



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

MATHEUS THIAGO DOMINGOS DA SILVA

**VIESES EM ALGORITMOS DE RECONHECIMENTO FACIAL: UM
MAPEAMENTO SISTEMÁTICO DA LITERATURA**

CAMPINA GRANDE - PB

2024

MATHEUS THIAGO DOMINGOS DA SILVA

**VIESES EM ALGORITMOS DE RECONHECIMENTO FACIAL: UM
MAPEAMENTO SISTEMÁTICO DA LITERATURA**

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador : Tiago Massoni

CAMPINA GRANDE - PB

2024

MATHEUS THIAGO DOMINGOS DA SILVA

**VIESES EM ALGORITMOS DE RECONHECIMENTO FACIAL: UM
MAPEAMENTO SISTEMÁTICO DA LITERATURA**

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA:

Tiago Massoni

Orientador – UASC/CEEI/UFCG

Hyggo Oliveira de Almeida

Examinador – UASC/CEEI/UFCG

Francisco Vilar Brasileiro

Professor da Disciplina TCC – UASC/CEEI/UFCG

Trabalho aprovado em: 15 de Maio de 2024.

CAMPINA GRANDE - PB

RESUMO

O reconhecimento facial é uma ferramenta bastante popular em aplicativos, utilizada principalmente na autenticação. Devido à sua popularidade, surgiu o questionamento se ela é justa em reconhecer faces de pessoas de diferentes gêneros e etnias. Então, foi desenvolvido um mapeamento sistemático para analisar viés em algoritmos de reconhecimento facial. Estudando diversos artigos científicos, foi possível identificar que os sistemas de uso comercial apresentam viés, evidenciando que a base de dados usada pelo algoritmo não é representativa o suficiente para pessoas de diferentes nacionalidades. Isso foi mostrado através de diversos experimentos, indicando que a solução mais eficiente para mitigar o viés é uma base de dados equilibrada. Este artigo serve como base para que estudos futuros busquem soluções práticas e eficientes.

BIASES IN FACIAL RECOGNITION ALGORITHMS: A SYSTEMATIC MAPPING OF THE LITERATURE

ABSTRACT

Facial recognition is a very popular tool in applications, mainly used in authentication. Due to its popularity, the question has arisen whether it is fair in recognizing faces of people of different genders and ethnicities. Then, a systematic mapping was developed to analyze bias in facial recognition algorithms. Studying several scientific articles, it was possible to identify that commercial use systems present bias, showing that the database used by the algorithm is not representative enough for people of different nationalities. This has been shown through several experiments, indicating that the most efficient solution to mitigate bias is a balanced database. This article serves as a basis for future studies to seek practical and efficient solutions.

Vieses em algoritmos de reconhecimento facial: Um mapeamento sistemático da literatura

Matheus Thiago Domingos da
Silva

Universidade Federal de Campina Grande
CEEI - Centro de Engenharia Elétrica e
Informática
Campina Grande, Brasil

matheus.silva@ccc.ufcg.edu.br

RESUMO

O reconhecimento facial é uma ferramenta bastante popular em aplicativos, utilizada principalmente na autenticação. Devido à sua popularidade, surgiu o questionamento se ela é justa em reconhecer faces de pessoas de diferentes gêneros e etnias. Então, foi desenvolvido um mapeamento sistemático para analisar vieses em algoritmos de reconhecimento facial. Estudando diversos artigos científicos, foi possível identificar que os sistemas de uso comercial apresentam vieses, evidenciando que a base de dados usada pelo algoritmo não é representativa o suficiente para pessoas de diferentes nacionalidades. Isso foi mostrado através de diversos experimentos, indicando que a solução mais eficiente para mitigar o viés é uma base de dados equilibrada. Este artigo serve como base para que estudos futuros busquem soluções práticas e eficientes.

Keywords

Reconhecimento Facial, Viés, Algoritmo, IA, Machine Learning, Discriminação, Mapeamento Sistemático, IA ética, reconhecimento facial ético.

1. INTRODUÇÃO

A inteligência artificial tornou-se centro de grande debate e discussão nos últimos anos devido à sua alta expansão e aplicabilidade, utilizada em áreas como saúde [22], carros autônomos [23] e segurança [24]. Com a evolução da IA, as relações entre humano e computação estão sofrendo uma mudança drástica, devido à automação de tarefas. As pessoas estão perdendo espaço para máquinas inteligentes, surgindo discussões sobre a substituição do ser humano no mercado de trabalho e o aumento do desemprego.

Uma das áreas em que o uso da IA está atuando diretamente é a segurança, que se destaca cada vez mais. Com a evolução de técnicas de hacking avançadas, surge a necessidade de usar ferramentas para reforçar a segurança. Um dos métodos mais populares é o uso do reconhecimento facial na autenticação em sistemas de segurança, o que permite garantir que terceiros não acessem contas bancárias de forma indevida. O reconhecimento facial não se limita apenas ao uso na segurança, mas se expandiu para outros setores, como em sistemas que selecionam candidatos para vagas de emprego [2].

Com a popularidade do uso do reconhecimento facial, surgiu o questionamento se esses sistemas são justos para pessoas de diferentes gêneros [1],[6],[7] e grupos étnicos [8],[9], validando se possuem algum viés em sua origem, que pode estar no algoritmo [21] ou na base de dados [8],[16], que não é representativa o suficiente para determinados grupos de pessoas. Esse tipo de viés no reconhecimento facial pode levar à criação de sistemas injustos e parciais, resultando em reconhecimento incorreto. Muitos desses sistemas são sub-sistemas de outros mais completos, gerando um problema em cadeia.

Para analisar com mais detalhes os estudos relacionados ao viés em algoritmos de reconhecimento facial, foi elaborado um mapeamento sistemático [20] com o objetivo de conhecer os trabalhos relacionados à área, analisando quais os tipos de vieses, se os principais sistemas comerciais (IBM, Microsoft e Face++) são tendenciosos, quais grupos de pessoas são afetados e quais as principais soluções que os pesquisadores propõem para mitigar esse problema. Para realizar esse mapeamento, foram usados sites de publicações acadêmicas voltadas para engenharia da computação, como IEEE, ACM e Google Scholar, utilizando strings de busca e métodos de busca snowball. No final, foram selecionados 16 artigos científicos de acordo com os critérios de inclusão e exclusão.

Os resultados do mapeamento sistemático mostraram que os principais sistemas de reconhecimento facial são tendenciosos para pessoas do gênero masculino e pele clara (caucasiana), apresentando uma maior taxa de sucesso no reconhecimento facial, enquanto grupos de pessoas do sexo feminino e etnia africana apresentam maiores taxas de erros. Para reduzir esse problema, é proposta a utilização de uma base de dados inclusiva, onde grupos de pessoas sejam representados de maneira igualitária e equilibrada, e o uso de algoritmos treinados em machine learning utilizando técnicas para reduzir o viés.

Ao finalizar este trabalho, podemos ver que os estudos sobre o viés em algoritmos que atuam na análise facial são recentes. Seu debate cresceu devido à popularidade da IA, e os estudos estão em busca de uma solução eficiente e prática. Este mapeamento serve como ponto de partida para que os pesquisadores desenvolvam métodos e ferramentas para os diversos problemas mostrados neste estudo, impactando positivamente nossa sociedade de maneira mais justa e inclusiva.

2. METODOLOGIA DE PESQUISA

Para fins científicos, foi desenvolvido um mapeamento sistemático com o objetivo de realizar um estudo da arte do viés em algoritmos de reconhecimento facial. Este estudo envolve a análise de trabalhos relacionados ao tema, analisando quais são os principais algoritmos e sistemas comerciais, verificando se são justos, parciais e se há discriminação, validando se as pessoas de diferentes gêneros e grupos étnicos são bem representadas em sistemas de reconhecimento facial. Ao final, os trabalhos são agrupados de acordo com sua área de estudo e o tipo de viés.

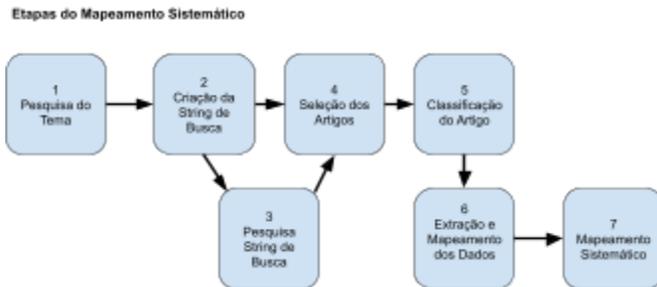


Figura 1: processo do Mapeamento Sistemático

Para realização desse mapeamento, foram seguidos os seguintes passos, conforme mostrado na Figura 1, como descrito abaixo:

1. Na primeira etapa, foi realizada uma pesquisa geral no Google Acadêmico, utilizando como busca o tema "viés em algoritmos de reconhecimento facial".
2. Os artigos retornados na etapa anterior foram submetidos a uma triagem, das quais foram extraídas as principais palavras-chave, que serviram de base para a criação da string de busca.
3. Utilização da string de busca da etapa 2 para pesquisar novos artigos.
4. Seleção dos artigos utilizados no mapeamento sistemático.
5. Classificação de todos os artigos de acordo com o tipo de estudo.
6. Realização da extração dos dados e informações importantes usados na análise deste trabalho.
7. Finalização do mapeamento sistemático.

Do passo 1 ao 4, são etapas referentes à obtenção dos artigos para estudo; do 4 ao 7, são referentes à análise e resultados.

2.1 String de Busca

Para realizar o mapeamento, precisamos buscar os possíveis artigos de estudo. Primeiro, foi estabelecido um método de busca, no qual foi criada uma string de busca para realizar pesquisa nos principais sites de artigos científicos disponíveis na internet.

Na criação da string de busca, é necessário fazer uma inferência dos possíveis termos da busca. Então, o tema do estudo foi estruturado em várias subáreas, sendo atribuídas as principais palavras-chave correspondentes ao assunto. Ao final desse processo, a seguinte string de busca foi definida, conforme mostrado abaixo:

("facial recognition" OR "facial expression transformations" OR "facial detection") AND ("Discrimination" OR "bias" OR "digital inclusion" OR "algorithmic bias") AND ("AI" OR "machine learning" OR "Deep learning" OR "Dataset") AND -("emotions" OR "education" OR "Emotion Recognition")

Figura 2: String de busca

A pesquisa foi realizada nos principais sites de publicação científica focados na área de engenharia de software e ciência da computação, que são: ACM Digital Library [17], IEEE Xplore [18] e Google Scholar [19].

Após a seleção dos possíveis artigos relevantes para o estudo, foi aplicada a técnica de snowballing, realizando novas pesquisas com base nas referências dos artigos selecionados.

2.2 Critérios de Inclusão e Exclusão

Para a escolha dos artigos, foram estabelecidos critérios de exclusão, pois ao realizar a pesquisa, muitos artigos não eram relevantes para o estudo. Para a exclusão, foram estabelecidos dois critérios, que são:

- Artigos cujo estudo principal não envolve "viés em algoritmos de reconhecimento facial", sendo apenas um subtópico do trabalho principal.
- Artigos que analisam "viés em algoritmos de reconhecimento facial" em aplicações que não têm impacto social relevante.

Da mesma forma, foram estabelecidos critérios para a inclusão, que são:

- Trabalhos que incluam em seu tema principal o estudo de IA, viés e reconhecimento facial; qualquer outro trabalho que deixe de abordar algum desses campos foi desconsiderado. Por exemplo, muitos trabalhos abordavam o viés em algoritmos de IA, mas foram excluídos porque não eram usados no reconhecimento facial.
- Artigos que apresentem alguma contribuição que ajude a mensurar ou reduzir o viés.

3. RESULTADOS

Neste trabalho, foi utilizada a string de busca mostrada na figura 2 nos principais sites de busca (ACM, IEEE e Google Scholar) mencionados anteriormente. E foram aplicados os critérios de inclusão e exclusão descritos no tópico anterior, como podemos ver a seguir:

ACM: aplicando a string de busca e pesquisando palavras-chave nos resumos dos artigos para encontrar possíveis trabalhos utilizados no estudo, foram retornados um total de 11 artigos, dos quais foram selecionados 3. IEEE: ao pesquisar usando a string de busca, foram retornados 68 resultados, dos quais foram selecionados 9 artigos. Diferente da pesquisa anterior, a busca por palavras-chave foi feita sem nenhuma restrição, buscando em todo o conteúdo do trabalho. Google Scholar: encontrar artigos neste site que se enquadrem nos requisitos estabelecidos no mapeamento de busca foi um grande desafio devido à alta quantidade de trabalhos retornados, mesmo usando a string de

busca (Figura 2), pois a grande maioria não era relevante. Por isso, utilizamos o método de busca chamado pesquisa em snowball¹ [20]. No final desse processo, foram selecionados 4 artigos.

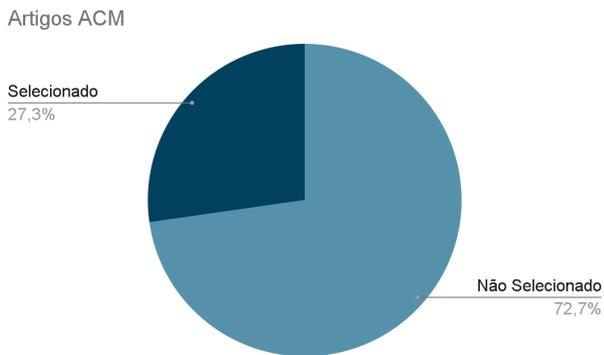


Gráfico 1: Artigos Pesquisados no ACM

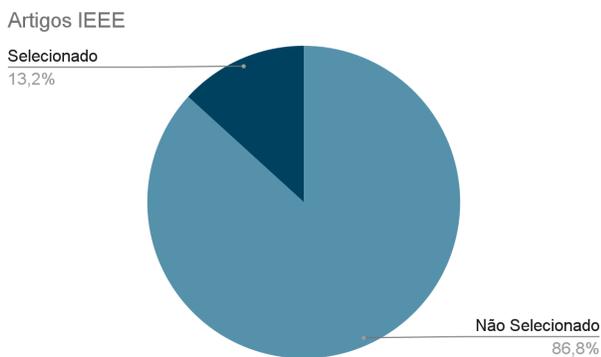


Gráfico 2: Artigos Pesquisados no IEEE

Ao terminar o processo de pesquisa, foram selecionados no total 16 artigos, conforme mostrado na tabela abaixo:

Evaluation of Gender Bias in Facial Recognition with Traditional Machine Learning Algorithms [1]
Ethical Considerations in AI-Based Recruitment [2]
Measuring Bias in AI Models: An Statistical Approach Introducing N-Sigma [3]
Reduction of Bias for Gender and Ethnicity from Face Images using Automated Skin Tone Classification [4]

¹ O método de pesquisa snowball é uma forma de rastrear trabalhos relacionados usando a bibliografia ou lista de referências no final de um artigo como ponto de partida.

Information-Theoretic Bias Assessment Of Learned Representations Of Pretrained Face Recognition [5]
Implicit Stereotypes in Pre-Trained Classifiers [6]
Gendered Differences in Face Recognition Accuracy Explained by Hairstyles, Makeup, and Facial Morphology [7]
FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation [8]
Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition [9]
Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure [10]
Face Mis-ID: An Interactive Pedagogical Tool Demonstrating Disparate Accuracy Rates in Facial Recognition [11]
How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services [12]
Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning [13]
Addressing Racial Bias in Facial Emotion Recognition [14]
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification [15]
Racial Faces in-the-Wild: Reducing Racial Bias by Information Maximization Adaptation Network [16]

Tabela 1: Artigos Selecionados

4. ANÁLISE DOS DADOS

O reconhecimento facial é realizado através de redes neurais convolucionais no contexto da IA. Como a ferramenta foi desenvolvida por pessoas, consequentemente, haverá viés, ou seja, o indivíduo pode rotular atributos de características faciais de acordo com sua visão e opinião, o que faz com que seja transmitido algum nível de preconceito ao sistema, realizando reconhecimentos tendenciosos [25].

Para estudar o viés em algoritmos de análise facial, foram selecionados 16 artigos, extraindo informações importantes sobre os principais pontos da área de estudo e classificando-os em duas grandes áreas de pesquisa: uma focada em mensurar o nível de viés e outra especializada na sua redução.

Estudos especializados em medir o viés procuram analisar os principais sistemas de uso comercial de reconhecimento facial, mensurando através de variáveis estatísticas [3],[9], se é justo

reconhecer pessoas de diferentes gêneros [1],[7],[12],[14] e grupos étnicos [2],[3],[5],[9],[11],[14].

O principal desafio encontrado é mensurar o viés em grandes sistemas comerciais de visão computacional [1], como Microsoft, IBM ou Face++, que são criticados pela sua precisão assimétrica em sub-demografias em estudos recentes, nos quais os algoritmos tiveram o pior desempenho em mulheres do que em homens [6][7]. Sistemas que usam classificação de gênero estão cada vez mais populares sendo usados para diversas finalidades, como vigilância [24], interação humano-computador e publicidade [1].

As principais técnicas usadas para realizar o reconhecimento facial são Redes Neurais Convolucionais (CNN) utilizando diversos modelos que incluem VGG, ResNet, InceptionNet e implementação de rede (Network Implementation) com ajuste fino (Fine-tuning) [1]. Utilizando como datasets o UTKFace e o FairFace [8]. Outro estudo [6] usa uma abordagem que vai além do tradicional, utilizando não apenas atributos protegidos como gênero e cor da pele, mas estereótipos ocultos como atração, riqueza e inteligência, verificando se o sistema demonstra algum viés significativo quando se consideram atributos ocultos.

Para encontrar o algoritmo com melhor desempenho na classificação de gênero, avaliando qual tem menos viés, foram selecionados cinco algoritmos de classificação [1],[16]: Support Vector Classifier (SVC), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Trees (DT) e Logistic Regression (LR). Utilizando uma base de dados faciais [1],[16] equilibrada e outras duas desequilibradas chamadas de feminino dominante (FD) e masculino dominante (MD), descobriu-se que são necessárias mais imagens faciais de mulheres para treinar algoritmos de machine learning, pois o maior número de imagens femininas não afeta negativamente a precisão. Em outras palavras, os algoritmos testados em dados equilibrados e MD apresentaram maior taxa de erro em mulheres em comparação com os conjuntos de dados FD. Algoritmos LR e LDA são vulneráveis e tendenciosos para indivíduos do sexo feminino, enquanto os algoritmos DT, SVC e KNN atenuam o preconceito de gênero [1],[16].

Viés de gênero também foi encontrado em outros estudos [6], [7], [12], nos quais a causa do viés não se limita apenas ao tamanho da amostra facial feminina, mas à própria fisiologia da face, pois existem diferenças físicas nítidas entre homens e mulheres. Devido às características únicas como tamanho do cabelo, maquiagem e tamanho da mandíbula (menores em mulheres), resulta em uma fração menor da imagem, gerando menos informação e dificultando a análise correta da imagem, causando maiores erros nas mulheres, como mostrado no artigo [7].

Outro tipo de viés encontrado foi étnico, causado por uma base de dados não representativa e equilibrada o suficiente para diferentes tipos de etnia e raça. Os datasets VGGFace2, CASIA-WebFace e MS-Celeb-1M possuem 70% das imagens faciais de pessoas de origem caucasiana, como mostrado nos artigos [8] e [16]. Esse desequilíbrio nas amostras faciais é um dos fatores da taxa de erro no reconhecimento facial ser maior em pessoas de origem africana, asiática e indiana.

Além de medir, é necessário encontrar uma variável eficiente e de fácil entendimento para avaliar o viés. No mapeamento realizado, foi proposto o uso de uma variável estatística N-Sigma, que é usada também na física e em áreas sociais, mas é inexplorada na

IA [3]. Foi avaliado em modelos com base em dados tendenciosos para etnias, mostrando ser mais interpretável e fácil de usar [3].

No mapeamento, também foi encontrada uma categoria de artigos que busca reduzir ao máximo o viés, tornando o reconhecimento facial mais justo com um nível de precisão igual para diferentes grupos étnicos e sociais, ou seja, um reconhecimento igual entre homens e mulheres [6],[10],[15], pessoas de pele clara e mais escura, asiáticas e africanas [4],[8],[10],[13],[15],[16].

O método principal para diminuir a taxa de erro e viés no reconhecimento facial é usar uma base de dados equilibrada para diferentes tipos de gênero e etnias, chamada RFW² [16]. Mesmo ao usar uma base de dados equilibrada, ainda é possível identificar um pior desempenho entre grupos étnicos não caucasianos, mostrando que a origem do viés não é apenas causada pela distribuição dos dados, mas também pelo algoritmo utilizado [6] [16].

Devido à dificuldade de eliminar o viés, resta apenas realizar estudos para mitigá-lo, principalmente os ocultos, que são desconhecidos e difíceis de serem identificados por cientistas nos dados de treinamento. Para isso, foram desenvolvidas algumas técnicas específicas, como o uso de algoritmos para maximização de informação [16], algoritmos de treinamento ajustáveis para reduzir o viés [10] e ferramentas avançadas para identificar a tonalidade da pele [4].

Para reduzir a porcentagem de viés em sistemas de reconhecimento facial, foi desenvolvido um algoritmo não supervisionado chamado Information Maximization Adaptation Network (IMAN), que usa pessoas de origem caucasiana (com melhor desempenho) como domínio de origem e outras raças (com pior desempenho) como domínios de destino. Ele usa pessoas caucasianas para aprender características profundas que serão usadas como atributos no domínio alvo, melhorando seu desempenho [16]. Utilizando em conjunto a base de dados RFW e o algoritmo IMAN, foi possível alcançar um desempenho de 94.78%, 94.15%, 91.15% e 91.42% para Caucasiano, Indiano, Asiático e Africano, respectivamente [16].

Outra técnica para mitigar o viés é a criação de um algoritmo de aprendizagem profunda [10], que aprende com a distribuição latente de maneira não supervisionada. Esse algoritmo busca identificar preconceitos ocultos ou implícitos nos conjuntos de dados de treinamento e utiliza variáveis latentes aprendidas para ajustar a amostra, tornando-a mais equilibrada [10]. Ao final do processo, observa-se uma melhora significativa no reconhecimento facial para grupos de pessoas menos representados, sem sacrificar o desempenho em outros grupos que possuem alta precisão.

Por último, uma ferramenta que mapeia a tonalidade da pele utiliza uma Rede Neural Convolucional (CNN) para identificar faces no banco de dados. Em seguida, emprega a segmentação da pele com base nos limites do espaço de cores HSV em imagens faciais para obter a pele do rosto e, por fim, aplica o algoritmo K-means para determinar a cor predominante da pele do rosto [4]. Ao ser usada em pessoas de diferentes grupos étnicos, observou-se um aumento de 3,0% a 8,0% no desempenho geral

² Racial Faces in-the-Wild é uma base de dados equilibrada com imagens faciais de pessoas caucasianas, asiáticas, africanas e indianas.

em comparação com outros classificadores comerciais (IBM, Microsoft e Face++).

Alguns trabalhos [6], [10], [15] e [16] não se limitam apenas a uma área de estudo; utilizam ambos os métodos descritos acima para solucionar o problema do viés. Primeiro, buscam medir o quanto o sistema é injusto e depois aplicam técnicas para reduzir o viés. Em seguida, validam se o novo método é mais eficiente que o anterior.

Analisando todos esses dados, podemos destacar que os principais sistemas comerciais de reconhecimento facial possuem viés de gênero. No total, seis artigos [1], [4], [6], [7], [12], [14] estudam técnicas a nível de dados ou algoritmo para tornar a análise mais justa entre homens e mulheres. Existe também o viés étnico, onde oito artigos [5], [6], [8], [10], [13], [14], [15], [16] demonstram, através de experimentos, que o reconhecimento facial apresenta maior taxa de sucesso para pessoas de pele mais clara.

E por último, a informação mais relevante do mapeamento são as soluções para amenizar o viés. No total, sete artigos [1], [7], [8], [10], [13], [14], [16] apresentam como solução um banco de dados equilibrado e representativo, usado pelo algoritmo na análise facial. A maioria das propostas se concentra na base de dados, devido à facilidade de implementação. Poucos artigos [10], [16] se concentram em eliminar o viés no algoritmo, apenas treinando para melhorar seu desempenho.

5. CONCLUSÃO

Analisando todos os dados e informações dos estudos e artigos usados no mapeamento sistemático, é possível perceber que o viés em algoritmos de reconhecimento facial ainda é uma área pouco explorada, apenas ganhando destaque nos últimos anos. Os estudos ainda estão em fase experimental, propondo métodos e técnicas para amenizar o problema do viés.

Os sistemas de reconhecimento facial disponibilizados pelas grandes empresas possuem viés e são tendenciosos, causando impactos sociais que muitas vezes passam despercebidos pelos engenheiros responsáveis pelo desenvolvimento do sistema.

Os desafios são grandes, mas é necessário que o tema seja melhor discutido e se torne mais popular no meio científico, para que grandes empresas desenvolvam ferramentas e técnicas, a fim de tornar seus sistemas mais justos e inclusivos, contribuindo para uma sociedade mais igualitária.

6. REFERÊNCIAS

- [1] M. Atay, H. Gipson, T. Gwyn and K. Roy, "Evaluation of Gender Bias in Facial Recognition with Traditional Machine Learning Algorithms," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 2021, pp. 1-7, doi: 10.1109/SSCI50451.2021.9660186.
- [2] D. F. Mujtaba and N. R. Mahapatra, "Ethical Considerations in AI-Based Recruitment," 2019 IEEE International Symposium on Technology and Society (ISTAS), Medford, MA, USA, 2019, pp. 1-7, doi: 10.1109/ISTAS48451.2019.8937920.
- [3] D. DeAlcala, I. Serna, A. Morales, J. Fierrez and J. Ortega-Garcia, "Measuring Bias in AI Models: An Statistical Approach Introducing N-Sigma," 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), Torino, Italy, 2023, pp. 1167-1172, doi: 10.1109/COMPSAC57700.2023.00176.
- [4] D. A. Molina, L. Causa and J. Tapia, "Reduction of Bias for Gender and Ethnicity from Face Images using Automated Skin Tone Classification," 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2020, pp. 1-5.
- [5] J. Li and W. Abd-Almageed, "Information-Theoretic Bias Assessment Of Learned Representations Of Pretrained Face Recognition," 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Jodhpur, India, 2021, pp. 1-8, doi: 10.1109/FG52635.2021.9667053.
- [6] N. Dehouche, "Implicit Stereotypes in Pre-Trained Classifiers," in IEEE Access, vol. 9, pp. 167936-167947, 2021, doi: 10.1109/ACCESS.2021.3136898.
- [7] V. Albiero, K. Zhang, M. C. King and K. W. Bowyer, "Gendered Differences in Face Recognition Accuracy Explained by Hairstyles, Makeup, and Facial Morphology," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 127-137, 2022, doi: 10.1109/TIFS.2021.3135750.
- [8] K. Kärkkäinen and J. Joo, "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation," 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021, pp. 1547-1557, doi: 10.1109/WACV48630.2021.00159.
- [9] I. Dominguez-Catena, D. Paternain and M. Galar, "Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2024.3361979.
- [10] Alexander Amini, Ava P. Soleimany, Wilko Schwarting, Sangeeta N. Bhatia, and Daniela Rus. 2019. Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19). Association for Computing Machinery, New York, NY, USA, 289–295. <https://doi.org/10.1145/3306618.3314243>
- [11] Daniella Raz, Corinne Bintz, Vivian Guetler, Aaron Tam, Michael Katell, Dharma Dailey, Bernease Herman, P. M. Krafft, and Meg Young. 2021. Face Mis-ID: An Interactive Pedagogical Tool Demonstrating Disparate Accuracy Rates in Facial Recognition. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21). <https://doi.org/10.1145/3461702.3462627>
- [12] Morgan Klaus Scheuerman, Jacob M. Paul, and Jed R. Brubaker. 2019. How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 144 (November 2019), 33 pages. <https://doi.org/10.1145/3359246>
- [13] Ignacio Serna, Aythami Morales, Julian Fierrez, Nick Obradovich, Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep

learning, *Artificial Intelligence*, Volume 305, 2022, 103682, ISSN0004-3702, <https://doi.org/10.1016/j.artint.2022.103682>.

- [14] Fan, Alex, Xingshuo Xiao, and Peter Washington. "Addressing Racial Bias in Facial Emotion Recognition." arXiv preprint arXiv:2308.04674 (2023).
- [15] Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." *Conference on fairness, accountability and transparency*. PMLR, 2018.
- [16] Wang, Mei, et al. "Racial faces in the wild: Reducing racial bias by information maximization adaptation network." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [17] ACM Digital Library. <https://dl.acm.org/>.
- [18] IEEE Xplore. <https://ieeexplore.ieee.org/Xplore/home.jsp>.
- [19] Google Scholar. <https://scholar.google.com/>.
- [20] Petersen, Kai, et al. "Systematic mapping studies in software engineering." *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development, 2008.
- [21] S. Mitchell, E. Potash, S. Barocas, A. D'Amour, and K. Lum, "Algorithmic Fairness: Choices, Assumptions, and Definitions," *Annual Review of Statistics and Its Application*, vol. 8, no. 1, pp. 141–163, Mar. 2021.
- [22] Maciej A Mazurowski, Piotr A Habas, Jacek M Zurada, Joseph Y Lo, Jay A Baker, and Georgia D Tourassi. 2008. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks* 21, 2-3 (2008), 427–436.
- [23] Alexander Amini, Guy Rosman, Sertac Karaman, and Daniela Rus. 2019. Variational End-to-End Navigation and Localization. In *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.
- [24] S. Noiret, J. Lumetzberger and M. Kampel, "Bias and Fairness in Computer Vision Applications of the Criminal Justice System," *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA, 2021, pp. 1-8, doi: 10.1109/SSCI50451.2021.9660177.
- [25] Hardt, Moritz, Eric Price, and Nati Srebro. "Equality of opportunity in supervised learning." *Advances in neural information processing systems* 29 (2016).

Sobre o autor:

Matheus Thiago Domingos da Silva, estudante do curso de graduação em ciência da computação pela universidade Federal de Campina Grande, Brasil.