Contributions and Limitations About the Use of Deep Learning for Skin Diagnosis: A Review

Eduardo L. L. Nascimento(✉) and Angel Freddy Godoy Viera

Postgraduate Program in Information Science - Federal University of Santa Catarina, Florianópolis, SC 88040-900, Brazil
nascimento.lln@gmail.com

Abstract. The aim of this study is to analyze the characteristics and applicability of Deep Learning (DL) models for the diagnosis of skin diseases. This study is characterized as a bibliographic review, exploratory-descriptive, qualitative in nature. Primary data was reported in the article databases. A total of 37 articles were analyzed to characterize the use of DL for the diagnosis of skin diseases. The survey results that public datasets access is mostly used in these surveys are (86%). The data collection that stood out was ISIC - International Skin Imaging Collaboration (54%). Greater commonly used data types in these models are images. Ultimately used model is the Convolutional Neural Network (CNN) and the uttermost used pre-trained model was ResNet. The most used techniques in the articles, in addition to classification (73%), focused on data segmentation (35%) and feature extraction (24%). The evaluation indicators that stand out are accuracy (89%), sensitivity (75%), and specificity (67%). The literature indicated that the approaches of studies that use DL for classification of skin diseases are very promising, however, practically all of the applied technologies have a greater need for interaction with clinical practices. As a suggestion for different works, studies that approach the task of DL work for diagnosis of different ethnic groups and their solutions for the democratization of such technologies.

Keywords: Deep learning · Skin lesion classification · Skin lesion diagnostics · Skin disease · Dermatopathology · Image recognition

1 Introduction

Emerging technological alternatives that have the ability to impact healthcare are Deep Learning (DL) applications. With the improvement of these practices in recent years models based on DL are already presented as possible allies for the identification of diseases such as COVID-19 [1, 2], in addition to possible applications in medical areas such as: psychiatry [3], ophthalmology [4], oncology [5, 6] and dermatology [7].

The traditional task of recognizing skin lesions consists of several steps, where the automated system is typically trained on pre-processed images with a known diagnosis, which allows the prediction of new specimens in defined classes [8].

A. L. Pinto and R. Arencibia-Jorge (Eds.): DIONE 2022, LNICST 452, pp. 133–149, 2022.
https://doi.org/10.1007/978-3-031-22324-2_11

In the development of DL models some tasks are performed to increase performance, such as: Feature Extraction, Data Segmentation, Data Augmentation, Transfer Learning, etc. Feature Extraction is the process of simplifying a large dataset to make it more suitable for further processing [9–11].

Data Augmentation (DA) is another widely used approach that consists of expanding the amount of training data [12]. In this way, AD seeks to solve the problem of a small dataset for training of skin diseases, in addition to improving the generalization capacity of the model. Some tasks such as: rotation, mirroring, and adding artifacts are tricks used to increase the amount of dataset elements [13, 14].

The use of pre-trained models (Transfer learning) is another technique applied in these models where characteristics of other models are incorporated into the model under development with similar objectives [15–18].

Despite promising Deep Learning systems to aid in skin diagnosis, several challenges are proposed for the adoption of these technologies in a safe and effective way, when it comes to the health of the population [19].

Another point that deserves attention is the availability of specialized professionals in remote or rural areas is a problem faced in several countries. When we combine technologies to support medical diagnosis based on techniques such as DL and Telehealth service this combination allows us to overcome obstacles such as time, cost and distance to provide health care in a supervised way regardless of how far these health services are from the population [20].

The research question that guided this study is: How is Deep Learning used in the process of diagnosing skin diseases?

The objective of this study was to analyze and describe the scientific literature on the subject of diagnosis of skin diseases that use deep learning identified in the bases of selected articles.

As specific objectives: 1) to observe if the article presents the structural criteria for the elaboration of the scientific production, 2) to analyze if the research presents the technical aspects for the construction of the experiment, and 3) to identify if the study presents/informs the context of application, users and/or integration with clinical practices.

2 Literature Review

Computer-aided diagnostic (CAD) systems allow accurate detection, segmentation and classification of skin lesions. Thus, in clinical applications a CAD system can be a great ally to support medical decisions [21].

CAD models developed for skin diseases are often image-trained, others also use metadata and clinical data [15, 22, 23]. Models based on convolutional neural network (CNN) architectures were developed with promising results revealing good diagnostic accuracy [24].

Shoieb, Youssef and Aly [25] present an improved CAD system using a hybrid model for skin lesion identification composed of CNN (designed as a feature extractor) combined with a multiclass Support Vector Machine (SVM) based classifier model.

The results showed superior performance compared to other models identified in the literature.

However, automatic segmentation of lesions on dermoscopic images is a challenge. Some studies compare the accuracy of these models with health professionals at a specialized level, others also compare it with non-specialized professionals [26].

Esteva et al. [7] suggest a model for classifying skin lesions using CNN full connect from images and later evaluate the performance of the model comparing it to 21 dermatologists.

Automated classification of skin lesions through images is a challenging task due to the variability in the appearance of skin lesions. It is extremely difficult due to the complexities of the skin tone, color, skin hair and the presence of other artifacts contained in the images [27].

As for the applied techniques, Saker et al. [28] point out that most skin lesion segmentation approaches use models with a large number of parameters configuring a major obstacle for practical application in the clinical environment.

Majtner, Yayilgan and Hardeberg [8] addressed the effective use of the feature set extracted from pre-trained deep learning models - ImageNet. In this study the authors propose the improvement of a melanoma detection method based on the combination of linear discriminant analysis (LDA).

Saker et al. [28] presented a lightweight and efficient generative adversarial network model called MobileGAN, which basically seeks to extract relevant features with multiscale convolutional networks evaluated in two datasets ISBI 2017 [29] and ISIC 2018 [30]. In addition, the experimental results demonstrate that the proposed model provides results comparable to the state of the art of skin lesion segmentation in terms of precision and similarity coefficients.

There are studies that highlight the use of machine learning classifiers in the diagnosis of skin lesions and point to their importance in clinical practice as a promising proposal [31].

Polat and Koc [32] propose a hybrid model for automatic classification of skin diseases divided into seven classes, combining CNN with One-versus-All, using data from the HAM10000 dataset.

3 Methodology

As for the methodological procedures, this is an exploratory-descriptive bibliographic study characterized as an integrative review. Thus, this research aims to generate knowledge about the use of deep learning for the diagnosis of skin diseases as well as the identification of gaps in the scientific field and the formulation of proposals for further research.

The primary data comes from the limitations imposed by the researchers in the searches in available databases. The approach is a qualitative study, in the production of critical analysis performed at the time of selection and analysis of the set of articles. As for the technical aspects, it is classified as a bibliography research in its form of integrative literature review.

This research was divided into three stages: 1) definitions and delimitations of research strategies, 2) data collection selection and 3) data analysis. In the first stage, a research protocol was elaborated to guide the delimitations of the theme and strategies for defining the objectives and research proposal.

In the second stage, a search was carried out on the CAPES journals portal, between the 22nd and 26th of March 2021, only articles written in english and with a publication date between 01/01/2016 to 06/30/ 2021 was selected. The period of publication selected was the result of search tests with better adherence to the problem that involves this research and the feasibility of the proposal.

As a search strategy, only articles that contained the search terms in the title and abstract were selected. The bases selected for collecting the studies were: Scopus (18), Web of Science (24), PubMed (11) and IEEE Xplore (7). The selection of the bases was the one that offered the best return on the works related to the techniques addressed in this study and their relevance in the scientific field.

The search expression used: "deep learning" AND (skin lesion* OR "skin lesion classification" OR "skin lesion segmentation" OR "skin cancer" OR "skin disease" OR "dermatopathology"), which resulted in 60 preliminary articles.

Then, titles, abstracts and keywords were read. Articles that did not contain the application of models based on Deep Learning for the diagnosis of dermatitis were excluded.

Duplicate studies, articles that were not practical (experimental) and literature reviews were excluded and resulted in 37 articles that were identified at the end of each reference in the last item of this research with "[PB]".

In the third step, the systematization and analysis of data in electronic spreadsheets were carried out using MS Excel 2010.

In this way it was verified whether the abstract presents and elements of the article's structure such as: introductory section, methodological procedures results obtained sources used in the studies.

Analysis criteria: clarity, justification and consistency in the methodology used, if there is a description of the elements/procedures, statistical analysis of the data, indicators used in the study, and sample size (Table 1).

Regarding the conclusions, it was observed whether the articles present results consistent with the problem analyzed, allowing the reproduction of the experiment (Table 1).

As for the relevance of the research the approaches used and their characteristics, and the results obtained and their integration with clinical practices were analyzed. This last element seeks to understand whether there is an integration and/or understanding of these studies with the health context, also identifying the users of these technologies such as: specialized and non-specialized health professionals and the population that can use these health services.

Table 1. Article evaluation criteria:

Block 1. Evaluation of the article structure:
Is it clear about the unambiguous presentation of the problem studied?
Can the methodology used be reproduced in another experiment?
Does the work present the form of data collection?
Does the work inform the origin of the data used in the experiment? (public and/or private access)
Does the work present the size of the sample used?
Block 2. Evaluation of the technical aspects of the articles:
Does the work objectively present a justification for choosing the techniques used in the research?
What models are used in the experiments?
Which software libraries are used?
Do they present the evaluation indicators? Which?
Do you use metadata (or other types of data) in the experiment?
Block 3. Assessment of aspects of the health context:
Are the models concerned with the application of clinical practices?
Does the article inform/cite the participation of the health professional?
Does the study characterize/cite the population that gave rise to the training and testing data of the models?

Source: survey data.

4 Results and Discussion

The scientific aspects (Block 1) raised based on the structures of the articles indicate that 26 studies presented the problem studied in a clear and unambiguous way. As for the reproduction of the experiment, in 31 articles the methodology used can be reproduced in another study. The form of data collection was another criterion observed in only 25 studies and in 22 they adopted statistics to analyze the data.

Regarding the technical aspects of the articles (Block 2), 78% of the articles presented a clear justification for the choice of techniques used.

According to the bibliographic survey carried out, there are fundamental steps to build a Deep Learning model for the diagnosis of skin diseases, which are: 1) definition of the data collection (Data collections), 2) data preprocessing step (they are techniques such as: Data Segmentation, Feature Extraction, Data Augmentation and Data Aggregation), 3) training and testing steps performed in a classifier model, and 4) evaluation and results step. Visually, these steps were described in Fig. 1:

The datasets most used by the authors for training and testing were the International Skin Imaging Collaboration (ISIC) and its editions (freq. 20), followed by the dataset from the Centro de Dermatologia do Hospital Pedro Hispano (PH2) (freq. 5), both are publicly accessible (Table 2).

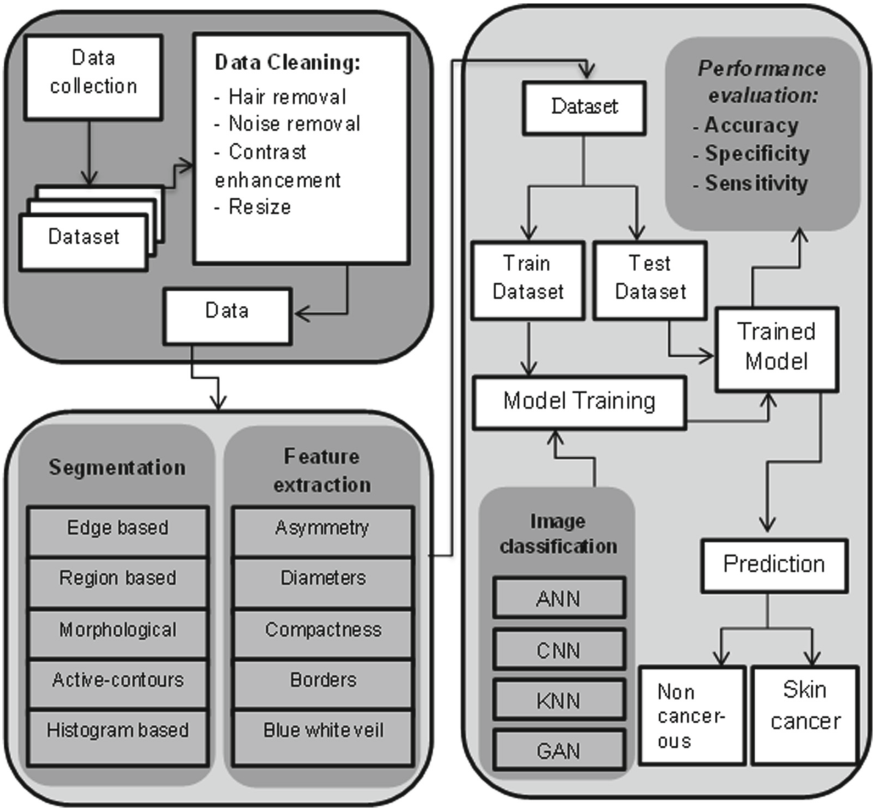

Fig. 1. The skin cancer diagnosis process. (Source: Adapted from Dildar (2021) [33]. ANN = Artificial neural network; CNN = Convolutional neural network; KNN = Kohonen self-organizing neural network; GAN = Generative adversarial neural network.)

Other data collections used by the authors were: HAM10000 (freq. 4) and with only 1 occurrence: Atlas of Dermoscopy Dataset (AtlasDerm), Dermnet Skin Disease Atlas dataset (Dermnet) and DreamQuest, all of which are publicly accessible (Table 3). The experiments that used data from public and private datasets (freq. 2) and only private (freq. 7).

As for the data collected in the studies, the research with data from the Taiwan National Health Insurance Research Database [34] stands out for the maximum amount, with 2 million images used in the experiment, while the minimum amount observed was 102 images and the average of 65,422 images. Only 2 articles did not report the amount of data collected.

As for the topics most addressed in the process of diagnosing skin diseases that use LBP, all studies had the final objective of classifying/predicting skin lesions. However, 27 studies were identified where the focus or main objective was only the classification step, aiming to promote the performance of the model.

Table 2. Studies that used the ISIC and PH2 data sources:

Authors	Article focus*	Source of data	Sample used
(Yap, Yolland, Tschand, 2018)	CLAS	ISIC	2.917
(Mohamed, Mohamed, Zekry, 2019)	CLAS	ISIC	600
(Kanani, Padole, 2019)	SEG/FEX/AUG	ISIC	10.000
(Seeja, Suresh, 2019)	SEG	ISIC	1.800
(Hekler et al., 2020)	CLAS	ISIC	1.188
(Wei, Ding, Hu, 2020)	SEG/FEX	ISIC(ISBI 2016)	1.289
(Ravikumar, Maier, 2018)	CLAS	ISIC(ISBI 2017)	2.750
(Li, Shen, 2018)	SEG/FEX/CLA	ISIC 2017	2.500
(Serte, Demirel, 2019)	CLAS/AUG	ISIC 2017	2.000
(Serte, Demirel, 2020)	CLAS/AUG	ISIC 2017	2.000
(Sherif, Mohamed, Mohra, 2019)	CLAS	ISIC 2018	700
(Almaraz-Damian et al., 2020)	FEX	ISIC 2018	10.015
(Pacheco, Krohling, 2021)	CLAS/AGG	ISIC 2019	35.867
(Molina-Molina et al., 2020)	SEG	ISIC 2019	25.331
(El-Khatib, Popescu, Ichim, 2020)	CLAS	ISIC 2019 + PH2	300
(Khan et al., 2021)	SEG/CLAS	ISIC (ISBI) + PH2	17.738
(Rodrigues et al., 2020)	FEX/CLAS	ISIC (ISBI) + PH2	6.570
(Goyal et al., 2020)	SEG	ISIC 2017 + PH2	2.950
(Tan, Zhang, Lim, 2019)	FEX	Dermofit + PH2	1.500
(Song, Lin, Wang, Wang, 2020)	CLAS/SEG	ISIC 2017 + ISBI 2016	13.750
(Abhishek, Kawahara, Hamarneh, 2021)	CLAS	ISIC (MClass-D) + Atlas	1.011
(Dascalu, David, 2019)	SEG/CLAS	ISIC + Private	5.161

Source: survey data. * Focus of articles - CLAS = Classification, SEG = Data Segmentation, FEX = Feature Extraction, AUG = Data Augmentation, AGG = Data Aggregation

Data segmentation was the subject of study in 13 articles, being also a technique widely used by the authors. The Feature Extraction task was 9 articles. In 4 studies the Data Augmentation approach was the main theme. Another technique identified Data Aggregation was also the object of study in only 1 article. All these techniques were oriented towards improving the efficiency of the models and/or the state of the art.

Other data sources that were used for training the pre-trained models cited by the authors were: 1 article used only the ImageNet dataset [24] and another study [47] that also used ImageNet with the MS COCO dataset (Microsoft Common Objects in Context).

Table 3. Studies that used other data sources

Authors	Article focus*	Source of data	Sample used
(Ameri,2020)	SEG	HAM10000	3.400
(Harangi, Baran, Hajdu, 2020)	CLAS	HAM10000	10.015
(Srinivasu et al., 2021)	CLAS	HAM10000	8.479
(Kadampur, Al Riyaee, 2020)	CLAS	HAM10000	55.043
(Olsen et al., 2018)	CLAS	Private	695
(Liu et al., 2019)	CLAS	Private	79.720
(Jinnai et al., 2020)	CLAS	Private	5.846
(Zhu et al., 2021)	CLAS	Private	13.603
(Jiang, Li, Jin, 2021)	CLAS/SEG	Private	1.167
(Wang et al., 2019)	CLAS	Private (Taiwan National Health)	2.000.000
(Thomsen et al., 2020)	CLAS /FEX /AUG	Private (AUH Dinamarca)	16.543
(Burlina et al., 2019)	CLAS	Private + Public	1.834
(Premaladha, Ravichandran, 2016)	SEG/FEX	SkinCancerAtlas + Dermnet	992
(Jafari, Mohammad et al., 2016)	SEG	Dermquest	126
(Bhavani, R. et al., 2019)	FEX	N/A	102

Source: survey data. * Focus of articles - CLAS = Classification, SEG = Data Segmentation, FEX = Feature Extraction, AUG = Data Augmentation, AGG = Data Aggregation, N/A = Not Presented by Author

CNN models are the most used and were identified in all articles in the portfolio, other models combined with CNN were also observed, such as: LSTM - Long Short Term Memory (freq. 1), Deconvolutional Network (freq. 1) and FCN - Fully Connected Network (freq. 1).

The Transfer Learning technique was observed in 25 articles (Tables 4 and 5) of which the most used models were based on ResNet [60] identified in 9 studies, followed by the Inceptions class [60] in 6 articles. The VGG [61] and U-Net [62] models are present in 3 researches each and the models that appeared in only 2 articles were FCRN [10], DenseNet [63] and MobileNet [64].

The most used software libraries are: Keras/TensorFlow (freq. 12), Pytorch (Table 4) and Caffe DL (Table 5). In 14 studies, the authors did not inform the libraries used to build the DL models. The most used development environments for implementation are: Python, followed by MatLab (freq. 8), and with only 1 article each, they used R Language and Deep Learning Studio (DLS). In 11 studies the development environment was not reported.

Table 4. Python software libraries:

Software libraries	Articles	Models
Keras and TensorFlow	(Li & Shen, 2018)	FCRN-88
	(Kanani & Padole, 2019)	CNN
	(Wei, Ding & Hu, 2020)	U-Net
	(Abhishek, Kawahara & Hamarneh, 2021)	InceptionV3
	(Liu et al., 2019)	Inception V4
	(Goyal et al., 2020)	CNN + ENC_DEC*
	(Thomsen et al., 2020)	VGG-16
	(Song et al., 2020)	ResNet50/ResNet101
	(Burlina et al., 2019)	ResNet50
	(Rodrigues et al., 2020)	VGG/Inception/ResNet/MobileNet/DenseNet
Scikit-learn and Scipy	(Hekler et al., 2020)	ResNet 50
Keras and scikit-learn	(Almaraz-Damian et al., 2020)	CNN
Chainer, ChainerCV and Cupy	(Jinnai et al., 2020)	FRCNN
OpenCV and Caffe DL	(Serte & Demirel, 2020)	ResNet 18 ResNet 50
PyTorch	(Srinivasu et al., 2021)	MobileNet V2
	(Zhu et al., 2021)	CNN
	(Pacheco & Krohling, 2021)	CNN

Source: survey data. * Model Encoder - Decoder.

In the classification step, and in addition to the Multilayer Perceptron (MLP), other most used algorithms were: Support Vector Machines (SVMs), Naïve Bayes, Extreme Learning Machine, multi-class SVM (MSVM), Random Forest (RF) and K-Nearest Neighbor. (KNN).

The evaluation indicators most used to evaluate models were: Accuracy (33), Sensitivity (28), Specificity (25), ROC Area under the curve (14). Other indicators observed in the studies with less frequency but of great importance: Dice coefficient (19%), Jaccard Index (16%), Precision (13%), F-Score (10%), MCC - Matthew Correlation Coefficient (8%), Kappa (5%), in only one F2-Score and Recall study. In two articles, the indicators of the evaluation stage were not presented.

The use of metadata and other types of data such as clinical reports for training and testing the models were used in only 7 articles, revealing that multimodal models are still approaches that are still little used by these researchers.

Regarding aspects of the health context (Block 3), almost half of the studies (18) analyzed did not have the proper concern with the improvement of techniques and assertiveness of the models, acting restricted to the experimental field without addressing the application context presenting distance from the clinical aspect (Table 6).

As for the inclusion of health professionals (specialized, residents, non-specialized) in the research and in addition to the practical application in LD, they were identified in six articles that, in the evaluation stage, compare results between man and machine.

Studies that characterize or cite the population that gave rise to training and testing data were identified in only 2 studies. Characteristics such as ethnicity/race were presented: 100% Caucasian [16] and 100% Asian [34].

This point has great relevance, due to the fact that many articles of this research revealed that the skin tone would be one of the biggest obstacles for the development of LD models for skin diagnosis by images. In this way, when analyzing the collections of images informed by the authors, it is observed the absence of images of skin color that are not Caucasian or Asian, as well as the origins of these collections are predominantly from countries in the northern hemisphere.

In order to solve this gap, only 1 study was found in the literature, proposed by Minagawa et al. [65] a Deep Learning model that, despite mentioning the effectiveness, accuracy and sensitivity of the proposed model, did not present specific results for each participating population of the study. Thus, the verification of the effectiveness of this proposal for ethnic/racial diversity in the diagnosis of skin can result in ambiguities.

5 Final Considerations

The objective of this study is to analyze the characteristics and use of Deep Learning (DL) models for the diagnosis of skin diseases and their integration with clinical practices. This research carried out a literature review of a qualitative nature. The primary data collected in the databases of scientific articles which resulted in 37 articles from the bases: Scopus, Web of Science, PubMed and IEEE Xplore.

The scientific aspects raised based on the structures of the articles indicate that 26 studies presented the problem studied in a clear and unambiguous way, in 31 experiments the methodology adopted can be replicated in another study and 29 articles presented a justification of the approaches used.

The data collection that stood out was ISIC - International Skin Imaging Collaboration. The most commonly used data types in these models are images. The most used model is the Convolutional Neural Network (CNN) and the most used pre-trained model was ResNet.

The most used techniques in the articles focused on the prediction of skin diseases, data segmentation and feature extraction. The assessment indicators that stand out are accuracy, sensitivity and specificity.

Integration with clinical practices was another theme analyzed in this research, a little less than half of the studies did not have any direction or mention the application in the

Table 5. Other software libraries used:

Software	Libraries	Articles	Models
MATLAB (MathWorks Inc.)	Caffe DL	(Jafari et al., 2016)	CNN
		(Olsen et al., 2018)	VGG
	Patternnet	(El-Khatib, Popescu & Ichim, 2020)	ResNet 101
	SGDM algorithm	(Ameri, 2020)	CNN
	N/A	(Sherif, Mohamed & Mohra, 2019)	CNN
		(Tan, Zhang & Lim, 2019)	Ensemble
		(Molina-Molina et al., 2020)	Densenet 201
		(Khan et al., 2021)	ResNet101 and DenseNet 201
R Language (3.4.4)	Keras/TensorFlow	(Wang et al., 2019)	CNN
Deep Learning Studio (DLS)	Model Driven Architecture Tool	(Kadampur & Al Riyaee, 2020)	ResNet 50, Densenet and Inception V3
N/A	N/A	(Premaladha & Ravichandran, 2016)	Ensemble
		(Ravikumar & Maier, 2018)	U-Net
		(Yap, Yolland & Tschand, 2018)	ResNet50
		(Serte & Demirel, 2019)	CNN
		(Mohamed, Mohamed & Zekry, 2019)	CNN
		(Bhavani et al., 2019)	Inception V3
		(Seeja & Suresh, 2019)	U-Net
		(Dascalu & David, 2019)	Inception V2
		(Harangi, Baran & Hajdu, 2020)	Inception V3
		(Jiang, Li & Jin, 2021)	CNN

Source: survey data. N/A - not informed by the authors.

clinical context being limited only to the development and improvement of algorithms. Only in 6 studies were health professionals participating in the experiment. Another

Table 6. Indicators for analysis of clinical practices:

Article	Was the study aimed at application in clinical practice?		Did any health professional participate in the study?		Does the study report patient characteristics?	
	YEA	NO	YEA	NO	YEA	NO
(Abhishek, Kawahara, Hamarneh, 2021)	√		√			√
(Almaraz-Damian et al., 2020)		√		√		√
(Ameri,2020)		√		√		√
(Bhavani, R. et al., 2019)	√			√		√
(Burlina et al., 2019)	√		√			√
(Dascalu, David, 2019)	√			√	√	
(El-Khatib, Popescu, Ichim, 2020)		√		√		√
(Goyal et al., 2020)	√			√		√
(Harangi, Baran, Hajdu, 2020)		√		√		√
(Hekler et al., 2020)		√	√			√
(Jafari, Mohammad et al., 2016)		√		√		√
(Jiang, Li, Jin, 2021)	√			√		√
(Jinnai et al., 2020)	√		√			√
(Kadampur, Al Riyaee, 2020)	√			√		√
(Kanani, Padole, 2019)		√		√		√
(Khan et al., 2021)		√		√		√
(Li, Shen, 2018)		√		√		√
(Liu et al., 2019)	√		√			√
(Mohamed, Mohamed, Zekry, 2019)	√			√		√
(Molina-Molina et al., 2020)	√			√		√
(Olsen, Thomas et al., 2018)	√			√		√
(Pacheco, Krohling, 2021)		√		√		√
(Premaladha, Ravichandran, 2016)	√			√		√
(Ravikumar, Maier, 2018)		√		√		√
(Rodrigues et al., 2020)		√		√		√
(Seeja, Suresh, 2019)		√		√		√
(Serte, Demirel, 2019)		√		√		√
(Serte, Demirel, 2020)		√		√		√
(Sherif, Mohamed, Mohra, 2019)		√		√		√
(Song, Lin, Wang, Wang, 2020)		√		√		√
(Srinivasu et al., 2021)	√			√		√
(Tan, Zhang, Lim, 2019)		√		√		√
(Thomsen et al., 2020)	√			√		√
(Wang et al., 2019)	√			√	√	
(Wei, Ding, Hu, 2020)	√			√		√
(Yap, Yolland, Tschand, 2018)	√			√		√
(Zhu et al., 2021)	√		√			√
Total (37)	19	18	6	31	2	35
(%)	51,3%	48,6%	16,2%	83,7%	5,4%	94,5%

Source: survey data.

important factor, identified in only 2 articles, was the characteristics or information of the patients.

Finally, the analyzed literature indicated that approaches that use DL to classify skin diseases are very promising, however, most studies using these technologies reveal the need for greater interaction with clinical practices.

As a suggestion for future work, studies that address the effectiveness of DL models for the diagnosis of skin lesions in different ethnic groups and their limitations for the democratization of such technologies. The aspects raised for analysis of integration with clinical practices can be expanded in greater depth on the subject.

As limitations of this research, only articles in the English language were collected and the criteria raised in this study are subject to interpretation and subjectivity, which may not converge with the opinion of the authors of the studies surveyed.

References

1. Elaziz, M.A., Hosny, K.M., Salah, A., Darwish, M.M., Lu, S., Sahlol, A.T.: New machine learning method for image-based diagnosis of COVID-19. PLoS ONE **15**(6), e0235187 (2020). https://doi.org/10.1371/journal.pone.0235187
2. Marques, G., Agarwal, D., de la Torre Díez, I.: Automated medical diagnosis of COVID-19 through efficientnet convolutional neural network. Appl. Soft Comput. J. **96** (2020). https://doi.org/10.1016/j.asoc.2020.106691
3. Su, C., Xu, Z., Pathak, J., et al.: Deep learning in mental health outcome research: a scoping review. Transl. Psychiatry **10**, 116 (2020). https://doi.org/10.1038/s41398-020-0780-3
4. Jaiswal, A.K., Tiwari, P., Kumar, S., Al-Rakhami, M.S., Alrashoud, M., Ghoneim, A.: Deep learning-based smart iot health system for blindness detection using retina images. IEEE Access **9**, 70606–70615 (2021). https://doi.org/10.1109/ACCESS.2021.3078241
5. Kuwahara, T., et al.: Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas. Clin. Transl. Gastroenterol. **10**(5) (2019). https://doi.org/10.14309/ctg.0000000000000045
6. Li, Y., Shen, L.: Deep learning based multimodal brain tumor diagnosis. In: Crimi, Alessandro, Bakas, Spyridon, Kuijf, Hugo, Menze, Bjoern, Reyes, Mauricio (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 149–158. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_13
7. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056
8. Majtner, T., Yildirim-Yayilgan, S., Hardeberg, J.Y.: Optimised deep learning features for improved melanoma detection. Multimed. Tools Appl. **78**(9), 11883–11903 (2018). https://doi.org/10.1007/s11042-018-6734-6
9. Rodrigues, D.D.A., Ivo, R.F., Satapathy, S.C., Wang, S., Hemanth, J., Filho, P.P.R.: A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system. Pattern Recogn. Lett. **136**, 8–15 (2020). https://doi.org/10.1016/j.patrec.2020.05.019
10. Li, Y., Shen, L.: Skin lesion analysis towards melanoma detection using deep learning network. Sensors **18**(2), 556 (2018). https://doi.org/10.3390/s18020556
11. Tan, T.Y., Zhang, L., Lim, C.P.: Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models. Appl. Soft Comput. **84**, 105725 (2019)
12. Serte, S., Demirel, H.: Gabor wavelet-based deep learning for skin lesion classification. Comput. Biol. Med. **113**, 103423 (2019)

13. Pollastri, F., Bolelli, F., Paredes, R., Grana, C.: Augmenting data with GANs to segment melanoma skin lesions. Multimed. Tools Appl. **79**(21–22), 15575–15592 (2019). https://doi.org/10.1007/s11042-019-7717-y
14. Serte, S., Demirel, H.: Wavelet-based deep learning for skin lesion classification. IET Image Process. **14**(4), 720–726 (2020)
15. Abhishek, K., Kawahara, J., Hamarneh, G.: Predicting the clinical management of skin lesions using deep learning. Sci. Rep. **11**(1), 1–14 (2021). https://doaj.org/article/7db5a0869123411895ea4b08b90c398b
16. Dascalu, A., David, E.O.: Skin cancer detection by deep learning and sound analysis algorithms: a prospective clinical study of an elementary dermoscopeResearch in context. EBioMedicine **43**, 107–113 (2019). https://doaj.org/article/ee249e1460e249f2965a73b06db9683a
17. Khan, M.A., Sharif, M., Akram, T., Damaševičius, R., Maskeliūnas, R., Ellinger, I.: Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. Diagnostics **11**(5) (2021). https://www.mdpi.com/2075-4418/11/5/811/htm
18. Srinivasu, P.N., Sivasai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J.: Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors (Basel, Switzerland) **21**(8), 2852 (2021). https://doaj.org/article/4b4420e8f6dc4239a3ee2e97ff34b3a3
19. Greenhalgh, T., et al.: Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care. Technol. J. Med. Internet Res. **19**(11), e367 (2017). https://doi.org/10.2196/jmir.8775
20. Coccia, M.: Deep learning technology for improving cancer care in society: new directions in cancer imaging driven by artificial intelligence. Technol. Soc. **60**, 1–11(2020). https://ssrn.com/abstract=3493655
21. Shrivastava, V.K., Londhe, N.D., Sonawane, R.S., Suri, J.S.: Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind. Comput. Meth. Programs Biomed. **126**, 98–109 (2016). https://doi.org/10.1016/j.cmpb.2015.11.013
22. Liu, Y., et al.: A deep learning system for differential diagnosis of skin diseases. Nature Med. **26**(6), 900–908 (2019). https://doi.org/10.1038/s41591-020-0842-3
23. Pacheco, A.G.C., Krohling, R.: An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. IEEE J. Biomed. Health Inf. **25**(9), 3554–3563 (2021)
24. Thomsen, K., Christensen, A.L., Iversen, L., Lomholt, H.B., Winther, O.: Deep learning for diagnostic binary classification of multiple-lesion skin diseases. Front. Med. **7**(September), 1–7 (2020). https://doi.org/10.3389/fmed.2020.574329
25. Shoieb, D.A., Youssef, S.M., Aly, W.M.: Computer-aided model for skin diagnosis using deep learning. J. Image Graph. **4**(2), 122–129 (2016). https://doi.org/10.18178/joig.4.2.122-129
26. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, … Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. Aug 1,29(8):1836–1842 (2018).https://doi.org/10.1093/annonc/mdy166. PMID: 29846502
27. Patnaik, S.K., Sidhu, M.S., Gehlot, Y., Sharma, B., Muthu, P.: Automated skin disease identification using deep learning algorithm. Biomed. Pharmacol. J. **11**(3), 1429 (2018)
28. Sarker, M.M.K., et al.: MobileGAN: skin lesion segmentation using a lightweight generative adversarial network (2019)

29. Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E.: RECOD titans at ISIC challenge 2017. In: International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI) (2017). https://arxiv.org/pdf/1703.04819.pdf
30. Codella, N., et al.: Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC) (2018). https://arxiv.org/abs/1902.03368
31. Tschandl, P., et al.: Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol. (2019).https://doi.org/10.1016/s1470-2045(19)30333-x
32. Polat, K., Koc, K.O.: Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. J. Artif. Intell. Syst. **2**(1), 80–97 (2020). https://doi.org/10.33969/AIS.2020.21006
33. Dildar, M., et al.: Skin cancer detection: a review using deep learning techniques. Int. J. Environ. Res. Public Health **18**, 5479 (2021). https://doi.org/10.3390/ijerph18105479
34. Wang, H., Wang, Y., Liang, C., Li, Y.: Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. JAMA Dermatol. **155**(11), 1277–1283 (2019). https://doi.org/10.1001/jamadermatol.2019.2335[PB]
35. Yap, J., Yolland, W., Tschandl, P.: Multimodal skin lesion classification using deep learning. Exp. Dermatol. **27**(11), 1261–1267 (2018)
36. Mohamed, A., Mohamed, W., Zekry, A.H.: Deep learning can improve early skin cancer detection. Int. J. Electron. Telecommun. **65**(3), 507–513 (2019). http://ijet.pl/index.php/ijet/article/download/10.24425-ijet.2019.129806/600
37. Kanani, P., Padole, M.: Deep learning to detect skin cancer using google colab. Int. J. Eng. Adv. Technol. **8**(6), 2176–2183 (2019). https://www.ijeat.org/wp-content/uploads/papers/v8i6/F8587088619.pdf
38. Seeja, R.D., Suresh, A.: Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). Asian Pac. J. Cancer Prev. **20**(5), 1555–1561 (2019). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6857898/pdf/APJCP-20-1555.pdf
39. Hekler, A., et al.: Effects of label noise on deep learning-based skin cancer classification. Front. Med. **7**, 117 (2020). https://doaj.org/article/7c8a3d919b4d4e7da94887fa5381fe8d
40. Wei, L., Ding, K., Hu, H.: Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE Access **8**, 99633–99647 (2020). https://doi.org/10.1109/ACCESS.2020.2997710[PB]
41. Ravikumar, N., Maier, A.: SkinNet: a deep learning framework for skin lesion segmentation (2018). ArXiv.Org. http://search.proquest.com/docview/2074059294/
42. Sherif, F., Mohamed, W.A., Mohra, A.S.: Skin lesion analysis toward melanoma detection using deep learning techniques. Int. J. Electron. Telecommun. **65**(4), 597–602 (2019). http://ijet.pl/index.php/ijet/article/view/10.24425-ijet.2019.129818%0A
43. Almaraz-Damian, J.A., Ponomaryov, V., Sadovnychiy, S., Castillejos-Fernandez, H.: Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. Entropy **22**(4), 484 (2020). https://doi.org/10.3390/e22040484
44. Molina-Molina, E.O., Solorza-Calderón, S., Álvarez-Borrego, J.: Classification of dermoscopy skin lesion color-images using fractal-deep learning features. Appl. Sci. **10**(5954), 5954 (2020). https://doi.org/10.3390/app10175954
45. El-Khatib, H., Popescu, D., Ichim, L.: Deep learning-based methods for automatic diagnosis of skin lesions. Sensors **20**(6), 1753 (2020). https://doi.org/10.3390/s20061753
46. Goyal, M., Oakley, A., Bansal, P., Dancey, D., Yap, M.H.: Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. IEEE Access **8**, 4171–4181 (2020)

47. Song, L., Lin, J., Wang, Z.J., Wang, H.: An end-to-end multi-task deep learning framework for skin lesion analysis. IEEE J. Biomed. Health Inf. **24**(10), 2912–2921 (2020)
48. Ameri, A.: A deep learning approach to skin cancer detection in dermoscopy images. J. Biomed. Phys. Eng. **10**(6), 801–806 (2020). https://doaj.org/article/dd64ab7df19b42938fe73a2ba2ebdd78
49. Harangi, B., Baran, A., Hajdu, A.: Assisted deep learning framework for multi-class skin lesion classification considering a binary classification support. Biomed. Signal Process. Control **62** (2020)
50. Kadampur, M.A., Al Riyaee, S.: Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Inf. Med. Unlocked **18** (2020). https://doi.org/10.1016/j.imu.2019.100282
51. Olsen, T., et al.: Diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology (Original Article). J. Pathol. Inf. **9**(1), 32 (2018)
52. Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., Hamamoto, R.: The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules **10**(8) (2020)
53. Zhu, C.-Y., et al.: A deep learning based framework for diagnosing multiple skin diseases in a clinical environment. Front. Med. **8** (2021). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8085301/
54. Jiang, S., Li, H., Jin, Z.A.: Visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis. IEEE J. Biomed. Health Inf. **25**(5), 1483–1494 (2021)
55. Burlina, P.M., Joshi, N.J., Ng, E., Billings, S.D., Rebman, A.W., Aucott, J.N.: Automated detection of erythema migrans and other confounding skin lesions via deep learning. Comput. Biol. Med. **105**, 151–156 (2019). https://doi.org/10.1016/j.compbiomed.2018.12.007[PB]
56. Premaladha, J., Ravichandran, K.S.: Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. J. Med. Syst. **40**(4), 1–12 (2016). https://doi.org/10.1007/s10916-016-0460-2
57. Jafari, M., Nasr-Esfahani, E., Karimi, N., Soroushmehr, S., Samavi, S., Najarian, K.: Extraction of skin lesions from non-dermoscopic images using deep learning. ArXiv.Orghttp://search.proquest.com/docview/2074987175/
58. Bhavani, R., Prakash, V., Kumaresh, R.V, Sundra Srinivasan, R.: Vision-based skin disease identification using deep learning. Int. J. Eng. Adv. Technol. **8**(6), 3784–3788 (2019). https://www.ijeat.org/wp-content/uploads/papers/v8i6/F9391088619.pdf
59. He, K., Zhang, X., Ren, S., Sun, J.: ResNet50 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
60. Zegedy, C., Ioffe, S., Vanhoucke, V. Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence, pp. 4278–4284 (2017)
61. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:1409.1556
62. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
63. Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K.: Densenet: implementing efficient convnet descriptor pyramids (2014). arXiv preprint arXiv:1404.1869

64. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications (2017). arXiv preprint arXiv:1704.04861
65. Minagawa, A., et al.: Dermoscopic diagnostic performance of Japanese dermatologists for skin tumors differs by patient origin: a deep learning convolutional neural network closes the gap. J. Dermatol. (2020). https://doi.org/10.1111/1346-8138.15640