

Modelo para Detecção e Reconhecimento de Tons de Pele baseado em Inteligência Artificial e Processamento de Imagens

Janaina Ferreira Santos¹, Gabriel dos Santos Mendonça¹, Alcides X. Benicasa¹

¹Departamento de Sistemas de Informação – Universidade Federal de Sergipe (UFS)
Caixa Postal – 49.506-036 – Itabaiana – SE – Brasil

janainaferreira53@hotmail.com, gabrieldossantosmendonca0@gmail.com,

alcides@academico.ufs.br

Abstract. *Computer Vision and Machine Learning have enabled the development of more accurate systems for detecting patterns in images, such as skin tone classification — a task relevant to product personalization and the analysis of racial diversity. However, low accuracy for darker skin tones and lack of diversity in datasets contribute to racial bias. This work proposes a model based on the Monk scale, combining facial detection, rule-based skin segmentation, and color extraction using K-means, comparing the results to a reference palette. Using an experimental and quantitative approach, the model achieved 99% accuracy in controlled environments and 80% in low-light conditions.*

Resumo. *A Visão Computacional e Aprendizado de Máquina têm possibilitado sistemas mais precisos para detecção de padrões em imagens, como a classificação de tons de pele, relevante para personalização de produtos e análise de diversidade racial. Contudo, a baixa acurácia para tons escuros e a falta de diversidade nos dados geram vieses raciais. Este trabalho propõe um modelo baseado na escala de Monk, combinando detecção facial, segmentação de pele por regras e extração de cor com K-means, comparando os resultados a uma paleta de referência. Com abordagem experimental e quantitativa, o modelo atingiu 99% de acurácia em ambientes controlados e 80% em baixa iluminação.*

1. Introdução

Nos últimos anos, o avanço em Visão Computacional e Aprendizado de Máquina possibilitou a criação de sistemas eficazes para detecção e reconhecimento de padrões em imagens, incluindo a identificação de tons de pele. Essa tecnologia é aplicada em diversas áreas, como personalização de produtos de beleza e análise de diversidade racial.

Trabalhos como os de [Borza et al. 2017] e [Manoel 2022] exploram a detecção de tons de pele para aplicações em visagismo e redução de viés em bases de dados. Estudos também demonstram a eficácia dos espaços de cores RGB, HSV e YCbCr na detecção de pixels de pele, enquanto [Borza et al. 2017] investiga modelos de mistura gaussiana e SVM, ressaltando a importância do espaço de cores Lab.

Embora a detecção automática de pele seja um tema recorrente na literatura [Ali et al. 2021], enfrenta desafios devido à variação de iluminação e características individuais, tornando a classificação subjetiva. Além disso, a precisão de modelos em

Visão Computacional é crítica, especialmente frente a relatos de viés racial em tecnologias de análise facial, como evidenciado em APIs da Microsoft, IBM e Face++ [Buolamwini and Gebru 2018].

Neste sentido, um importante ponto a ser observado, e destacado no trabalho de [Buolamwini and Gebru 2018], é que o padrão utilizado para categorizar tons de pele foi baseado em um estudo de 1975, desenvolvido pelo dermatologista de Harvard Thomas Fitzpatrick, a escala é composta por 6 cores, e tem a finalidade de avaliar a sensibilidade UV de diferentes tipos de pele para fins dermatológicos. [Buolamwini and Gebru 2018] destaca ainda que, como resultado, a escala se inclina para tons mais claros, que tendem a ser mais sensíveis aos raios UV, porém, por outro lado, sua utilização em aplicações de análise facial, seu desempenho foi substancialmente pior em imagens contendo faces mais escuras.

Tais problemas evidenciam a dificuldade recorrente na obtenção de bases de dados verdadeiramente representativas, o que acaba por gerar modelos enviesados, com desempenho superior para tons de pele mais claros. Sendo assim, propomos aqui o desenvolvimento de um modelo inteligente para detecção e reconhecimento de tons de pele, baseado na escala Monk e, a partir de entradas de imagens reais e estáticas, obter resultados equilibrados de acurácia para diferentes tons de pele.

2. Revisão Bibliográfica

2.1. Escala de tons de pele

A escala de tons de pele mais comumente utilizada na área da saúde e da ciência da computação é a escala Fitzpatrick, desenvolvida em 1975 pelo dermatologista Harvard Thomas Fitzpatrick, visando originalmente a análise de sensibilidade UV para determinação de doses de fototerapia ultravioleta A, ela incluía quatro tipos de pele, sendo atualizada em 1988, com a adição de mais dois tipos para incluir indivíduos com tons de pele mais escuros [Fitzpatrick 1988, *apud* [Howard et al. 2021]].

No entanto, apesar do seu amplo uso, estudos recentes, como o de [Howard et al. 2021], mostram que essa escala não é confiável nem preditiva, e não deve ser usada em avaliações de aplicações de visão computacional. Em especial, a escala não foi projetada para avaliar o tom da pele, em vez disso, a pele, o cabelo, os olhos e histórico de queimaduras solares foram considerados para avaliar a probabilidade de queimaduras na pele durante a fototerapia [Schumann et al. 2023].

Visando uma melhor representatividade dos tons de pele, a Escala de Tons de Pele Monk (MST) foi apresentada em um trabalho recentemente proposto por [Monk 2022], em parceria com o Google, sendo composta por 10 tons.

2.2. Visão Computacional e Processamento de Imagens

Os procedimentos em Visão Computacional frequentemente precisam de uma fase de pré-processamento que engloba o processamento de imagens. Em alguns casos, as imagens das quais desejamos extrair informações precisam ser ajustadas para um formato ou tamanho específico, além de serem filtradas para eliminar ruídos provenientes do processo de aquisição da imagem [Marengoni and Stringhini 2010].

Para [Gonzalez and Woods 2000] uma imagem pode ser definida como uma função bidimensional $f(x, y)$, composta por um determinado número de elementos denominados *pixels*, tais elementos possuem valores para x e y que são coordenadas de localizações e o valor de f associado é chamado de intensidade ou nível de cinza, o processamento dessas imagens por um computador é definido como o campo de Processamento de Imagens Digitais.

Normalmente, uma imagem capturada por uma câmera de vídeo é apresentada em cores, e cada pixel da imagem é representado por um conjunto de valores, conhecidos como canais de cores, que geralmente variam entre três e quatro componentes, apresentados à seguir. Um modelo de cores é um termo que denota um modelo abstrato para descrever a representação de cores como conjunto de números, chamados componentes de cores ou coordenadas de cores.

De acordo com [Kolkur et al. 2017], o espaço RGB, é composto por três componentes: vermelho (*red*), verde (*green*) e azul (*blue*). Todas as outras cores são derivadas da combinação desses três. O YCbCr é um sinal RGB não linear codificado. Neste formato, as informações de luminância são armazenadas como um único componente (Y) e as informações de crominância são armazenadas como dois componentes de diferença de cor (Cb e Cr). O espaço de cores HSV é mais intuitivo na forma como as pessoas experimentam as cores do que o espaço de cores RGB, todos os seus componentes variam de 0 a 1.

A diferença entre duas cores representadas no modelo RGB pode ser calculada usando a distância Euclidiana, contudo, quando duas cores são metaméricas, o resultado pode não evidenciar tal similaridade. Para este tipo de cores é preciso que quando a distância seja calculada as identifique como cores próximas, dessa forma surge, em 1976, o sistema CIE-Lab que estabelece uniformidade em um espaço de cor tridimensional [Scuri 1999].

O modelo de cor uniforme CIELab foi definido pela CIE na tentativa de aumentar a uniformidade das cores percebidas pelo sistema visual humano. O valor de luminosidade L varia de branco a preto, o valor de a pode variar de verde a vermelho e o valor de b de azul a amarelo, de maneira análoga à percepção das cores pelo cérebro, ou seja, cores-opostas. O modelo de cor CIELab funciona como um tradutor universal de línguas entre os dispositivos, permitindo controlar as cores que passam de um dispositivo para outro, correlacionando os valores em RGB com os valores em Lab [Leão et al. 2005]. Diante disso, passou-se a buscar maior uniformidade nas diferenças de cores através da melhoria nas fórmulas de diferenças de cores no espaço CIELab, ao longo dos anos várias foram propostas [Ridolf 2012]. Abaixo é exibida a Equação 1, que define a fórmula de diferença perceptual de cores CIEDE2000, usada para aferir pequenas diferenças de cores.

$$\Delta E = \sqrt{\left(\frac{\Delta L}{S_L}\right)^2 + \left(\frac{\Delta C}{S_C}\right)^2 + \left(\frac{\Delta H}{S_H}\right)^2 + R_T \left(\frac{\Delta C}{S_C}\right) \left(\frac{\Delta H}{S_H}\right)}, \quad (1)$$

onde, ΔL , ΔC e ΔH : são as diferenças nos componentes de luminosidade, croma e tonalidade entre as duas cores, S_L , S_C e S_H são os fatores de escala ajustados dependendo do valor da luminosidade e da saturação das cores e R_T é o fator de correção de interação

entre croma e tonalidade, que ajusta a fórmula para levar em conta a interação perceptiva entre as diferenças de croma e tonalidade.

2.3. Detecção de Face e Segmentação de Pele

A detecção facial é um passo importante para os algoritmos de análise facial. Para [Kumar et al. 2019] a detecção facial como processo computacional age identificando a posição e o tamanho de um rosto em uma imagem digital. Ela foca nas características visíveis do rosto, ignorando outros elementos. A detecção facial pode ser considerada uma versão mais ampla da localização facial, onde a tarefa é identificar as posições e tamanhos de um número conhecido de rostos, geralmente apenas um. Essa técnica pode ser vista como um subconjunto da detecção de objetos.

Diversos métodos foram propostos ao longo dos anos, com avanços notáveis nas últimas décadas [Minaee et al. 2021]. Nesta seção, com base em estudos na área, será abordado o Método de Detecção Facial baseado em Redes Neurais Convolucionais (MTCNN, do inglês *Multi-Task Cascaded Convolutional Neural Networks*) [Zhang et al. 2016].

O Método de Detecção Facial baseado em Redes Neurais Convolucionais, proposta por [Zhang et al. 2016], é um método de detecção facial baseado em CNN, implementado em múltiplas etapas. O modelo se divide em três etapas consecutivas, localizando inicialmente a região facial de maneira não refinada e, em seguida, refinando essa detecção nas etapas subsequentes, incluindo a identificação de pontos de referência faciais como olhos, nariz e boca.

A arquitetura MTCNN, inicia com a *Proposal Network* (P-Net), uma *Fully Convolutional Network* (FCN), uma rede convolucional porém sem as camadas densas, responsável por detectar rapidamente janelas candidatas e vetores de regressão de caixas delimitadoras. Em seguida, as janelas altamente sobrepostas são combinadas e inseridas na *Refine Network* (R-Net), uma rede que possui uma camada densa que refina e rejeita janelas sem faces. A última etapa, similar à R-Net, descreve o rosto com mais detalhes e gera posições dos pontos de referência faciais [Zhang et al. 2016].

A segmentação de imagem é o processo de dividir uma imagem em regiões distintas com base em características específicas. O nível de detalhamento da segmentação pode variar conforme a tarefa e a resolução da imagem disponível [Marengoni and Stringhini 2010]. Um método comum é a limiarização, um algoritmo intuitivo e rápido no processamento de imagens, que classifica cada pixel dentro de uma faixa específica de cor. A segmentação é baseada na estimativa da média das cores de interesse, que pode ser obtida através de um conjunto representativo de amostras dessas cores, esse método separa pixels de interesse do fundo, criando uma imagem binária [Gonzalez and Woods 2000].

A segmentação de pele é uma aplicação específica da segmentação de imagens, que visa identificar *pixels* correspondentes à pele humana. No entanto, a segmentação de pele enfrenta desafios como variações de iluminação, diferentes dispositivos de captura e a ampla variação na cor da pele.

3. Metodologia

De um modo geral, o modelo proposto inicia com a recepção da imagem, seguida pela detecção de face. Caso uma face seja identificada, o processo continua com a segmentação da pele na área detectada e, a partir do segmento gerado é realizada a extração das cores predominantes. Essas cores são então comparadas com uma paleta de referência (baseada na escala *Monk*) através do cálculo de distâncias, e, finalmente, o resultado da classificação dos tons de pele é apresentado. Se nenhuma face for detectada, o processo é interrompido, sinalizando a ausência de uma face na imagem analisada.

3.1. Aprendizado e geração de padrões de referências da Escala Monk

O conjunto de dados de acesso aberto proposto por [Schumann et al. 2023] foi composto por 1.515 imagens e 31 vídeos de 19 pessoas, incluindo indivíduos de diferentes etnias e gêneros, para ajudar a separar o conceito de tom de pele de raça, abrangendo a escala de 10 pontos, sendo que cada pessoa foi fotografada em diferentes poses e expressões faciais. É importante notar que, as imagens deste *benchmark* foram revisadas e anotadas usando a escala de *Monk*, idealizada pelo próprio autor, Dr. Ellis Monk, especialista em percepção social e desigualdade. Considerando sua validade, tal *benchmark* será amplamente utilizado nos experimentos apresentados na seção seguintes.

A paleta de cores *Monk* (Seção 2.1) foi utilizada como referência, entretanto, após diversos experimentos e análises, percebeu-se que o tom de pele de uma pessoa pode variar, significativamente, mesmo entre pessoas classificadas como de uma mesma escala.

Neste sentido, a escala *Monk*, mesmo que contemplando uma diversidade de 10 tons, cada tom é representado por um único padrão específico referente à cada tonalidade de pele. Tal consideração nos levou a uma abordagem mais diversificada, de forma a propormos aqui uma variação da escala de *Monk* com uma maior representatividade referente a cada tonalidade.

Assim, para a construção da paleta de cores, consideramos uma variedade de 10 padrões pertencentes a uma mesma tonalidade, extraídos do conjunto de dados mencionado anteriormente. Para cada cor com maior predominância extraída da segmentação da pele na imagem de entrada foi inserida um novo padrão de variação para a tonalidade em questão. Neste caso, a representatividade aumentou de 1 para 10 padrões de representatividade por tonalidade de cor.

3.2. Detecção e segmentação facial

Para o reconhecimento de pele, é necessário analisar apenas regiões de interesse que apresentem informações suficientes, como a face. Portanto, a detecção facial e a subsequente segmentação são realizadas inicialmente para eliminar possíveis ruídos presentes na imagem e facilitar os procedimentos subsequentes, além disso, partiu-se do pressuposto de que a face estará na posição frontal e com boa iluminação.

Sendo assim, para a detecção de faces será aplicado o método baseado em uma CNN, o MTCNN, disponível na biblioteca de mesmo nome. Para sua utilização, foi definida uma função que recebe uma imagem contendo uma face, inicializa um objeto MTCNN, cujos atributos padrões são, *min_face_size*, que define o tamanho mínimo da face a ser detectada (o padrão é 20), *steps_threshold*, que define os limiares de confiança

para as três redes (P-Net, R-Net, O-Net, com padrão $[0.6, 0.7, 0.7]$), *scale_factor*, que é o fator de escala entre níveis de imagem (o padrão é 0.709), e *thresholds*, que são os limiares de confiança para a detecção de faces (o padrão é $[0.6, 0.7, 0.7]$).

Em seguida, é utilizada a funcionalidade *detect_faces* para localizar a face na imagem recebida e retornar suas coordenadas. Além disso, a função retorna as coordenadas dos olhos esquerdo e direito, nariz e cantos da boca. Com essas coordenadas, a face é recortada em uma nova imagem e retornada para a próxima etapa, a segmentação da pele, descrita na seção seguinte.

Após a detecção da face, a próxima etapa é a segmentação da pele presente na região facial identificada. Esta segmentação visa eliminar *pixels* que não apresentem informações de pele, para que se possa isolar as regiões de pele da face para facilitar a análise subsequente de tons de pele.

Para realizar esta segmentação, utilizou-se o método baseado em regras, que aplica uma série de condições específicas aos valores RGB para identificar os *pixels* que correspondem à pele. Para prevenir a ocorrência de falsos negativos em peles com tons mais escuros, onde muitos *pixels* referentes à pele acabam sendo retirados da imagem, os valores limites para RGB da regra 1 foram diminuídos, saindo de $R > 95$, $G > 40$ e $B > 20$ para $R > 50$, $G > 30$ e $B > 20$. Com isso, obtém-se um *range* maior de valores.

Definimos a função *segmentar_pele_regras*, que implementa este método. De maneira descritiva, primeiramente a imagem contendo uma face é convertida do espaço de cores BGR, que é o padrão para imagens carregadas utilizando a biblioteca OpenCV, para o RGB. Em seguida, são aplicadas quatro regras baseadas nos valores RGB para identificar *pixels* que representam pele. A primeira delas define limites mínimos para os valores RGB, a próxima verifica a diferença máxima entre os canais de cores, a regra seguinte calcula a diferença absoluta entre os canais vermelho e verde e a última regra compara os valores dos canais vermelho, verde e azul. Por fim, as regras são combinadas para criar uma máscara final que indica os *pixels* de pele. A máscara resultante é então aplicada na imagem original para isolar as regiões de pele da face.

3.3. Extração das cores predominantes

Com a segmentação de pele realizada, o próximo passo é a extração das cores predominantes na região segmentada. Para isso, utilizamos o algoritmo de clusterização *K-means*, que é adequado para agrupar *pixels* de cores similares em *clusters*.

O algoritmo *K-means* é aplicado diretamente nos *pixels* da região segmentada, especificando um número de *clusters*, k , correspondente ao número de cores que desejamos identificar. Após a execução do *K-means*, obtemos os centros dos *clusters*, que representam as cores predominantes na imagem segmentada.

As cores dos *clusters* são ordenadas pela frequência de ocorrência, de forma que a cor com maior ocorrência fique na primeira posição da lista, contendo os valores RGB e a porcentagem de ocorrência na imagem.

3.4. Comparaçao com paleta de referência

Por fim, após a extração das cores predominantes, o passo seguinte é comparar essas cores com a paleta de referência baseada na escala *Monk*. Para tanto, foi aplicada a métrica ΔE ,

que quantifica a diferença perceptual entre duas cores. A métrica ΔE é calculada entre cada cor predominante extraída e cada tom da paleta de referência. A cor da paleta que resultar na menor distância ΔE em relação à cor predominante é selecionada como o tom de pele classificado.

Após detalhar as soluções desenvolvidas para os componentes do modelo, assim como as técnicas empregadas para atingir os objetivos propostos, a próxima seção apresentará os experimentos realizados

4. Resultados

Nesta seção, será demonstrado como o modelo opera quando todas as funcionalidades são executadas em conjunto.

Aqui, cada imagem selecionada deverá passar por uma série de etapas em sequência, sendo as seguintes: redimensionamento da imagem para a resolução de 480x640px, detecção facial, segmentação da pele e extração e classificação das três cores predominantes da área segmentada.

Sobre esta última etapa, para cada uma das cores predominantes será calculada a similaridade com as cores de uma escala específica de tons de pele, composta por 12 exemplos para cada tom, coletados a partir das próprias pessoas representadas na escala. Por fim, a partir das distâncias calculadas torna-se possível a identificação do tom de pele com maior similaridade, apresentando então a predição do modelo.

Sobre este experimento, serão analisadas as previsões para cada tom de pele, sendo utilizadas 10 imagens por pessoa, sendo que a maioria dos tons são representados com imagens de 2 pessoas, totalizando então 20 imagens por tonalidade. Apenas os tons 7 e 10 apresentam apenas 1 pessoa como exemplo, tendo então 10 imagens. Ao todo o experimento fará uso de 180 imagens. Além disto, a fim de observar o comportamento do modelo para imagens com baixa iluminação, serão consideradas 3 imagens por pessoa nestas condições.

De uma maneira mais detalhada, tomando como exemplo uma imagem pertencente ao Tom 1, apresentada na Figura 1, pode ser observada a imagem de entrada (à direita), a face com a pele segmentada e as 3 cores predominantes, ordenadas por suas respectivas proporções. Dentre elas, a com menor proporção é descartada, visto que na maioria dos casos traz informações de sombra, ou *pixels* que não foram descartados durante a etapa de segmentação.

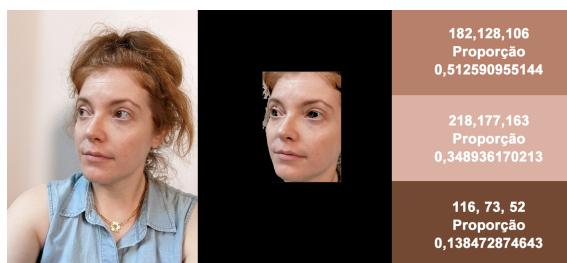


Figura 1. Entrada, segmento e cores predominantes.

Na Tabela 1 são apresentados os resultados de similaridades das duas cores com maiores predominâncias. Vale observar que, conforme previamente descrito, não existe

Tabela 1. Similaridades com a Escala Monk para as Cores Predominantes

Monk ID	Monk 01	Monk 02	Monk 03	Monk 04	Monk 05	Monk 06	Monk 07	Monk 08	Monk 09	Monk 10
Cor Pred. 1	0.1310	0.2588	1.7584	1.5707	1.2276	1.9804	4.9907	8.5137	9.4095	9.9956
Cor Pred. 2	0.0000	2.9770	1.1656	1.6862	3.8681	11.1631	14.7194	20.3125	21.5603	21.9967

apenas uma cor para cada tom, mas sim 12 cores. Por fim, a tonalidade que apresentar a menor distância (ou maior similaridade) será a tonalidade de classificação.

Os experimentos foram realizados de maneira similar para todas as imagens de todos os tons de pele. A matriz de confusão deste experimento é apresentada na Figura 2(a), onde é possível visualizar todos os resultados obtidos.

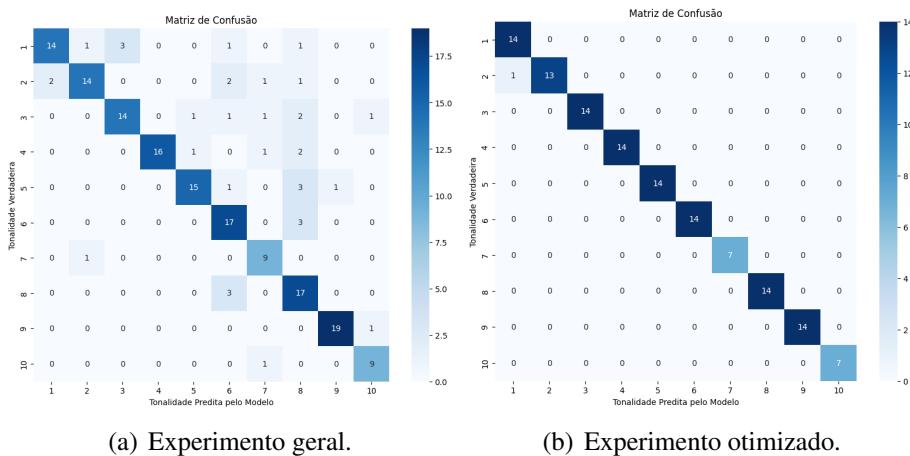


Figura 2. Matriz de Confusão.

Neste experimento geral, o modelo alcançou uma acurácia de 80%, indicando que, em média, 80% das previsões realizadas foram corretas. A precisão do modelo foi de 82%, refletindo uma boa habilidade em identificar corretamente as classes, especialmente aquelas com maior representatividade. Além disto, pode ser observado que na matriz de confusão a maioria das tonalidades foi corretamente classificada, com a maior parte dos erros concentrada em classes adjacentes, sugerindo que o modelo, apesar do erro, os falsos positivos estão localizados próximos à classificação correta.

Ademais, os resultados indicam que o modelo alcançou, satisfatoriamente, mesmo em condições ambientais desfavoráveis. Neste caso, vale destacar que a iluminação exerce uma influência significativa no desempenho do modelo, uma vez que, quando a imagem apresenta uma iluminação muito intensa, há uma tendência de identificar o tom de pele como mais claro. Por outro lado, em ambientes com pouca luz, o modelo tende a classificar o tom de pele como mais escuro. Como consequência, o número de acertos para os tons de pele mais escuro neste experimento foi maior, uma vez que consideramos também imagens com baixa iluminação.

Considerando o comportamento do modelo em um ambiente possível de ser controlado, o experimento foi executado novamente para avaliar seu desempenho em condições mais favoráveis, excluindo imagens com iluminação baixa ou muito alta. Com essa filtragem, restaram 14 imagens para cada tom, exceto para os tons 7 e 10, que no-

vamente ficaram com apenas 7 imagens cada, devido à limitação do banco de dados, que possui imagens de apenas uma pessoa para esses tons.

Uma nova matriz de confusão foi construída para este cenário, na qual o modelo demonstrou um desempenho significativamente aprimorado, classificando corretamente a maioria das imagens (ver Figura 2(b)). O único erro ocorreu ao classificar uma imagem do tom de pele 2 como sendo do tom 1. Considerando que esses tons são visualmente muito próximos, até mesmo para observadores humanos, esse erro é compreensível e reflete a dificuldade natural em diferenciar nuances tão similares.

Neste cenário, o modelo alcançou uma acurácia e precisão de 99%. O único erro cometido foi na classificação de uma imagem em uma classe adjacente, o que ressalta a robustez do modelo em condições otimizadas, sugerindo que a iluminação é um fator importante para o seu desempenho.

5. Conclusões

O presente projeto de pesquisa propôs o desenvolvimento de um modelo inteligente capaz de identificar e classificar de maneira justa os diferentes tons de pele com base na escala Monk, e a partir de imagens reais e estáticas. Para isto foi utilizada uma rede neural que detecta faces, complementada por métodos de segmentação de pele para reduzir ruídos e técnicas de extração de cores, incorporando conceitos de clusterização e similaridade de cores para a classificação.

Os resultados, detalhados nas seções anteriores, demonstraram que o modelo foi eficaz na detecção das faces e na classificação dos tons de pele, tanto em ambientes gerais quanto em ambientes controlados, tendo alcançado os objetivos propostos nos dois cenários. A eficácia do modelo foi condicionada pela qualidade e representatividade dos exemplos utilizados para o treinamento, evidenciando que a precisão é altamente dependente da representatividade dos dados.

Contudo, algumas limitações foram identificadas durante a implementação e avaliação do modelo. A detecção facial, embora funcional, requer condições ambientais ideais, e a extração de cores mostrou um aumento na taxa de erro sob condições de iluminação inadequada. Além disso, a utilização da escala como única referência para o aprendizado mostrou-se pouco representativa quando comparada a diversidade das peles reais.

Uma possível melhoria para resolver este problema pode ser o uso somente de imagens padronizadas com iluminação adequada, sem nenhum tipo de adereço que interfira na extração das cores. Por outro lado, caso o objetivo seja tornar o modelo mais robusto, uma possível solução está em modificar o algoritmo que extrai as cores predominantes, para que seja identificado o problema da iluminação e sua correção.

Para trabalhos futuros, propõe-se a adaptação do modelo para condições ambientais variadas, incluindo diferentes níveis de luminosidade. A exploração de redes neurais para comparação de cores é também uma direção promissora.

Referências

Ali, H., Alnafaakh, H., Ghazali, R., El abbadi, N., and El Abbadi, N. (2021). A review of human skin detection applications based on image processing. Bulletin of Electrical

Engineering and Informatics, 10:129–137.

- Borza, D., Borza, D., Nistor, S. C., Nistor, S. C., Dărăbant, A. S., and Darabant, A. S. (2017). Towards automatic skin tone classification in facial images. International Conference on Image Analysis and Processing.
- Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency, pages 77–91. PMLR.
- Fitzpatrick, T. B. (1988). The Validity and Practicality of Sun-Reactive Skin Types I Through VI. Archives of Dermatology, 124(6):869–871.
- Gonzalez, R. C. and Woods, R. E. (2000). Processamento de imagens digitais. Editora Blucher.
- Howard, J. J., Sirotin, Y. B., Tipton, J. L., and Vemury, A. R. (2021). Reliability and validity of image-based and self-reported skin phenotype metrics. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(4):550–560.
- Kolkur, S., Kalbande, D., Shimpi, P., Bapat, C., and Jatakia, J. (2017). Human skin detection using rgb, hsv and ycbr color models. arXiv preprint arXiv:1708.02694.
- Kumar, A., Kaur, A., and Kumar, M. (2019). Face detection techniques: a review. Artificial Intelligence Review, 52:927–948.
- Leão, A. C., Araújo, A., and Souza, L. A. C. (2005). Implementação de sistema de gerenciamento de cores para imagens digitais. Teixeira A, Barrére E, Abrão IC. Web e multimídia: desafios e soluções. Poços de Caldas: PUC-Minas.
- Manoel, L. A. V. (2022). Reduzindo viés em classificação de tons de pele em bases de dados de imagens. PhD thesis, Universidade de São Paulo.
- Marengoni, M. and Stringhini, S. (2010). Tutorial: Introdução à visão computacional usando opencv. Revista de Informática Teórica e Aplicada, 16(1):125–160.
- Minaee, S., Luo, P., Lin, Z., and Bowyer, K. W. (2021). Going deeper into face detection: A survey. CoRR, abs/2103.14983.
- Monk, E. (2022). Monk skin tone scale. <https://skintone.google/>.
- Ridolf, L. F. G. G. M. (2012). Construo de Espaços de Cor Euclidianos e Perceptualmente Uniformes com base na fórmula CIEDE2000. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- Schumann, C., Olanubi, G. O., Wright, A., Monk Jr, E., Heldreth, C., and Ricco, S. (2023). Consensus and subjectivity of skin tone annotation for ml fairness. arXiv preprint arXiv:2305.09073.
- Scuri, A. E. (1999). Fundamentos da imagem digital. Pontifícia Universidade Católica do Rio de Janeiro, page 13.
- Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503.