Análise do Impacto da Variação da Boca e do Nariz em Modelos de Reconhecimento Facial

Antonio Crislân Da C. Torres[1], Cornélia Janayna Pereira Passarinho[1]

[1]Universidade Estadual do Piauí (UESPI)

Piripiri, Piauí

Brasil

crislantorrespr@gmail.com, janainapassarinho@prp.uespi.br

Abstract. *Explainability in AI can help increase the reliability and transparency of these systems, improving decision-making quality. This article aims to explore the influence of mouth and nose variation on facial recognition errors in computer vision systems and propose solutions to mitigate these issues using explainable AI (XAI) techniques. It is common for these systems to exhibit detection errors, particularly in specific groups, due to bias inherent in facial recognition algorithms. To achieve the proposed objective, detailed data analyses were conducted, aaccompanied by a thorough comparison of feature similarity among different individuals, in conjunction with XAI methods were employed to clearly and comprehensibly explain the decisions made by machine learning models. In doing so, valuable insights were gained to improve the accuracy and fairness of these systems, contributing to the development of more just and reliable facial recognition systems.*

Resumo. *A explicabilidade em IA pode ajudar a aumentar a confiabilidade e a transparência desses sistemas, melhorando a qualidade da tomada de decisão. Este artigo visa explorar a influência da variação da boca e do nariz nos erros de reconhecimento facial em sistemas de visão computacional, bem como propor soluções para minimizar esses problemas, utilizando técnicas de XAI (Explicabilidade em Inteligência Artificial). É comum que esses sistemas apresentem erros de detecção, especialmente em grupos específicos, devido ao viés presente nos algoritmos de reconhecimento facial. Para alcançar o objetivo proposto, foram realizadas análises de dados detalhadas, uma comparação profunda de similaridade das características entre diferentes indivíduos, juntamente com métodos de XAI, a fim de explicar de forma clara e compreensível as decisões tomadas pelos modelos de aprendizado de máquina. Com isso, foram obtidas informações valiosas para aprimorar a precisão e equidade desses sistemas e contribuir para o desenvolvimento de sistemas de reconhecimento facial mais justos e confiáveis.*

1. Introdução

A tecnologia de reconhecimento facial tem sido amplamente adotada em diversos setores, incluindo segurança, varejo, serviços financeiros, entre outros. Entretanto, esses sistemas têm apresentado erros na detecção, o que pode ser prejudicial para as pessoas como, por exemplo, em situações onde uma decisão pode limitar ou restringir sua liberdade de vir e ir. Tais erros ocorrem especialmente em grupos específicos, tais como pessoas do gênero feminino, de pessoas entre 46 e 85, além dos grupos de tons de pele relacionados com pele escura, conforme o estudo realizado por [Hanna F. Menezes, et al. 2020]. Esses erros podem resultar em consequências significativas, como a exclusão de indivíduos em situações críticas ou um tratamento injusto. É imprescindível abordar esses desafios e buscar soluções para mitigar esses erros, aumentando a precisão e a equidade dos sistemas de reconhecimento facial.

Tais erros muitas vezes são causados por vieses em modelos de aprendizado de máquina, agravados pela variação de características faciais. A detecção de características faciais pode ser particularmente desafiadora em ambientes de baixa qualidade, como aqueles com pouca iluminação. Além disso, a variação nas formas dos lábios, aberturas e proporções nasais, assim como oclusões parciais, podem levar a erros de reconhecimento, especialmente em condições não controladas. A iluminação inadequada, as oclusões parciais e as mudanças de ângulo também podem impactar negativamente em uma detecção precisa do modelo de reconhecimento facial [Klare et al. 2012]. Esses desafios ressaltam a importância de desenvolver abordagens que possam mitigar essas dificuldades e aprimorar a eficiência e a confiabilidade dos sistemas de reconhecimento facial.

Uma abordagem que tem se mostrado promissora para mitigar esses problemas é a utilização de XAI (*Explicabilidade em Inteligência Artificial*). A explicabilidade em IA possibilita que os usuários entendam como um modelo de aprendizado de máquina toma decisões, fornecendo percepções sobre o processo de tomada de decisão e ajudando a identificar possíveis vieses e erros de reconhecimento [Gunning et al. 2019]. Com a capacidade de fornecer explicações claras e compreensíveis sobre como um modelo de reconhecimento facial funciona, a explicabilidade em IA pode ajudar a aumentar a confiabilidade e a transparência desses sistemas, melhorando a qualidade da tomada de decisão e fortalecendo a confiança dos usuários.

Assim, o objetivo principal deste artigo foi realizar uma análise detalhada do impacto da variação da boca e do nariz em modelos de reconhecimento facial em sistemas de visão computacional. Além disso, este trabalho propõe soluções utilizando técnicas de XAI (Explicabilidade em Inteligência Artificial) para minimizar os erros decorrentes dessa variação. Com este estudo, foram obtidas informações valiosas que podem ser utilizadas para aprimorar a precisão e equidade desses sistemas,

contribuindo, assim, para o desenvolvimento de sistemas de reconhecimento facial mais robustos e confiáveis.

2. **Trabalhos Relacionados**

O objetivo de um sistema de Inteligência Artificial Explicável (XAI) é tornar o comportamento dos sistemas de Inteligência Artificial mais compreensível para os seres humanos, fornecendo explicações claras e coerentes [Gunning et al. 2019]. Para alcançar esse objetivo, existem princípios gerais que ajudam a criar sistemas de IA eficazes e que possam ser facilmente compreendidos pelos humanos. Um sistema XAI deve conseguir explicar as capacidades e decisões do próprio sistema de Inteligência Artificial, fornecendo informações relevantes sobre as ações que está realizando. Além disso, é importante que o sistema explique o que foi feito no passado, o que está sendo realizado no presente e quais serão os próximos passos em sua atuação. Esses princípios visam proporcionar transparência e confiança aos usuários e facilitar a interpretação das ações do sistema de Inteligência Artificial.

Em [Rajpal et. al 2022] foram avaliados 4 modelos de reconhecimento facial baseados em DNN (*Deep Neural Networks*), utilizando a técnica *Lime*, para avaliar o desempenho dos modelos, eles selecionaram aleatoriamente uma imagem do banco de dados *AT & T* para o qual todos os quatro modelos previram corretamente. Os resultados mostram as pontuações de previsão dos seis melhores resultados para cada modelo, onde a pontuação da previsão correta é significativamente maior em comparação com outras imagens. Cada sub-figura apresenta as explicações geradas para os seis melhores resultados. É interessante notar que, embora cada modelo preveja a etiqueta correta, eles se concentram em características ligeiramente diferentes para gerar suas previsões. Além disso, os autores apresentam instâncias onde os modelos preveem incorretamente, observando que a pontuação da etiqueta verdadeira é menor do que a do melhor resultado que resultou em uma previsão incorreta.

Em [Fábio L. D. M. 2020] o autor enfatiza que a maioria dos algoritmos de aprendizado de máquina é frequentemente visto como "caixas pretas", o que pode levar a uma falta de confiança e aceitação por parte dos usuários. O trabalho em questão avalia a compreensibilidade das explicações fornecidas por diferentes técnicas de XAI, com o foco na perspectiva dos especialistas no domínio, como oncologistas. Eles desenvolvem um sistema de diagnóstico de câncer de alta precisão usando métodos de ensemble e aplicam três técnicas de XAI para gerar explicações. Em seguida, essas explicações são apresentadas aos especialistas, e entrevistas semiestruturadas são conduzidas para avaliar a compreensão, confiança e aceitação das explicações. Como resultados do artigo, a técnica de XAI *LIME*, tem a vantagem em relação as outras técnicas testadas, pois o resultado (gráfico) é apresentado com as características de maior/menor influência na explicação, permitindo que os usuários visualizem as

características de maior risco e menor risco numa única visualização, aumentando assim o grau de explicabilidade.

Em [Ye Yu et. al 2020] é discutido a aplicação de técnicas de balanceamento de dados, aprimoramento e fusão de informações para melhorar a justiça no reconhecimento facial. Os autores conduzem uma análise cuidadosa dos métodos utilizados e suas implicações na equidade do reconhecimento facial. Eles exploram técnicas de balanceamento de dados para abordar vieses étnicos e de gênero, bem como melhorias no processo de aquisição e pré-processamento de imagens faciais. Além disso, o artigo discute a fusão de informações a partir de várias fontes, a fim de aprimorar a precisão e justiça do sistema de reconhecimento facial. O artigo é valioso não apenas por suas descobertas, mas também por sua contribuição para a discussão sobre ética e equidade em tecnologias de reconhecimento facial. Ele destaca a importância de abordar preocupações de justiça e vieses étnicos em sistemas de IA e reconhecimento facial, o que é fundamental em um mundo cada vez mais dependente dessas tecnologias.

Em [Baah, G. S 2013] o PCA (*Principal Component Analysis*) é utilizando como uma técnica de redução de dimensionalidade para o reconhecimento facial. O método PCA é aplicado para extrair as principais características faciais que são mais discriminativas e representativas das variações presentes nas imagens de rosto. Isso é realizado através do cálculo dos componentes principais das imagens de treinamento, os quais são vetores que representam as direções de máxima variação dos dados. Em seguida, os coeficientes de projeção são calculados para cada imagem de teste, mapeando-a no espaço das faces principais.

Em [Fabian P. et. al 2011] os autores descrevem o *framework Scikit-lean* de maneira abrangente, destacando suas funcionalidades e a facilidade de uso que a torna acessível para pesquisadores, cientistas de dados e desenvolvedores. O trabalho realizado destaca como o framework oferece suporte para algoritmos de classificação, regressão, agrupamento, pré-processamento de dados e validação de modelos. Além disso, aborda a importância das práticas de código limpo e documentação clara, promovendo a transparência e colaboração na comunidade de aprendizado de máquina. O *Scikit-learn*, com sua ampla base de usuários e contribuidores, continua a ser uma ferramenta indispensável para aqueles que desejam explorar e aplicar técnicas de aprendizado de máquina com Python.

3. Proposta do Trabalho

A proposta deste trabalho é compreender a influência da variação da boca e nariz na precisão dos sistemas e propor soluções eficazes para mitigar esses erros. Foram investigados diversos fatores que contribuem para a variação, como iluminação, expressões faciais e posicionamento da câmera, a fim de obter uma compreensão mais

aprofundada de suas influências no processo de reconhecimento facial. Para alcançar esse objetivo, foi desenvolvido um modelo de reconhecimento facial específico, dedicado à análise detalhada do impacto dessas características para a predição. Esse modelo permitiu investigar como variações nessas regiões podem impactar a precisão do reconhecimento facial e identificar os principais desafios associados.

3.1. Materiais e Métodos

Para este estudo, foram utilizadas três bases de dados amplamente reconhecidas no campo do reconhecimento facial: LFW (*Labeled Faces in the Wild*), *Olivetti faces* dataset e *FairFace Database*. A base de dados LFW contém mais de 13.000 imagens de rostos de celebridades coletadas em condições não controladas, incluindo variações na iluminação, pose e expressões faciais. A *Olivetti faces dataset* é uma base de dados que contém um conjunto de 400 imagens em escala de cinza de 40 indivíduos diferentes, cada um com 10 imagens capturadas em condições controladas. Já a *FairFace Datset* foi criado para abordar questões de equidade, fornecendo uma base de dados masi equiotativa e inclusiva. A mesma conta com 108.501 mil imagens de 7 diferentes grupos (Pretos, brancos, indianos, leste asiático, sudeste asiático, oriente médio e Latino). A figura 1 é uma tabela do estudo realizado por Kimmo K. e Jungseock J. al 2021, onde a mesma apresenta a precisão de classificação de gênero em conjuntos de dados de validação entre raças e faixa etárias. Expondo assim, a importância de se ter uma base diversificada e que todas as pessoas estejam representadas.

		Mean across races	SD across races	Mean across ages	SD across ages
Model trained on	FairFace	**94.89%**	**3.03%**	**92.95%**	**6.63%**
	UTKFace	89.54%	3.34%	84.23%	12.83%
	LFWA+	82.46%	5.60%	78.50%	11.51%
	CelebA	86.03%	4.57%	79.53%	17.96%

Figura 1. Precisão da classificação de gênero em conjuntos de dados de validação externa, entre raças e faixas etárias.

As bases de dados foram divididas em conjuntos de treinamento e teste, garantindo uma avaliação precisa do modelo proposto e assegurando a representatividade e generalização dos resultados obtidos. Durante a preparação das imagens, serão realizados vários passos de pré-processamento para garantir a qualidade e a padronização dos dados. Essas etapas visam melhorar a consistência das imagens e otimizar a eficácia dos algoritmos de reconhecimento facial. A tabela 1, resume os passos de pré-processamento realizados nas imagens.

Tabela 1. Etapas do Pré-Processamento das imagens.

Passo de Pré-processamento	**Descrição**
Redimensionamento das imagens	Todas as imagens foram redimensionadas para que todas tenham o mesmo tamanho.
Achatamento das imagens	Foi utilizado para reduzir o tamanho das imagens, melhorando o desempenho do modelo, principalmente em velocidade.

Conversão para escalas de Cinza	Foi utilizado para simplificar a manipulação e análise de imagens, pois reduz a complexidade do processamento de múltiplos canais de cores.
Normalização das imagens	Foi utilizado para que os dados estejam em escalas semelhantes.

Essas etapas de pré-processamento têm um impacto significativo no trabalho final, pois as mesmas garantem que as imagens estejam em um formato adequado e apresentem características mais relevantes e distintas para o reconhecimento facial. Ao redimensionar as imagens para um tamanho específico, é possível garantir a consistência e a comparabilidade entre elas. O achatamento das imagens contribui com a padronização do tamanho, independente das dimensões originais, proporciona uma comparação mais precisa, melhora o desempenho do sistema, principalmente em termos de velocidade, elimina distrações, como partes irrelevantes das imagens, melhora a compatibilidade dos dados aos modelos de aprendizado de máquina, o processo é feito de forma cautelosa, sem a distorção das características faciais. Já a normalização, envolve dimensionar os valores dos dados para um determinado intervalo.

Após o pré-processamento, Foi realizada uma extração de características, que desempenha um papel crucial no desenvolvimento do sistema de reconhecimento facial. Nessa fase, foram extraídas informações relevantes das imagens faciais que foram utilizadas para diferenciar e identificar os indivíduos. Foi também utilizado o método de análise de componentes principais (PCA, do inglês *Principal Component Analysis*) para a extração de características [Yonghong Z. 2016]. Esse método identificou as principais características faciais que contêm a maior variação nos dados. Ele realizou uma transformação linear nas imagens, projetando-as em um novo espaço de características, onde os componentes principais são ortogonais entre si e capturam a maior quantidade possível de informação relevante.

Além disso, foram realizadas outras etapas como a leitura das imagens e a classificação das mesmas, a tabela 2 mostra algumas das tecnologias e bibliotecas utilizadas na criação do modelo.

Tabela 2. Etapas do Pré-Processamento das imagens.

Ferramenta	**Versão**
Python	3.10.0
NumPy	1.26
Matplotlib	3.8
scikit-learn	1.3.2
lime	0.1.1.37

Todos os métodos e técnicas foram implementados na linguagem de programação Python, a NumPy é uma biblioteca para computação numérica em Python. Ela forneceu suporte para arrays multidimensionais e funções matemáticas essenciais para análise de dados e processamento das matrizes das imagens. Já a biblioteca Matplotlib foi utilizada para a visualização dos resultados obtidos com as explicações via Lime, que é abordando posteriormente neste artigo. O pacote sckikit-lean forneceu as ferramentas para o pré-processamento das imagens e classificação.

Em [Jia Uddin et. al 2022] O *Gradient Boosting Classifier* é empregado para classificação. Isso implica que o algoritmo é usado para determinar a categoria ou classificação dos dados com base nas características extraídas. O *Gradient Boosting* é uma técnica de aprendizado de máquina que melhora o desempenho do modelo de classificação. A mesma foi utilizada no modelo juntamente com PCA como parte do *pipeline* de aprendizado de máquina para classificar dados de imagens após a execução das etapas de pré-processamento das imagens. A figura 1 ilustra a execução do pipeline passo a passo.

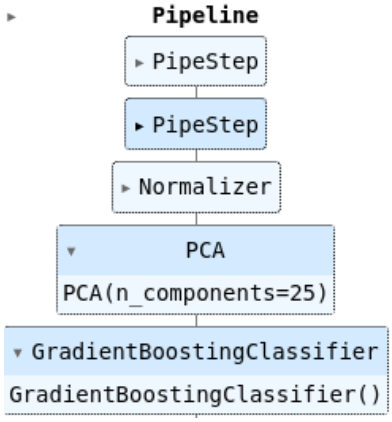

Figura 2. Pipeline de classificação dos dados.

Esse pipeline é uma representação típica de um fluxo de trabalho de aprendizado de máquina que realiza pré-processamento de imagens, normalização e redução de dimensionalidade antes de aplicar um algoritmo de classificação, neste caso, o *GradientBoostingClassifier*, para tarefas de classificação de imagens.

O pipeline é criado usando a classe Pipeline do scikit-lean. Ele consiste nas seguintes etapas: *Make Gray* (etapada de conversão de imagens coloridas para escalas de cinza; *Flatten Image* (A etapa de achatamento das imagens); *Normalize* (normalização dos dados usando o Normalizer do scikit-learn. Isso é comum quando os dados de entrada precisam ser escalados para uma escala específica); PCA (Análise de Componentes Principais (PCA) é aplicada para reduzir a dimensionalidade dos dados para 25 componentes principais) e XGBoost (O *GradientBoostingClassifier* é usado para classificar os dados após todas as etapas de pré-processamento). No *pipeline*, cada etapa é executada sequencialmente, e o resultado de uma etapa é passado para a próxima etapa. O *GradientBoostingClassifier* é a etapa final que realiza a classificação dos dados.

Para a avaliação do modelo, foi utilizada a função *classification_report* da biblioteca *scikit-learn* (*sklearn*) para gerar um relatório de classificação com base nas

previsões do modelo. Esse relatório de classificação contém várias métricas de avaliação de desempenho, incluindo precisão, recall, pontuação F1 e acurácia. A avaliação foi realizada após os testes com o conjunto de imagens de treinamento, possibilitando assim, uma verificação do desempenho e taxa de erro, proporcionando informações para ajustes de parâmetros e melhorias no modelo antes de aplicá-lo no conjunto de dados de testes, onde o mesmo irá fornecer uma estimativa realista do desempenho do modelo em condições reais de uso. O relatório de classificação é útil para entender como o modelo está se saindo em termos de classificação de diferentes classes ou categorias e pode ser valioso na avaliação do desempenho do modelo.

Além da meticulosa avaliação do desempenho do modelo de reconhecimento facial, a integração de técnicas de Explicabilidade em Inteligência Artificial (XAI, do inglês *Explainable Artificial Intelligence*) emerge como um componente de importância fundamental. Essas técnicas possibilitam não apenas a compreensão profunda do processo decisório adotado pelo modelo, mas também fornecem justificativas claras e compreensíveis para as decisões resultantes [Rajpal et al., 2022]. Nesse contexto, destaca-se a escolha estratégica do método LIME. Esse método se distingue por sua notável flexibilidade, sendo capaz de elucidar modelos de aprendizado de máquina com diversas arquiteturas e complexidades. A aplicação do LIME não apenas amplia a transparência do processo decisório, mas também contribui para uma compreensão mais profunda das nuances envolvidas nas predições do modelo. Essa abordagem mais abrangente não apenas valida o desempenho do modelo, mas também confere uma camada adicional de confiança e interpretabilidade, sendo crucial para aplicativos de reconhecimento facial em contextos sensíveis.

Além disso, o LIME fornece interpretações locais, ou seja, explicações específicas para cada instância de dados. Isso permite entender quais os aspectos dos dados foram mais influentes na tomada de decisão do modelo para cada caso individualmente. Essas interpretações locais podem ajudar a identificar quais características faciais ou regiões específicas (como boca e nariz) contribuíram significativamente para as predições do modelo. A figura 3 ilustra as interpretações locais do LIME para uma imagem.

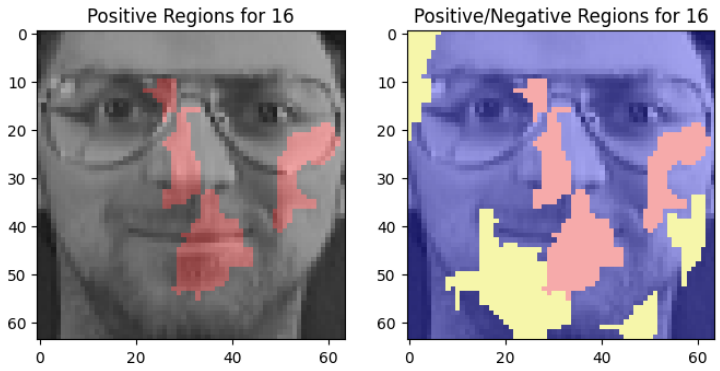

Figura 3. Explicações geradas pelo Lime na detecção de uma face.

Como demonstrado na figura 3, é possível criar uma explicação local de cada imagem, com as regiões que demonstraram ser mais importantes para o modelo positivamente

(imagem da esquerda), quanto as regiões importantes em aspectos positivos e negativos (imagem da direita).

Com o Lime, é possível gerar também explicações para cada imagem em predições do modelo, o que seria isso, o *Gradient Boosting Classifier* gera uam árvore sequencial e atribui pesos a cada nó dessa árvore, ou seja, para cada imagem da base de dados é criada uma árvore com as imagens e os filhos dessas imagens de maior peso, são as imagens, com a maior probabilidade de serem a mesma pessoa. As previsões de todas as árvores individuais são combinadas para criar a previsão final do modelo, as árvores contribuem com base nos seus pesos atribuídos. A figura 4 mostra o exemplo de predição de uma imagem para as 6 imagens com maior probabilidade de serem preditas pelo modelo, proporcionando assim, não uma explicação individual, mas sim uma explicação onde quais características tiveram relevância em predições reais do modelo, tanto em situações de verdadeiro positivo ou falso positivo.

Figura 4. Explicações geradas pelo Lime em predições do modelo.

Como demonstrado nas figuras 3 e 4, o Lime fornece explicações que sejam facilmente compreensíveis para os seres humanos. Ao invés de fornecer uma explicação complexa baseada nos parâmetros internos do modelo, o LIME gera explicações em termos mais simples e intuitivos, como a importância relativa das características faciais para a predição do modelo. Em alguns casos, como o da figura 4 quatro, é importante mostrar que algumas das vezes, a imagem não apresentará nenhuma característica relevante, o que significa que não houve nenhuma característica em específico que pesou para a decisão do modelo naquela circunstância.

4. Resultados Obitidos

Os resultados obtidos destacam o avanço significativo da explicabilidade em Inteligência Artificial, proporcionando uma substancial melhoria na confiabilidade e transparência dos sistemas de reconhecimento facial em visão computacional. Ao realizar análises detalhadas dos dados e empregar métodos de Explicabilidade em Inteligência Artificial (XAI), foi possível investigar como a variação das características faciais, especialmente da boca e do nariz, influenciam nos erros de reconhecimento facial. Esta abordagem aprofundada permitiu uma comparação meticulosa da

similaridade das características entre diferentes indivíduos, explicando de maneira clara e compreensível as decisões adotadas pelos modelos de aprendizado de máquina. As informações obtidas a partir das bases de dados analisadas oferecem uma valiosa contribuição para a obtenção de precisão e equidade nos sistemas de reconhecimento facial. Essa compreensão mais profunda não apenas aprimora a confiabilidade dos sistemas, mas também impulsiona o desenvolvimento de soluções mais justas e confiáveis, promovendo assim avanços significativos no campo da visão computacional.

4.1. Olivetti faces

A base de dados *Olivetti Faces* compreende um conjunto de 400 imagens em escala de cinza representando 40 indivíduos distintos. Devido ao número limitado de amostras, foi optado por utilizar 25% dessas imagens para a extração final dos dados, totalizando 100 imagens. Cada uma dessas imagens foi minuciosamente analisada quanto às regiões faciais de impacto por meio da ferramenta LIME. Para cada imagem, foram geradas seis predições distintas do modelo em questão. Priorizamos as imagens com maior probabilidade de predição, resultando em um total de 600 predições submetidas à análise detalhada nesta base de dados. Essa abordagem visa otimizar a representatividade dos dados selecionados, permitindo uma investigação mais aprofundada das características faciais que influenciam as decisões do modelo.

A figura 5 ilustra a análise de impacto das características faciais da base de dados Olivetti faces, revelando que o nariz apresentou 96 casos de predições com impacto positivo e 88 casos com impacto negativo. A discrepância de 8 casos evidencia um impacto global mais positivo do que negativo nesta região facial. Essa tendência é também observada na boca, com 128 casos de impacto positivo, superando os casos negativos em 16. Da mesma forma, os olhos, que fornecem um parâmetro para avaliar o impacto relativo do nariz e da boca em relação ao restante do rosto, registraram 112 casos de impacto positivo e 96 casos de impacto negativo.

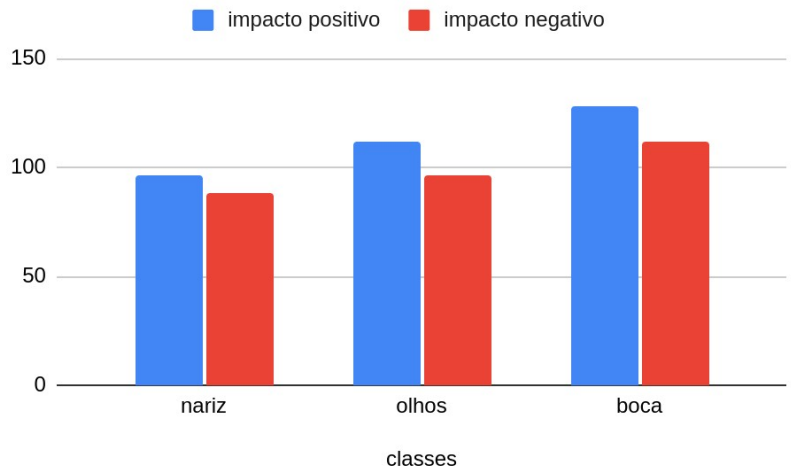

Figura 5. Gráfico com o total de casos com impacto positivo e negativo de cada região facial Olivetti faces.

Os dados obtidos revelam que, em regiões controladas, as características faciais analisadas demonstram uma similaridade notável quanto ao número de casos em que cada região teve impacto, seja positivo ou negativo. A boca emergiu como a região com o maior número de resultados, enquanto o nariz, por sua vez, foi a região com menor impacto entre as três na análise conduzida nesta base de dados específica.

A Figura 6 exibe um gráfico que compara o número total de predições em que uma ou mais das regiões estudadas foram identificadas pela ferramenta LIME como áreas de impacto, em contraste com os casos em que a ferramenta não destacou nenhuma dessas regiões como sendo crucial para a predição do modelo.

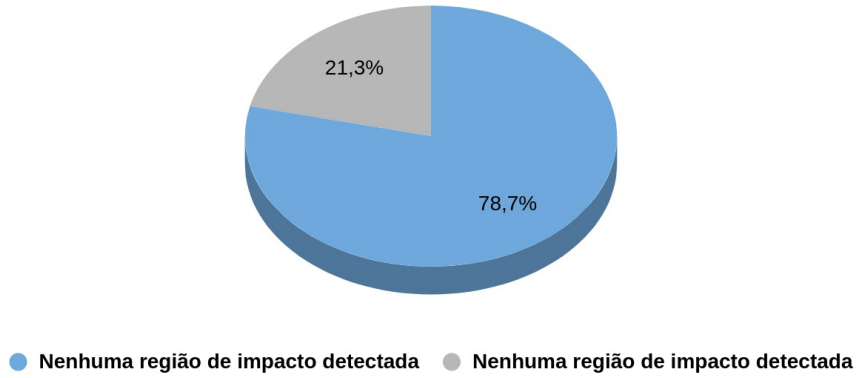

Figura 6. Gráfico com o total de casos com região de impacto detectada x nehuma região de impacto detectada.

Com base nos resultados obtidos, observa-se que em 78,7% das predições do modelo nesta base de dados, a ferramenta identificou uma ou mais regiões de impacto. Em 21,3% dos casos, nenhuma região foi destacada como de maior relevância para as decisões do modelo. Esses valores indicam uma eficácia notável, especialmente considerando que todas as imagens foram capturadas em situações controladas. Essa condição controlada não apenas contribui para uma melhor precisão na predição, mas também aprimora a capacidade da ferramenta em identificar as partes específicas do rosto que impactam as decisões do modelo.

4.2. LFW Dataset

O conjunto de dados *LFW Dataset* é composto por 13.233 imagens em escala de cinza, representando 5.749 indivíduos únicos. Nos experimentos com este conjunto de dados, optamos por utilizar a divisão padrão de treinamento e teste fornecida pela biblioteca Scikit-learn. Essa decisão foi baseada na recomendação da própria biblioteca, que não apenas disponibiliza a base de dados, mas também oferece ferramentas para essa divisão. A distribuição dos dados foi estabelecida conforme a orientação da *Scikit-learn*, resultando em uma divisão de 70% para treinamento e 30% para teste. Esses 30% representam um total de 3.961 imagens. Para a construção final dos dados, foram consideradas 6 predições do modelo para cada imagem, totalizando 23.766 casos analisados após passarem pela explicação do LIME. Essa abordagem proporciona uma

avaliação abrangente e representativa do desempenho do modelo em uma variedade de cenários.

A análise de impacto das características faciais na base de dados LFW é representada na figura 7. Nota-se que o nariz apresentou 96 casos de predições com impacto positivo e 88 casos com impacto negativo, resultando em uma discrepância de 8 casos.. Esta diferença sugere um impacto global mais favorável do que desfavorável nessa região facial específica. Essa tendência positiva é também evidente na boca, onde foram observados 128 casos de impacto positivo, superando os casos negativos em 16. De maneira semelhante, os olhos, que proporcionam um parâmetro para avaliar o impacto relativo do nariz e da boca em relação ao restante do rosto, apresentaram 112 casos de impacto positivo e 96 casos de impacto negativo.

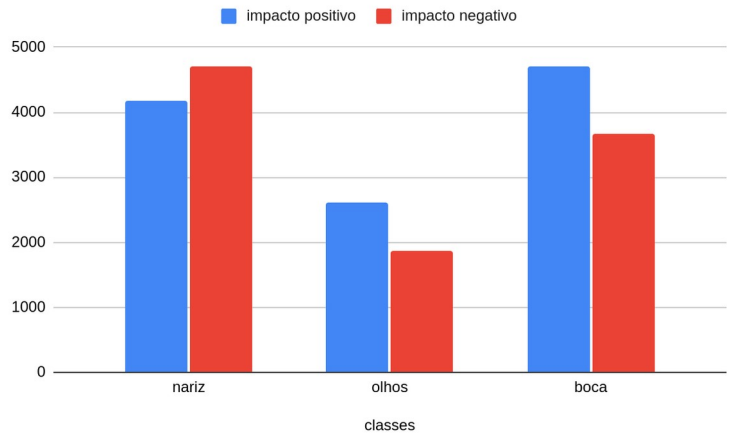

Figura 7. Gráfico com o total de casos com impacto positivo e negativo de cada região facial LFW Dataset.

Os resultados obtidos destacam que, em imagens não controladas, as características faciais mais proeminentes na análise foram o nariz e a boca. Isso sugere que, em situações de ambientação menos controlada, essas características tornam-se ainda mais influentes para o desempenho do modelo. Especificamente, nesta base de dados, o nariz emergiu como uma região de notável destaque, principalmente no que diz respeito ao impacto negativo. Por outro lado, a boca demonstrou ter um impacto positivo mais significativo do que o negativo, indicando sua extrema importância nas decisões corretas do modelo.

A Figura 8 apresenta um gráfico que contrasta o número total de predições em que a ferramenta LIME identificou uma ou mais das regiões analisadas como áreas de impacto na base de dados LFW, em comparação com os casos em que a ferramenta não destacou nenhuma dessas regiões como sendo crucial para a predição do modelo.

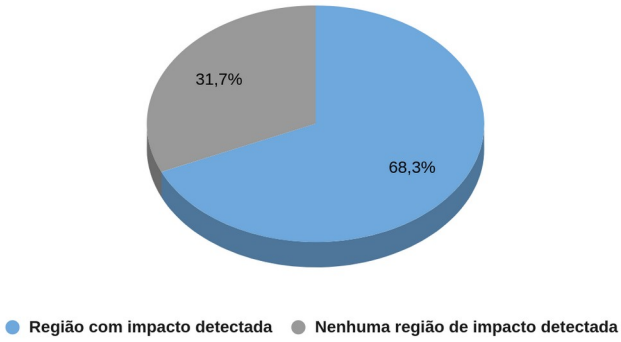

Figura 8. Gráfico com o total de casos com região de impacto detectada x nehuma região de impacto detectada na base de dados LFW.

Com base nos resultados, constata-se que em 68,3% das predições do modelo nesta base de dados, a ferramenta identificou uma ou mais regiões de impacto. Em 31,7% dos casos, nenhuma região foi destacada como sendo de maior relevância para as decisões do modelo. Notavelmente, em comparação com a base de dados Olivetti Faces, os dados obtidos com a LFW apresentaram uma discrepância significativa nos casos em que nenhuma região foi destacada. Isso ocorre principalmente devido às imagens não terem sido capturadas em situações controladas, o que pode dificultar tanto a predição do modelo quanto a detecção da face, assim como as explicações geradas pela ferramenta.

4.3. FairFace Dataset

FairFace é um conjunto de dados de imagens faciais com equilibrio racial. Ele contém 108.501 imagens de 7 grupos raciaias diferentes: Branco, Negro, Indiano, Leste Asiático, Sudeste Asiático, Oriente Médio e Latino. As imagens foram coletadas do conjunto de dados YFCC-100M Flickr e rotuladas com raça, sexo e faixas etárias. Nos experimentos realizados com essa base de dados, foi trabalhada com a própria divisão de dados da base, sendo 89,9% para testes, o que representa 97.547 imagens, já para validação, ficaram 10,1% da base de dados, representado 10954 imagens. Para a construção final dos dados, foram consideradas 6 predições do modelo para cada imagem, totalizando 65724 casos analisados após passarem pela explicação do LIME. Até o presente momento foram contabilizados 48623 casos, dando assim, uma análise abrangente e representativa do desempenho do modelo em uma variedade de cenários.

A análise de impacto das características faciais na base de dados LFW é representada na figura 9. Nota-se que o nariz apresentou 10056 casos de predições com impacto positivo e 5133 casos com impacto negativo, resultando em uma discrepância de 4923 casos. Esta diferença sugere um impacto global mais favorável do que desfavorável nessa característica especifica, muito mais notório que em outras bases de dados. Essa têndencia positiva é também vista na boca, onde foram observados 9876 casos com impacto positivo, superando o impacto positivo em 5134 casos analisados. De maneira semelhante, os olhos, que proporcionam um parâmetro para avaliar o impacto relativo do nariz e da boca em relação ao restante do rosto, apresentaram 4836 casos de impacto positivo e 3214 casos de impacto negativo.

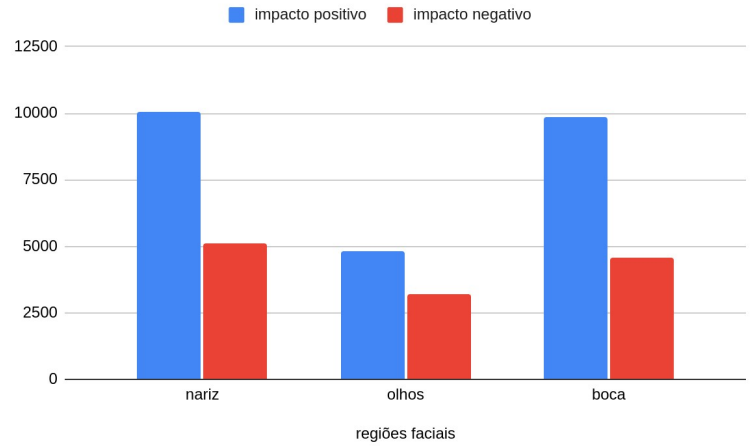

Figura 9. Gráfico com o total de casos com impacto positivo e negativo de cada região facial FairFace Dataset.

Os resultados obtidos nesta base de dados específica, apresentam uma diferença sigificativa em relação as outras bases de dados abordadas neste artigo, principalmente pelo fato da mesma não mostrar uma similaridade dos dados de impacto positivo e negativo, tendo assim, uma discrepância maior nessa comparação. Isso mostra que, mesmo em situações não controladas, quando se tem uma base de dados mais diversificada, essas características acabam impactando muito mais positvamente nas decisões do modelo. Os resultados também apontam uma maior importância da boca e do nariz em relação aos olhos, apontando assim, para uma maior relevância dessas regiões para as predições do modelo de reconhecimento facial.

A Figura 10 apresenta um gráfico que contrasta o número total de predições em que a ferramenta LIME identificou uma ou mais das regiões analisadas como áreas de impacto na base de dados FairFace, em comparação com os casos em que a ferramenta não destacou nenhuma dessas regiões como sendo crucial para a predição do modelo.

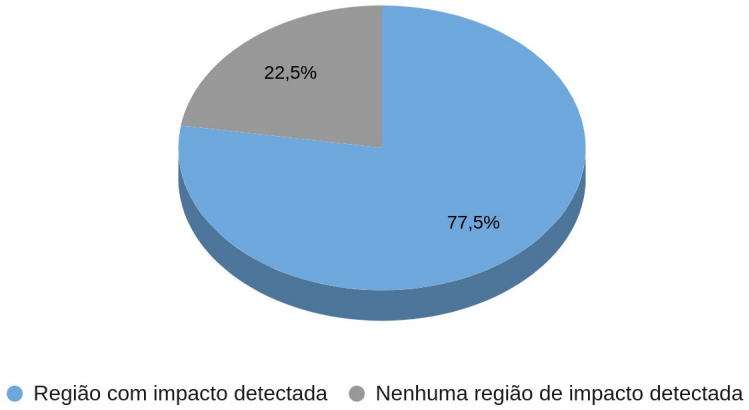

Figura 10. Gráfico com o total de casos com região de impacto detectada x nehuma região de impacto detectada na base de dados FairFace.

Com base nos resultados, constata-se que em 77,5% das predições do modelo nesta base de dados, a ferramenta identificou uma ou mais regiões de impacto. Em 22,5% dos casos, nenhuma região foi destacada como sendo de maior relevância para as decisões do modelo. Notavelmente, em comparação com a base de dados LFW, os dados obtidos com a FairFace apresentaram uma discrepância significativa nos casos em que uma ou mais de uma região foi destacada. Mesmo que ambas as bases de dados estejam es siuações controladas, o ocorrido se deve muito possivelmente pela maior diversidade apresentada da FairFace, apresentando uma menor diversidade de dados, o modelo tende a reconhecer características como similares, tornando assim, o impacto dessa região, pouco significativo.

5. Conclusão

Após analisar os resultados deste estudo, destaca-se que, acima de todas as variáveis na criação de um modelo de reconhecimento facial, a seleção de uma base de dados integralmente diversificada é de extrema importância. Isso garante a representação de todos os grupos, contribuindo significativamente para a melhoria do desempenho do modelo. Os resultados indicam que, em relação ao impacto na predição de um modelo de reconhecimento facial, as regiões da boca e do nariz têm uma influência substancialmente maior do que os olhos, que foram os parâmetros focados na pesquisa. Isso sugere uma maior variação nessas características entre os indivíduos, sublinhando a necessidade de atenção especial a essas regiões. A Figura 11 apresenta a diferença percentual entre o impacto positivo e negativo de cada região em três bases de dados distintas. Observa-se que o *FaiFace Dataset* obteve melhores rendimentos nas três características em comparação com as outras bases de dados. Por outro lado, a LFW, em relação ao nariz, demonstrou um desempenho negativo, indicando que esta região teve um impacto mais adverso do que benéfico.

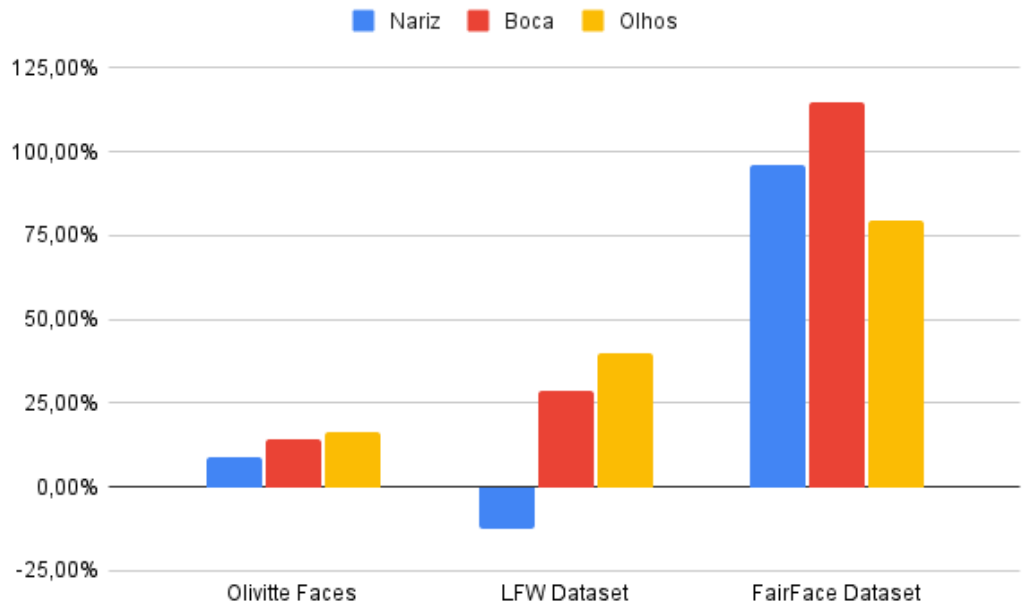

Figura 11. Diferença percentual entre impacto positivo e negativo.

Essas constatações reforçam a importância de uma abordagem abrangente na escolha da base de dados e na consideração das regiões específicas do rosto ao desenvolver modelos de reconhecimento facial. A compreensão diferenciada do impacto dessas características pode orientar futuros desenvolvimentos na melhoria da precisão e eficácia dos modelos, destacando a relevância crítica da diversidade e atenção detalhada às características específicas na construção de modelos mais robustos e eficientes. É crucial assegurar a qualidade das imagens, mesmo aquelas capturadas em ambientes não controlados. Além disso, observou-se que uma parcela significativa das três bases de dados não apresentou retorno da ferramenta LIME quanto a uma área de impacto identificada. A Figura 12 oferece uma representação do número de casos nos quais o LIME identificou pelo menos uma característica e o número de casos nos quais nenhuma característica foi identificada.

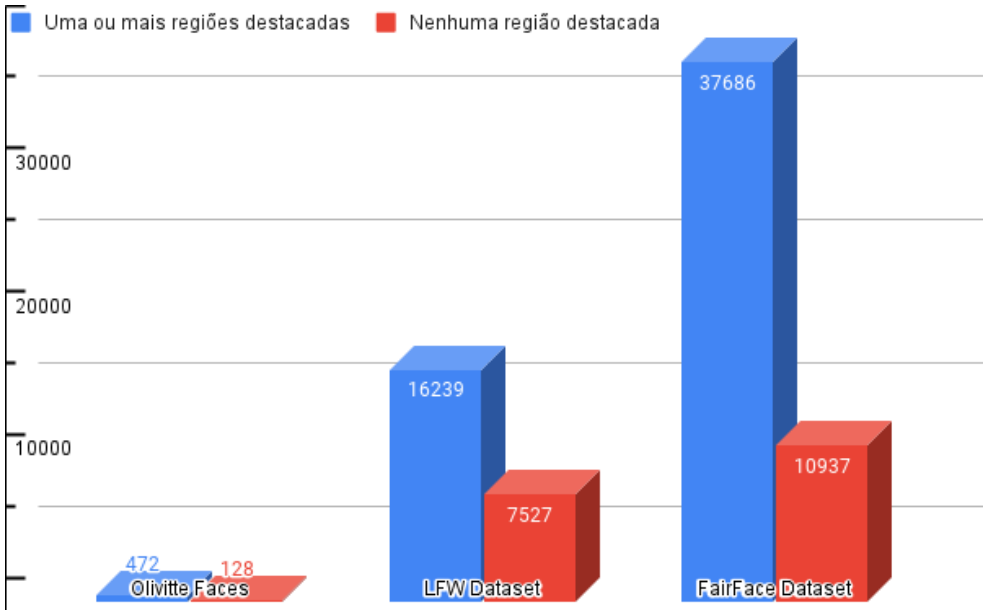

Figura 12. Diferença de vezes em que uma ou mais características foram destacadas pela ferramentea x quando nenhuma característica foi destacada.

Evidentemente, as bases de dados *FairFace* e *Olivetti Faces* demonstraram um desempenho superior na frequência em que pelo menos uma característica foi destacada, em comparação com as situações em que nenhuma característica foi realçada. Por outro lado, a base de dados LFW teve um desempenho significativamente inferior em relação às outras duas.

Em estudos prospectivos, seria altamente benéfico aprofundar a análise do impacto dessas características identificadas. Recomenda-se não apenas a exploração de conjuntos de dados mais diversificados, mas também o aprimoramento da metodologia para examinar o impacto dessas características considerando variáveis específicas, como gênero e faixa etária. Essa abordagem refinada permitiria uma investigação mais minuciosa, proporcionando percepções específicas e dados mais direcionados a diferentes grupos demográficos. A consideração cuidadosa dessas variáveis demográficas contribuiria não apenas para uma compreensão mais abrangente, mas também para insights mais precisos sobre o impacto dessas características em contextos específicos.

References

Menezes, H. F. et al. (2021). Bias and Fairness in Face Detection. In 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Campina Grande, PB, Brazil: Academic Unity of Systems and Computing, Federal University of Campina Grande.

Gunning, D. et al. (2019). XAI - Explainable artificial intelligence. Science Robotics, 4(37), eaay7120. Doi: 1 0.1126/scirobotics.aay7120.

Shetty, A. B. et al. (2021). Facial recognition using Haar cascade and LBP classifiers. Dept. Computer Science and Engineering, Shri Madhwa Vadiraja Institute of Technology and Management (Affiliated to VTU) Udupi, 574115 India.

Singh, S., Singh, D., & Yadav, V. (2020). Face Recognition Using HOG Feature Extraction and SVM Classifier. International Journal of Computer Vision and Image Processing, 10(2), 150-165.

Baah, G. S. (B.Sc. Mathematics) (2013). Face Recognition Using Principal Component Analysis. (Master's thesis, Kwame Nkrumah University of Science and Technology, Institute of Distance Learning).

DAS, Arun; RAD, Paul (2021). Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 3, pp. 2986-3005.

YANG, Yu et al. (2022). Enhancing Fairness in Face Detection in Computer Vision Systems by Demographic Bias Mitigation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022, p. 10527-10536.

INGH, R. et al. (2022). Anatomizing Bias in Facial Analysis. IIT Jodhpur, IIIT-Delhi. The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22).

KLARE, B. F. et al. (2012).Face Recognition Performance: Role of Demographic Information. To appear: IEEE Transactions on Information Forensics and Security. M.J. Burge, and J. Klontz are with The MITRE Corporation, McLean, VA, U.S.A.

Rajpal, A. et al. (2022). XAI-FR: Explainable AI-Based Face Recognition Using Deep Neural Networks..

Rajpal, A. et al. (2022). Figura 3: Explicações geradas pelo Lime na detecção de uma face. XAI-FR: Explainable AI-Based Face Recognition Using Deep Neural Networks.

Coutinho B. (2019). Modelos de Predição | SVM: Aprenda a criar seu primeiro algoritmo de classificação com SVM. Turing Talks, June, 2019.

Rodrigues V. (2018). Entenda o que é AUC e ROC nos modelos de Machine Learning. Bio Data Blog, October, 2018.

Yonghong, Z., & Yanchao, X. (2016). A Study on Cost Drivers Based on Principal Component Analysis. School of Management, Wuhan University of Technology, Wuhan, P.R.China, 430070.

De Jesus, R. A., Komati, K. S., & Simões, S. N. (2020). Comparação das Técnicas de Extração de Características HOG e LBP para Detecção de Glaucoma em Retinografias. Programa de Pós-Graduação em Computação Aplicada (PPComp), Campus Serra, Instituto Federal do Espírito Santo (Ifes), Serra - ES - Brasil.

Henrik K., Camille L., & Jakob E. S. (2018). Children And Gender Inequality: Evidence From Denmark. National Bureau Of Economic Research, Massachusetts Avenue Cambridge.

Tomás S., et al. (2020). FairFace Challenge at ECCV 2020: Analyzing Bias in Face Recognition, Czech Technical University in Prague, Czech Republic, Universitat Oberta de Catalunya, Spain, Computer Vision Center, Spain, Universitat de Barcelona, Spain, The Queen's University of Belfast, United Kingdom, Anyvision, United Kingdom, Barcelona – Espanha.

Karamizadeh, S. et al. (2013). A. An Overview of Principal Component Analysis. Journal of Signal and Information Processing, v. 4, p. 173-175, Universiti Teknologi Malayisa, Malásia.

Mei Wang, Yaobin Zhang, Weihong Deng, (2015). Meta Balanced Network for Fair Face Recognition. National Natural Science Foundation of China, Excellent Ph.D. Students Foundation, China.

Kristijonas C. et al. (2021). Argumentative XAI: A Survey. Department of Computing, Imperial College London, UK.

Kimmo Kärkkäinen, Jungseock Joo, (2021). FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. Universidade da Califórnia, Estados Unidos da América.

Rafael Menezes Barreto, (2013). Aprendizagem de Métrica baseada na Distância Euclidiana aplicada ao Reconhecimento de Faces. Universidade Federal de Pernambuco, Recife, Pernambuco – Brasil.

Wagner Oliveira de Araujo, Clarimar Jose Coelho, (2009). Análise de Componentes Principais (PCA). Centro Universitário de Anápolis, Goiás – Brasil.

Fábio A. D., et al. (2013). RedFace: um sistema de reconhecimento facial baseado em técnicas de análise de componentes principais e autofaces: comparação comdiferentes classificadores. Revista Brasileira de Computação Aplicada(ISSN 2176-6649), Passo Fundo, v.5, n. 1, p. 42-54, abr. 2013.

Mostofa K. N., et al. (2022). Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. Journal of King Saud University - Computer and Information Sciences. Volume 34, Issue 8, Part A, September 2022, Pages 4773-4781

Fabian P. et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12 (2011).

Yaniv T., et al. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. Conference on Computer Vision and Pattern Recognition (CVPR).

Anirudha B S., et al. (2021). Facial recognition using Haar cascade and LBP classifiers. Dept. Computer Science and Engineering, Shri Madhwa Vadiraja Institute of Technology and Management (Affiliated to VTU) Udupi, 574115 India.

Swarnima S., Durgesh S., Vikash Y., (2020). Face Recognition Using HOG Feature Extraction and SVM Classifier. International Journal of Emerging Trends in Engineering Research.

Fábio L. D. M. (2020). Técnicas de Inteligência Artificial explicáveis agnósticas: uma análise qualitativa sob a perspectiva do especialista de domínio oncológico. Rio de Janeiro, RJ – Brasil.

Ye Yu, et al. (2020). Fair Face Recognition Using Data Balancing, Enhancement and Fusion. Computer Vision – ECCV 2020 (pp.492-505)Edition: ECCV 2020Chapter: ECCV 2020Publisher: ECCV 2020

Tabela 4. Cronograma de atividades do Trabalho de Conclusão de Curso.

Atividades	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov
Introdução	✔	✔							
Trabalhos Relacionados	✔	✔	✔						
Proposta do Trabalho		✔	✔	✔		✔	✔		
Desenvolvimento					✔	✔	✔	✔	
Resultados esperados								✔	✔
Conclusão								✔	✔
Escrita do artigo	✔	✔	✔	✔	✔	✔	✔	✔	✔

✔: Tarefas concluídas; ✘: Tarefas pendentes.