Skin Type Diversity: a Case Study in Skin Lesion Datasets

Neda Alipour (D21124371@mytudublin.ie)
Technological University Dublin
Ted Burke
Technological University Dublin
Jane Courtney
Technological University Dublin

Research Article

Keywords: skin lesion datasets, race bias, Fitzpatrick skin type, skin cancer detection

Posted Date: July 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3160120/v1

Abstract

Inadequate skin type diversity, leading to racial bias, is a widespread problem in datasets involving human skin. For example, skin lesion datasets used for training deep learning-based models can lead to low accuracy for darker skin types, which are typically under-represented in these datasets. This issue has been discussed in previous works; however,skin type diversity of datasets and reporting of skin types have not been fully assessed. Frequently, ethnicity is used instead of skin type, but ethnicity and skin type are not the same, as many ethnicities can have diverse skin types. Some works define skin types, but do not attempt to assess skin type diversity in datasets. Others, focusing on skin lesions, identify the issue, but also do not measure skin type diversity in the datasets examined. Building on previous works in the area of skin lesion datasets, this review explores the general issue of skin type diversity in datasets by investigating and evaluating skin lesion datasets specifically. The main contributions of this work are: an evaluation of all publicly available skin lesion datasets and their metadata to assess frequency and completeness of reporting of skin type and an investigation into the diversity and representation of specific skin types within these datasets.

1 Introduction

Diversity is an important feature in datasets used for training artificial intelligence (AI) based models, as the performance of AI is only as good as its data. Although AI has brought many advantages to our daily lives, when it comes to human skin, the findings show racial bias and low accuracy for dark-skinned people (Kostick-Quenet et al. 2022; Obermeyer et al. 2019). This can potentially lead to the exclusion of this group by AI-based models. The effect of inadequate skin type diversity and under-representation of dark-skinned people in datasets can be seen in many AI-based technologies. For example, AI systems that judge beauty pageant winners are biased against darker-skinned contestants (Fuchs 2018). In a beauty contest run by Beauty.ai, of the 44 finalists judged by the algorithms as the most attractive, except for six who were described as “Asian”, all were described as “white”. Only one finalist was dark-skinned (Khalil et al. 2020; Jordan 2016).

Another study investigating performance of object detection systems on pedestrians with different skin types showed higher precision on lighter skin types than darker skin types (Wilson et al. 2019). In other work, bias in face verification applications and datasets was evaluated with respect to different skin types and found that recognition accuracy was reduced for darker-skinned people (Lu et al. 2019). Howard and Borenstein (2018) assessed the effect of racial bias on the performance of robotic systems such as a robot peacekeeper, a self-driving car, and a medical robot. Their work shows how bias has been infused into current AI and robotic systems.

In AI in healthcare sectors, there are consumer wearable devices that are used for tracking activity, sleep, and other health-related purposes, but due to some limitations, these health products may only be useful for light-skinned people. Findings show that these devices are inaccurate, and even may not work at all for dark-skinned people (Shcherbina et al. 2017; Fallow et al. 2013). In another example, a recent skin cancer detection study trained a model with a dataset where only 5% of images were of dark-skinned individuals (Zou and Schiebinger 2018).

All mentioned examples show that the needs of some population groups are not well-represented (Myers West 2020), which can potentially lead to exclusion of, or reduced accuracy for, these groups by deep learning-based

models. A number of factors play a role in biased performance of these models towards dark-skinned people. Most significant among them is the lack of skin type diversity in datasets used for training AI-based models (Kamulegeya et al. 2019; Wen et al. 2022).

There are many reasons for not having enough data from dark-skinned people in datasets used for AI applications. For example, in the case of skin lesion datasets, reasons include: low incidence of skin cancer in dark-skinned people (Diepgen and Mahler 2002; Gloster Jr and Neal 2006), unequal access to healthcare (Hudson et al. 2015), poor quality images due to poor quality of care (Betancourt et al. 2019; FitzGerald and Hurst 2017) and racial bias encoded in the algorithms used in digital cameras as well as computer software (Benjamin 2020; Kraehe and Herman 2020). Due to these reasons, dark-skinned people are under-represented in datasets from health services as well as research datasets (Hudson et al. 2015). Deep learning-based models trained on lighter-skinned subjects are at risk of poor performance for people with darker skin (Marcus and Davis 2019).

Due to the problems mentioned above, it is necessary to evaluate skin type diversity to detect under-representation in datasets before using them for training AI systems. Doing this helps to prevent models having bias toward specific groups of people. The Fitzpatrick scale provides a skin tone classification based on reaction to exposure to sunlight (Fitzpatrick 1997). It is used in dermatology to classify skin tones into six numbered categories as shown in Fig. 1. The Fitzpatrick skin type scale has previously been used to evaluate skin type diversity in datasets (Groh et al. 2021).

Although the issue of inadequate skin type diversity has been discussed in previous works, these have not attempted to evaluate skin type diversity for datasets. For instance, in the Gender Shades study (Buolamwini and Gebru 2018), the Fitzpatrick scale is used to evaluate the PPB, IJB-A and Adience datasets. However, rather than measuring skin type diversity over six separate Fitzpatrick skin type categories, the authors instead classify the images in these datasets using two aggregate groups - darker and lighter.

To mitigate discrimination toward certain groups of people, Karkkainen and Joo (2021) created the FairFace dataset, which is a balanced face image dataset for seven race groups that provides more accurate and consistent modeling across different race and gender groups. However, they focus on ethnic diversity and do not report skin type diversity. McDuff et al. (2018) proposed a new method using computer simulations to detect biases in face detection using Bayesian parameter search in high dimensional feature space. Although they consider the Fitzpatrick scale for identification of demographic biases in commercial face application programming interfaces (APIs), they do not measure skin type diversity.

Xu et al. (2022) introduced a new method for human skin detection, not using colour information, but rather using a U-Net-based segmentation network. Their method was tested on two datasets containing face images: ECU (Edith Cowan University) and RFW (Racial Faces in the Wild). ECU is an imbalanced dataset created based on six different Fitzpatrick skin types and RFW is a balanced dataset with only the annotation of ethnicity, based on four test subsets: “Caucasian”, “Asian”, “Indian”, and “African”. In the case of the RFW dataset, it is not evaluated based on Fitzpatrick skin type, but just based on ethnicity.

Porgali et al. (2023) created Casual Conversations, which is a fair and diverse dataset of videos collected from seven countries for AI applications, and labelled the dataset based on the two skin tone scales of Monk (Monk

Jr 2014) and Fitzpatrick (Fitzpatrick 1997). Nonetheless, the authors do not report any measurement of skin type diversity for their dataset. Daneshjou et al. (2022) created the SkinCon dataset for training models related to skin diseases, which contains labels for different skin types. This dataset was constructed from two skin disease image datasets: Fitzpatrick 17k (Groh et al. 2021) and Diverse Dermatology Images (DDI) (Daneshjou et al. 2022). Although the Fitzpatrick skin type scale is mentioned in this work, no measurement of skin type diversity is presented.

Wen et al. (2022) assessed skin lesion image datasets for diversity based on their metadata including age, gender, ethnicity, and skin type. The authors mentioned that there is limited reporting on skin type in the metadata and also less representation of darker-skinned people in skin lesion datasets. However, they did not measure skin type diversity in any of the skin lesion datasets.

In order to measure skin type diversity and detect bias in datasets used for training deep learning-based models, Fitzpatrick skin type metadata should be included in the datasets. Accessing this information is a crucial step to not only detect under-representation in datasets, but also help to avoid training models on biased datasets, and as a result prevent models performing poorly for specific groups of people. According to our investigation, three available skin lesion datasets provide Fitzpatrick scale skin type metadata, labelled by dermatologists: PAD-UFES-20 (Pacheco et al. 2020), Fitzpatrick 17k (Groh et al. 2021), and DDI (Daneshjou et al. 2022). To investigate the issue of inadequate skin type diversity in datasets used for training deep learning models, just two datasets - PAD-UFES-20 and Fitzpatrick 17k - are utilized as examples in this review. DDI was not used because it is a balanced dataset (albeit for three aggregate skin type groups, rather than for all six Fitzpatrick skin types). Sample images from the PAD-UFES-20 and Fitzpatrick 17k datasets are shown in Fig. 2 and Fig. 3 respectively.

Investigation of metadata in these two datasets is helpful to assess skin type diversity and check to what extent the lack of diversity in the datasets potentially leads to discrimination by models trained on the datasets. The main contributions of this study are an investigation into reporting of skin type information in available skin lesions datasets, a significant extension of the work by Wen et al. (2022), and an investigation into the diversity and representation of specific skin types within these datasets. Previous similar work by Daneshjou et al. (2021) discussed the lack of transparency in medical skin datasets and the necessity of demographic descriptions such as ethnicity and Fitzpatrick skin type for further analysis and deep learning applications, but those authors only assessed the metadata of a small group of skin lesion datasets, such as: number of images, type of skin disease, Fitzpatrick skin type, and ethnicity. To our knowledge, this is the first time that a comprehensive investigation of reporting of skin type in publicly available skin lesion datasets has been provided. Furthermore, for those datasets that report skin type, an evaluation of skin type diversity is presented. Our results indicate that there is low diversity in these skin lesion datasets. The distinction between ethnicity and skin type is also discussed in the results section below. This review emphasizes the danger of implementing algorithms on datasets without transparency and not having diversity, as this issue can result in inequalities (Musselwhite et al. 2016; Williams and Cooper 2019; Yu et al. 2016).

2 Methodology

The selection process used in our review to identify papers that used publicly available skin lesion datasets was based on the PRISMA statement (Page 2021). The databases of PubMed, Elsevier, Springer, Google Scholar, and IEEE Xplore were searched. In our initial search, the following search terms were used: “skin cancer detection”, “skin lesion segmentation”, “skin lesion datasets”, “Fitzpatrick skin lesion”, “Fitzpatrick skin type metadata skin lesion”, “Fitzpatrick skin typology angle”, and “race bias skin lesion images” to identify papers on skin type diversity that make use of skin lesion image datasets. Table 1 provides a summary of which datasets are used in each of the selected papers. Section 3 includes a review of a subset of the identified datasets that match the following criteria: gender, age, ethnicity, and skin type.

3 Results

Our initial search (using the search terms listed in Section 2) returned over 1,300 publications as shown in Fig. 4. In the first screening, more than 700 duplicate papers were eliminated, leaving 577 papers to be assessed. In the second step, a further 420 publications were excluded due to lack of relevance (did not use skin lesion datasets), or being unavailable (including those not accessible without payment in Technological University Dublin), leaving 157 papers to be assessed for eligibility. Of these, 36 were excluded due to not being peer reviewed. Ultimately, 121 publications were included in the systematic review.

The 121 papers identified from the search process made use of one or more publicly available skin lesion datasets. Table 1 shows a subset of these papers[1].

Table 1 A subset of papers identified through the PRISMA process that used publicly available skin lesion datasets. We have attempted to select a subset that spans the majority of the skin lesion datasets used in the full list of identified papers.

Author	Year	Dataset
(Mendonça et al. 2013)	2013	PH2
(Saez et al. 2014)	2014	Interactive Atlas of Dermoscopy
(Giotis et al. 2015a)	2015	MED-NODE
(Sun et al. 2016)	2016	SD-198
(Sun et al. 2016)	2016	SD-128
(Liao et al. 2016)	2016	AtlasDerm / Danderm / DermIS / Dermnet / Derma / DermQuest (Derm101)
(Liao H 2016)	2016	Dermnet / OLE
(Kawahara et al. 2016)	2016	Dermofit Image Library
(Ge et al. 2017)	2017	MoleMap / ISBI-2016
(Lopez et al. 2017)	2017	Dermofit Image Library / Dermnet / ISBI 2016 Challenge
(Gu et al. 2017)	2017	Edinburgh Dermofit Image Library / PH2
(Kawahara et al. 2018)	2018	7-point checklist
(Han et al. 2018)	2018	Asan Dataset / MED-NODE
(Khan et al. 2018)	2018	PH2 / ISIC / ISBI 2017
(Gutman et al. 2016)	2018	ISIC-MSK-2
(Han et al. 2018)	2018	Edinburgh Dermofit Image Library / Hallym
(Tschandl et al. 2018)	2018	HAM10000
(Luo and Yang 2018)	2018	DermQuest
(Shoieb and Youssef 2018)	2018	DermQuest / MED-NODE / DermIS
(França 2018)	2018	ISIC
(Goyal et al. 2018)	2018	ISBI 2017 / PH2 / HAM10000
(Mendes and da Silva 2018)	2018	MED-NODE / Atlas / Edinburgh
(Gonzalez-Diaz 2018)	2018	2017 ISBI challenge / EDRA / ISIC Archive
(Pham et al. 2018)	2018	ISBI challenge / PH2 / ISIC Archive / ISIC 2017 challenge
(Yang et al. 2019)	2019	SD-198 / SD-260
(Pacheco et al. 2019)	2019	ISIC challenge 2019
(Brinker TJ et al. 2019)	2019	MClass-D
(Brinker TJ et al. 2019)	2019	MClass-ND
(Ünver and Ayan 2019)	2019	PH2 / ISBI 2017

(Combalia et al. 2019)	2019	BCN20000
(Xie et al. 2019)	2019	XiangyaDerm
(Sae-Lim et al. 2019)	2019	HAM10000
(He et al. 2019)	2019	Skin-10 / skin-100
(Jiang et al. 2019)	2019	ISIC 2017
(Nunnari and Sonntag 2019)	2019	ISIC 2019
(Aldwgeri and Abubacker 2019)	2019	ISBI 2017
(Bisla et al. 2019)	2019	ISIC 2017 challenge / PH2 / Edinburgh
(Gu et al. 2019)	2019	HAM10000 / MoleMap
(Tan et al. 2019)	2019	Dermofit Image Library / PH2 / ISIC 2017
(Goyal et al. 2020)	2020	Dermnet NZ / Derm7pt / The Cancer Genome Atlas / Hallym /
(Lucius et al. 2020)	2020	ISIC archive
(Banerjee et al. 2020)	2020	PH2 / ISIC 2019 / ISBI 2017
(Pacheco et al. 2020)	2020	PAD-UFES-20
(Han et al. 2020)	2020	SNU / Edinburgh
(Nedelcu et al. 2020)	2020	7-point checklist (EDRA)
(Han et al. 2020)	2020	Normal / Web
(Milantev et al. 2020)	2020	SD-198 / MED-NODE / PH2 / SKINL2v2 / Seven-Point / Light Field Image
(Andrade et al. 2020)	2020	SMARTSKINS / Dermofit Image Library
(Akram et al. 2020)	2020	PH2 / ISIC MSK / ISIC UDA / ISBI-2017
(Pour and Seker 2020)	2020	ISBI 2016, 2017
(Mahajan et al. 2020)	2020	ISIC 2018 / Derm7pt / SD-198
(Valle et al. 2020)	2020	ISIC Challenge 2017 / ISIC Archive / PH2 / EDRA Interactive Atlas of Dermoscopy
(Bissoto et al. 2020)	2020	ISIC Archive / Interactive Atlas of Dermoscopy
(Hosny et al. 2020)	2020	MED-NODE / Derm (IS & Quest) / ISIC
(Ashraf et al. 2020)	2020	DermIS / DermQuest
(Fisher et al. 2020)	2020	Edinburgh Dermofit
(Waweru et al. 2020)	2020	HAM10000 / ISIC 2018 challenge
(Zhang et al. 2020)	2020	Skin-Cancer-Detection (SCD) / ISIC 2018
(Shah et al. 2020)	2020	ISIC 2020

(Wu et al. 2020)	2020	ISBI2017 / ISIC2018
(Hasan et al. 2021)	2021	Skin Cancer' Benign vs. Malignant
(Abhishek et al. 2021)	2021	Interactive Atlas of Dermoscopy / MClass-D
(Maron et al. 2021)	2021	HAM10000 / PH2 / SKINL2 / BCN20000/ PROP
(Rotemberg et al. 2021)	2021	2020 SIIM-ISIC Melanoma Classification challenge
(Cano et al. 2021)	2021	ISIC Training Challenge 2019
(Coronado-Gutiérrez et al. 2021)	2021	ISIC Archive
(Sun et al. 2021)	2021	ISIC 2018, 2019 / MED-NODE / seven-point / PH2
(Yu et al. 2021)	2021	ISIC 2020 / PH2
(Sayed et al. 2021)	2021	ISIC 2020
(Barata et al. 2021)	2021	ISIC 2017 and 2018
(Rahman et al. 2021)	2021	HAM10000 / ISIC 2019
(Chaturvedi et al. 2021)	2021	HAM10000
(Steppan and Hanke 2021)	2021	ISIC 2019 / PH2 / SD-198 / MED-NODE / 7-point criteria evaluation / Light Field Image
(Krohling et al. 2021)	2021	PAD-UFES-20
(Yao et al. 2021)	2021	ISIC 2018 / Seven-Point Criteria Evaluation (7-PT)
(Groh et al. 2021)	2021	Fitzpatrick 17k
(Ahmad et al. 2021)	2021	HAM10000
(Khan et al. 2021)	2021	ISBI 2016, 2017 / ISIC 2018 / PH2 / HAM10000
(Jiang et al. 2021)	2021	SIIM-ISIC Melanoma Classification
(Maiti et al. 2021)	2021	HAM10000
(Zhao et al. 2021)	2021	ISIC 2019
(Ren et al. 2021)	2021	BCN20000 / HAM10000
(Perez et al. 2021)	2021	UDA-1,2 / PH2 / HAM10000 / ISBI 2016, 2017 / MED-NODE / MSK-1,2,3,4
(Begum et al. 2021)	2021	Dermnet
(Abbas et al. 2021)	2021	Yonsei University Hospital
(Moataz et al. 2021)	2021	HAM10000
(Milczarski et al. 2021)	2021	PH2 / Derm7pt / ISIC
(Bagheri et al. 2021)	2021	ISBI 2016, 2017, 2018 / PH2 / DermQuest
(Zhang et al. 2021)	2021	In-house / DermNet / DermNet NZ / AtlasDerm / DermIS / SD-

		260 / DanDerm / Kaggle
(Hasan et al. 2022)	2022	ISIC-2016, ISIC-2017, ISIC-2018
(Shorfuzzaman 2022)	2022	ISIC
(Ali et al. 2022)	2022	Monkeypox Skin Lesion Dataset (MSLD)
(Alenezi et al. 2023)	2023	ISIC-2019, 2020

As shown in Table 1, there are overlaps between papers using the same groups of skin lesion datasets. Through the process, 54 different skin lesion datasets were identified from these papers. Table 2 summarizes each dataset's reporting of the following metadata: age, gender, ethnicity, and Fitzpatrick skin type. The number of images is also shown.

Table 2 54 different publicly available skin lesion datasets used in publications and their reporting of four main metadata, showing a lack of reporting of skin type information to cover skin type diversity in datasets.

Skin lesion datasets	No. Images	Metadata			
		Gender	Age	Ethnicity	Skin type
7-point criteria evaluation dataset (Kawahara et al. 2018)	> 2,000	√	-	-	-
Asan (Han et al. 2018)	120,780	√	√	√	-
Atlas (Mendes and da Silva 2018)	3,816	-	-	-	-
AtlasDerm (Zhang et al. 2021)	9,503	-	-	√	-
BCN20000 (Combalia et al. 2019)	19,424	√	√	-	-
Cancer Genome Atlas (Argenziano et al. 2000)	2,860	-	-	-	-
Clinical Atlas (Tschandl et al. 2018)	839	-	-	-	-
DanDerm (Liao H 2016)	1,110	-	-	√	-
Derm7pt (Kawahara et al. 2018)	> 2000	-	-	-	-
Derm101 (Boer and Nischal 2007)	107,656	-	-	√	-
Dermatology Dataset (G√venir et al. 1998)	336	-	√	-	-
DermIS (Hosny et al. 2019; Mikołajczyk and Grochowski 2018)	7,172	-	√	√	-
Dermnet (Liao et al. 2016)	19,500	-	-	√	-
DermNet NZ (Zhang et al. 2021)	246	-	-	-	-
Dermofit Image Library (Fisher 2016)	1300	-	-	√	-
Dermoscopic Atlas (Tschandl et al. 2018)	872	-	-	-	-
Dermoscopy Skin Lesion Multispectral Image Database (Lézoray et al. 2014)	30	-	-	-	-
DermQuest (Hosny et al. 2019)	137	-	-	-	-
DDI (Daneshjou et al. 2022)	656	√	√	-	√
Edinburgh (Ballerini et al. 2013)	1,300	√	√	√	-
EDRA Interactive Atlas of Dermoscopy (Argenziano et al. 2000)	1,000	-	-	-	-
Fitzpatrick 17k (Groh et al. 2021)	16,577	-	-	-	√

Hallym (Han et al. 2018)	152	√	√	√	-
HAM10000 (Tschandl et al. 2018)	10,015	√	√	-	-
Interactive Atlas of Dermoscopy (IAD) (Argenziano et al. 2000)	> 2, 000	-	-	-	-
ISBI 2016 (Gutman et al. 2016)	1,279	-	-	-	-
ISBI 2017 (Codella et al. 2018)	2,750	-	-	-	-
ISIC challenge 2020 (Rotemberg et al. 2021)	33,126	√	√	√	-
ISIC-MSK (Gutman et al. 2016)	225	√	√	-	-
ISIC-UDA (Gutman et al. 2016)	557	-	-	-	-
Kaggle (Zhang et al. 2021)	367	-	-	-	-
Light Field Image (de Faria et al. 2019)	250	√	√	-	-
MClass (Brinker et al. 2019)	100	-	-	-	-
MED-NODE (Giotis et al. 2015a)	170	-	-	-	-
MoleMap (Gu et al. 2019; Mikołajczyk and Grochowski 2018)	102,451	-	-	-	-
Monkeypox Skin Lesion Dataset (MSLD) (Ali et al. 2022)	228	-	-	√	-
Normal (Han et al. 2020)	48,271	√	√	√	-
OLE (Liao et al. 2016)	1,300	-	-	-	-
PAD-UFES-20 (Pacheco et al. 2020)	2,299	√	√	√	√
PH2 (Mendonça et al. 2013)	200	-	-	-	-
SD-128 (Sun et al. 2016)	5,619	-	-	√	-
SD-198 (Sun et al. 2016; Yang et al. 2018)	6,584	√	√	√	-
SD-260 (Yang et al. 2019)	20,600	√	√	√	-
SIIM-ISIC Melanoma (Rotemberg et al. 2021)	33,126		√	-	-
Skin-10 (He et al. 2019)	10,218	-	-	-	-
Skin-100 (He et al. 2019)	19,807	-	-	-	-
Skin Cancer' Malignant vs. Benign (Ashim et al. 2021; Fanconi 2019)	6,594	-	-	-	-
SkinCon (Daneshjou et al. 2022)	3230	-	-	-	-
SkinL2 (de Faria et al. 2019)	376	-	-	-	-
SMARTSKINS (Vasconcelos et al. 2014)	-	√	√	-	-

SNU (Han et al. 2020)	2,201	√	√	√	-
Web (Han et al. 2020)	51,459	√	√	√	-
XiangyaDerm (Xie et al. 2019)	107,565	-	-	-	-
Yonsei University Health System South Korea (Yu et al. 2018)	724	-	-	-	-

Ideally, skin lesion datasets should achieve skin type diversity as well as have transparency in their metadata. As a result not only would their diversity be easily measured, but also any bias will be detected before training models using these datasets. As seen in Table 2, only three datasets: PAD-UFES-20, Fitzpatrick 17k, and DDI provide metadata on skin type. They have skin type labels based on the Fitzpatrick rating system (Wen et al. 2022). Fig. 5 also shows the breakdown of reporting in the metadata for gender, age, ethnicity, and skin type. As shown, skin type metadata are the least frequently provided, being included in just 3 of 54 datasets (5.56%). Age metadata were the most frequently provided, being included in 35.19% of the datasets.

Although the PAD-UFES-20 and Fitzpatrick 17k datasets provide skin type metadata, they contain far fewer images of darker skin types (e.g. only 635 out of 16,577 images in Fitzpatrick 17k are of skin type VI and only one image of skin type VI in PAD-UFES-20). Thus, apart from the lack of reporting of skin type metadata, even if datasets cover skin type information, there is not any guarantee that they have enough representation for darker-skinned groups. Fig. 6 and Fig. 7 show the distributions of skin types in the PAD-UFES-20 and Fitzpatrick 17k datasets respectively. It can be seen that skin type VI accounts for the lowest percentage in both datasets: 0.07% in PAD-UFES-20 and 3.97% in Fitzpatrick 17k. Note that in the Fitzpatrick 17k dataset, the full number of images is 16,577, but 565 images were excluded, because they had unknown Fitzpatrick skin type (they were labelled “-1”).

Fig. 8 shows the distribution of three groups of Fitzpatrick skin types in the DDI dataset. As shown, the distribution percentages of these three groups are close to each other.

As shown in Fig. 8, the DDI dataset metadata classify images into three skin type groups, rather than providing exact information for each of the six individual Fitzpatrick skin types. Therefore, although the dataset is balanced with respect to these three groups, it does not guarantee that for each individual skin type group it is balanced. More importantly, due to its small size, it is not suitable for generalizing deep learning models for all skin types. In the case of ethnicity label, it should be noted that ethnicity is different from skin type. To a significant degree, shared ethnicity reflects shared ancestry, but people of the same ethnic group can have a wide range of skin types.

4 Conclusions

This study is the first review to date which investigates publicly available skin lesion datasets and their metadata in detail for the crucial issue of skin type diversity. As these datasets are used for training deep learning models, inadequate skin type diversity within the datasets could affect the performance of the models, in terms of having low accuracy and bias toward specific groups of people. To overcome this issue, it is

important that, firstly, information about skin type distribution be provided for datasets, and secondly that skin type diversity in datasets be evaluated prior to using them for training models.

The issue of inadequate skin type diversity has been discussed in previous works, but without reporting a measurement for each skin type. For example, in the Gender Shades study (Buolamwini and Gebru 2018), although the authors used the Fitzpatrick skin type descriptions for their facial image datasets, they just divided the datasets into two skin type groups: darker and lighter. In (Karkkainen and Joo 2021), a balanced dataset, FairFace, was created according to different ethnicities, rather than different skin types. Also, Wen et al. (2022) discussed the issue for skin lesion datasets, but did not measure skin type diversity for those datasets. Failure to report the distribution of skin types used in a dataset raises concerns about the extent to which different populations are represented in that dataset, and also about generalizability of machine learning algorithms that have been trained using it.

Our results showed a lack of skin type reporting in all identified skin lesion datasets, with the exception of three: PAD-UFES-20, Fitzpatrick 17k, and DDI. Of the skin lesion datasets used in the papers identified in our review, these three are the only ones that provide information about skin type using the Fitzpatrick scale. However, as was shown in the results, two of those datasets - PAD-UFES-20 and Fitzpatrick 17k - have considerably less representation of darker skin. The DDI dataset reports skin tone distribution in three aggregate groups, rather than for each of the six Fitzpatrick skin types; therefore, exact information about the number of images belonging to each individual skin type is not available. Furthermore, it is too small for training a generalized model that works for all skin types. Also, the distinction between ethnicity and skin type should be restated as one ethnicity can include different skin types.

Finally, our review highlights that deep learning-based models should be developed with inclusion of all skin tones to mitigate algorithmic biases toward darker skin types. They should ideally include dermatologist-assigned Fitzpatrick skin type labels, which would allow the skin type diversity within the dataset to be evaluated.

Declarations

Acknowledgments

This publication has emanated from research supported by a grant from Science Foundation Ireland under Grant number 18/CRT/6222

References

1. Abbas Q, Ramzan F, Ghani MU (2021) Acral melanoma detection using dermoscopic images and convolutional neural networks. Visual Computing for Industry, Biomedicine, and Art, 4, 1-12.
2. Abhishek K, Kawahara J, Hamarneh G (2021) Predicting the clinical management of skin lesions using deep learning. Scientific Reports, 11(1), 1-14
3. Ahmad B, Jun S, Palade V, You Q, Mao L, Zhongjie M (2021) Improving skin cancer classification using heavy-tailed Student t-distribution in generative adversarial networks (TED-GAN). Diagnostics, 11(11),

2147

4. Akram T, Lodhi HMJ, Naqvi SR, Naeem S, Alhaisoni M, Ali M, Haider SA, Qadri NN (2020) A multilevel features selection framework for skin lesion classification. Human-centric Computing and Information Sciences, 10, 1-26
5. Aldwgeri A, Abubacker NF (2019) Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images. Advances in Visual Informatics: 6th International Visual Informatics Conference, IVIC 2019, Bangi, Malaysia, November 19–21, 2019, Proceedings 6,
6. Alenezi F, Armghan A, Polat K (2023) A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimization-based decision support in dermoscopy images. Expert Systems with Applications, 215, 119352
7. Ali SN, Ahmed M, Paul J, Jahan T, Sani S, Noor N, Hasan T (2022) Monkeypox skin lesion detection using deep learning models: A feasibility study. arXiv preprint arXiv:2207.03342
8. Andrade C, Teixeira LF, Vasconcelos MJM, Rosado L (2020) Data Augmentation Using Adversarial Image-to-Image Translation for the Segmentation of Mobile-Acquired Dermatological Images. Journal of Imaging, 7(1), 2
9. Argenziano G, Soyer HP, De Giorgio V, Piccolo D, Carli P, Delfino M, Ferrari A, Hofmann-Wellenhof R, Massi D, Mazzocchetti G (2000) Interactive atlas of dermoscopy. Dermoscopy: a tutorial
10. Ashim LK, Suresh N, Prasannakumar C (2021) A Comparative Analysis of Various Transfer Learning Approaches Skin Cancer Detection. 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)
11. Ashraf R, Afzal S, Rehman AU, Gul S, Baber J, Bakhtyar M, Mehmood I, Song OY, Maqsood M (2020) Region-of-interest based transfer learning assisted framework for skin cancer detection. IEEE Access, 8, 147858-147871
12. Bagheri F, Tarokh MJ, Ziaratban M (2021) Skin lesion segmentation based on mask RCNN, Multi Atrous Full-CNN, and a geodesic method. International Journal of Imaging Systems and Technology, 31(3), 1609-1624
13. Ballerini L, Fisher RB, Aldridge B, Rees J (2013) A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In Color medical image analysis (pp. 63-86). Springer
14. Banerjee S, Singh SK, Chakraborty A, Das A, Bag R (2020) Melanoma diagnosis using deep learning and fuzzy logic. Diagnostics, 10(8), 577
15. Barata C, Celebi ME, Marques JS (2021) Explainable skin lesion diagnosis using taxonomies. Pattern Recognition, 110, 107413
16. Begum M, Dhivya A, Krishnan AJ, Keerthana S (2021) Automated Detection of skin and nail disorders using Convolutional Neural Networks. 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)
17. Benjamin R (2020) Race after technology: Abolitionist tools for the new jim code. In: Oxford University Press
18. Betancourt JR, Tan-McGrory A, Flores E, López D (2019) Racial and ethnic disparities in radiology: a call to action. Journal of the American College of Radiology, 16(4), 547-553

19. Bisla D, Choromanska A, Berman RS, Stein JA, Polsky D (2019) Towards automated melanoma detection with deep learning: Data purification and augmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops
20. Bissoto A, Valle E, Avila S (2020) Debiasing skin lesion datasets and models? not so fast. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops
21. Boer A, Nischal K (2007) Get set for the net-www. derm101. com: A growing online resource for learning dermatology and dermatopathology
22. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Schilling B, Haferkamp S, Schadendorf D, Fröhling S (2019) A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. European Journal of Cancer, 111, 148-154
23. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Schilling B, Haferkamp S, Schadendorf D, Holland-Letz T (2019) Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47-54
24. Brinker TJ, Hekler A, Hauschild A, Berking C, Schilling B, Enk AH, Haferkamp S, Karoglan A, von Kalle C, Weichenthal M (2019) Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. European Journal of Cancer, 111, 30-37
25. Buolamwini J, Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on Fairness, Accountability and Transparency
26. Cano E, Mendoza-Avilés J, Areiza M, Guerra N, Mendoza-Valdés JL, Rovetto CA (2021) Multi skin lesions classification using fine-tuning and data-augmentation applying NASNet. PeerJ Computer Science, 7, e371
27. Chaturvedi SS, Gupta K, Prasad PS (2021) Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020
28. Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H (2018) 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). 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
29. Combalia M, Codella NC, Rotemberg V, Helba B, Vilaplana V, Reiter O, Carrera C, Barreiro A, Halpern AC, Puig S (2019) Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288
30. Coronado-Gutiérrez D, López C, Burgos-Artizzu XP (2021) Skin cancer high-risk patient screening from dermoscopic images via Artificial Intelligence: An online study. medRxiv, 2021.2002. 2004.21251132
31. Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J (2021) Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. JAMA dermatology, 157(11), 1362-1369
32. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, Ko J, Swetter SM, Bailey EE, Gevaert O (2022) Disparities in dermatology AI performance on a diverse, curated clinical image set. Science Advances, 8(31), eabq6147
33. Daneshjou R, Yuksekgonul M, Cai ZR, Novoa R, Zou JY (2022) Skincon: A skin disease dataset densely annotated by domain experts for fine-grained debugging and analysis. Advances in Neural Information

Processing Systems, 35, 18157-18167

34. de Faria SM, Filipe JN, Pereira PM, Tavora LM, Assuncao PA, Santos MO, Fonseca-Pinto R, Santiago F, Dominguez V, Henrique M (2019) Light field image dataset of skin lesions. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
35. Diepgen TL, Mahler V (2002) The epidemiology of skin cancer. British Journal of Dermatology, 146(s61), 1-6
36. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639), 115-118
37. Fallow BA, Tarumi T, Tanaka H (2013) Influence of skin type and wavelength on light wave reflectance. Journal of clinical monitoring and computing, 27, pp.313-317
38. Fanconi C (2019) https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign
39. Fisher R (2016) Dermofit Image Library. https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library
40. Fisher RB, Rees J, Bertrand A (2020). Classification of ten skin lesion classes: Hierarchical KNN versus deep net. Medical Image Understanding and Analysis: 23rd Conference, MIUA 2019, Liverpool, UK, July 24–26, 2019, Proceedings 23
41. FitzGerald C, Hurst S (2017) Implicit bias in healthcare professionals: a systematic review. BMC medical ethics, 18(1), 1-18
42. Fitzpatrick T (1997) The validity and practicality of sunreactive skin types I through VI. Arehives of Dermatology, 124, 868
43. França LDR (2018) Neuroevolution of Augmenting Topologies Applied to the Detection of Cancer in Medical Images. Dissertation, Universidade Federal Rural de Pernambuco – UFRPE
44. Fuchs DJ (2018) The dangers of human-like bias in machine-learning algorithms. Missouri S&T's Peer to Peer, 2(1), 1
45. Ge Z, Demyanov S, Bozorgtabar B, Abedini M, Chakravorty R, Bowling A, Garnavi R (2017) Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)
46. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N (2015a) MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert Systems with Applications, 42(19), 6578-6585
47. Gloster Jr HM, Neal K (2006) Skin cancer in skin of color. Journal of the American Academy of Dermatology, 55(5), 741-760
48. Gonzalez-Diaz I (2018) Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. IEEE Journal of Biomedical and Health Informatics, 23(2), 547-559
49. Goyal M, Hassanpour S, Yap MH (2018) Region of interest detection in dermoscopic images for natural data-augmentation. arXiv preprint arXiv:1807.10711
50. Goyal M, Knackstedt T, Yan S, Hassanpour S (2020) Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. Computers in Biology and Medicine, 127, 104065

51. Groh M, Harris C, Soenksen L, Lau F, Han R, Kim A, Koochek A, Badri O (2021) Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
52. Gu Y, Ge Z, Bonnington CP, Zhou J (2019) Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. IEEE journal of biomedical and health informatics, 24(5), 1379-1393
53. Gu Y, Zhou J, Qian B (2017) Melanoma detection based on mahalanobis distance learning and constrained graph regularized nonnegative matrix factorization. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 797-805). IEEE
54. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) 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)
55. Güvenir HA, Demiröz G, Ilter N (1998) Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. 13(3), 147-165
56. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE (2018) Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. Journal of Investigative Dermatology, 138(7), 1529-1538
57. Han SS, Park GH, Lim W, Kim MS, Na JI, Park I, Chang SE (2018) Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. PlOS ONE, 13(1), e0191493
58. Han SS, Park I, Chang SE, Lim W, Kim MS, Park GH, Chae JB, Huh CH, Na J-I (2020) Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. Journal of Investigative Dermatology, 140(9), 1753-1761
59. Hasan MK, Elahi MTE, Alam MA, Jawad MT, Martí R (2022) DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. Informatics in Medicine Unlocked, 28, 100819
60. Hasan MR, Fatemi MI, Monirujjaman Khan M, Kaur M, Zaguia A (2021) Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. Journal of Healthcare Engineering, 2021
61. He X, Wang S, Shi S, Tang Z, Wang Y, Zhao Z, Dai J, Ni R, Zhang X, Liu X (2019) Computer-aided clinical skin disease diagnosis using CNN and object detection models. 2019 IEEE International Conference on Big Data
62. Hosny KM, Kassem MA, Foaud MM (2019) Classification of skin lesions using transfer learning and augmentation with Alex-net. 14(5), e0217293
63. Hosny KM, Kassem MA, Foaud MM (2020) Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. Multimedia Tools and Applications, 79, 24029-24055
64. Howard A, Borenstein J (2018) The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. Science and Engineering Ethics, 24, 1521-1536

65. Hudson K, Lifton R, Patrick-Lake B, Burchard EG, Coles T, Collins R, Conrad A (2015) The precision medicine initiative cohort program—Building a Research Foundation for 21st Century Medicine. Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, ed
66. Jiang F, Zhou F, Qin J, Wang T, Lei B (2019) Decision-augmented generative adversarial network for skin lesion segmentation. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
67. Jiang Y, Huang R, Shi J (2021) EfficientNet-Based model with test time augmentation for cancer detection. 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)
68. Jordan P (2016) Why an AI-judged beauty contest picked nearly all white winners. https://www.vice.com/en/article/78k7de/why-an-ai-judged-beauty-contest-picked-nearly-all-white-winners. 05 September 2016
69. Kamulegeya LH, Okello M, Bwanika JM, Musinguzi D, Lubega W, Rusoke D, Nassiwa F, Börve A (2019) Using artificial intelligence on dermatology conditions in Uganda: A case for diversity in training data sets for machine learning. BioRxiv, 826057
70. Karkkainen K, Joo J (2021) Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision
71. Kawahara J, BenTaieb A, Hamarneh G (2016) Deep features to classify skin lesions. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)
72. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G (2018) Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE Journal of Biomedical and Health Informatics, 23(2), 538-546
73. Khalil A, Ahmed SG, Khattak AM, Al-Qirim N (2020) Investigating bias in facial analysis systems: A systematic review. IEEE Access, 8, 130751-130761
74. Khan MA, Akram T, Sharif M, Shahzad A, Aurangzeb K, Alhussein M, Haider SI, Altamrah A (2018) An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification. BMC Cancer, 18, 1-20
75. Khan MA, Sharif M, Akram T, Damaševičius R, Maskeliūnas R (2021) Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. Diagnostics, 11(5), 811
76. Kostick-Quenet KM, Cohen IG, Gerke S, Lo B, Antaki J, Movahedi F, Njah H, Schoen L, Estep JE, Blumenthal-Barby J (2022) Mitigating racial bias in machine learning. Journal of Law, Medicine & Ethics, 50(1), 92-100
77. Kraehe AM, Herman Jr (2020) Racial encounters, ruptures, and reckonings: Art curriculum futurity in the wake of Black Lives Matter. In (Vol. 73, pp. 4-7): Taylor & Francis
78. Krohling B, Castro PB, Pacheco AG, Krohling RA (2021) A smartphone based application for skin cancer classification using deep learning with clinical images and lesion information. arXiv preprint arXiv:2104.14353
79. Lézoray O, Revenu M, Desvignes M (2014) Graph-based skin lesion segmentation of multispectral dermoscopic images. 2014 IEEE International Conference on Image Processing (ICIP)

80. Liao H (2016) A deep learning approach to universal skin disease classification. University of Rochester Department of Computer Science, CSC
81. Liao H, Li Y, Luo J (2016) Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks. 2016 23rd International Conference on Pattern Recognition (ICPR)
82. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. 2017 13th IASTED International Conference on Biomedical Engineering (BioMed)
83. Lu B, Chen J-C, Castillo CD, Chellappa R (2019) An experimental evaluation of covariates effects on unconstrained face verification. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(1), 42-55
84. Lucius M, De All J, De All JA, Belvisi M, Radizza L, Lanfranconi M, Lorenzatti V, Galmarini CM (2020) Deep neural frameworks improve the accuracy of general practitioners in the classification of pigmented skin lesions. Diagnostics, 10(11), 969
85. Luo W, Yang M (2018) Fast skin lesion segmentation via fully convolutional network with residual architecture and CRF. 2018 24th International Conference on Pattern Recognition (ICPR)
86. Mahajan K, Sharma M, Vig L (2020) Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops
87. Maiti R, Agarwal P, Kumar RR, Bhat A (2021) Detection of skin cancer using neural architecture search with model quantization. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)
88. Marcus G, Davis E (2019) Rebooting AI: Building artificial intelligence we can trust. Vintage
89. Maron RC, Haggenmüller S, von Kalle C, Utikal JS, Meier F, Gellrich FF, Hauschild A, French LE, Schlaak M, Ghoreschi K (2021) Robustness of convolutional neural networks in recognition of pigmented skin lesions. European Journal of Cancer, 145, 81-91
90. McDuff D, Cheng R, Kapoor A (2018) Identifying bias in AI using simulation. arXiv preprint arXiv:1810.00471
91. Mendes DB, da Silva NC (2018) Skin lesions classification using convolutional neural networks in clinical images. arXiv preprint arXiv:1812.02316
92. Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J (2013) PH 2-A dermoscopic image database for research and benchmarking. 2013 35th annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
93. Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW)
94. Milantev S, Olyunin V, Bykov I, Milanteva N, Bessmertny I (2020) Skin Lesion Analysis Using Ensemble of CNN with Dermoscopic Images and Metadata. MICSECS
95. Milczarski P, Beczkowski M, Borowski N (2021) Enhancing dermoscopic features classification in images using invariant dataset augmentation and convolutional neural networks. Neural Information Processing:

28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28

96. Moataz L, Salama GI, Abd Elazeem MH (2021) Skin cancer diseases classification using deep convolutional neural network with transfer learning model. Journal of physics: Conference Series
97. Monk Jr EP (2014). Skin tone stratification among Black Americans, 2001–2003. Social Forces, 92(4), 1313-1337
98. Musselwhite LW, Oliveira CM, Kwaramba T, de Paula Pantano N, Smith JS, Fregnani JH, Reis RM, Mauad E, de Lima Vazquez F, Longatto-Filho A (2016) Racial/ethnic disparities in cervical cancer screening and outcomes. Acta Cytologica, 60(6), 518-526
99. Myers West S (2020) Discriminating Systems: Gender, Race and Power in Artificial Intelligence
100. Nedelcu T, Vasconcelos M, Carreiro A (2020) Multi-dataset training for skin lesion classification on multimodal and multitask deep learning. Proceedings of the 6th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'20), Prague, Czech Republic
101. Nunnari F, Sonntag D (2019) A CNN toolbox for skin cancer classification. arXiv preprint arXiv:1908.08187
102. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453
103. Pacheco AG, Ali A-R, Trappenberg T (2019) Skin cancer detection based on deep learning and entropy to detect outlier samples. arXiv preprint arXiv:1909.04525
104. Pacheco AG, Lima GR, Salomao AS, Krohling B, Biral IP, de Angelo GG, Alves Jr FC, Esgario JG, Simora AC, Castro PB (2020) PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in brief, 32, 106221
105. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. International journal of surgery, 88, p.105906
106. Perez E, Reyes O, Ventura S (2021) Convolutional neural networks for the automatic diagnosis of melanoma: An extensive experimental study. Medical image analysis, 67, 101858
107. Pham T-C, Luong C-M, Visani M, Hoang V-D (2018) Deep CNN and data augmentation for skin lesion classification. Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part II 10
108. Porgali B, Albiero V, Ryda J, Ferrer CC, Hazirbas C (2023) The Casual Conversations v2 Dataset. arXiv preprint arXiv:2303.04838
109. Pour MP, Seker H (2020) Transform domain representation-driven convolutional neural networks for skin lesion segmentation. Expert Systems with Applications, 144, 113129
110. Rahman Z, Hossain MS, Islam MR, Hasan MM, Hridhee RA (2021) An approach for multiclass skin lesion classification based on ensemble learning. Informatics in Medicine Unlocked, 25, 100659
111. Ren Z, Guo Y, Stella XY, Whitney D (2021) Improve Image-based Skin Cancer Diagnosis with Generative Self-Supervised Learning. 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)

112. Rotemberg V, Kurtansky N, Betz-Stablein B, Caffery L, Chousakos E, Codella N, Combalia M, Dusza S, Guitera P, Gutman D (2021) A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data, 8(1), 34
113. Sae-Lim W, Wettayaprasit W, Aiyarak P (2019) Convolutional neural networks using MobileNet for skin lesion classification. 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)
114. Saez A, Serrano C, Acha B (2014) Model-based classification methods of global patterns in dermoscopic images. IEEE Transactions on Medical Imaging, 33(5), pp.1137-1147
115. Sayed GI, Soliman MM, Hassanien AE (2021) A novel melanoma prediction model for imbalanced data using optimized SqueezeNet by bald eagle search optimization. Computers in Biology and Medicine, 136, 104712
116. Shah V, Autee P, Sonawane P (2020) Detection of melanoma from skin lesion images using deep learning techniques. 2020 International Conference on Data Science and Engineering (ICDSE)
117. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, Wheeler MT, Ashley EA (2017) Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. Journal of Personalized Medicine, 7(2), 3
118. Shoieb DA, Youssef SM (2018) An enhanced hybrid model for skin diagnosis using deep convolution neural network. 2018 9th Cairo International Biomedical Engineering Conference (CIBEC)
119. Shorfuzzaman M (2022) An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. Multimedia Systems, 28(4), 1309-1323
120. Steppan J, Hanke S (2021) Analysis of skin lesion images with deep learning. arXiv preprint arXiv:2101.03814
121. Sun Q, Huang C, Chen M, Xu H, Yang Y (2021) Skin lesion classification using additional patient information. BioMed Research International, 2021
122. Sun X, Yang J, Sun M, Wang K (2016) A benchmark for automatic visual classification of clinical skin disease images. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14
123. Tan TY, Zhang L, Lim CP, Fielding B, Yu Y, Anderson E (2019) Evolving ensemble models for image segmentation using enhanced particle swarm optimization. IEEE Access, 7, pp.34004-34019
124. Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5(1), 1-9
125. Ünver HM, Ayan E (2019) Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. Diagnostics, 9(3), 72
126. Valle E, Fornaciali M, Menegola A, Tavares J, Bittencourt FV, Li LT, Avila S (2020) Data, depth, and design: Learning reliable models for skin lesion analysis. Neurocomputing, 383, 303-313
127. Vasconcelos MJM, Rosado L, Ferreira M (2014) Principal axes-based asymmetry assessment methodology for skin lesion image analysis. International Symposium on Visual Computing
128. Waweru AK, Ahmed K, Miao Y, Kawan P (2020) Deep Learning in Skin Lesion Analysis Towards Cancer Detection. 2020 24th International Conference Information Visualisation (IV)

129. Wen D, Khan SM, Xu AJ, Ibrahim H, Smith L, Caballero J, Zepeda L, de Blas Perez C, Denniston AK, Liu X (2022) Characteristics of publicly available skin cancer image datasets: a systematic review. The Lancet Digital Health, 4(1), e64-e74
130. Williams DR, Cooper LA (2019) Reducing racial inequities in health: using what we already know to take action. International Journal of Environmental Research and Public Health, 16(4), 606
131. Wilson B, Hoffman J, Morgenstern JJ (2019) Predictive inequity in object detection. arXiv preprint arXiv:1902.11097
132. Wu H, Pan J, Li Z, Wen Z, Qin J (2020) Automated skin lesion segmentation via an adaptive dual attention module. IEEE Transactions on Medical Imaging, 40(1), 357-370
133. Xie B, He X., Zhao S, Li Y, Su J, Zhao X, Kuang Y, Wang Y, Chen X (2019) XiangyaDerm: a clinical image dataset of asian race for skin disease aided diagnosis. Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention: International Workshops, LABELS 2019, HAL-MICCAI 2019, and CuRIOUS 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 4
134. Xu H, Sarkar A, Abbott AL (2022) Color invariant skin segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
135. Yang J, Sun X, Liang J, Rosin PL (2018) Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
136. Yang J, Wu X, Liang J, Sun X, Cheng M-M, Rosin PL, Wang L (2019) Self-paced balance learning for clinical skin disease recognition. IEEE Transactions on Neural Networks and Learning Systems, 31(8), 2832-2846
137. Yao P, Shen S, Xu M, Liu P, Zhang F, Xing J, Shao P, Kaffenberger B, Xu RX (2021) Single model deep learning on imbalanced small datasets for skin lesion classification. IEEE Transactions on Medical Imaging, 41(5), 1242-1254
138. Yu C, Tang M, Yang S, Wang M, Xu Z, Yan J, Chen H, Yang Y, Zeng X-J, Li X (2021) Towards better dermoscopic image feature representation learning for melanoma classification. Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part IV 28
139. Yu C, Yang S, Kim W, Jung J, Chung K-Y, Lee SW, Oh B (2018) Acral melanoma detection using a convolutional neural network for dermoscopy images. PlOS ONE, 13(3), e0193321
140. Yu L, Chen H, Dou Q, Qin J, Heng P-A (2016) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging, 36(4), 994-1004
141. Zhang J, Petitjean C, Ainouz S (2020) Kappa loss for skin lesion segmentation in fully convolutional network. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
142. Zhang L, Mishra S, Zhang T, Zhang Y, Zhang D, Lv Y, Lv M, Guan N, Hu S, Chen DZ, Han X (2021) Design and assessment of convolutional neural network based methods for vitiligo diagnosis. 1901
143. Zhao C, Shuai R, Ma L, Liu W, Hu D, Wu M (2021) Dermoscopy image classification based on StyleGAN and DenseNet201. IEEE Access, 9, 8659-8679
144. Zou J, Schiebinger L (2018) AI can be sexist and racist—it's time to make it fair. In: Nature Publishing Group UK London

Footnotes

1. The full list of 121 publications is available here: https://arrow.tudublin.ie/engschelecon/15/

Figures

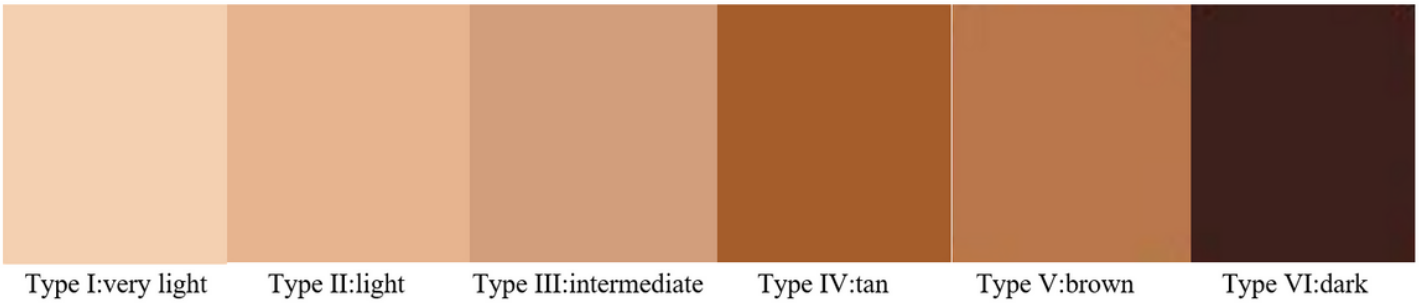

Figure 1

Range of skin tones in Fitzpatrick skin type scale which classifies skin tones to six types.

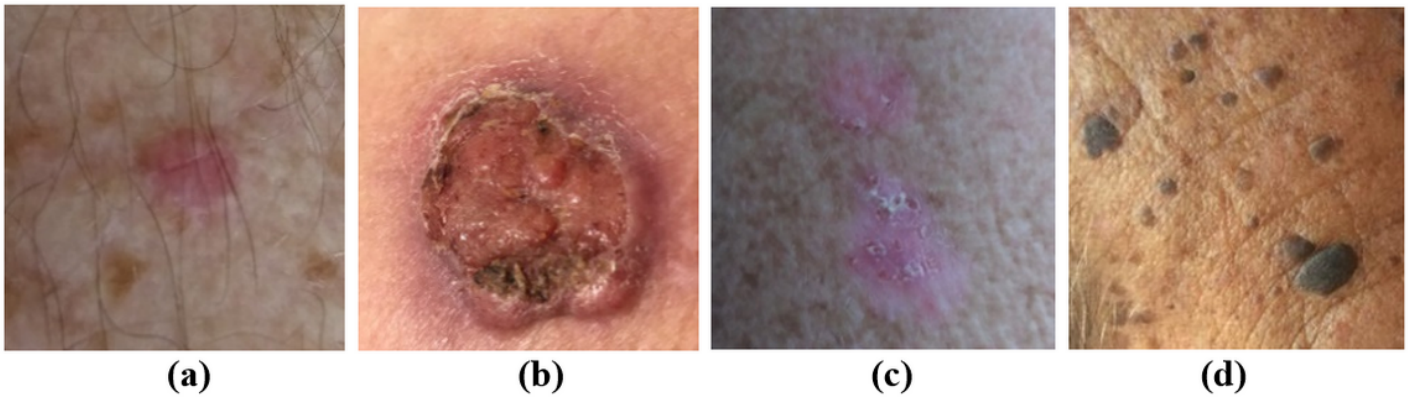

Figure 2

Some sample images from PAD-UFES-20 dataset. (a) Skin type I. (b) Skin type II. (c) Skin type III. (d) Skin type IV.

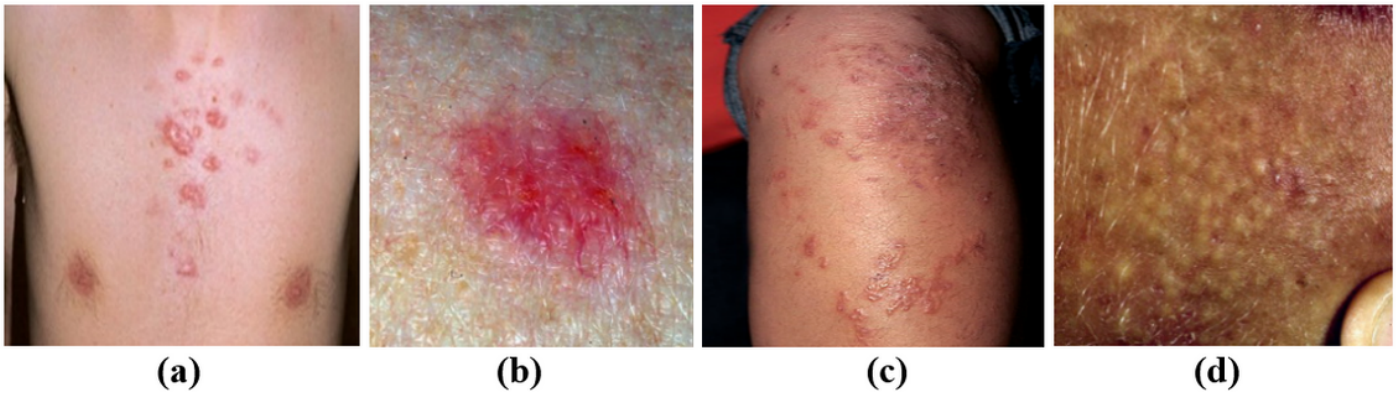

Figure 3

Some sample images from Fitzpatrick 17k dataset. (a) Skin type I. (b) Skin type II. (c) Skin type III. (d) Skin type IV.

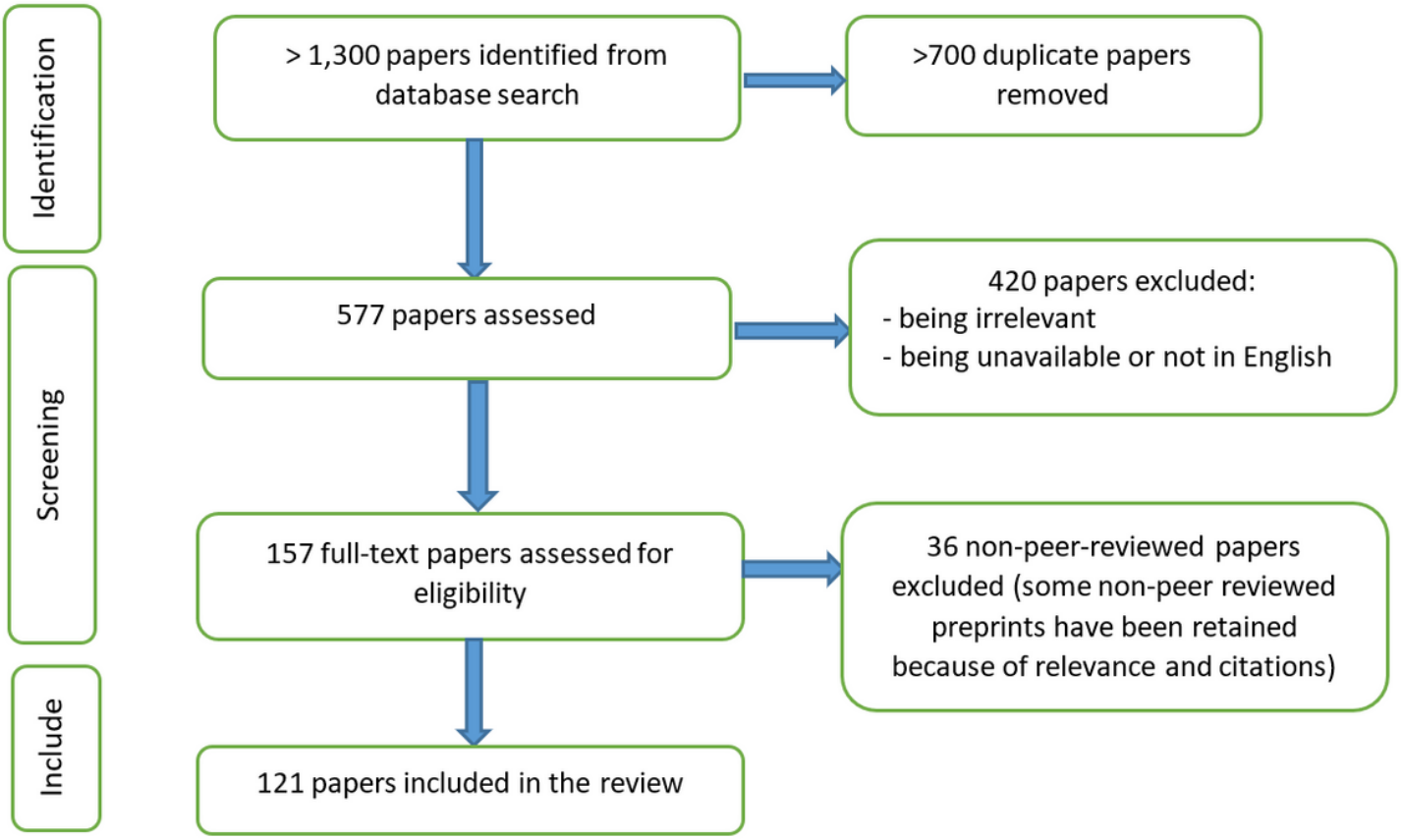

Figure 4

PRISMA flow chart of study selection.

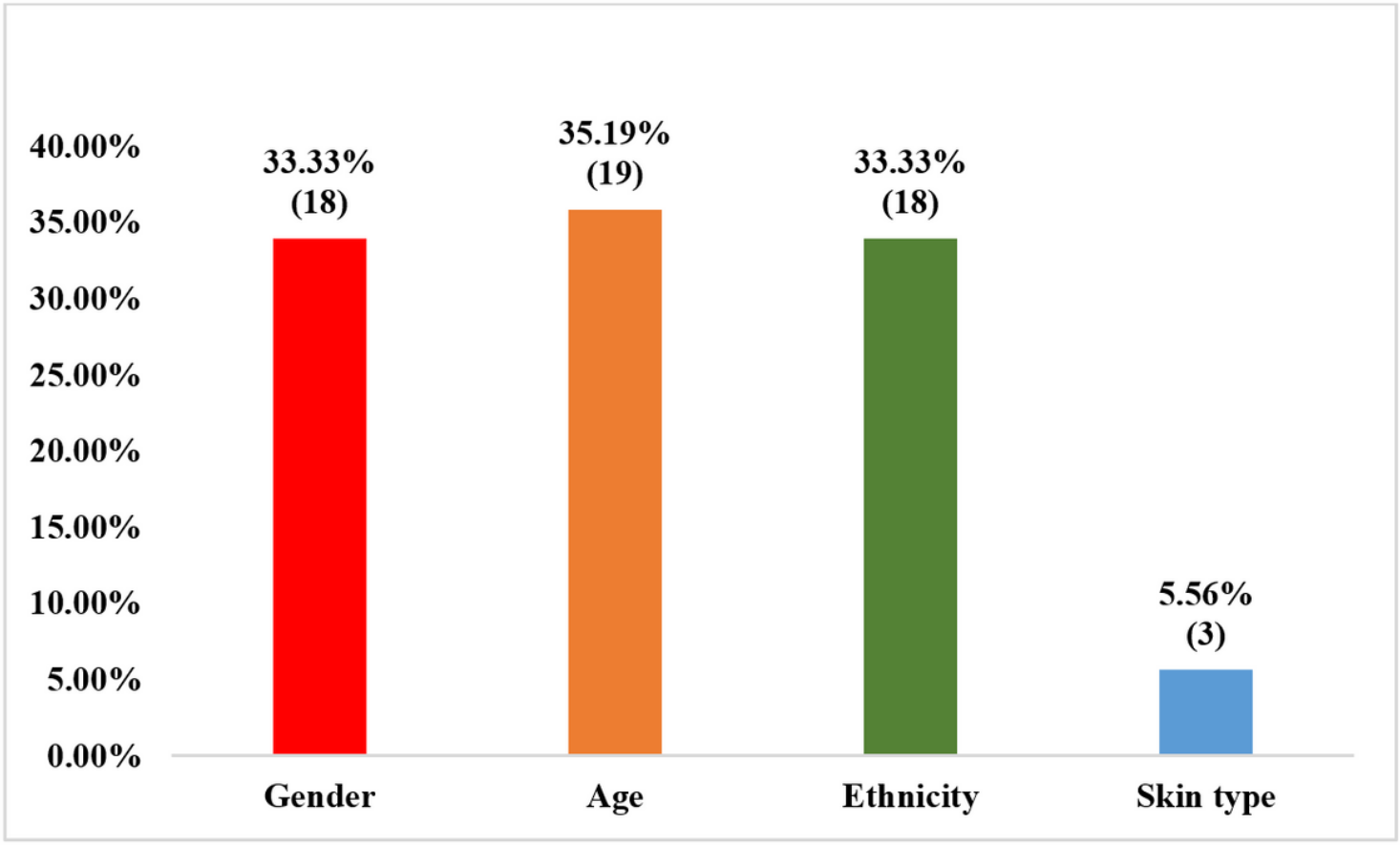

Figure 5

Percentage of the 54 skin lesion datasets that provide metadata for gender, age, ethnicity, and skin type respectively.

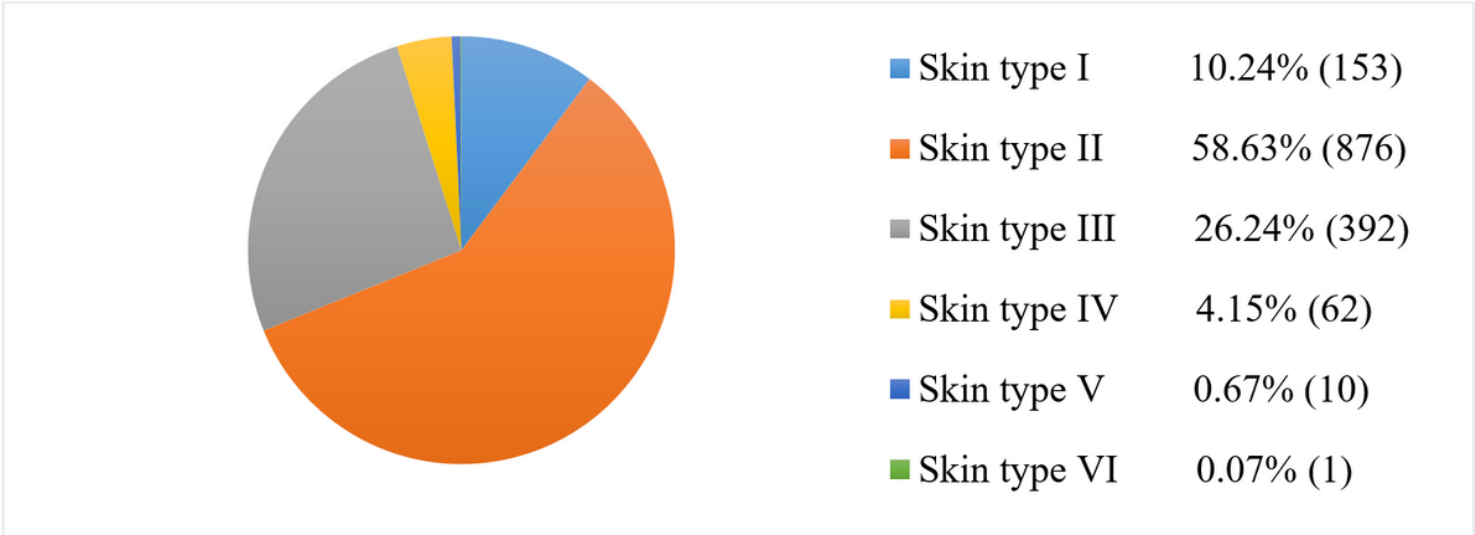

Figure 6

Skin type distribution for 1,494 images in the PAD-UFES-20 dataset (Pacheco et al. 2020), according to dermatologist-assigned Fitzpatrick scale labels.

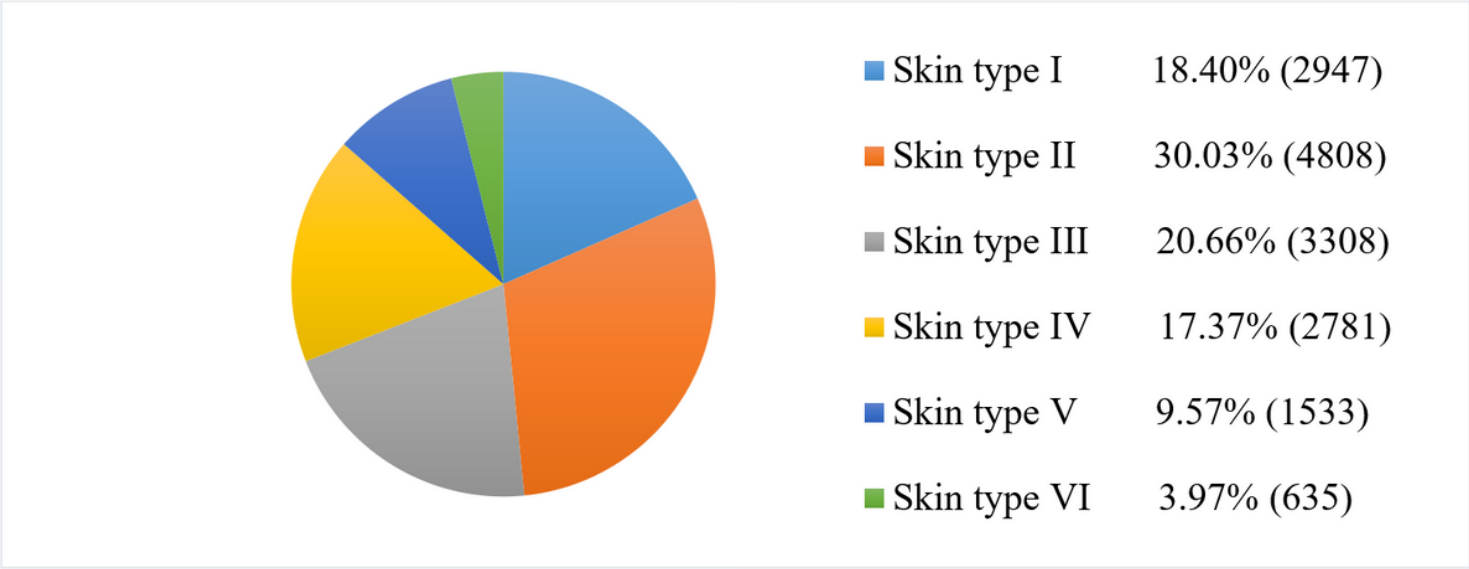

Figure 7

Skin type distribution for 16,012 images in the Fitzpatrick 17k dataset (Groh et al. 2021), according to dermatologist-assigned Fitzpatrick scale labels. The original number of images was 16,577, but 565 images had unknown Fitzpatrick skin type.

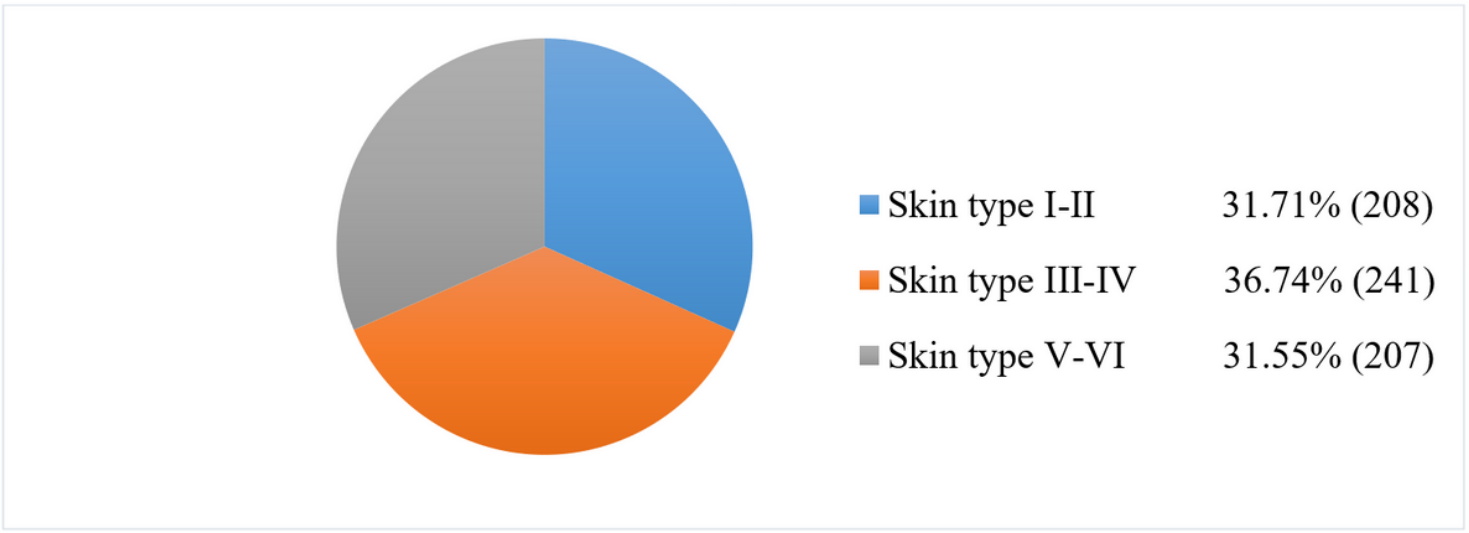

Figure 8

Skin type distribution for the 656 images in the DDI dataset (Daneshjou et al. 2022), according to dermatologist-assigned Fitzpatrick scale labels.