Fairness of Classifiers Across Skin Tones in Dermatology

Newton M. Kinyanjui[1,4], Timothy Odonga[1,4], Celia Cintas[1], Noel C. F. Codella[2][0000−0001−6735−9067], Rameswar Panda[3][0000−0003−4359−2475], Prasanna Sattigeri[2][0000−0003−4435−0486], and Kush R. Varshney[1,2][0000−0002−7376−5536]

[1] IBM Research – Africa, Nairobi 00100, Kenya
[2] IBM Research – T. J. Watson Research Center, Yorktown Heights NY 10598, USA
[3] IBM Research – Cambridge, Cambridge MA 02142, USA
[4] Carnegie Mellon University Africa, Kigali, Rwanda

Abstract. Recent advances in computer vision have led to breakthroughs in the development of automated skin image analysis. However, no attempt has been made to evaluate the consistency in performance across populations with varying skin tones. In this paper, we present an approach to estimate skin tone in skin disease benchmark datasets and investigate whether model performance is dependent on this measure. Specifically, we use individual typology angle (ITA) to approximate skin tone in dermatology datasets. We look at the distribution of ITA values to better understand skin color representation in two benchmark datasets: 1) the ISIC 2018 Challenge dataset, a collection of dermoscopic images of skin lesions for the detection of skin cancer, and 2) the SD-198 dataset, a collection of clinical images capturing a wide variety of skin diseases. To estimate ITA, we first develop segmentation models to isolate non-diseased areas of skin. We find that the majority of the data in the two datasets have ITA values between 34.5° and 48°, which are associated with lighter skin, and is consistent with under-representation of darker skinned populations in these datasets. We also find no measurable correlation between accuracy of machine learning models and ITA values, though more comprehensive data is needed for further validation.

Keywords: Algorithmic fairness · Dermatology image analysis · Medical imaging

1 Introduction

As machine learning is becoming more frequently applied to support consequential decisions, there is increasing interest in accurately measuring latent dataset characteristics and demographic representation to prevent the potential negative consequences of dataset imbalances [27], henceforth referred to as "dataset bias". Dataset bias is a critical issue because it is one of the causes of machine learning-based systems placing certain groups at a systematic disadvantage [3]. Recognition and mitigation of unwanted bias is necessary to build machine learning systems that are trustworthy [31].

Skin diseases continue to bear significant negative impacts on human health. Skin diseases contribute 1.79% to the global burden of disease [19] and skin cancer accounts for about 7% of new cancer cases [4]. Within skin cancer, there is evidence of some outcome disparities with respect to ethnicity: although people of color are roughly 20 to 30 times less likely to develop melanoma than lighter skinned individuals, for certain melanoma sub-types they have been found to have lower [23, 34, 22] or higher [22] survival rates. Some studies have found that for people of color, the diagnosis of skin cancer may occur at a more advanced stage, leading to lower rates of survival and poorer outcomes [14, 21]. However, increased screening also carries risks, such as unnecessary surgeries, disfigurement, disability, morbidity, and over-diagnosis [23].

Computer vision has been studied in the context of dermatology image analysis for decades [28, 20, 1]. The success of deep learning models has led to studies applying the technology to dermatological use cases [7, 8]. Models using convolutional neural networks (CNNs) have been applied to problems such as skin cancer diagnosis and were found to outperform trained dermatologists in controlled settings and datasets [11, 15, 13]. However, as most of the publicly available datasets of skin images come from lighter skinned populations, due to the extreme disparities in disease prevalence, there are concerns about how to best collect data, train, and evaluate models for darker skinned populations [2, 27]. Also, because of the significant risks of harm from over-diagnosis with increased screening in low-risk dark skin populations, there is a need to better discriminate between life-threatening and stable presentations of disease [23, 27].

In this paper, we work towards quantifying skin tone distributions in datasets where this information is currently unavailable, and measuring downstream effects on classifier performance. Specifically, our contributions are as follows:

- We propose a pipeline to automatically estimate skin tone for images in two public benchmark skin disease datasets using the individual typology angle (ITA), which has been used previously as a measure of skin tone in absence of manually curated information [24].
- We create manually-labeled segmentation masks and automatically generated masks for non-diseased skin in both public benchmarks.
- We quantitatively confirm that the two benchmark skin disease datasets under-represent ITA values correlated with darker skin populations.
- No correlation between model performance and ITA value is measurable at this time, though more data is needed for conclusive results.

2 Related Work

Recent years have seen significant advances in automated skin lesion analysis, with hundreds of deep learning models implemented for skin cancer diagnosis. Much of this work has been enabled by the International Skin Imaging Collaboration (ISIC) [17, 10, 30], which has organized a public repository of annotated dermoscopic images, and hosted 4 years of public challenge benchmarks. In 2016, the first work demonstrating classification accuracy higher than the average of

expert dermatologists was described [11], employing an ensemble of methods that included hand-coded feature extraction, sparse coding methods, support vector machines, CNNs for skin lesion classification, and fully convolutional networks for skin lesion segmentation. Other models have also been implemented by researchers, such as a computationally efficient skin lesion classification model that uses the MobileNet architecture implemented by [9], and an Inception architecture trained on a large dataset of over 100,000 images [13].

Outside of dermatology, there has been work on evaluating fairness in computer vision with respect to skin type. Recent studies evaluated bias in automated facial analysis models with respect to phenotypic groups [5, 26]. They found poor accuracy for darker females compared to lighter females, darker males, and lighter males in gender classification systems. A related study revealed that well-performing gender classification systems are already invariant to skin type and thus the skin type by itself has a minimal effect on classification disparities [25]. Another study investigated equitable performance in state-of-the-art object detection systems on pedestrians with different skin types, finding higher precision on lighter skin than darker skin [33].

3 Datasets

Public benchmark datasets, in addition to fostering direct comparisons among various algorithms to facilitate advancement in terms of classification performance, are also capable of supporting detailed analysis of that performance with respect to various characteristics of the dataset [27]. Therefore, we focus our analysis on two of the most widely used dermatology datasets in the computer vision literature: the ISIC 2018 Challenge dataset the SD-198 dataset.

ISIC2018. This collection of dermoscopic images is separated into datasets for image segmentation (Task 1), clinical feature detection (Task 2), and disease classification (Task 3). Dermoscopic images are acquired through a digital dermatoscope, with relatively low levels of noise and consistent background illumination. The training dataset for Task 3 is the largest among the tasks and used in this work. It consists of 10,015 dermoscopic images [10, 30], falling into one of 7 skin diseases: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma and vascular lesion.

SD-198. The SD-198 dataset contains 6,548 clinical images from 198 skin disease classes, varying according to scale, color, shape and structure downloaded from DermQuest [29]. Clinical images are collected via various devices, most of which are digital cameras and mobile phones [29]. Higher levels of noise and varying illumination in clinical images makes segmentation more challenging than the dermoscopic images in ISIC2018.

In this work, we preprocess the SD-198 dataset to exclude classes containing images with no observable non-diseased skin. Eventually 136 disease classes are retained, and henceforth the pre-processed dataset is referred to as "SD-136".

Some of the classes excluded from the SD-136 dataset include classes of lesions inside the mouth, such as fibroma, geographic tongue, and stomatitis. Other diseases such as arsenical keratosis, pustular psoriasis, and mal perforans contain images of lesions on palms and soles of the feet from which it is difficult to determine the individual's skin tone. Other disease classes such as stasis ulcer and eccrine poroma contain images of severely scarred skin from which it is visually impossible to differentiate non-diseased and diseased skin. This preprocessing step was done manually and eventually 4,467 images were retained from the original 6,548 images.

Since there are no existing ground truth segmentation masks for the SD-136 dataset, we manually segmented a subset of 343 images. We were particularly interested in segmenting regions with non-diseased skin from other regions of the image containing diseased skin, shadows, and other artifacts. We used these ground-truth masks in training the segmentation model for the SD-136 dataset. The data is split into 90%/10% training/validation partitions.

4 Methods

Our proposed method is summarized in Fig. 1. First, we train a model to segment skin disease images to obtain the non-diseased skin in the image, this model returns a set of pairs with the image and the mask associated to it $(I_i, Mask_i)$ for all images in any of the datasets D_1 or D_2. Second, we use the provided $Mask_i$ to select and compute the metric to stratify the non-diseased skin into a skin tone category $(tone_j)$ from the categorization scheme (S_m). After that, a classification model is trained to classify skin images into one of the skin diseases in the dataset. Finally, the performance of the classifier on samples in each skin tone category is evaluated.

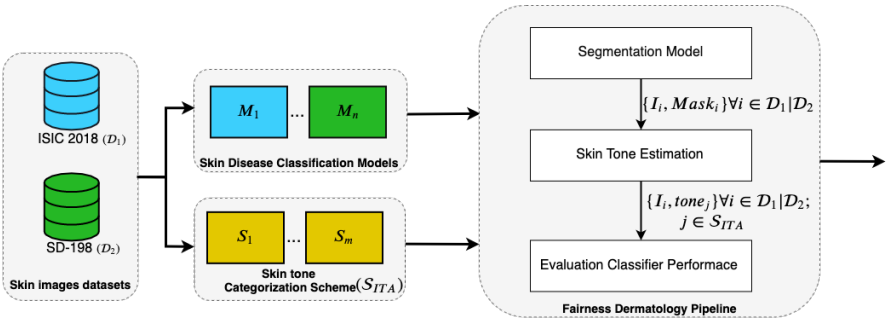


Fig. 1. Block diagram of methodology, where D_1 and D_2 correspond to the datasets ISIC 2018 and SD-198, M_i is a trained model for skin disease classification (e.g. Densenet201) over the previously mentioned datasets, S_m is a categorization scheme (e.g. ITA ranges) and $tone_j$ is a skin tone under the ITA ranges.

4.1 Quantification of Representation of Skin Tone Categories

Segmentation of the skin lesion from the non-diseased skin is done using a Mask R-CNN model [16]. Mask R-CNN was selected because it was one of the top performing skin lesion boundary segmentation models in the 2018 ISIC challenge [17] and also because it has been shown to be highly effective and efficient in performing semantic segmentation [18].

To obtain a segmentation model for the ISIC2018 dataset, a Mask R-CNN pretrained on the COCO dataset is finetuned with the lesion boundary segmentation data from ISIC 2018 Challenge Task 1. The data is split into training and validation data using 90% to 10% train validation split. The images are resized to 600×450 pixels to correspond to the size used for classification. Random horizontal flips are done on the images for data augmentation during training. The segmentation model for the ISIC2018 dataset is trained for 25 epochs. This model is used to predict segmentation masks for all the classification data. Finally, thresholding of the predicted masks is performed via contour extraction.

The segmentation model from the ISIC2018 dataset is finetuned with the 343 manually-segmented images from SD-136. All other steps are the same as for ISIC2018 except the image size is 450×450 and the number of epochs is 50.

The quality of segmentation is evaluated using accuracy and false negative rate. False negative rate is considered because it is worse to wrongly classify a diseased region as non-diseased in our analysis. To further evaluate the quality of segmentation, mean absolute error is computed between the ITA estimates from ground truth masks and ITA estimates from predicted masks.

With the segmentation masks obtained from the previous step, we obtain pixels in the non-diseased region for each image and use these pixels to categorize the skin tone. There is no universal method for characterizing skin type or skin tone among dermatologists. The Fitzpatrick skin type, used in [5, 26, 25], is a dermatologist's determination of a person's risk of sunburn. It is by definition, however, a subjective human determination [12]. In contrast, the melanin index is measured objectively via reflectance spectrophotometry, and has a strong correlation with the Fitzpatrick type and is useful in assigning it [32]. The metric we use to quantify skin tone in this work is the ITA (in degrees) because it has strong (anti-)correlation to the melanin index [32], and can be simply computed from images, making it a practical method for categorizing skin color [24].

The pixels from the non-diseased region are examined in CIELab-space using luminance (L) and the amount of yellow (b). To prevent the effect of outliers, we only consider L and b values within one standard deviation of their mean values in the region. The ITA value is calculated as [24]:

$$\mathrm{ITA} = \arctan\left(\frac{\mathrm{L} - 50}{\mathrm{b}}\right) \times \frac{180^\circ}{\pi}. \tag{1}$$

We bin the mean ITA value using a scheme similar to [6], which uses 5 skin tone categories: Very Light, Light, Intermediate, Tanned, and Dark. We further subdivide the Light, Intermediate, and Tanned categories into two equal ranges, giving a total of 8 ITA categories. The scheme is summarized in Table 1.

Table 1. Skin tone categorization scheme.

ITA Range	Skin Tone Category	Abbreviation
ITA $> 55^\circ$	Very Light	very_lt
$48^\circ <$ ITA $\leq 55^\circ$	Light 2	lt2
$41^\circ <$ ITA $\leq 48^\circ$	Light 1	lt1
$34.5^\circ <$ ITA $\leq 41^\circ$	Intermediate 2	int2
$28^\circ <$ ITA $\leq 34.5^\circ$	Intermediate 1	int1
$19^\circ <$ ITA $\leq 28^\circ$	Tanned 2	tan2
$10^\circ <$ ITA $\leq 19^\circ$	Tanned 1	tan1
ITA $\leq 10^\circ$	Dark	dark

4.2 Evaluation of Classification Performance Across Skin Tones

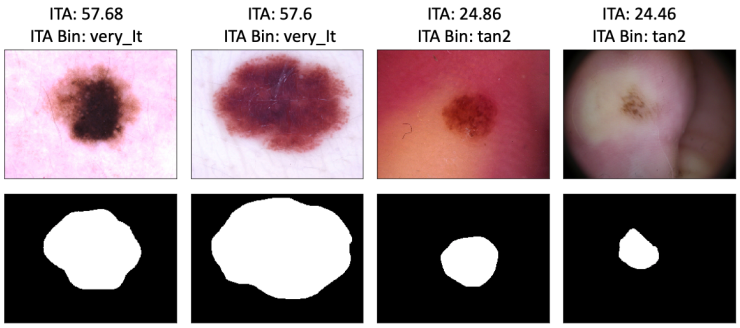
A Densenet201 model pretrained on ImageNet is finetuned using our training data. The Densenet201 model is chosen because it was one of the best performing single models for lesion classification in the ISIC 2018 challenge [17]. During training, the early layers up to and including the first Dense block are frozen and all successive layers have their weights updated. Each classification model is trained for 300 epochs with a patience of 100 epochs at which early stopping would be applied to prevent overfitting.

The ISIC2018 dataset images are maintained at 600×450 pixels. Additional transformations such as random horizontal flipping are applied to augment the data. The samples in each batch are normalized using the mean and standard deviation computed on all samples in the dataset to ensure fast convergence during training. The data is split into training and validation data using an 80%/20% split. A weighted cross entropy loss function and an Adam optimizer are used for training. The weights for the loss function are obtained from the inverse of each disease class frequency. This loss function is chosen because it accounts for class imbalance.

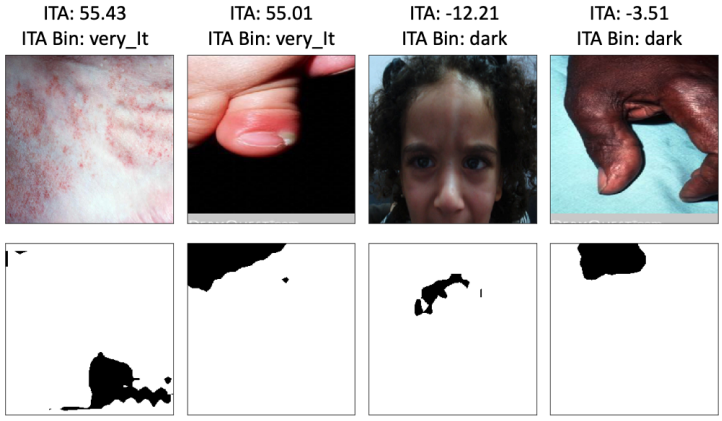
The SD-136 dataset images are resized to 450×450 pixels and center-cropped to 360×360 pixels. Transformations including random horizontal flipping and random rotation between -90° and 90° are applied to augment the data. All other details are the same as in ISIC2018.

5 Results

The Mask R-CNN model used for segmentation on the ISIC2018 dataset yields an accuracy of 0.956, a false negative rate of 0.024, and a mean absolute error in ITA computation of 0.428 degrees. The segmentation model on the SD-136 dataset yield an accuracy of 0.802, a false negative rate of 0.076, and a mean absolute error in ITA computation of 3.572 degrees. These are all fairly good results and sufficient for further analysis. Examples with segmented mask and ITA values for both datasets are shown in Fig. 2.



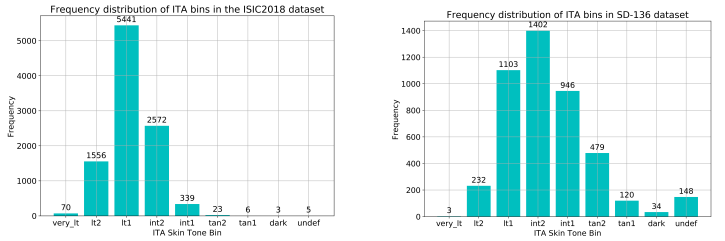
(a)



(b)

Fig. 2. Sample images (top row) and corresponding masks predicted by model (bottom row) for (a) ISIC2018 and (b) SD-136 datasets; ITA is computed on the non-diseased region which is colored black.

Fig. 3 shows the distributions of the ITA values estimated from the non-diseased skin regions of the images in the entire ISIC2018 and SD-136 datasets. Both datasets are found to predominantly lie in the Light category.



(a) (b)

Fig. 3. Skin tone distribution for (a) ISIC2018, and (b) SD-136 entire datasets.

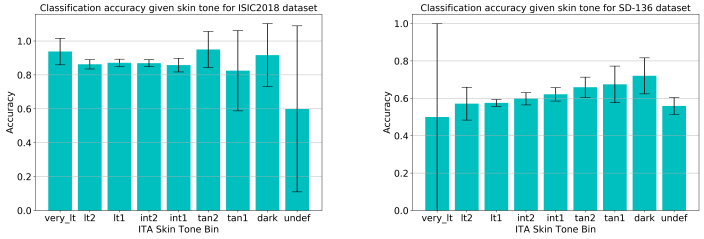
On the ISIC2018 dataset, the Densenet201 model achieves an accuracy 0.869 and a balanced accuracy score of 0.814 on our internal validation partition after training approximately 140 epochs when early stopping occurred. On a separate held-out test set used for the challenge leaderboard, our model achieves a balanced accuracy score of 0.760, placing its percentile ranking around 62%. This indicates that the model scored higher balanced accuracy than 62% of the Top 200 entries in the ISIC 2018 challenge. Unfortunately, since the challenge held-out set is unavailable to us, we cannot disaggregate this result by skin tone.

The model trained on SD-136 achieves an accuracy of 0.604 and a balanced accuracy score of 0.601. The benchmark model for SD-198 achieves an accuracy 0.52, as reported in [29]. However, since we dropped the number of classes from 198 to 136, we do not have a benchmark model for comparison. Nonetheless, we are confident that we have a well-performing model.

Importantly, on evaluating the classification performance with respect to skin tone category, our results do not show a clear trend in the performance of the model. Fig. 4 plots classification accuracy versus ITA for the validation set for the two datasets. The error bars indicate the standard error estimated through ten runs with random splits. The slope of the least squares line of best fit of the mean accuracy versus the midpoint ITA value of the bin for ISIC2018 is -0.000 (per degree) with a 95% confidence interval of $(-0.001, 0.001)$, whereas that for SD-136 is -0.002 (per degree) with a 95% confidence interval of $(-0.003, -0.001)$, which indicate that there are no particular trends in both datasets.

6 Conclusion

In this work, we implemented an approach to measure approximate skin tone distributions in public dermatology image datasets using ITA as an estimator, and evaluated the performance of dermatology classification models with respect



(a) (b)

Fig. 4. Accuracy versus ITA for (a) ISIC2018, and (b) SD-136 validation sets.

to the resultant ITA values. The distribution of ITA values across both ISIC2018 and SD-136 datasets are consistent with under-representation of darker skin tones. The results from the evaluation of the accuracy of the skin classification model for each skin tone category in the validation data shows that there is no observable trend in the performance of the model with respect to ITA value, which is contrary to other studies of skin color and computer vision systems. Although we have not found any evidence of model performance bias under the influence of dataset bias in this particular study, further investigation is needed on datasets with more comprehensive representation.

References

1. Abedini, M., Chen, Q., Codella, N., Garnavi, R., Sun, X., Celebi, M.E., Mendonca, T., Marques, J.S.: Accurate and scalable system for automatic detection of malignant melanoma. In: Celebi, M.E., Mendonca, T., Marques, J.S. (eds.) Dermoscopy Image Analysis. CRC Press (2015)
2. Adamson, A.S., Smith, A.: Machine learning and health care disparities in dermatology. JAMA Dermatol. **154**(11), 1247–1248 (Nov 2018)
3. Barocas, S., Selbst, A.D.: Big data's disparate impact. Calif. Law Rev. **104**(3), 671–732 (Jun 2016)
4. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018. CA-Cancer J. Clin. **68**(6), 394–424 (Nov/Dec 2018)
5. Buolamwini, J., Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Proc. Conf. Fair. Account. Transp. pp. 77–91 (Feb 2018)
6. Casale, G.R., Siani, A.M., Diémoz, H., Agnesod, G., Parisi, A.V., Colosimo, A.: Extreme UV index and solar exposures at Plateau Rosà (3500 m a.s.l.) in Valle d'Aosta Region, Italy. Sci. Total Environ. **512–513**, 622–630 (Apr 2015)
7. Celebi, M.E., Codella, N., Halpern, A.: Dermoscopy image analysis: Overview and future directions. IEEE J. Biomed. Health **23**(2), 474–478 (Mar 2019)
8. Celebi, M.E., Codella, N., Halpern, A., Shen, D.: Guest editorial: Skin lesion image analysis for melanoma detection. IEEE J. Biomed. Health **23**(2), 479–480 (Mar 2019)
9. Chaturvedi, S.S., Gupta, K., Prasad, P.: Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet. arXiv:1907.03220 (Aug 2019)

10. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv:1902.03368 (Mar 2019)
11. Codella, N.C.F., Nguyen, Q.B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. IBM J. Res. Dev. **61**(4/5), 5 (Jul/Sep 2016)
12. Eilers, S., Bach, D.Q., Gaber, R., Blatt, H., Guevara, Y., Nitsche, K., Kundu, R.V., Robinson, J.K.: Accuracy of self-report in assessing Fitzpatrick skin phototypes I through VI. JAMA Dermatol. **149**(11), 1289–1294 (Nov 2013)
13. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (Feb 2017)
14. Gohara, M.: Skin cancer: An African perspective. Brit. J. Dermatol. **173**(Suppl. 2), 17–21 (Jul 2015)
15. Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Ben Hadj Hassen, A., Thomas, L., Enk, A., Uhlmann, L.: Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann. Oncol. **29**(8), 1836–1842 (Aug 2018)
16. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask R-CNN. arXiv:1703.06870 (Jan 2018)
17. International Skin Imaging Collaboration: ISIC 2018: Skin lesion analysis towards melanoma detection (2018), https://challenge2018.isic-archive.com/, available: https://challenge2018.isic-archive.com/
18. Johnson, J.W.: Automatic nucleus segmentation with Mask-RCNN. In: Proc. Comput. Vis. Conf. pp. 399–407 (Apr 2019)
19. Karimkhani, C., Dellavalle, R.P., Coffeng, L.E., Flohr, C., Hay, R.J., Langan, S.M., Nsoesie, E.O., Ferrari, A.J., Erskine, H.E., Silverberg, J.I., Vos, T., Naghavi, M.: Global skin disease morbidity and mortality: An update from the global burden of disease study 2013. JAMA Dermatol. **153**(5), 406–412 (May 2017)
20. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: A review. Artif. Intell. Med. **56**(2), 69–90 (Oct 2012)
21. Kundu, R.V., Patterson, S.: Dermatologic conditions in skin of color: Part i. special considerations for common skin disorders. Am. Fam. Physician **87**(12), 850–856 (Jun 2013)
22. Mahendraraj, K., Sidhu, K., Lau, C.S.M., McRoy, G.J., Chamberlain, R.S., Smith, F.O.: Malignant melanoma in African–Americans: A population-based clinical outcomes study involving 1106 African–American patients from the surveillance, epidemiology, and end result (SEER) database (1988—2011). Medicine **96**(15), e6258 (Apr 2017)
23. Marchetti, M.A., Chung, E., Halpern, A.C.: Screening for acral lentiginous melanoma in dark-skinned individuals. JAMA Dermatol. **151**(10), 1055–1056 (Oct 2015)
24. Merler, M., Ratha, N., Feris, R.S., Smith, J.R.: Diversity in faces. arXiv:1901.10436 (Apr 2019)
25. Muthukumar, V.: Color-theoretic experiments to understand unequal gender classification accuracy from face images. In: Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops (Jun 2019)
26. Raji, I.D., Buolamwini, J.: Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In: Proc. AAAI/ACM Conf. Artif. Intell. Ethics Soc. pp. 429–435 (Jan 2019)

27. Rotemberg, V., Halpern, A., Dusza, S.W., Codella, N.C.F.: The role of public challenges and data sets towards algorithm development, trust, and use in clinical practice. Semin. Cutan. Med. Surg. **38**(1), E38–E42 (Mar 2019)
28. Stoecker, W.V., Moss, R.H.: Editorial: Digital imaging in dermatology. Comput. Med. Imag. Grap. **16**(3), 145–150 (May–Jun 1992)
29. Sun, X., Yang, J., Sun, M., Wang, K.: A benchmark for automatic visual classification of clinical skin disease images. In: Proc. Eur. Conf. Comput. Vis. pp. 206–222 (Oct 2016)
30. Tschandl, P., Rosendahl, C., Kittler, H.: Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**, 180161 (Aug 2018)
31. Varshney, K.R.: Trustworthy machine learning and artificial intelligence. ACM XRDS **26**(3), 26–29 (Spring 2019)
32. Wilkes, M., Wright, C.Y., du Plessis, J.L., Reeder, A.: Fitzpatrick skin type, individual typology angle, and melanin index in an African population. JAMA Dermatol. **151**(8), 902–903 (Aug 2015)
33. Wilson, B., Hoffman, J., Morgenstern, J.: Predictive inequity in object detection. arXiv:1902.11097 (Feb 2019)
34. Wu, X.C., Eide, M.J., King, J., Saraiya, M., Huang, Y., Wiggins, C., Barnholtz-Sloan, J.S., Martin, N., Cokkinides, V., Miller, J., Patel, P., Ekwueme, D.U., Kim, J.: Racial and ethnic variations in incidence and survival of cutaneous melanoma in the United States, 1999-2006. J. Am. Acad. Dermatol. **65**(5), S26.e1–S26.e13 (Nov 2011)