Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 32, Num. 1 (2025) 250-256

RESEARCH ARTICLE

Leveraging Optimal Methods for Diverse Skin Types Classification in Images for Reduced Bias

Utilizando Métodos Ótimos para Classificação de Diferentes Tipos de Pele em Imagens visando Redução de Viés

Luiz Augusto Vieira Manoel[1], Moacir Antonelli Ponti[1,2]*

Abstract: This paper addresses the bias in automatic skin tone classification, highlighting how technologies can mirror societal issues such as inadequate representation of diverse skin tones. The objective is to evaluate methods that ensure balanced performance across skin types despite a scarcity of annotated databases with ethnic-racial details. Using the Fitzpatrick Skin Type classification system, we assess various skin color labeling algorithms, selecting optimal approaches for each type based on f-score. Results show that a single method biases results towards certain skin tones, whereas combining methods enhances accuracy across diverse types. We leverage the LFW and Fitzpatrick17k databases and apply image processing techniques like gamma transformation, CLAHE, histogram equalization, and non-linear order statistics filters. By tailoring processes to specific skin tone ranges, our auto-labeling approach better mirrors manual labeling, aiming for more equitable technology.
Keywords: Skin Tone Classification — Image Processing — Representativeness

Resumo: Este artigo aborda o viés na classificação automática de tons de pele, destacando como as tecnologias podem refletir problemas sociais, como a representação inadequada de diferentes tons de pele. O objetivo é avaliar métodos que garantam desempenho equilibrado entre tipos de pele, apesar da falta de bancos de dados anotados com detalhes étnico-raciais. Utilizando o sistema de classificação de Tipos de Pele de Fitzpatrick, avaliamos vários algoritmos de rotulagem de cores de pele, selecionando as melhores abordagens para cada tipo com base no f-score. Os resultados mostram que um único método tende a gerar viés para certos tons de pele, enquanto a combinação de métodos melhora a precisão em diferentes tipos. Usamos os bancos de dados LFW e Fitzpatrick17k e aplicamos técnicas de processamento de imagens, como transformação gamma, CLAHE, equalização de histograma e filtros de estatísticas de ordem não linear. Ao ajustar os processos para faixas específicas de tons de pele, nossa abordagem de rotulagem automática se aproxima mais da rotulagem manual, buscando uma tecnologia mais justa.
Palavras-Chave: Classificação de Tons de Pele — Processamento de Imagens — Representatividade

[1] *Instituto de Ciências Matemáticas e de Computação. (ICMC), Universidade de São Paulo (USP), Brazil*
[2] *Mercadolibre Inc.*
***Corresponding author**: ponti@usp.br
DOI: http://dx.doi.org/10.22456/2175-2745.143537 • **Received:** 24/10/2024 • **Accepted:** 02/10/2024

1. Introduction

There is a contradiction in society between profit, work and scientific and technological advancement. According to [1], those who have more money have more power and they use this power in search of more money; therefore, scientific and technological advancement can only be understood within the history and concrete conditions of reality.

In [2] a survey shows the impact of cases where Artificial Intelligence and Machine Learning reflect the injustices of the system in which we live. Many papers just apply machine learning techniques without a clear guideline of the practical applicability, rendering unuseful [3]. A clear example of bias is Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), software that measures a person's risk of being a criminal repeat offender, which is used by judges in the US to support the decision of judgments. An investigation into this software found that it points to a higher risk of recidivism for black people than for white people[1].

[1] http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Also in the same survey [2] there are multiple examples of selection bias in word embedding algorithms that make classifiers sexist and more cases of bias related to skin color, such as in the identification of skin cancer, since the databases used in this context have few examples of darker skins.

One of the sources of problems is data. Systems increasingly rely on the availability of data from which it is possible to learn the concepts necessary for a given task. This learning of concepts through examples is known as Machine Learning [4]. However, there is a notable difficulty in obtaining databases that are considered ethnic-racial aspects. In [5] the authors report that there are few image databases that contain annotations by skin color. These labels would be important for studying the impact of selection bias on technologies.

To analyze the skin tone attribute, most studies use the Fitzpatrick Skin Type classification system [6] or the Individual Typology Angle (ITA) [7] to estimate skin type and allow analysis of representativeness in large databases.

The Individual Typology Angle (ITA) is a metric for categorizing human skin tones based on light reflection in the LAB color space˜[8]. Calculated from the L* (lightness) and b* (blue-yellow) values, ITA provides a numerical classification of skin tone, aligning with scales like the Fitzpatrick scale. This standardization enhances consistency and objectivity in skin color analysis for applications in dermatology, cosmetology, and computer vision. The Fitzpatrick scale is a numerical scale of human skin color, initially used to measure the correct dose of ultraviolet radiation for dermatology.

The 2012 survey [9] recollects the most used methods and techniques for the task of automatic skin detection and their numerical evaluation results. As in [10], it is pointed out that the lighting in the image capture interferes a lot in this task and that, for this reason, it is important to develop correction techniques through image processing. In both articles, most of the approaches raised are pixel-based: the algorithm walks through each pixel and makes a binary classification between skin and non-skin. In [10], a detection process is described that takes into account a pre-processing of color correction and face detection so that only the face pixels are considered. The two articles highlighted in this paragraph identify the RGB color system as a leading method for automatic skin color detection. Despite advancements in deep learning for computer vision, significant improvements in automatic skin detection methods remain scarce.

The goal of this work is to develop an automatic skin tone classification method that closely matches manual classification in large image databases, while ensuring consistent performance across various skin types.

2. Method

2.1 Processing steps

We begin by assuming manual skin color labeling is superior. Our aim is an automatic method that aligns closely with manual results. The main steps, illustrated in Figure 1, include:

(a) Use of an algorithm for automatic skin color classification, which will be detailed shortly;

(b) Pre-processing images with variable gamma transformations followed by the algorithm from step (a);

(c) Use CLAHE [11] and histogram equalization on images, followed by the algorithm from step (a);

(d) Conduct triple weightings of the results from steps (a), (b), and (c);

(e) Employ non-linear order statistics filters with varied parameters as a post-processing on pixel-based skin color classification;

(f) Evaluate the f-score for each method to identify the best approach for each skin color range and apply a refined algorithm based on these results.

Fitzpatrick	Color
Type I	Palest
Type II	Light colored but darker than fair
Type III	Golden honey or olive
Type IV	Moderate brown
Type V	Dark brown
Type VI	Deeply pigmented dark brown to darkest brown

Table 1. Fitzpatrick types and definitions

2.2 Detailed Processing Steps

The algorithm in step (a) involves iterating through each pixel of an image, labeling them as skin or non-skin following [9]. Skin pixels are converted to LAB color space to calculate the Individual Typology Angle (ITA) as per [12]. The ITA is normalized and grouped into six Fitzpatrick-based skin tone ranges (ITA1 to ITA6), serving as labels for the images. This procedure applies to the original and pre-processed images from steps (b) and (c).

In step (d), labels from (a), (b), and (c) are weighted to improve accuracy. For each labeled image, if the gamma transformation label matches the CLAHE and histogram equalization label, it is retained; otherwise, the original label is kept.

Step (e) refines the method in (a) by converting images to LAB, calculating ITA values for each skin pixel, and forming a grayscale image. Non-skin pixels are assigned a value of ”0” and normalized. A median filter is applied to the face region, discretizing pixel values into six skin tone ranges. A second median filter may be applied, and the mode of the pixel matrix establishes the image’s final label. An ”absurd” classification assigns all images a label of ”3” for comparison.

In step (f), a new classification method applies previous best F-Score methods for each skin tone range. The method prioritizes skin color ranges $x, y, z, w, v, u \in [1, 2, 3, 4, 5, 6]$. It starts with the highest priority method and proceeds down the list. At the last level, if the classification mismatches the priority range, we use the method with the best average F-Score for darker skin tones (ranges 4, 5, and 6).

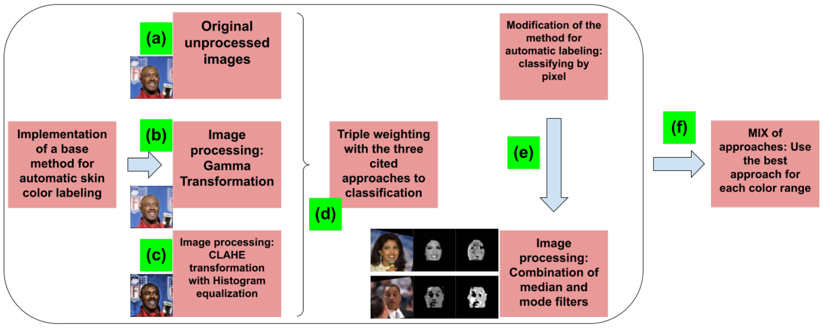

Figure 1. General flow: Auto skin color classification methods.

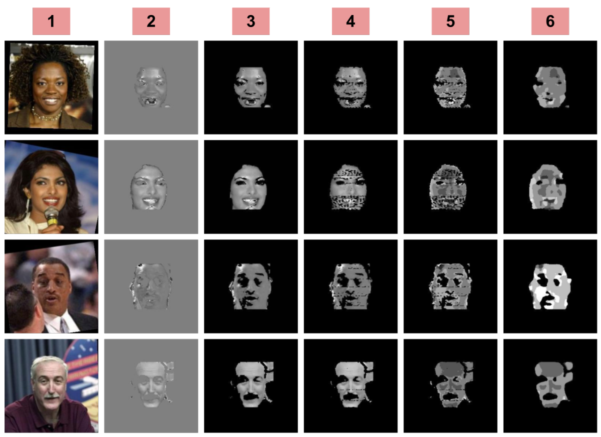

Figure 2. 1) Original image; 2) Mask per pixel; 3) Normalization; 4) Application of the first median; 5) ITA Discretization and 6) Application of a second median.

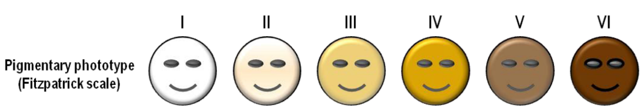

Figure 3. Pigment phototype representation for Fitzpatrick Scale [13].

2.3 Evaluation

We use the Fitzpatrick [6] scale with a division into 6 types, as detailed in Table 1. To evaluate how well each automatic skin color classification approach aligns with manual classification, we used the following metrics:

– Number of Examples Ranked: Counts how many unique images each approach assigns to each skin color ranges;
– HIT Rate: Tallies instances where automatic classification matches the manual sorting for each image;
– Recall for Each Skin Color Range: For a given skin color range, recall is calculated by dividing the number of HITs by the total number of HITs plus false negatives;
– Precision for Each Skin Color Range: For each skin color range, precision is determined by dividing the number of HITs by the total number of HITs plus false positives;
– F-Score for Each Skin Color Range: Combines precision and recall to provide a single performance measure for each skin color range.

2.4 Images Databases

2.4.1 Labeled Faces in the Wild (LFW)

This publicly available database features images of 5749 unique people, totaling 13,233 images. All are aligned and cover a variety of conditions commonly encountered in real life [14]. The database is not annotated by gender, age or skin color. Therefore, we performed a manual labeling of 150 examples to use as ground truth.

#img p.person	N (%)	#imgs (%)
1	4069 (70.8)	4096 (30.7)
2-5	1369 (23.8)	3739 (28.3)
6-10	168 (2.92)	1251 (9.45)
11-20	86 (1.50)	1251 (9.45)
21-30	25 (0.43)	613 (4.63)
31-80	27 (0.47)	1170 (8.84)
≥81	5 (0.09)	1140 (8.61)
Total	**5749**	**13233**

Table 2. Images and people distribution at LFW.

2.4.2 Fitzpatrick17k

This publicly available database features 16,577 clinical skin images from two dermatological databases "DermaAmin" and "*Atlas Dermatologico*" with Fitzpatrick skin color band labels. The annotated images represent 114 skin conditions with at least 53 images and a maximum of 653 images per skin condition [15].

3. Experimental setup

Our method is applied in LFW and Fitzpatrick17k datasets. Table 3 presents the distribution of each Fitzpatrick scale level among the labeled examples. Figure 4 illustrates examples of manual labeling derived from individual interpretations of the Fitzpatrick scale, with each number on the upper left corner of the images indicating the assigned scale level.

# Fitzpatrick scale	sample size
1	19
2	31
3	30
4	20
5	28
6	22

Table 3. Distribution of manually labeled examples by scale.

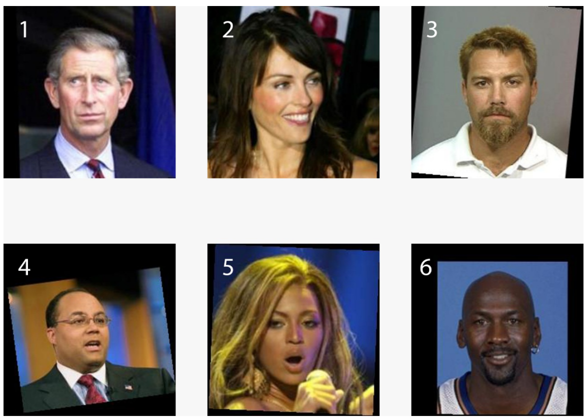

Figure 4. Examples of manual labeling from the interpretation of the Fitzpatrick scale.

For experiments on the LFW dataset, we initiated by detecting faces in all images using the HaarCascades method [16]. The bounding boxes of each detected face were saved as part of our preprocessing pipeline for automatic skin color classification algorithms. This step is crucial because some images contain multiple faces, but our focus is on the skin color of the largest face in each image.

In conducting experiments on both LFW and Fitzpatrick17k, we employed different methods for automatic skin color classification. Each method allowed for multiple variations by adjusting parameters, resulting in 14 distinct approaches. These approaches were evaluated using the metrics outlined in Section 2.3:

1) Method A: no preprocessing (baseline);
2) Method A: gamma transformation, $\gamma = 1.2$;
3) Method A: gamma transformation, $\gamma = 1/1.22$;
4) Method A: CLAHE and histogram equalization ;
5) Method B: weighting between 1), 2) and 4);
6) Method B: weighting between 1), 3) and 4) ;
7) Method C: median filter (3×3 squared);
8) Method C: median filters (3×3 squared then 3×3 disk);
9) Method C: median filters (3×3 squared then 5×5 disk);
10) Method C: median filters (3×3 squared then 9×9 disk);
11) Method C: median filters (3×3 squared then 13×13 dsk);
12) Mode: all images are assigned skin color 3;
13) Method D: following the priority range from darkest to lightest skin color for classification;

LFW - # of examples ranked for each skin color range							
Met.	**ITA1**	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**	**HIT**
##01	42	44	182	115	62	27	103
##02	34	31	148	169	81	19	85
##03	48	74	158	100	56	36	93
##04	31	37	202	144	41	18	84
##05	37	39	177	133	61	27	105
##06	44	38	199	110	56	27	98
##07	25	54	333	28	26	6	133
##08	21	51	343	29	23	4	115
##09	22	60	343	19	24	2	119
##10	20	68	346	16	20	0	124
##11	21	77	338	19	14	0	121
##12	0	0	474	0	0	0	105
##13	15	46	193	133	58	29	112
##14	21	70	181	116	57	29	122
GT	89	146	105	42	54	38	474

Table 4. LFW: Number of ranked examples for each skin color range and HIT.

14) Method D: following the following priority from least represented to most represented skin color range for classification

"Method A" comprises the application of steps (a), (b) and (c) of processing; "Method B" refers to (d), "Method C" to (e) and "Method D" to (f).

Tables 4 through 9 present all the metrics extracted for the 14 approaches discussed earlier. These metrics are applied to subsets of the manually labeled LFW images and the FITZ17K (Fitzpatrick17k) database. To enhance visualization, the best performances for each metric are highlighted in green, with the most vibrant green indicating the top results. Conversely, the poorest performances are highlighted in red, with the most vibrant red indicating the lowest outcomes. Instances where values could not be calculated due to division by zero are marked in black. Notably, the standout results from the "absurd" approach (12) are highlighted in orange.

Figures 5 and 6 illustrate the data from the six central columns of Tables 4 and 7, showing the proportional distribution of classified images for each approach across the six skin color bands in the LFW and FITZ17K datasets, respectively. Here, "GT" stands for ground truth, referring to the verified true information obtained from manual labeling, as opposed to the automatically classified data.

In the graphs in the Figures 7 and 8, the f-score values per skin color range of each approach are added, referring to the application in LFW and FITZ17K respectively.

4. Discussion

During the manual labeling of LFW samples across skin color ranges 1 to 6, identifying individuals with darker skin tones, particularly Black women, was challenging. This is evidenced by the lower number of images for darker-skinned individuals compared to lighter-skinned individuals in the labeled dataset.

The results of Tables 4 to 6 (LFW metrics) and Tables 7 to 9 (Fitzpatrick17k metrics) reveal that no single approach

RECALL						PRECISION						
ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	Met.
0,079	0,082	0,495	0,357	0,204	0,158	0,167	0,273	0,282	0,130	0,177	0,222	##01
0,067	0,062	0,356	0,429	0,204	0,105	0,176	0,290	0,250	0,107	0,155	0,211	##02
0,101	0,200	0,308	0,310	0,056	0,184	0,188	0,392	0,203	0,130	0,054	0,194	##03
0,056	0,048	0,452	0,381	0,093	0,105	0,161	0,189	0,233	0,111	0,122	0,222	##04
0,067	0,082	0,495	0,405	0,222	0,158	0,162	0,308	0,294	0,128	0,197	0,222	##05
0,079	0,048	0,505	0,357	0,167	0,184	0,159	0,184	0,266	0,136	0,161	0,259	##06
0,135	0,151	0,724	0,048	0,019	0,000	0,480	0,407	0,227	0,071	0,038	0,000	##07
0,124	0,151	0,743	0,095	0,000	0,000	0,524	0,431	0,226	0,133	0,000	0,000	##08
0,135	0,192	0,743	0,024	0,000	0,000	0,545	0,459	0,226	0,050	0,000	0,000	##09
0,112	0,226	0,733	0,024	0,000		0,500	0,478	0,221	0,059	0,000		##10
0,101	0,247	0,714	0,024	0,000		0,429	0,462	0,221	0,050	0,000		##11
		1,000						0,222				##12
0,056	0,137	0,486	0,405	0,222	0,184	0,333	0,435	0,264	0,128	0,207	0,241	##13
0,09	0,212	0,457	0,381	0,222	0,184	0,381	0,443	0,265	0,138	0,211	0,241	##14
Recall and Precision - LFW												

Table 5. LFW Metrics: Recall and Precision

f-score						
ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	Met.
0,121	0,170	0,380	0,233	0,190	0,189	##01
0,119	0,165	0,301	0,248	0,179	0,156	##02
0,143	0,289	0,253	0,213	0,055	0,189	##03
0,106	0,114	0,334	0,231	0,107	0,161	##04
0,112	0,184	0,387	**0,251**	0,209	0,189	##05
0,118	0,112	0,375	0,237	0,164	**0,220**	##06
0,285	0,266	0,434	0,059	0,028	0,000	##07
0,294	0,276	**0,439**	0,114	0,000	0,000	##08
0,309	0,312	**0,439**	0,037	0,000	0,000	##09
0,277	0,340	0,433	0,041	0,000	0,000	##10
0,244	**0,346**	0,426	0,037	0,000	0,000	##11
0,000	0,000	0,517	0,000	0,000	0,000	##12
0,178	0,269	0,366	**0,251**	0,214	0,212	##13
0,218	0,317	0,354	0,248	**0,216**	0,212	##14
F-Score - LFW						

Table 6. LFW Metrics: F-Score

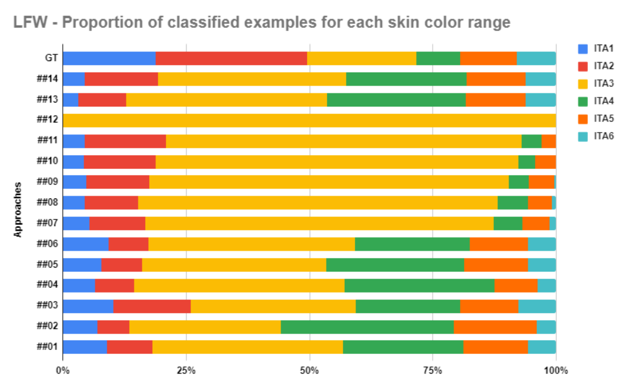

Figure 5. LFW - Ground truth (GT) and proportion of classified examples for each skin color range in each classification approach

FITZ17K - # of examples ranked for each skin color range							
Met.	ITA1	ITA2	ITA3	ITA4	ITA5	ITA6	HIT
##01	1580	2071	5449	4428	1984	988	3812
##02	1394	1669	4663	5385	2276	1109	3878
##03	1812	2602	5176	3792	2001	1120	3329
##04	2041	2620	4075	3751	2675	1344	3068
##05	1532	2085	5193	4704	2042	944	3897
##06	1638	2119	5446	4410	1925	962	3782
##07	670	2126	9890	2491	1080	239	3889
##08	626	2203	9897	2473	1074	203	3906
##09	591	2288	9856	2504	1055	164	3913
##10	573	2404	9824	2481	1010	127	3943
##11	549	2519	9787	2445	981	101	3981
##12	0	0	16525	0	0	0	3299
##13	336	1637	4679	4056	2587	3163	3709
##14	346	1637	4636	4056	2587	3163	3705
GT	2940	4796	3299	2776	1527	628	16525

Table 7. FITZ17K: Number of ranked examples for each skin color range and HIT.

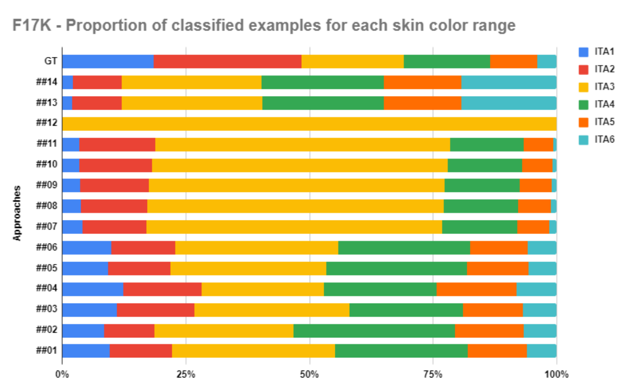

Figure 6. FITZ17K - Ground truth (GT) and proportion of classified examples for each skin color range in each classification approach

RECALL						PRECISION						
ITA1	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**	**ITA1**	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**	**Met.**
0,118	0,152	0,428	0,352	0,184	0,109	0,230	0,370	0,267	0,226	0,146	0,073	##01
0,113	0,134	0,350	0,477	0,221	0,147	0,251	0,405	0,254	0,251	0,154	0,089	##02
0,125	0,180	0,379	0,217	0,121	0,096	0,344	0,251	0,163	0,096	0,056	##03	
0,084	0,187	0,168	0,325	0,261	0,108	0,127	0,359	0,142	0,244	0,152	0,052	##04
0,116	0,159	0,393	0,411	0,194	0,107	0,235	0,383	0,257	0,248	0,150	0,075	##05
0,118	0,155	0,422	0,349	0,180	0,096	0,221	0,367	0,264	0,225	0,148	0,066	##06
0,092	0,185	0,714	0,127	0,016	0,003	0,411	0,427	0,246	0,152	0,023	0,009	##07
0,088	0,191	0,714	0,127	0,015	0,003	0,420	0,427	0,246	0,152	0,022	0,010	##08
0,086	0,197	0,711	0,128	0,013	0,002	0,434	0,424	0,245	0,150	0,019	0,006	##09
0,085	0,207	0,709	0,128	0,013	0,002	0,448	0,421	0,245	0,150	0,019	0,008	##10
0,085	0,218	0,707	0,127	0,011	0,002	0,463	0,424	0,245	0,152	0,018	0,010	##11
		1,000						0,207				##12
0,080	0,170	0,366	0,334	0,288	0,149	0,336	0,417	0,230	0,233	0,146	0,089	##13
0,081	0,171	0,364	0,334	0,289	0,149	0,335	0,417	0,230	0,233	0,146	0,089	##14
Recall and Precision - FITZ17K												

Table 8. FITZ17K Metrics: Recall and Precision

f-score						
ITA1	**ITA2**	**ITA3**	**ITA4**	**ITA5**	**ITA6**	**Met.**
0,171	0,252	0,343	0,286	0,165	0,091	##01
0,178	0,255	0,300	**0,355**	0,187	0,117	##02
0,165	0,257	0,312	0,189	0,108	0,076	##03
0,105	0,267	0,155	0,283	0,204	0,079	##04
0,172	0,261	0,322	0,325	0,172	0,091	##05
0,167	0,252	0,338	0,284	0,164	0,081	##06
0,231	0,295	**0,443**	0,139	0,019	0,006	##07
0,232	0,298	**0,443**	0,139	0,018	0,006	##08
0,236	0,301	0,441	0,139	0,016	0,004	##09
0,240	0,305	0,441	0,139	0,016	0,005	##10
0,246	**0,313**	0,440	0,139	0,014	0,006	##11
0,000	0,000	0,505	0,000	0,000	0,000	##12
0,194	0,282	0,294	0,282	**0,213**	**0,118**	##13
0,195	0,282	0,294	0,282	**0,213**	**0,118**	##14
F-Score - FITZ17K						

Table 9. FITZ17K Metrics: F-Score

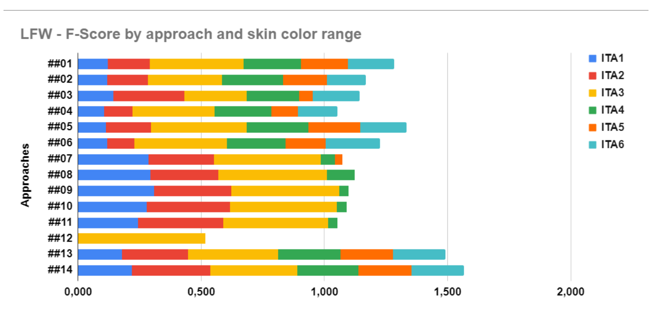

Figure 7. LFW - Sum of F-Scores by skin color range and classification approach

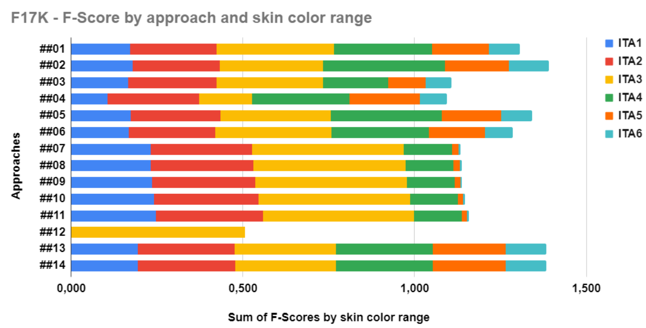

Figure 8. FITZ17K - Sum of F-Scores by skin color range and classification approach

excels in all metrics and ITA bands. Figures 5 and 6 indicate that all approaches tend to underestimate lighter and darker skin tone examples (ITA1, 2, 5, 6) and overestimate those in the medium range (ITA3 and 4). In terms of recall and precision, no approach is superior across all skin color ranges. However, "Method D" (approaches 13 and 14) consistently ranks among the best across both datasets. As shown in Figure 7, Method D achieves the highest f-score on the LFW dataset and ranks among the top three in Fitzpatrick17k (Figure 8).

The underestimation of lighter and darker tones comes from the fact that the rule-based methods have a larger coverage of types 3-4, which further motivates the need for specialized preprocessing methods for each type. Furthermore, our results indicate that lighter skin tones are favored by post-processing methods (Method C) using median filters to smooth the pixels, while darker tone classification is improved by pre-processing, in particular combining Gamma adjustment with histogram equalization (Method B).

Generalizing the proposed method involves: selecting distinct image pre-processing methods and post-processing classification algorithms for automatic skin color labeling, then prioritizing and applying the best approach for each skin color band based on the highest f-score on a validation set.

5. Conclusion

We demonstrate that combining the best approaches for each skin color range significantly improves automatic labeling in large datasets, aligning results more closely with manual labeling. Previous research shows that automatic skin color labeling often yields higher detection rates for lighter-skinned images, likely due to methodologies that favor lighter tones.

Towards reduced bias, our approach prioritizes effective methods across a diverse spectrum of skin types, aiming for a more equitable system that reflects human diversity. This strategy enhances performance and allows straightforward implementation with validation sets for real-world scenarios.

The improved performance in the LFW dataset is likely linked to our pre-processing step involving face detection, which helps identify skin color pixels by minimizing background noise and lighting issues. This approach enables more accurate labeling, especially for darker-skinned individuals, who are often underrepresented.

Overall, our findings confirm the need for inclusive methods in automatic skin tone classification, contributing to fairer technologies that serve diverse populations and promote equity in artificial intelligence applications.

Future work may focus on expanding the diversity of datasets for skin tone classification by collecting more annotated images of individuals with darker skin tones and other underrepresented groups. Implementing real-time testing in practical applications is also important. Additionally, engaging with communities will ensure our technologies address their specific needs and promote fair representation in automated systems.

Acknowledgements

This work was supported by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) scholarship, and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) fellowship #315158/2023-9.

Author contributions and Disclaimer

L.A.V. Manoel contributed to the conception of the research project, data organization and labeling, execution of experiments, and writing of the manuscript. M.A. Ponti contributed to the conception of the research project, supervised the experiments, and participated in writing and revising the final version of the manuscript. This paper was published in the context of the University of São Paulo (USP) and does not reflect the opinions of MercadoLibre on the subject.

References

[1] MOROZOV, E.; MARCONDES, C. *Big Tech: a ascensão dos dados e a morte da política*. [S.l.]: UBU, 2018. ISBN 9788571260122.

[2] MEHRABI, N. et al. A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*, 2019.

[3] SILVA, J. M. C. da; REZENDE, P. M. B.; PONTI, M. A. Detecting and mitigating issues in image-based COVID-19 diagnosis. In: PMLR. *ICML Workshop on Healthcare AI and COVID-19*. Baltimore, USA, 2022. p. 127–135.

[4] MELLO, R. F.; PONTI, M. A. *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Cham: Springer, 2018.

[5] MERLER, M. et al. Diversity in faces. *arXiv preprint arXiv:1901.10436*, 2019.

[6] FITZPATRICK, T. B. Soleil et peau. *J Med Esthet*, v. 2, p. 33–34, 1975.

[7] CHARDON, A.; CRETOIS, I.; HOURSEAU, C. Skin colour typology and suntanning pathways. *International journal of cosmetic science*, Wiley Online Library, v. 13, n. 4, p. 191–208, 1991.

[8] PONTI, M.; NAZARÉ, T. S.; THUMÉ, G. S. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. *Neurocomputing*, Elsevier, v. 173, p. 385–396, 2016.

[9] PREMA, C.; MANIMEGALAI, D. Survey on skin tone detection using color spaces. *International Journal of Applied Information Systems*, Citeseer, v. 2, n. 2, p. 18–26, 2012.

[10] HARVILLE, M. et al. Consistent image-based measurement and classification of skin color. In: *IEEE Int. Conf. on Image Processing*. Genoa, Italy: [s.n.], 2005. p. II–374.

[11] PIZER, S. M. et al. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, Elsevier, v. 39, n. 3, p. 355–368, 1987.

[12] KINYANJUI, N. M. et al. Estimating skin tone and effects on classification performance in dermatology datasets. *arXiv preprint arXiv:1910.13268*, 2019.

[13] D'ORAZIO, J. et al. Uv radiation and the skin. *International journal of molecular sciences*, Multidisciplinary Digital Publishing Institute, v. 14, n. 6, p. 12222–12248, 2013.

[14] HUANG, G. B.; LEARNED-MILLER, E. *Labeled Faces in the Wild: Updates and New Reporting Procedures*. Amherst, USA, 2014.

[15] GROH, M. et al. Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. *arXiv preprint arXiv:2104.09957*, 2021.

[16] VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Computer Vision and Pattern Recognition*. Kauai, HI, USA: [s.n.], 2001. p. 1063–6919.