UK	S and C	Dermofit, PH2, ISIC 2017	flipped and rotated	yes	CNN, HLPSO	HLPSO	–	Jaccard score, ACC	Matlab	–	Knowledge-Based Systems	8.038	–
[36]	2020	China	S and C	ISBI 2016, ISIC 2017	flipping images horizontally and vertically, affine transformation, multiplying each image with a random value between 0.8 and 1.5, and Gaussian blurring	yes	Feature Pyramid Network, Region Proposal Network, Convolution Subnets, ResNet50, ResNet101	Focal Loss, cross entropy, Convolution Subnets Loss, the sum of the RPN loss	SGD	average precision (AP), area under the ROC curve (AUC), sensitivity (SE) and specificity (SP), Jaccard index (JA), Dice coefficient (DI), pixel accuracy (AC)	Tensorflow	–	IEEE journal of biomedical and health informatics	5.772	Data Preparation

TABLE 6. *(Continued.)* Review papers with impact factor greater than 2 (S = segmentation; C = classification).

TABLE 6 continue

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[85]	2020	Turkey	C	ISIC 2017	rotated, crops	yes	ResNet-18 and ResNet-50	–	–	AUC, ACC,SE, SP	Caffe	NVIDIA GeForce GTX 1080Ti	IET Image Processing	2.373	Ensemble of the DL-based methods
[168]	2020	Brazil	C	ISIC2016, PH2	rotation, horizontal and vertical rotation, gamma, logarithmic adjustment, and contrast	yes	VGG, Inception, ResNet, Inception-ResNet, Xception, MobileNet, DenseNet, NASNet; five classifiers: Bayes, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF).	–	–	Acc, F1-Score, Recall, Precision	TensorFlow and Keras	–	Pattern Recognition Letters	3.756	–
[169]	2020	China	C	ISIC2018	GAN-based data augmentation	yes	ResNet50	–	–	accuracy, sensitivity, specificity, average precision and balanced multiclass accuracy(for classification), IS, FID, Precision and recall (for GAN)	–	NVIDIA Quadro P4000	Computer Methods and Programs in Biomedicine	5.428	Based Gan, data augmentation
[98]	2020	Taiwan	C	ISIC2019	shear, rotate, contrast, equalize, posterize, contrast, sharpness, and cut out, translate, auto-invert, solarize, color, brightness, and scaled, color contrast	–	EfficientNet, DPI model	–	Adam, Bayesian optimization	AUC, Acc, SE, F1-Score	–	Tesla V100	IEEE Access	3.367	–
[109]	2020	Italy	S	ISIC2017	rotated, flipped, shifted, and scaled, color contrast	–	U-Net	Cross Entropy	Adam	Acc, Jaccard Index	–	–	Multimedia Tools and Applications	2.757	GAN
[126]	2020	Vietnam	C	HAM10000, ISIC 2018,ISIC 2019	Rotation, flip, shift, zoom, shear	yes	DenseNet169,EfficientNet(CLF)	Hloss(CLF)	Adam	Acc, mRecall, Recall, Precision, Sidev	Keras	–	IEEE Access	3.367	resolving imbalanced datasets have been applied, such as 1) Data level methods, 2) Algorithm level methods, and 3) Hybrid methods.
[69]	2020	Mexico	C	ISIC2019	–	–	Densenet-201; Esemble KNN, SVM-L and SVM with a linear plurality vote	–	–	Acc, Precision, SE, SP	MatLab	–	Applied Sciences	2.679	Ensemble Classifier
[225]	2020	US	–	ISIC2017	–	–	–	–	–	ROC, SE, SP	–	–	Journal of the American Academy of Dermatology	11.527	Accumulating evidence suggests that deep neural networks can classify skin images of melanoma and its benign mimickers with high accuracy and potentially improve human performance.
[52]	2020	Austria	S	ISIC 2017,HAM10000	horizontal and vertical flipping, rotations	yes	EfficientNet,LinkNet-152,U-Net, ResNet34,	combination of binary cross-entropy and Jaccard loss; combination of dice loss and focal loss	Adam	AUC	Keras, Pytorch, MATLAB, MedCalc	TITIAN V NVIDIA	Computer Methods and Programs in Biomedicine	5.428	The impact of segmentation on classification
[127]	2020	Spain	C	HAM10000	zoom, rotated, horizontal and vertical flipping, transposing, shear and brightness	yes	VGG16 and VGG19, ResNet34, ResNet50, ResNet101, SEResNet50, EcientNetB5, MobileNet	–	Adam	ACC	–	4 GeForce GTX 1080	Diagnostics	3.706	–
[60]	2020	China	S	ISBI2017 and ISIC2018	vertical and horizontal flipping, scaling, rotations, random cropping, sharpening, adjustments of color distributions, and adding noises	–	dual attention module, ResNet, U-Net	dice loss, bce loss	Adam	Jaccard similarity index (JSI), Dice similarity coefficient (DSC), sensitivity (SE), specificity(SP), accuracy (ACC), and threshold Jaccard index (TJI)	Pytorch	Nvidia GeForce GTX 1080Ti	IEEE Transactions on Medical Imaging	10.048	–
[159]	2020	Canada	C	ISIC 2017	random resize cropping, random horizontal and vertical flipping, random rotation and normalization	yes	ResNet and DenseNet	binary focal loss and dice loss	Adam optimization	Accuracy (ACC) and Area Under Curve (AUC)	Keras	Nvidia GeForce GTX 1080Ti	Computerized Medical Imaging and Graphics	4.79	5-fold cross-validation

TABLE 6. *(Continued.)* Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[170]	2020	Poland	C	ISIC	rotation, width and height shift, horizontal and vertical flip, and zooming	yes	VGG networks with NAS algorithm	binary cross-entropy loss	SGD	ACC, AUC	-	-	IEEE Access	3.367	5-fold cross-validation
[61]	2020	Poland	C	ISIC 2017	Rotation and Jigsaw	yes	ResNet50	binary cross-entropy loss	SGD	ROC AUC, ACC, PR AUC	Pytorch	-	Electronics	2.397	self-supervised learning
[171]	2020	Pakistan	C	ISBI2016 and ISIC2017	-	yes	DenseNet201 with faster rcnn	FRCNN(loss)	-	sensitivity (Sen), precision (Prec), F1-Score (F1-Score), AUC, FPR, accuracy, false negative rate (FNR), and time	-	-	Pattern Recognition Letters	3.756	-
[37]	2020	India	C	ISIC	geometric transformations	yes	Inception V3, VGG19, ResNet50 and SqueezeNet	-	-	precision, recall, F1-score, and accuracy	Tensorflow	-	Transactions on Emerging Telecommunications Technologies	2.638	deep learning internet of health and things (IoHT)
[86]	2020	Turkey	S	ISIC 2017	-	yes	FCN-AlexNet, FCN-8s, FCN-16s, and FCN-32s	-	-	Acc, dice, Jaccard	Caffe	NVIDIA GTX Titan X	Expert Systems with Applications	6.954	FCN
[70]	2020	Egypt	C	ISIC2019	vertical horizontal-shift, vertical and horizontal flip, and rotation	yes	GoogleNet	-	-	accuracy, sensitivity, specificity, precision and f1 score	Matlab	GeForce MX150 NVIDIA graphics card	IEEE Access	3.367	Transfer Learning
[183]	2020	Japan	C	private	horizontal flip, distort, random rotations, and zoom cropping	-	FRCNN, VGG-16	-	SGD	accuracy, sensitivity, specificity, false negative, false positive, positive predictive value	Chainer	-	Biomolecules	4.879	-
[226]	2020	India	S and C	ISBI 2016 and ISIC 2017	-	-	FCNs based on VGG-16 and GoogLeNet	-	SGD	sensitivity (SE), specificity (SP), accuracy (AC), Jaccard index (JA), and dice coefficient (DI)	-	NVIDIA TI-TAN X	International Journal of Imaging Systems and Technology	2	-
[128]	2020	Germany	C	HAM10000 and ISIC archive	-	yes	ResNet-50	-	-	Accuracy, sensitivity and specificity, ROC	Fastai	-	Frontiers in Medicine	5.091	4-fold cross validation
[172]	2020	Spain	S	ISIC 2017 and PH2	geometric augmentations such as rotation, zooming, shifting, and flipping	yes	U-Net, FCN8,DSNet	the sum of binary cross-entropy and IoU	-	mIoU, mSn, mSp, AUC	Keras	GeForce GTX 1080	Computers in Biology and Medicine	4.589	-
[71]	2020	UK	C	ISIC, PH2, Der-mIS, Der-mQuest, DermNZ	-	yes	AlexNet	-	-	accuracy, precision, sensitivity, and specificity	Matlab	-	Expert Systems with Applications	6.954	traditional machine learning approach and DL(3 levels)
[149]	2020	Germany	C	MSK-1, ISIC 2018	-	-	Inception v4	-	-	sensitivity, specificity, accuracy	-	-	Annals of Oncology	32.976	comparison with dermatologists
[62]	2020	Australia	C	MoleMap and HAM10000	rotation, cropping and random mirroring, cycleGAN	yes	cycleGAN	Cycle consistency loss	-	accuracy (AC), sensitivity (SEN), specificity, (SPC), and AUC (Area under ROC curve) score	Pytorch	NVIDIA TI-TAN X	IEEE journal of biomedical and health informatics	5.772	cycle-GAN
[38]	2020	UK	S	ISIC 2017 and PH2	No	yes	Xception-65, Mask R-CNN and DeeplabV3+	cross-entropy	momentum	Dice Similarity Coefficient (Dice), Sensitivity, Specificity, Accuracy and Matthew Correlation Coefficient (MCC)	Tensorflow	NVIDIA TI-TAN X	IEEE Access	3.367	ensemble methods
[227]	2020	Germany	C	Private	-	-	CNN	-	-	sensitivity, specificity and DOR, CI	-	-	Journal of the European Academy of Dermatology and Venereology	6.166	comparison with dermatologists
[72]	2020	Morocco	C	PH2 and ISIC 2017	-	yes	AlexNet, VggNet, GoogLeNet, and ResNet	-	-	Recall, Specificity, Precision, Accuracy, F-Measure and Kappa index	Matlab	NVIDIA GeForce GTX 1660 Ti	Multimedia Tools and Applications	2.757	Fusion of handcrafted and pre-train features
[73]	2020	Romania	C	PH2 and ISIC 2019	-	yes	GoogleNet, ResNet-101, and NasNet-Large	-	SGD	accuracy, specificity, sensitivity, and Dice coefficient	Matlab	-	Sensors	3.576	-
[74]	2020	India	-	PH2	-	-	RNN	-	-	accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1-score, and MCC	MATLAB 2018	-	International Journal of Imaging Systems and Technology	2	RNN
[129]	2020	India	C	HAM10000	-	yes	InceptionV3, ResNeXt101, InceptionResNeTV2, Xception, NASNetLarge	categorical cross-entropy	SGD and Adam	Accuracy, recall, precision and F1-score	Keras	-	Multimedia Tools and Applications	2.757	four ensemble models

TABLE 6. *(Continued.)* Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[228]	2020	India	S and C	PH2 ISIC 2017 and ISIC 2019	adjusting the contrast	–	Yolo v3	–	–	sensitivity (Sen), specificity (Spe), the dice coefficient (Dice), the Jaccard index (Jac) and accuracy(Acc)	–	–	Diagnostics	3.706	YOLO
[87]	2020	Morocco	C	HAM10000	rotation, translation and reflection	yes	VGG-16, ResNet-18 and DenseNet-121	cross-entropy	Adam	Accuracy, sensitivity, specificity, precision, F1-score and ROC-AUC	caffe	NVIDIA GeForce GTX 1060	Multimedia Tools and Applications	2.757	Feature fusion
[115]	2020	Germany	C	DermNet and ISIC Archive	randomly flipped horizontally, cropped	–	ResNet-152, DenseNet-161, SE-ResNeXt-101, and NASNet	–	-	Precision, Sensitivity, Specificity, and F1-Score	–	–	Applied Sciences	2.679	5 cross validation: ensemble, used average of individual predictions of four best performing CNNs to output final prediction. Transfer Learning
[75]	2020	Pakistan	C	DermIS and DermQuest	cropping, mirroring, flipping, and rotation	yes	AlexNet	–	–	Accuracy	MATLAB 2018b	6GB NVIDIA GPU	IEEE Access	3.367	Transfer Learning
[42]	2020	South Korea	S and C	ISIC 2016, 2017, and 2018	rotations, horizontal and vertical flipping	yes	full resolution convolutional network (FrCN); Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201	–	SGD	sensitivity (recall or true positive rate), specificity (the true negative rate), F1-score (also known as Dice index), and overall accuracy, ROC	Keras and Tensorflow	NVIDIA GeForce GTX 1080	Computer methods and programs in biomedicine	5.428	5 fold cross validation
[29]	2020	Mexico	C	ISIC 2018	–	–	VGG19, VGG16, ResNET-50, Inception v3, Mobilenet v1, Mobilenet v2, DenseNET-201, Xception	–	–	accuracy, sensitivity, specificity, precision, f-score, and the Matthews correlation coefficient	Keras	NVIDIA GeForce 1080Ti	Entropy	2.524	Fusion of handcrafted and deep learning features
[106]	2020	UK	C	ISIC 2018	rotating, and flipping horizontally and vertically	–	FUZZY MULTILAYER PERCEPTRON; U-Net	–	–	accuracy	–	Tesla P100 GPU	Frontiers in medicine	5.091	–
[141]	2020	US	S and C	MED-NODE	reflection, rotation, noise, brightness, and contrast	yes	Inception V3	–	–	TPR, TNR, PPV, NPV, ACC, F1-Score, MCC	–	–	IEEE Access	3.367	10-fold cross validation; ensembles handcrafted and deep learning features Transfer learning
[173]	2020	Pakistan	C	PH2, ISIC, MSK, ISIC UDA, and ISIC-2017	–	yes	DenseNet 201, Inception, ResNet-v2, and Inception-V3	–	–	OA(Overall accuracy), Recall, Precision FNR FPR, AUC, Time	–	–	Human-centric Computing and Information Sciences	5.9	Transfer learning
[39]	2020	China	C	Private	random rotation and distortion	yes	ResNet152 and InceptionResNet-V2	triplet loss	SGD	accuracy, sensitivity (recall) and specificity, ROC	Tensorflow	NVIDIA GTX 980 Ti	IEEE Access	3.367	–
[130]	2020	South Africa	S and C	HAM10000, ISBI 2018, ISIC 2017 and PH2	Rescaling, Shear range, Zoom, Horizontal flip. Rotation	–	encoder-decoder FCN, CRF, DenseNet, ResNet152, Inception V3, VGG19	–	Adam, SGD, RMSprop	DSC, SE,SP, ACC, PRE, F1-score	–	–	IEEE Access	3.367	–
[184]	2019	China	C	ISIC 2017 and PH2	random flipping, rotation, whitening, and two extra generated images	–	Encoder-Decoder Network	dice loss	–	DSC, SE,SP, ACC	Keras	NVIDIA Tesla K40c	IEEE Access	3.367	–
[174]	2019	China	C	ISIC 2017, ISIC Archive	random rotation, zoom, horizontal and vertical flips	yes	ResNet14, ARLCNN14, ResNet50 and ARL-CNN50; RAN14, SENET14, RAN50 AND SENET50	–	SGD	accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC)	–	NVIDIA GTX Titan XP	IEEE transactions on medical imaging	10.048	Attention Mechanism
[40]	2019	China	S	ISBI 2017 and PH2	random flipping, rotation, whitening, and two extra generated images	yes	CRF, U-Net, multi-scale	binary cross-entropy and dice loss	Adam	accuracy (AC), sensitivity (SE), jaccard index (JA), and dice coef cient (DI)	Tensorflow	NVIDIA GTX 1080	IEEE Access	3.367	multi-scale network, averaged results
[175]	2019	Germany	C	Private	–	yes	Inception-v4	–	–	Sensitivity, specificity, and ROC AUC	–	–	JAMA dermatology	10.282	–
[229]	2019	Taiwan	C	Private	–	–	CNN	–	–	Sensitivity, specificity, and area under the receiver operating characteristic (AU-ROC) curve	Keras	–	JAMA dermatology	10.282	5-fold cross validation
[88]	2019	Georgia	C	ISIC 2017	flips, rotations, and crops	–	Inception V2	–	–	ROC AUC	Caffe	–	EBioMedicine	8.143	visual and audio

TABLE 6. *(Continued.)* Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Libary	Graphics card	Journal	Impact factor 2020	Highlight
[176]	2019	Turkey	S	PH2 and ISBI 2017	No	yes	YOLO and GrabCut	–	–	accuracy, specificity, Dice coefficient, and Jaccard index	–	NVIDIA GTX 1080Ti	Diagnostics	4.879	–
[177]	2019	Turkey	C	ISIC 2017	rotated	yes	Gabor wavelet-based CNN, AlexNet, and ResNet-18	–	–	AUC, ACC, SE, SP	–	–	Computers in biology and medicine	4.589	fusion
[230]	2019	Australia	C	Private	–	–	Deep Ensemble model	–	–	AUROC, specificity, predictive values, false-positive rates, and false-negative rates	–	–	JAMA network open	8.483	10-fold cross-validation
[113]	2019	Pakistan	S and C	ISIC 2016	–	yes	RCNN, Fuzzy C-mean (FCM) clustering	RMSE	–	Di, Jc, ACC, SE, SP, ROC	–	Titan X Pascal	International journal of medical informatics	4.046	RCNN
[178]	2019	US	C	HAM10000	flipped, shifted, distorted, rotated	yes	VGG-16, ResNet-50, and DenseNet-169	weighted cross-entropy	Adam	Uncertainty Accuracy (UA)	TensorFlow and Keras	NVIDIA GeForce GTX 1080	Journal of clinical medicine	4.241	Bayesian deep learning; 5-fold cross-validation
[77]	2019	Norway	C	ISBI 2016	–	yes	AlexNet,VGG, linear discriminant analysis (LDA)	–	–	ACC, AP, AUC, TPR, SPC	MatConvNet	–	Multimedia Tools and Applications	2.757	–
[30]	2019	Austria	C	ISIC 2016 and ISIC 2017	rotation and horizontal flipping	yes	AlexNet, VGGNet, ResNet-18 and ResNet-101	cross-entropy	SGDM, RMSProp and Adam	AUC	MatLab	NVIDIA GTX 1070	Computerized Medical Imaging and Graphics	4.79	Ensembling deep features
[43]	2019	Canada	C	Atlas	flips, rotations, zooms, and height and width shifts	yes	Inception V3	multi-task loss	–	accuracy, sensitivity, specificity, precision and AUROC	Keras and TensorFlow	NVIDIA Titan X	IEEE journal of biomedical and health informatics	5.772	7-point checklist
[24]	2019	US	C	ISIC 2018	horizontal and vertical flips, random contrast and random brightness	yes	Xception, AlexNet, VGGNet, ResNet	–	–	TPR, TNR, PPV, ACC, F-measure	Keras	NVIDIA GTX 1080 TI	Tissue and Cell	2.466	–
[121]	2019	US	C	Private one and HAM10000	–	yes	ResNet-50	cross-entropy	–	AUC	–	–	IEEE journal of biomedical and health informatics	5.772	Fusion
[221]	2019	Israel	C	ISIC 2017	–	yes	Inception V2	–	–	AUC	–	–	EBioMedicine	8.143	sonified, A dermoscopy image is acquired by a smartphone and conveyed to cloud computing
[117]	2019	Germany	C	ISIC archive	–	yes	ResNet50 CNN	–	–	ROC, sensitivity and specificity, confidence intervals (CI)	–	–	European Journal of Cancer	9.162	–
[118]	2019	Brazil	C	ISIC Archive and Atlas	random horizontal and vertical flips, resized random crops, random rotations, hue	yes	Inception4	–	SGD	ROC Curve (AUC)	–	–	CVPR*	–	–
[231]	2019	US	S and C	ISIC 2017, ISIC 2018	GAN	yes	U-Net, ResNet-50	binary cross-entropy	Adam	ROC AUC	–	–	CVPR*	–	–
[78]	2019	Australia	S	ISBI 2016, ISBI 2017 and PH2	random cropping and flipping	yes	ResNet based FCN, probability based step-wise integration	cross-entropy	SGD	dice similarity coefficient (DSC), Jaccard index (Jac.), sensitivity (Sen.), specificity (Spec.) and accuracy (Acc.)	MatConvNet	Nvidia Maxwell Titan X	Pattern recognition	7.74	–
[79]	2019	Australia	S and C	PH2, ISBI 2017	rotation and shuffling	yes	encoder-decoder, RNN, VGG-16	–	–	Accuracy, Specificity, Sensitivity, Dice, Jaccard Index, AUC	MATLAB2017a	NVIDIA TAN X TI	Computer methods and programs in biomedicine	5.428	recurrent neural networks; using weakly labelled data
[107]	2019	UK	S	ISIC 2018	rotation and shifting horizontally and vertically	–	U-Net	–	adam, sgd, tanh, selu	Dice coefficient and Jaccard Index	–	Tesla P100	Canadian Conference on Artificial Intelligence*	–	Supervised Versus Unsupervised
[119]	2019	Saudi Arabia	C	ISIC archive	–	–	CNN	–	–	AUC-ROC	–	–	IEEE Access	3.367	–
[80]	2019	Saudi Arabia	C	Skin-EDRA, ISIC, DermNet, PH2	–	–	Stack-based auto-encoder; RNN	cross-entropy	SGD	sensitivity(SE), specificity (SP) and area under the receiver operating characteristics (AUC)curve	MATLAB 2016	–	Multimedia Tools and Applications	2.757	fusion
[232]	2018	US	C	Private	–	–	CNN	–	–	Precision, Recall, F-measure	–	–	BMC medical informatics and decision making	2.796	–
[81]	2018	Pakistan	S and C	PH2, ISIC	–	–	A method based on probabilistic distribution and best features selection	entropy-variance	–	Execution time, Sensitivity, Precision, Specificity, FNR, FPR, Accuracy	MATLAB 2017a	–	BMC cancer	4.43	Fusion
[185]	2018	SwitzerlandS	–	ISBI 2017	rotation, flipped horizontally	–	Multi-scale Fully Convolutional DenseNets	–	SGD	Dice coefficient, accuracy	–	–	International Conference Image Analysis and Recognition*	–	–

TABLE 6. *(Continued.)* Review papers with impact factor greater than 2 (S = segmentation; C = classification).

Paper	Year	Country	task	Dataset	Data augmentations	Transfer learning	Applying models	Loss	Optimization algorithms	Metrics	Library	Graphics card	Journal	Impact factor 2020	Highlight
[233]	2018	Canada	C	private	–	–	ResNet-50	–	–	mean average precision (mAP), Top-1 Accuracy (Top-1 Acc) and the area under the ROC curve	–	–	Experimental dermatology	3.96	combines multiple imaging modalities together with patient metadata
[50]	2018	China	C and S	ISIC 2017	rotation, randomly flipped	–	FCRN-88,	–	SGD	accuracy(AC), Jaccard Index (JA), Dice coefficient (DI), sensitivity (SE) and specificity(SP)	MatConvNet, Keras	GeForce GTX TITAN X	Sensors	3.576	Superpixel Extraction
[82]	2018	Hungary	C	ISBI 2017	random cropping samples from the images or horizontal flipping or rotating them with different angles	yes	GoogLeNet, ResNet, AlexNet, VGGNet	–	–	ACC, SE, SP, AUC	MatConvNet	NVIDIA TI-TAN X	Journal of biomedical informatics	6.317	ensemble of deep convolutional neutral networks
[89]	2018	South Korea	C	Asan dataset, MED-NODE, Edinburgh, atlas, Hallym	–	yes	ResNet-152	–	–	the area under the curve (AUC), sensitivity, and specificity	Caffe	–	Journal of Investigative Dermatology	8.551	–
[83]	2018	South Korea	C	–	–	yes	AlexNet	–	–	accuracy, sensitivity, and specificity	Matlab R2016b	NVIDIA GeForce GT 540 M	Multimedia Tools and Applications	2.757	–
[76]	2018	China	C	ISBI 2016	resize and normalization	yes	CNN and feature encoding strategy (FV encoding); ResNet-50 and ResNet-101	hinge loss	stochastic dual coordinate ascent algorithm (SDCA solver)	mean average precision (mAP), accuracy (Acc), area under receive operation curve (AUC)	MatConvNet	NVIDIA Quadro K4000	IEEE Transactions on Biomedical Engineering	4.538	Fusion
[44]	2018	US	S	ISIC Archive	–	yes	VGG-16, InceptionV3	binary cross entropy	SGD, RMSProp	Accuracy, Sensitivity, Precision, AUC	Keras, Matlab,Theano, Tensorflow	–	Journal of digital imaging	4.056	–
[76]	2018	China	C	ISBI 2016	resized, rotated, shifted	yes	ResNet-50, ResNet-101, AlexNet ,VGGNet, Fisher Vector Encoding	hinge loss	stochastic dual coordinate ascent	mean average precision (mAP), accuracy (Acc), area under receive operation curve (AUC)	MatConvNet	NVIDIA Quadro K4000	IEEE Transactions on Biomedical Engineering	4.538	–
[51]	2018	South Korea	S	ISIB 2017 and PH2	HSV, rotating,	yes	full resolution convolutional networks, VGG-16	cross entropy	–	SEN, SPE, DIC, JAC, ACC, MCC	Theano and Keras	NVIDIA GeForce GTX 1080.	Computer methods and programs in biomedicine	5.428	–
[92]	2017	US	S	ISBI 2016 and PH2	flipped horizontally, flipped vertically, rotation, rescaled, shifted; randomly normalizing the contrast of each channel	yes	19-layer deep FCN	a novel loss function based on Jaccard distance	Adam	accuracy (AC), sensitivity (SE), specificity (SP), dice coefficient (DI), and Jaccard index (JA)	Theano and Lasagne	Nvidia GeForce GTX 1060	IEEE transactions on medical imaging	10.048	5-fold cross validation
[84]	2017	Iran	S	–	–	–	CNN	–	SGD	ACC, SE, SP	MATLABand Caffe	GTX Titan X	International journal of computer assisted radiology and surgery	2.924	–
[91]	2017	US	S	ISBI 2017	added the three channels from HSV color space and the lightness channel (L) from CIELAB color space; randomly flipping, shifting, rotating as well as scaling, randomly normalizing the contrast of each channels	–	CDNN	a new loss function based on Jaccard distance	Adam	accuracy (AC), sensitivity (SE), specificity (SP), dice coefficient (DI), and Jaccard index (JA)	Theano and Lasagne	NVIDIA Geforce GTX 1060	IEEE journal of biomedical and health informatics	5.772	5-fold cross validation
[41]	2017	US	C	ISIC Archive, Dermofit, and data from the Stanford Hospital	rotated, cropped, and flipped	yes	Inception v3	–	RMSProp	sensitivity and specificity	TensorFlow	–	nature	49.962	–
[90]	2016	Hong Kong	C and S	ISBI 2016	cropped, rotation, translation and adding random noise	–	fully convolutional residual network; VGG-16, GoogleNet, FCRN-38, FCRN-50, FCRN-101	–	SGD	sensitivity (SE), specificity (SP), accuracy (AC), Jaccard index (JA) and Dice coefficient (DI), average precision (AP), ROC	Caffe	NVIDIA TI-TAN X	IEEE transactions on medical imaging	10.048	–

limited medical resources; improving interobserver reliability issues; and expanding the diagnostic toolbox of physicians, then we can say that AI in dermatological health care is yet in its infancy. Indeed, specific task-driven algorithms are only beginning to be introduced. Compared to the predecessor forms of computing, these new methods are dynamically changing systems that improve with continuous data exposure, and therefore performance is dependent on the quality and generalizability of the training datasets.

Artificial intelligence in dermoscopy is not replacing specialists or placing decision making into the hands of nonexperts. Developments shortly will follow what is already happening in radiology, where AI is proving to be useful for triaging and improving workflow efficiency by helping to prioritize tasks, which is the current direction for the most significant research efforts.

We project that in the next 5 years, clinicians will become increasingly involved in training and testing large-scale validation as well as monitoring narrow AI in clinical trials. At this point, CNNs have shown in very few cases that they make physicians better at diagnosing skin cancer with respect to available real-world clinical data. Only in the future, when large, standardized training datasets and, above all, validation with prospective clinical trials will be completed, will DL truly improve dermatological workflow, for example by providing computer-aided triage (e.g., through scanning which pigmented lesion might need prompt evaluation by a dermatologist) and supporting young professionals in classification tasks.

APPENDIX

A supplementary appendix presents a list of review papers with impact factors over 2, (see table 6, S is segmentation, C is classification).

REFERENCES

[1] American Institute for Cancer Research. *Skin Cancer Statistics-Melanoma of the Skin is the 19th Most Common Cancer Worldwide*. Accessed: Aug. 10, 2020. [Online]. Available: https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics

[2] WHO Cancer Today. *Skin Melanoma Heatmap of the World in 2020*. Accessed: May 11, 2021. [Online]. Available: https://gco.iarc.fr/

[3] American Cancer Society. *Cancer Facts and Figures 2021*. Accessed: Feb. 22, 2021. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf

[4] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA, Cancer J. Clinicians*, vol. 71, no. 1, pp. 7–33, Jan. 2021.

[5] G. A. Holmes, J. M. Vassantachart, B. A. Limone, M. Zumwalt, J. Hirokane, and E. S. Jacob, "Using dermoscopy to identify melanoma and improve diagnostic discrimination," *Federal Practitioner*, vol. 35, no. 4, p. S39, 2018.

[6] O. Yélamos, R. P. Braun, K. Liopyris, Z. J. Wolner, K. Kerl, P. Gerami, and A. A. Marghoob, "Usefulness of dermoscopy to improve the clinical and histopathologic diagnosis of skin cancers," *J. Amer. Acad. Dermatol.*, vol. 80, no. 2, pp. 365–377, Feb. 2019.

[7] P. Carli, E. Quercioli, S. Sestini, M. Stante, L. Ricci, G. Brunasso, and V. D. E. Giorgi, "Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology," *Brit. J. Dermatol.*, vol. 148, no. 5, pp. 981–984, May 2003.

[8] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art," *Artif. Intell. Rev.*, vol. 54, pp. 811–841, Jun. 2020.

[9] M. A. Kassem, K. M. Hosny, R. Damaševičius, and M. M. Eltoukhy, "Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review," *Diagnostics*, vol. 11, no. 8, p. 1390, Jul. 2021.

[10] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. design Implement. (OSDI)*, 2016, pp. 265–283.

[11] F. Chollet, "Keras: Deep learning library for Theano and TensorFlow," *Keras*, vol. 7, no. 8, p. T1, 2015.

[12] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[13] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 689–692.

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.

[15] R. Al-Rfou *et al.*, "Theano: A Python framework for fast computation of mathematical expressions," 2016, *arXiv:1605.02688*.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.

[17] Y. LeCun, "LeNet-5, convolutional neural networks," *Lenet*, vol. 20, no. 5, p. 14, 2015.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[22] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and Li Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[24] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue Cell*, vol. 58, pp. 76–83, Jun. 2019.

[25] G. I. Sayed, M. M. Soliman, and A. E. Hassanien, "A novel melanoma prediction model for imbalanced data using optimized SqueezeNet by bald eagle search optimization," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104712.

[26] E. Pérez, O. Reyes, and S. Ventura, "Convolutional neural networks for the automatic diagnosis of melanoma: An extensive experimental study," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101858.

[27] D. Kim and B.-W. Hong, "Unsupervised feature elimination via generative adversarial networks: Application to hair removal in melanoma classification," *IEEE Access*, vol. 9, pp. 42610–42620, 2021.

[28] L. Wei, K. Ding, and H. Hu, "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," *IEEE Access*, vol. 8, pp. 99633–99647, 2020.

[29] J.-A. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and H. Castillejos-Fernandez, "Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures," *Entropy*, vol. 22, no. 4, p. 484, 2020.

[30] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Med. Imag. Graph.*, vol. 71, pp. 19–29, Jan. 2019.

[31] A. Aldwgeri and N. F. Abubacker, "Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images," in *Proc. Int. Vis. Inform. Conf.* Cham, Switzerland: Springer, 2019, pp. 214–226.

[32] A. Mahbod, G. Schaefer, C. Wang, G. Dorffner, R. Ecker, and I. Ellinger, "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105475.

[33] L. Tognetti *et al.*, "A new deep learning approach integrated with clinical data for the dermoscopic differentiation of early melanomas from atypical nevi," *J. Dermatolog. Sci.*, vol. 101, no. 2, pp. 115–122, Feb. 2021.

[34] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. de Oliveira Marinho, J. Gallegos, S. Gabriele, and V. Gupta, "A deep learning system for differential diagnosis of skin diseases," *Nature Med.*, vol. 26, no. 6, pp. 900–908, 2020.

[35] S.-Q. Wang, X.-Y. Zhang, J. Liu, C. Tao, C.-Y. Zhu, C. Shu, T. Xu, and H.-Z. Jin, "Deep learning-based, computer-aided classifier developed with dermoscopic images shows comparable performance to 164 dermatologists in cutaneous disease diagnosis in the Chinese population," *Chin. Med. J.*, vol. 133, no. 17, p. 2027, 2020.

[36] L. Song, J. Lin, Z. J. Wang, and H. Wang, "An end-to-end multi-task deep learning framework for skin lesion analysis," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2912–2921, Oct. 2020.

[37] A. Khamparia, P. K. Singh, P. Rani, D. Samanta, A. Khanna, and B. Bhushan, "An Internet of Health Things-driven deep learning framework for detection and classification of skin cancer using transfer learning," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 7, p. e3963, 2021.

[38] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020.

[39] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative feature learning for skin disease classification using deep convolutional neural network," *IEEE Access*, vol. 8, pp. 39025–39033, 2020.

[40] G. Zhang, X. Shen, S. Chen, L. Liang, Y. Luo, J. Yu, and J. Lu, "DSM: A deep supervised multi-scale network learning for skin cancer segmentation," *IEEE Access*, vol. 7, pp. 140936–140945, 2019.

[41] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[42] M. A. Al-masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Methods Programs Biomed.*, vol. 190, Jul. 2020, Art. no. 105351.

[43] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multi-modal neural nets," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 538–546, Mar. 2019.

[44] J. Burdick, O. Marques, J. Weinthal, and B. Furht, "Rethinking skin lesion segmentation in a convolutional classifier," *J. Digit. Imag.*, vol. 31, no. 4, pp. 435–440, Aug. 2018.

[45] S. Sreena and A. Lijiya, "Skin lesion analysis towards melanoma detection," in *Proc. 2nd Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, vol. 1, Jul. 2019, pp. 32–36.

[46] Y. Ren, L. Yu, S. Tian, J. Cheng, Z. Guo, and Y. Zhang, "Serial attention network for skin lesion segmentation," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 2, pp. 799–810, 2021.

[47] J. Räsänen, M. Salmivuori, I. Pölönen, M. Grönroos, and N. Neittaanmäki, "Hyperspectral imaging reveals spectral differences and can distinguish malignant melanoma from pigmented basal cell carcinomas: A pilot study," *Acta Dermato Venereologica*, vol. 101, no. 2, 2021, Art. no. adv00405.

[48] S. Qamar, P. Ahmad, and L. Shen, "Dense encoder–decoder–based architecture for skin lesion segmentation," *Cognit. Comput.*, vol. 13, no. 2, pp. 583–594, Mar. 2021.

[49] D. N. A. Ningrum, S.-P. Yuan, W.-M. Kung, C.-C. Wu, I.-S. Tzeng, C.-Y. Huang, J. Y.-C. Li, and Y.-C. Wang, "Deep learning classifier with patient's metadata of dermoscopic images in malignant melanoma detection," *J. Multidisciplinary Healthcare*, vol. 14, p. 877, Apr. 2021.

[50] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018.

[51] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.

[52] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, Art. no. 105725.

[53] D. Moldovan, "Transfer learning based method for two-step skin cancer images classification," in *Proc. E-Health Bioeng. Conf. (EHB)*, Nov. 2019, pp. 1–4.

[54] Y. Yan, J. Kawahara, and G. Hamarneh, "Melanoma recognition via visual attention," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 793–804.

[55] D. Zhuang, K. Chen, and J. M. Chang, "CS-AF: A cost-sensitive multi-classifier active fusion framework for skin lesion classification," 2020, *arXiv:2004.12064*.

[56] C. Zhao, R. Shuai, L. Ma, W. Liu, D. Hu, and M. Wu, "Dermoscopy image classification based on StyleGAN and DenseNet201," *IEEE Access*, vol. 9, pp. 8659–8679, 2021.

[57] C.-Y. Zhu, Y.-K. Wang, H.-P. Chen, K.-L. Gao, C. Shu, J.-C. Wang, L.-F. Yan, Y.-G. Yang, F.-Y. Xie, and J. Liu, "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment," *Frontiers Med.*, vol. 8, Apr. 2021, Art. no. 626369.

[58] Q. Zhou, Y. Shi, Z. Xu, R. Qu, and G. Xu, "Classifying melanoma skin lesions using convolutional spiking neural networks with unsupervised STDP learning rule," *IEEE Access*, vol. 8, pp. 101309–101319, 2020.

[59] Q. Zhou, C. Ren, and S. Qi, "An imbalanced R-STDP learning rule in spiking neural networks for medical image classification," *IEEE Access*, vol. 8, pp. 224162–224177, 2020.

[60] H. Wu, J. Pan, Z. Li, Z. Wen, and J. Qin, "Automated skin lesion segmentation via an adaptive dual attention module," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 357–370, Jan. 2021.

[61] A. Kwasigroch, M. Grochowski, and A. Mikołajczyk, "Self-supervised learning to increase the performance of skin lesion classification," *Electronics*, vol. 9, no. 11, p. 1930, Nov. 2020.

[62] Y. Gu, Z. Ge, C. P. Bonnington, and J. Zhou, "Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 5, pp. 1379–1393, May 2020.

[63] R. Vani, J. C. Kavitha, and D. Subitha, "Novel approach for melanoma detection through iterative deep vector network," *J. Ambient Intell. Humanized Comput.*, vol. 2021, pp. 1–10, Apr. 2021.

[64] M. A. Khan, T. Akram, Y.-D. Zhang, and M. Sharif, "Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework," *Pattern Recognit. Lett.*, vol. 143, pp. 58–66, Mar. 2021.

[65] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," *Artif. Intell. Med.*, vol. 102, Jan. 2020, Art. no. 101756.

[66] Z. Yu, F. Jiang, F. Zhou, X. He, D. Ni, S. Chen, T. Wang, and B. Lei, "Convolutional descriptors aggregation via cross-net for skin lesion recognition," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106281.

[67] M. A. Khan, M. Sharif, T. Akram, R. Damaševičius, and R. Maskeliūnas, "Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization," *Diagnostics*, vol. 11, no. 5, p. 811, Apr. 2021.

[68] T. Y. Tan, L. Zhang, and C. P. Lim, "Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104807.

[69] E. O. Molina-Molina, S. Solorza-Calderón, and J. Álvarez-Borrego, "Classification of dermoscopy skin lesion color-images using fractal-deep learning features," *Appl. Sci.*, vol. 10, no. 17, p. 5954, Aug. 2020.

[70] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.

[71] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112961.

[72] Y. Filali, H. E. Khoukhi, M. A. Sabri, and A. Aarab, "Efficient fusion of handcrafted and pre-trained CNNs features to classify melanoma skin cancer," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31219–31238, Nov. 2020.

[73] H. El-Khatib, D. Popescu, and L. Ichim, "Deep learning–based methods for automatic diagnosis of skin lesions," *Sensors*, vol. 20, no. 6, p. 1753, Mar. 2020.

[74] D. Divya and T. R. Ganeshbabu, "Fitness adaptive deer hunting-based region growing and recurrent neural network for melanoma skin cancer detection," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 731–752, Sep. 2020.

[75] R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber, M. Bakhtyar, I. Mehmood, O.-Y. Song, and M. Maqsood, "Region-of-interest based transfer learning assisted framework for skin cancer detection," *IEEE Access*, vol. 8, pp. 147858–147871, 2020.

[76] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, Apr. 2019.

[77] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Optimised deep learning features for improved melanoma detection," *Multimedia Tools Appl.*, vol. 78, no. 9, pp. 11883–11903, May 2019.

[78] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognit.*, vol. 85, pp. 78–89, Jan. 2019.

[79] M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, and A. Yazd-abadi, "Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture," *Comput. Methods Programs Biomed.*, vol. 177, pp. 17–30, Aug. 2019.

[80] Q. Abbas and M. E. Celebi, "DermoDeep—A classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network," *Multimedia Tools Appl.*, vol. 78, no. 16, pp. 23559–23580, Aug. 2019.

[81] M. A. Khan, T. Akram, M. Sharif, A. Shahzad, K. Aurangzeb, M. Alhussein, S. I. Haider, and A. Altamrah, "An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification," *BMC Cancer*, vol. 18, no. 1, pp. 1–20, Dec. 2018.

[82] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *J. Biomed. Inform.*, vol. 86, pp. 25–32, Oct. 2018.

[83] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools Appl.*, vol. 77, pp. 9909–9924, Feb. 2018.

[84] M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. M. R. Soroushmehr, S. Samavi, and K. Najarian, "Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 6, pp. 1021–1030, Jun. 2017.

[85] S. Serte and H. Demirel, "Wavelet-based deep learning for skin lesion classification," *IET Image Process.*, vol. 14, no. 4, pp. 720–726, Mar. 2020.

[86] R. Kaymak, C. Kaymak, and A. Ucar, "Skin lesion segmentation using fully convolutional networks: A comparative experimental study," *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113742.

[87] I. Bakkouri and K. Afdel, "Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images," *Multimedia Tools Appl.*, vol. 79, nos. 29–30, pp. 20483–20518, Aug. 2020.

[88] B. N. Walker, J. M. Rehg, A. Kalra, R. M. Winters, P. Drews, J. Dascalu, E. O. David, and A. Dascalu, "Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: Laboratory and prospective observational studies," *EBioMedicine*, vol. 40, pp. 176–183, Feb. 2019.

[89] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *J. Investigative Dermatol.*, vol. 138, no. 7, pp. 1529–1538, 2018.

[90] L. Q. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Dec. 2017.

[91] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 519–526, Mar. 2019.

[92] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017.

[93] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Sci. Data*, vol. 8, no. 1, pp. 1–8, 2021.

[94] A. S. Qureshi and T. Roos, "Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets," 2021, *arXiv:2103.12068*.

[95] M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "BCN20000: Dermoscopic lesions in the wild," 2019, *arXiv:1908.02288*.

[96] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data," *MethodsX*, vol. 7, Jan. 2020, Art. no. 100864.

[97] S. A. A. Ahmed, B. Yanikoglu, C. Zor, M. Awais, and J. Kittler, "Skin lesion diagnosis with imbalanced ECOC ensembles," in *Proc. Int. Conf. Mach. Learn., Optim., Data Sci.* Cham, Switzerland: Springer, 2020, pp. 292–303.

[98] T. A. Putra, S. I. Rufaida, and J.-S. Leu, "Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation," *IEEE Access*, vol. 8, pp. 40536–40546, 2020.

[99] F. Nunnari, C. Bhuvaneshwara, A. O. Ezema, and D. Sonntag, "A study on the fusion of pixels and patient metadata in CNN-based classification of skin lesion images," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*. Cham, Switzerland: Springer, 2020, pp. 191–208.

[100] A. Gong, X. Yao, and W. Lin, "Dermoscopy image classification based on StyleGANs and decision fusion," *IEEE Access*, vol. 8, pp. 70640–70650, 2020.

[101] S. M. Alizadeh and A. Mahloojifar, "Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 2, pp. 695–707, Jun. 2021.

[102] M. Cullell-Dalmau, S. Noé, M. Otero-Viñas, I. Meić, and C. Manzo, "Convolutional neural network for skin lesion classification: Understanding the fundamentals through hands-on learning," *Frontiers Med.*, vol. 8, p. 213, Mar. 2021.

[103] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, *arXiv:1902.03368*.

[104] A. Bissoto, F. Perez, E. Valle, and S. Avila, "Skin lesion synthesis with generative adversarial networks," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 294–302.

[105] A. C. F. Gouabou, J.-L. Damoiseaux, J. Monnier, R. Iguernaissi, A. Moudafi, and D. Merad, "Ensemble method of convolutional neural networks with directed acyclic graph using dermoscopic images: Melanoma detection application," *Sensors*, vol. 21, no. 12, p. 3999, Jun. 2021.

[106] A.-R. Ali, J. Li, S. Kanwal, G. Yang, A. Hussain, and S. Jane O'Shea, "A novel fuzzy multilayer perceptron (F-MLP) for the detection of irregularity in skin lesion border using dermoscopic images," *Frontiers Med.*, vol. 7, p. 297, Jul. 2020.

[107] A.-R. Ali, J. Li, and T. Trappenberg, "Supervised versus unsupervised deep learning based methods for skin lesion segmentation in dermoscopy images," in *Proc. Can. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2019, pp. 373–379.

[108] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[109] F. Pollastri, F. Bolelli, R. Paredes, and C. Grana, "Augmenting data with GANs to segment melanoma skin lesions," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15575–15592, Jun. 2020.

[110] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "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)," 2016, *arXiv:1605.01397*.

[111] N. C. Codella, Q.-B. Nguyen, S. Pankanti, D. A. Gutman, B. Helba, A. C. Halpern, and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM J. Res. Develop.*, vol. 61, nos. 4–5, pp. 1–5, 2017.

[112] C. N. Vasconcelos and B. N. Vasconcelos, "Experiments using deep learning for dermoscopy image analysis," *Pattern Recognit. Lett.*, vol. 139, pp. 95–103, Nov. 2020.

[113] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering," *Int. J. Med. Informat.*, vol. 124, pp. 37–48, Apr. 2019.

[114] *ISIC Project-ISIC Archive*. Accessed: Nov. 20, 2021. [Online]. Available: https://www.isic-archive.com

[115] M. N. Bajwa, K. Muta, M. I. Malik, S. A. Siddiqui, S. A. Braun, B. Homey, A. Dengel, and S. Ahmed, "Computer-aided diagnosis of skin diseases using deep neural networks," *Appl. Sci.*, vol. 10, no. 7, p. 2488, Apr. 2020.

[116] R. C. Maron *et al.*, "Robustness of convolutional neural networks in recognition of pigmented skin lesions," *Eur. J. Cancer*, vol. 145, pp. 81–91, Mar. 2021.

[117] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, T. Holland-Letz, J. S. Utikal, and C. von Kalle, "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *Eur. J. Cancer*, vol. 113, pp. 47–54, May 2019.

[118] A. Bissoto, M. Fornaciali, E. Valle, and S. Avila, "(De)Constructing bias on skin lesion datasets," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.

[119] M. A. Albahar, "Skin lesion classification using convolutional neural network with novel regularizer," *IEEE Access*, vol. 7, pp. 38306–38313, 2019.

[120] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Dec. 2018.

[121] J. R. Hagerty, R. J. Stanley, H. A. Almubarak, N. Lama, R. Kasmi, P. Guo, R. J. Drugge, H. S. Rabinovitz, M. Oliviero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images," *IEEE J. Biomed. Health*, vol. 23, no. 4, pp. 1385–1391, Jul. 2019.

[122] O. Sevli, "A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation," *Neural Comput. Appl.*, vol. 33, no. 18, pp. 12039–12050, 2021.

[123] R. C. Maron, A. Hekler, E. Krieghoff-Henning, M. Schmitt, J. G. Schlager, J. S. Utikal, and T. J. Brinker, "Reducing the impact of confounding factors on skin cancer classification via image segmentation: Technical model study," *J. Med. Internet Res.*, vol. 23, no. 3, Mar. 2021, Art. no. e21695.

[124] I. S. A. Abdelhalim, M. F. Mohamed, and Y. B. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113922.

[125] K. Thurnhofer-Hemsi and E. Dominguez, "A convolutional neural network framework for accurate skin cancer detection," *Neural Process. Lett.*, pp. 1–21, 2020.

[126] T.-C. Pham, A. Doucet, C.-M. Luong, C.-T. Tran, and V.-D. Hoang, "Improving skin-disease classification based on customized loss function combined with balanced mini-batch logic and real-time image augmentation," *IEEE Access*, vol. 8, pp. 150725–150737, 2020.

[127] M. Lucius, J. De All, J. A. De All, M. Belvisi, L. Radizza, M. Lanfranconi, V. Lorenzatti, and C. M. Galmarini, "Deep neural frameworks improve the accuracy of general practitioners in the classification of pigmented skin lesions," *Diagnostics*, vol. 10, no. 11, p. 969, Nov. 2020.

[128] A. Hekler *et al.*, "Effects of label noise on deep learning-based skin cancer classification," *Frontiers Med.*, vol. 7, p. 177, May 2020.

[129] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin cancer classification using deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 39–40, pp. 28477–28498, Oct. 2020.

[130] A. A. Adegun and S. Viriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images," *IEEE Access*, vol. 8, pp. 150377–150396, 2020.

[131] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH²—A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.

[132] G. Argenziano *et al.*, "Interactive atlas of dermoscopy," 2000.

[133] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*. Dordrecht, The Netherlands: Springer, 2013, pp. 63–86.

[134] X. Fan, M. Dai, C. Liu, F. Wu, X. Yan, Y. Feng, Y. Feng, and B. Su, "Effect of image noise on the classification of skin lesions using deep convolutional neural networks," *Tsinghua Sci. Technol.*, vol. 25, no. 3, pp. 425–434, Jun. 2020.

[135] *DermNet NZ*. Accessed: May 2, 2021. [Online]. Available: https://dermnetnz.org/

[136] S. Ranjana, R. Manimegala, and K. Priya, "Deep learning system for skin disorder segmentation using neural network," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 9, p. 2020, 2020.

[137] D. Choudhury, A. Naug, and S. Ghosh, "Texture and color feature based WLS framework aided skin cancer classification using MSVM and ELM," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–6.

[138] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Syst. Appl.*, vol. 42, no. 19, pp. 6578–6585, Nov. 2015.

[139] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, S. Fröhling, J. S. Utikal, and C. von Kalle, "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *Eur. J. Cancer*, vol. 111, pp. 148–154, Apr. 2019.

[140] H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk, and L. Uhlmann, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018.

[141] B. A. Albert, "Deep learning from limited training data: Novel segmentation and ensemble algorithms applied to automatic melanoma diagnosis," *IEEE Access*, vol. 8, pp. 31254–31269, 2020.

[142] J. Parker, "Rank and response combination from confusion matrix data," *Inf. fusion*, vol. 2, no. 2, pp. 113–120, 2001.

[143] J. Stojkovic-Filipovic, D. Tiodorovic, A. Lallas, B. N. Akay, C. Longo, C. Rosendahl, D. Dobrosavljevic, G. Nazzaro, G. Argenziano, I. Zalaudek, I. Tromme, P. Tschandl, S. Puig, S. Lanssens, and H. Kittler, "Dermatoscopy of combined blue nevi: A multicentre study of the international dermoscopy society," *J. Eur. Acad. Dermatol. Venereol.*, vol. 35, no. 4, pp. 900–905, Apr. 2021.

[144] K. Sies, J. K. Winkler, C. Fink, F. Bardehle, F. Toberer, T. Buhl, A. Enk, A. Blum, A. Rosenberger, and H. A. Haenssle, "Past and present of computer-assisted dermoscopic diagnosis: Performance of a conventional image analyser versus a convolutional neural network in a prospective data set of 1,981 skin lesions," *Eur. J. Cancer*, vol. 135, pp. 39–46, Aug. 2020.

[145] R. C. Maron *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *Eur. J. Cancer* vol. 119, pp. 57–65, Sep. 2019.

[146] A. Hekler, J. S. Utikal, A. H. Enk, W. Solass, M. Schmitt, J. Klode, D. Schadendorf, W. Sondermann, C. Franklin, F. Bestvater, M. J. Flaig, D. Krahl, C. von Kalle, S. Fröhling, and T. J. Brinker, "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images," *Eur. J. Cancer*, vol. 118, pp. 91–96, Sep. 2019.

[147] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal, R. C. Maron, C. Berking, S. Haferkamp, J. Klode, D. Schadendorf, B. Schilling, T. Holland-Letz, B. Izar, C. Von Kalle, S. Fröhling, and T. J. Brinker, "Superior skin cancer classification by the combination of human and artificial intelligence," *Eur. J. Cancer*, vol. 120, pp. 114–121, Oct. 2019.

[148] H. A. Haenssle *et al.*, "Skin lesions of face and scalp—Classification by a market-approved convolutional neural network in comparison with 64 dermatologists," *Eur. J. Cancer*, vol. 144, pp. 192–199, Feb. 2021.

[149] H. A. Haenssle *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.*, vol. 31, no. 1, pp. 137–143, Jan. 2020.

[150] T. J. Brinker, A. Hekler, A. Hauschild, C. Berking, B. Schilling, A. H. Enk, S. Haferkamp, A. Karoglan, C. von Kalle, M. Weichenthal, E. Sattler, D. Schadendorf, M. R. Gaiser, J. Klode, and J. S. Utikal, "Comparing artificial intelligence algorithms to 157 German dermatologists: The melanoma classification benchmark," *Eur. J. Cancer*, vol. 111, pp. 30–37, Apr. 2019.

[151] L. R. Soenksen, T. Kassis, S. T. Conover, B. Marti-Fuster, J. S. Birkenfeld, J. Tucker-Schwartz, A. Naseem, R. R. Stavert, C. C. Kim, M. M. Senna, J. Avilés-Izquierdo, J. J. Collins, R. Barzilay, and M. L. Gray, "Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images," *Sci. Transl. Med.*, vol. 13, no. 581, Feb. 2021, Art. no. eabb3652.

[152] M. Y. Sikkandar, B. A. Alrasheadi, N. Prakash, G. Hemalakshmi, A. Mohanarathinam, and K. Shankar, "Deep learning based an automated skin lesion segmentation and intelligent classification model," *J. Ambient Intell. Humanized Comput.*, vol. 12, pp. 3245–3255, Sep. 2020.

[153] S. Maniraj and P. Sardarmaran, "Classification of dermoscopic images using soft computing techniques," *Neural Comput. Appl.*, vol. 33, no. 19, pp. 13015–13026, 2021.

[154] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Med. Imag. Graph.*, vol. 88, Mar. 2021, Art. no. 101843.

[155] B. Zhang, Z. Wang, J. Gao, C. Rutjes, K. Nufer, D. Tao, D. D. Feng, and S. W. Menzies, "Short-term lesion change detection for melanoma screening with novel Siamese neural network," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 840–851, Mar. 2021.

[156] J. K. Winkler, K. Sies, C. Fink, F. Toberer, A. Enk, T. Deinlein, R. Hofmann-Wellenhof, L. Thomas, A. Lallas, A. Blum, W. Stolz, M. S. Abassi, T. Fuchs, A. Rosenberger, and H. A. Haenssle, "Melanoma recognition by a deep learning convolutional neural network—Performance in different melanoma subtypes and localisations," *Eur. J. Cancer*, vol. 127, pp. 21–29, Mar. 2020.

[157] S. Fekri-Ershad, M. Saberi, and F. Tajeripour, "An innovative skin detection approach using color based image retrieval technique," 2012, *arXiv:1207.1551*.

[158] H. Zunair and A. Ben Hamza, "Melanoma detection using adversarial training and deep transfer learning," *Phys. Med. Biol.*, vol. 65, no. 13, Jul. 2020, Art. no. 135005.

[159] L. Liu, L. Mou, X. X. Zhu, and M. Mandal, "Automatic skin lesion classification based on mid-level feature learning," *Computerized Med. Imag. Graph.*, vol. 84, Sep. 2020, Art. no. 101765.

[160] A. Ng. (2017). *Machine Learning Yearning*. [Online]. Available: http://www.mlyearning.org/

[161] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep learning-based transfer learning for classification of skin cancer," *Sensors*, vol. 21, no. 23, p. 8142, Dec. 2021.

[162] Y. Wang, D. C. Louie, J. Cai, L. Tchvialeva, H. Lui, Z. Jane Wang, and T. K. Lee, "Deep learning enhances polarization speckle for *in vivo* skin cancer detection," *Opt. Laser Technol.*, vol. 140, Aug. 2021, Art. no. 107006.

[163] Y. Wang, J. Cai, D. C. Louie, Z. J. Wang, and T. K. Lee, "Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104812.

[164] X. Wang, X. Jiang, H. Ding, Y. Zhao, and J. Liu, "Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108075.

[165] W. Barhoumi and A. Khelifa, "Skin lesion image retrieval using transfer learning-based approach for query-driven distance recommendation," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104825.

[166] S. Alzahrani, B. Al-Bander, and W. Al-Nuaimy, "A comprehensive evaluation and benchmarking of convolutional neural networks for melanoma diagnosis," *Cancers*, vol. 13, no. 17, p. 4494, Sep. 2021.

[167] M. A. Khan, M. Sharif, T. Akram, S. Kadry, and C.-H. Hsu, "A two-stream deep neural network-based intelligent system for complex skin cancer types classification," *Int. J. Intell. Syst.*, 2021.

[168] D. D. A. Rodrigues, R. F. Ivo, S. C. Satapathy, S. Wang, J. Hemanth, and P. P. R. Filho, "A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system," *Pattern Recognit. Lett.*, vol. 136, pp. 8–15, Aug. 2020.

[169] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105568.

[170] A. Kwasigroch, M. Grochowski, and A. Mikolajczyk, "Neural architecture search for skin lesion classification," *IEEE Access*, vol. 8, pp. 9061–9071, 2020.

[171] M. A. Khan, M. Sharif, T. Akram, S. A. C. Bukhari, and R. S. Nayak, "Developed Newton-Raphson based deep features selection framework for skin lesion recognition," *Pattern Recognit. Lett.*, vol. 129, pp. 293–303, Jan. 2020.

[172] M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí, "DSNet: Automatic dermoscopic skin lesion segmentation," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103738.

[173] T. Akram, H. M. J. Lodhi, S. R. Naqvi, S. Naeem, M. Alhaisoni, M. Ali, S. A. Haider, and N. N. Qadri, "A multilevel features selection framework for skin lesion classification," *Hum.-Centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–26, Dec. 2020.

[174] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2092–2103, Sep. 2019.

[175] J. K. Winkler, C. Fink, F. Toberer, A. Enk, T. Deinlein, R. Hofmann-Wellenhof, L. Thomas, A. Lallas, A. Blum, W. Stolz, and H. A. Haenssle, "Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition," *JAMA Dermatol.*, vol. 155, no. 10, pp. 1135–1141, 2019.

[176] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of Yolo and GrabCut algorithm," *Diagnostics*, vol. 9, no. 3, p. 72, Jul. 2019.

[177] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Comput. Biol. Med.*, vol. 113, Oct. 2019, Art. no. 103423.

[178] A. Mobiny, A. Singh, and H. Van Nguyen, "Risk-aware machine learning classifier for skin lesion diagnosis," *J. Clin. Med.*, vol. 8, no. 8, p. 1241, Aug. 2019.

[179] K. M. Hosny, M. A. Kassem, and M. M. Fouad, "Classification of skin lesions into seven classes using transfer learning with AlexNet," *J. Digit. Imag.*, vol. 33, no. 5, pp. 1325–1334, 2020.

[180] M. Nawaz, M. Masood, A. Javed, I. Iqbal, T. Nazir, A. Mehmood, and R. Ashraf, "Melanoma localization and classification through faster region-based convolutional neural network and SVM," *Multimedia Tools Appl.*, vol. 80, no. 19, pp. 28953–28974, 2021.

[181] O. Attallah and M. Sharkas, "Intelligent dermatologist tool for classifying multiple skin cancer subtypes by incorporating manifold radiomics features categories," *Contrast Media Mol. Imag.*, vol. 2021, pp. 1–14, Sep. 2021.

[182] P. Tang, Q. Liang, X. Yan, S. Xiang, and D. Zhang, "GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2870–2882, Oct. 2020.

[183] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," *Biomolecules*, vol. 10, no. 8, p. 1123, Jul. 2020.

[184] A. A. Adegun and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020.

[185] G. Zeng and G. Zheng, "Multi-scale fully convolutional DenseNets for automated skin lesion segmentation in dermoscopy images," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 513–521.

[186] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 24029–24055, Sep. 2020.

[187] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis.* Cham, Switzerland: Springer, 2018, pp. 303–311.

[188] K. Kato, K. Nemoto, Y. Kimura, Y. Kiyohara, H. Koga, N. Yamazaki, G. Christensen, C. Ingvar, K. Nielsen, A. Nakamura, T. Sota, and T. Nagaoka, "Performance improvement of automated melanoma diagnosis system by data augmentation," *Adv. Biomed. Eng.*, vol. 9, pp. 62–70, Jan. 2020.

[189] K. Thurnhofer-Hemsi and E. Domínguez, "A convolutional neural network framework for accurate skin cancer detection," *Neural Process. Lett.*, vol. 53, no. 5, pp. 3073–3093, 2021.

[190] S. Oukil, R. Kasmi, K. Mokrani, and B. García-Zapirain, "Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images," *Skin Res. Technol.*, vol. 28, no. 2, pp. 203–211, Mar. 2022.

[191] B. Ahmad, S. Jun, V. Palade, Q. You, L. Mao, and M. Zhongjie, "Improving skin cancer classification using heavy-tailed student T-distribution in generative adversarial networks (TED-GAN)," *Diagnostics*, vol. 11, no. 11, p. 2147, Nov. 2021.

[192] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[193] H. Inoue, "Data augmentation by pairing samples for images classification," 2018, *arXiv:1801.02929*.

[194] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.

[195] K. W. Lee and R. K. Y. Chin, "The effectiveness of data augmentation for melanoma skin cancer prediction using convolutional neural networks," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Eng. Technol. (IICAIET)*, Sep. 2020, pp. 1–6.

[196] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[197] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*.

[198] T. DeVries and D. Ramachandram, "Skin lesion classification using deep multi-scale convolutional neural networks," 2017, *arXiv:1703.01402*.

[199] A. H. Shahin, A. Kamal, and M. A. Elattar, "Deep ensemble learning for skin lesion classification from dermoscopic images," in *Proc. 9th Cairo Int. Biomed. Eng. Conf. (CIBEC)*, Dec. 2018, pp. 150–153.

[200] A. G. C. Pacheco, A.-R. Ali, and T. Trappenberg, "Skin cancer detection based on deep learning and entropy to detect outlier samples," 2019, *arXiv:1909.04525*.

[201] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1229–1233.

[202] X. Cao, J.-S. Pan, Z. Wang, Z. Sun, A. U. Haq, W. Deng, and S. Yang, "Application of generated mask method based on mask R-CNN in classification and detection of melanoma," *Comput. Methods Programs Biomed.*, vol. 207, Aug. 2021, Art. no. 106174.

[203] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection," *Multimedia Syst.*, vol. 28, pp. 1309–1323, Apr. 2021.

[204] L. P. Cordella, C. De Stefano, F. Fontanella, and A. S. Di Freca, "A weighted majority vote strategy using Bayesian networks," in *Proc. Int. Conf. Image Anal. Process.* Berlin, Germany: Springer, 2013, pp. 219–228.

[205] Y. Guo, Z. Zhang, J. Jiang, W. Wu, C. Zhang, B. Cui, and J. Li, "Model averaging in distributed machine learning: A case study with Apache Spark," *VLDB J.*, vol. 30, no. 4, pp. 693–712, 2021.

[206] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, "Averaging weights leads to wider optima and better generalization," 2018, *arXiv:1803.05407*.

[207] T. Garipov, P. Izmailov, D. Podoprikhin, D. Vetrov, and A. G. Wilson, "Loss surfaces, mode connectivity, and fast ensembling of DNNs," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 8803–8812.

[208] D. Vincent, S. Christoph, M. Martina, K. Harald, and T. Philipp, "Accuracy of computer-aided diagnosis of melanoma: A meta-analysis," *JAMA Dermatol.*, vol. 155, no. 11, pp. 1291–1299, 2019.

[209] Leaderboards. *Isic Challenge*. Accessed: Oct. 10, 2021. [Online]. Available: https://challenge.isic-archive.com/

[210] Q. Dong, S. Gong, and X. Zhu, "Imbalanced deep learning by minority class incremental rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1367–1381, Jun. 2019.

[211] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[212] S. Khouloud, M. Ahlem, T. Fadel, and S. Amel, "W-Net and inception residual network for skin lesion segmentation and classification," *Appl. Intell.*, vol. 52, no. 4, pp. 3976–3994, 2021.

[213] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[214] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[215] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for robust semantic pixel-wise labelling," 2015, *arXiv:1505.07293*.

[216] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[217] A. G. C. Pacheco and R. A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103545.

[218] A. Hekler, J. S. Utikal, A. H. Enk, C. Berking, J. Klode, D. Schadendorf, P. Jansen, C. Franklin, T. Holland-Letz, D. Krahl, C. von Kalle, S. Fröhling, and T. J. Brinker, "Pathologist-level classification of histopathological melanoma images with deep neural networks," *Eur. J. Cancer*, vol. 115, pp. 79–83, Jul. 2019.

[219] K. Freeman, J. Dinnes, N. Chuchu, Y. Takwoingi, S. E. Bayliss, R. N. Matin, A. Jain, F. M. Walter, H. C. Williams, and J. J. Deeks, "Algorithm based smartphone apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies," *bmj*, vol. 368, p. m127, Feb. 2020.

[220] I. Pölönen, S. Rahkonen, L. Annala, and N. Neittaanmäki, "Convolutional neural networks in skin cancer detection using spatial and spectral domain," *Proc. SPIE*, vol. 10851, Feb. 2019, Art. no. 108510B.

[221] A. Dascalu and E. O. David, "Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope," *EBioMedicine*, vol. 43, pp. 107–113, May 2019.

[222] European Union, "Regulation (EU) 2017/745 of the European parliament and of the council of 5 April 2017 on medical devices, amending directive 2001/83/EC, regulation (EC) No 178/2002 and regulation (EC) No 1223/2009 and repealing council directives 90/385/EEC and 93/42/EEC," *Off. J. Eur. Union*, vol. 50, pp. 1–175, 2017.

[223] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[224] J. Bian, S. Zhang, S. Wang, J. Zhang, and J. Guo, "Skin lesion classification by multi-view filtered transfer learning," *IEEE Access*, vol. 9, pp. 66052–66061, 2021.

[225] M. A. Marchetti, K. Liopyris, S. W. Dusza, N. C. F. Codella, D. A. Gutman, B. Helba, A. Kalloo, A. C. Halpern, H. P. Soyer, C. Curiel-Lewandrowski, L. Caffery, and J. Malvehy, "Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the international skin imaging collaboration 2017," *J. Amer. Acad. Dermatol.*, vol. 82, no. 3, pp. 622–627, Mar. 2020.

[226] K. Jayapriya and I. J. Jacob, "Hybrid fully convolutional networks-based skin lesion segmentation and melanoma detection using deep feature," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 2, pp. 348–357, Jun. 2020.

[227] C. Fink, A. Blum, T. Buhl, C. Mitteldorf, R. Hofmann-Wellenhof, T. Deinlein, W. Stolz, L. Trennheuser, C. Cussigh, D. Deltgen, J. K. Winkler, F. Toberer, A. Enk, A. Rosenberger, and H. A. Haenssle, "Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas," *J. Eur. Acad. Dermatol. Venereol.*, vol. 34, no. 6, pp. 1355–1361, Jun. 2020.

[228] S. Banerjee, S. K. Singh, A. Chakraborty, A. Das, and R. Bag, "Melanoma diagnosis using deep learning and fuzzy logic," *Diagnostics*, vol. 10, no. 8, p. 577, Aug. 2020.

[229] H.-H. Wang, Y.-H. Wang, C.-W. Liang, and Y.-C. Li, "Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer," *JAMA Dermatol.*, vol. 155, no. 11, pp. 1277–1283, 2019.

[230] M. Phillips, H. Marsden, W. Jaffe, R. N. Matin, G. N. Wali, J. Greenhalgh, E. McGrath, R. James, E. Ladoyanni, A. Bewley, G. Argenziano, and I. Palamaras, "Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions," *JAMA Netw. Open*, vol. 2, no. 10, Oct. 2019, Art. no. e1913436.

[231] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, "Towards automated melanoma detection with deep learning: Data purification and augmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.

[232] X. Zhang, S. Wang, J. Liu, and C. Tao, "Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge," *BMC Med. Informat. Decis. Making*, vol. 18, no. S2, pp. 69–76, Jul. 2018.

[233] J. Yap, W. Yolland, and P. Tschandl, "Multimodal skin lesion classification using deep learning," *Exp. Dermatol.*, vol. 27, no. 11, pp. 1261–1267, 2018.

MATTEO FERRO (Member, IEEE) was born in Salerno, Italy, in 1986. He received the M.S. degree in information engineering and the Ph.D. degree from the University of Salerno, Italy, in 2016 and 2020, respectively.

He is currently with the Department of Industrial Engineering, University of Salerno, as a Research Fellow of Electronic Measurements. He collaborates as a Representative of the Order of Engineers of Salerno with the Agency for Digital Italy in the commission for digital health. His current research interests include smart meters, wireless sensor networks, embedded systems, software medical device, big data analysis, and cyber security.

YALI NIE (Member, IEEE) received the B.S. degree in computer science from the Jiangxi University of Finance and Economics, China, in 2015, and the M.S. degree in engineering from Chonbuk National University, Republic of Korea, in 2018. She is currently pursuing the Ph.D. degree with Mid Sweden University. Her research interests include computer vision, medical image processing, machine learning, and deep learning.

PAOLO SOMMELLA (Member, IEEE) was born in Salerno, Italy, in 1979. He received the M.S. degree in electronic engineering and the Ph.D. degree in information engineering from the University of Salerno, Italy, in 2004 and 2008, respectively.

In 2015, he joined the Department of Industrial Engineering (DIIn), University of Salerno, as an Assistant Professor in electrical and electronic measurements, where he teaches the courses of automatic controls and vision-based measurement systems at the Department of Industrial Engineering. His research interests include instrument fault detection and isolation, measurement in software engineering, and biomedical image processing.

MATTIAS O'NILS received the B.Sc. degree in electronics design from Mid Sweden University, in 1993, and the Ph.D. degree in electronics system design from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1999. He is currently a Professor and leads the research group in embedded systems design with the Department of Electronics Design, Mid Sweden University.

JAN LUNDGREN received the Ph.D. degree from Mid Sweden University, Sundsvall, Sweden, in 2007.

After receiving his Ph.D. degree, he worked as an Assistant Professor. Since 2013, he has been an Associate Professor at Mid Sweden University. Currently, he leads a research group at Mid Sweden University focusing on AI-supported sensor systems, including industrial sound and imaging measurements and also medical sound and imaging measurements. His research interests include industrial sound measurements, medical sound and imaging measurements, and industrial soft sensors.

MARCO CARRATÙ (Member, IEEE) was born in Brugg, Switzerland, in 1989. He received the M.S. degree in electronic engineering and the Ph.D. degree from the University of Salerno, Italy, in 2015 and 2019, respectively. He is currently with the Department of Industrial Engineering, University of Salerno, as a Research Fellow of Electronic Measurements. His current research interests include instrument fault detection and isolation, sensor data fusion, digital signal processing for advanced instrumentation, ANN, and embedded systems.

. . .