INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS TERESINA CENTRAL
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

LIVIA TAINÁ ALVES DE BRITO

CLASSIFICAÇÃO DE GÊNERO COM REDES NEURAIS CONVOLUCIONAIS: uma análise comparativa aplicada a um dataset piauiense.

TERESINA, PIAUÍ
2025

LIVIA TAINÁ ALVES DE BRITO

CLASSIFICAÇÃO DE GÊNERO COM REDES NEURAIS CONVOLUCIONAIS: uma análise comparativa aplicada a um dataset piauiense.

Trabalho de Conclusão de Curso (monografia) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central.

Orientador: Prof. M^{e}. Ely da Silva Miranda

Coorientador: Prof. Dr^{e}. Pablo de Abreu Vieira

TERESINA, PIAUÍ
2025

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Brito, Livia Tainá Alves de

B862c Classificação de gênero com redes neurais convolucionais : uma análise comparativa aplicada a um dataset Piauiense / Livia Tainá Alves de Brito. - 2025.

51 f.: il. color.

Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas) - Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central, 2025.

Orientador : Prof Me. Ely da Silva Miranda.

1. Reconhecimento facial. 2. Rede neural. 3. Convolucional. 4. Gênero. I.Título.

CDD - 004.21

Elaborado por Tanize Maria Sales CRB 3/1024

LIVIA TAINÁ ALVES DE BRITO

CLASSIFICAÇÃO DE GÊNERO COM REDES NEURAIS CONVOLUCIONAIS: uma análise comparativa aplicada a um dataset piauiense.

Trabalho de Conclusão de Curso (monografia) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central.

Aprovada em 29/11/2025.

BANCA EXAMINADORA:

Prof. M^{e}. Ely da Silva Miranda (Orientador)
Universidade Federal do Piauí (UFPI)

Prof. Dr^{e}. Pablo de Abreu Vieira (Coorientador)
Universidade Federal de Rondonópolis (UFR)

Prof. Dr^{e}. Ricardo Martins Ramos
Universidade Luterana do Brasil (ULBRA)

Dedico este trabalho a todos que acreditaram que ele sairia.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me abençoado todo o processo de desenvolvimento, embora com muito percalços no decorrer, mas por meio Dele a entrega foi concluída.

Agradeço aos meus orientadores que me acompanharam nesta caminhada, ao Pablo que esteve comigo desde o inicio e ao Professor Ely que me aceitou como sua orientada, agradeço a vocês em especial pela paciência e carinho ao me acolherem.

Agradeço aos meus familiares e amigos humanos e inumanos que, também, me acompanharam e me apoiaram, seja com palavras de conforto e até corrigindo minha monografia, vocês são 10.

Por fim, agradeço ao meu namorado que esteve comigo e foi quem mais me auxiliou e me apoiou nos momentos mais complicados e me ajudou a continuar mesmo quando quis desistir, obrigada!

"A ciência não é nada mais que o refinamento do pensamento do dia a dia".
Albert Einstein

RESUMO

A classificação de sexo biológico a partir de imagens faciais é uma tarefa desafiadora e frequentemente comprometida por vieses demográficos que afetam a precisão dos modelos em populações específicas. Modelos treinados com conjuntos de dados internacionais, por exemplo, apresentam desempenho reduzido quando aplicados a populações com características faciais distintas, como a brasileira. Esta monografia propõe uma análise comparativa de cinco arquiteturas de Redes Neurais Convolucionais (CNNs) distintas para esta tarefa. Para mitigar os vieses de origem demográfica, a metodologia foi fundamentada na técnica de *transfer learning*, aplicada a um acervo do governo estadual de 33.033 recortes faciais de indivíduos piauienses, com o objetivo de determinar a arquitetura de maior eficácia para a classificação de sexo. Os resultados obtidos destacaram a superioridade do modelo ResNet50, que alcançou os melhores índices de performance, com valores de 0,8957 (Acurácia), 0,7905 (Kappa), 0,8958 (*F1-Score* Ponderado) e 0,9609 (AUC-ROC). A análise qualitativa dos erros, no entanto, revelou que os modelos desenvolveram uma forte correlação entre a presença de características cosméticas e a classificação de sexo, indicando a persistência de vieses relacionados à própria tarefa e originados a partir dos dados. Estes resultados demonstram o potencial da abordagem para criar sistemas mais adaptados à diversidade local e ressaltam a importância da seleção criteriosa da arquitetura e da contínua curadoria dos dados para o desenvolvimento de tecnologias de reconhecimento facial mais precisas e equitativas.

Palavras-chaves: reconhecimento facial; rede neural convolucional; gênero.

ABSTRACT

Classifying biological sex from facial images is a challenging task, often compromised by demographic biases that affect the accuracy of models in specific populations. Models trained with international datasets, for example, show reduced performance when applied to populations with distinct facial characteristics, such as the Brazilian population. This monograph proposes a comparative analysis of five different Convolutional Neural Network (CNN) architectures for this task. To mitigate demographic biases, the methodology was based on the transfer learning technique, applied to a state government database of 33,033 facial cutouts of individuals from Piauí, with the aim of determining the most effective architecture for sex classification. The results obtained highlighted the superiority of the ResNet50 model, which achieved the best performance indices, with values of 0.8957 (Accuracy), 0.7905 (Kappa), 0.8958 (Weighted F1-Score), and 0.9609 (AUC-ROC). However, the qualitative error analysis revealed that the models developed a strong correlation between the presence of cosmetic features and gender classification, indicating the persistence of biases related to the task itself and originating from the data. These results demonstrate the potential of the approach to create systems more adapted to local diversity and emphasize the importance of careful architecture selection and continuous data curation for the development of more accurate and equitable facial recognition technologies.

Keywords: facial recognition; convolutional neural network; gender.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AUC	Area Under the Curve
CAP	Credit Assignment Paths
CLAHE	Equalização de Histograma Adaptativa com Limitação de Contraste
CNN	Convolutional Neural Network
DL	Deep Learning
FPR	Taxa de Falsos Positivos
GPU	Graphics Processing Unit
IA	Inteligência Artificial
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
KNN	K-Nearest Neighbors
LGPD	Lei Geral de Proteção de Dados
MIT	Massachusetts Institute of Technology
ML	Machine Learning
ReLU	Rectified Linear Unit
ResNet50	Residual Network-50
RGB	Red, Green, Blue
RNA	Rede Neural Artificial
ROC	Receiver Operating Characteristic
TPR	Taxa de Verdadeiros Positivos
VGG16	Visual Geometry Group 16
ViT	Vision Transformers
YOLO	You Only Look Once

SUMÁRIO

1 INTRODUÇÃO

O reconhecimento facial se consolidou como uma das mais proeminentes tecnologias biométricas, sendo amplamente adotado por sua natureza menos invasiva em comparação a outras modalidades, como a análise de impressão digital ou de íris (ZHAO *et al.*, 2003). Sua aplicação se estende por áreas como segurança, autenticação e visão computacional, onde a capacidade de identificar indivíduos a partir de imagens digitais impulsionou inovações significativas (KODANDARAM *et al.*, 2015).

Dentro desse campo, a classificação de gênero emerge como uma tarefa fundamental, com aplicações práticas que vão desde a personalização de serviços em plataformas de e-commerce e streaming até o auxílio em sistemas de monitoramento para segurança pública (JAIN; ROSS; NANDAKUMAR, 2011; MALTONI *et al.*, 2009). No entanto, embora o cérebro humano execute essa tarefa com aparente facilidade, sua implementação computacional enfrenta desafios consideráveis. A distinção algorítmica entre padrões faciais masculinos e femininos depende de sutilezas morfológicas que são facilmente afetadas por fatores como iluminação, pose, expressões faciais e oclusões (uso de acessórios como óculos e chapéus) (SATO; SHAH; AGGARWAL, 1998; MORENO *et al.*, 2016; TARAWNEH *et al.*, 2019).

O desafio técnico é agravado pela alta variabilidade nos dados do mundo real. Fatores como idade e etnia, por exemplo, degradam significativamente o desempenho dos modelos. Estudos demonstram que rostos de adultos tendem a ser classificados com maior precisão do que os de indivíduos mais jovens (GUO *et al.*, 2009), e que a diversidade étnica, quando não representada adequadamente nos dados de treinamento, introduz vieses que comprometem a confiabilidade dos sistemas (BENABDELKADER; GRIFFIN, 2005).

Essa questão do viés é particularmente crítica no contexto brasileiro. Modelos de reconhecimento facial, frequentemente treinados com bases de dados predominantemente europeias ou norte-americanas, apresentam um desempenho inferior quando aplicados à população brasileira, cuja diversidade étnica é acentuada. A subrepresentação de determinados grupos nos dados de treinamento leva a taxas de erro mais altas, especialmente na identificação de mulheres e pessoas negras (RAMALHO, 2024; OLHAR DIGITAL, 2016). Tal limitação tem consequências graves, como falsas acusações geradas por sistemas de segurança pública que falham em identificar corretamente indivíduos de grupos minorizados (DURÃES, 2024). A utilização de modelos de inteligência artificial comprimidos, uma prática comum para otimizar o desempenho, pode intensificar ainda mais esse viés racial, comprometendo a equidade na aplicação da tecnologia (NETO *et al.*, 2023).

Para mitigar esses vieses e desenvolver sistemas mais justos e precisos, é

fundamental a utilização de bases de dados que reflitam a diversidade local (CEIA *et al.*, 2022). Nesse sentido, este trabalho utiliza uma base de dados composta por imagens de indivíduos do Piauí, buscando avaliar o desempenho de modelos de classificação em um contexto regional específico, contribuindo para a criação de soluções mais adaptadas à realidade brasileira.

Diante desses desafios, a complexidade técnica da classificação de gênero e o impacto dos vieses demográficos, as Redes Neurais Convolucionais (CNNs) surgem como uma abordagem eficaz, graças à sua capacidade de aprender e extrair características hierárquicas diretamente das imagens (LECUN; BENGIO; HINTON, 2015). Assim, para enfrentar o problema apresentado, este trabalho propõe uma análise comparativa do desempenho de cinco arquiteturas de CNNs de referência AlexNet, InceptionV3, ResNet50, VGG16 e Xception na tarefa de classificação de gênero. O objetivo é identificar a arquitetura mais eficaz quando treinada e avaliada em uma base de dados local, visando contribuir para o avanço de sistemas de reconhecimento facial mais acurados e equitativos para a população piauiense.

1.1 OBJETIVO GERAL

Este trabalho tem como objetivo geral avaliar comparativamente o desempenho de diferentes arquiteturas de CNNs (AlexNet, InceptionV3, ResNet50, VGG16 e Xception) na tarefa de classificação de gênero. A pesquisa busca, por meio da utilização de uma base de dados piauiense, mitigar os vieses demográficos e raciais presentes em modelos tradicionais, que resultam em altas taxas de erro e falsas acusações. O objetivo final é tornar o sistema de classificação mais preciso e equitativo, identificando a arquitetura mais adequada para o contexto regional e analisando o desempenho com métricas como acurácia, precisão, F1-score e matriz de confusão.

1.2 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo geral delineado na introdução, esta seção detalha as metas intermediárias que guiarão a execução da pesquisa. Portanto, faz-se necessário atingir os seguintes objetivos específicos:

- Desenvolver e utilizar uma base de dados, composta por imagens faciais de piauienses, que atenda a critérios específicos de qualidade e uniformidade, como datasets balanceados em relação a gênero e etnia, com imagens de alta resolução;
- Realizar o ajuste-fino de cinco arquiteturas de CNNs pré-treinadas (AlexNet, InceptionV3, ResNet50, VGG16 e Xception) para a tarefa de classificação de gênero a partir de recortes faciais;

- Avaliar o desempenho das arquiteturas com base em métricas confiáveis, como acurácia, precisão, *F1-score*, e análise da matriz de confusão, para identificar o modelo que oferece a melhor performance e menor viés.

1.3 CONTRIBUIÇÕES

Este trabalho visa agregar valor científico e prático ao campo do reconhecimento facial por meio de resultados específicos e relevantes. A seguir, são destacadas as principais inovações da pesquisa.

- Análise comparativa de múltiplas arquiteturas de Redes Neurais Convolucionais (CNNs): O estudo oferece uma avaliação detalhada do desempenho das arquiteturas VGG16, ResNet50, Xception, Inception V3 e AlexNet, na tarefa de classificação de gênero pelos recortes faciais;

- A abordagem foca na classificação de gênero usando exclusivamente recortes faciais, com um modelo otimizado para a diversidade brasileira. Para isso, utiliza-se uma base de dados local Piauiense, adaptando o sistema às características específicas desta população. Essa estratégia visa mitigar vieses comuns encontrados em dados de outras regiões, resultando em um modelo mais consistente, preciso e equitativo.

1.4 ORGANIZAÇÃO DO TRABALHO

A fim de apresentar a pesquisa de forma clara e estruturada, este documento foi organizado em capítulos que seguem uma progressão lógica. A seção a seguir descreve a estrutura do trabalho, detalhando o conteúdo de cada capítulo, desde a fundamentação teórica que embasa o estudo até a metodologia aplicada, a análise dos resultados obtidos e as conclusões finais:

- Capítulo 2 - Fundamentação Teórica: Este capítulo discute os conceitos fundamentais da pesquisa, como visão computacional, *machine learning*, *deep learning* e CNNs;

- Capítulo 3 - Trabalhos Relacionados: Neste capítulo, são apresentados e discutidos os trabalhos relacionados à pesquisa, com foco na classificação de gênero e nas técnicas empregadas por esses estudos;

- Capítulo 4 - Metodologia: Este capítulo detalha a metodologia utilizada para o modelo de classificação, descrevendo desde a aquisição das imagens até as métricas de classificação;

- Capítulo 5 - Resultados: Aqui são apresentados e analisados os resultados obtidos com o modelo;
- Capítulo 6 - Conclusão: Por fim, este capítulo apresenta as considerações finais, destacando as contribuições do trabalho, suas limitações e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo detalha os tópicos essenciais para a compreensão das técnicas empregadas na elaboração dos métodos propostos. Nas seções subsequentes, discutimos conceitos relacionados a visão computacional, *machine learning*, *deep learning*, aumento de dados e as métricas de desempenho utilizadas para validar os resultados experimentais.

2.1 VISÃO COMPUTACIONAL

A visão computacional é uma área da ciência que confere às máquinas a capacidade de “enxergar” e interpretar o mundo. Ela se concentra em como um computador pode entender o seu entorno, extraindo informações relevantes de imagens e vídeos obtidos por câmeras, sensores e scanners (MILANO; HONORATO, 2014). Essas informações são a base para que a máquina possa reconhecer, manipular e até mesmo raciocinar sobre os objetos em uma imagem (SZELISKI, 2022).

O campo da visão computacional é relativamente novo. As primeiras ideias surgiram em 1955, quando o pesquisador Selfridge sugeriu a criação de "olhos e ouvidos para o computador". Nas décadas seguintes, as pesquisas avançaram, principalmente a partir dos anos 1970, com os primeiros trabalhos que uniam a visão computacional à Inteligência Artificial. No entanto, percebeu-se que a complexidade da visão humana era muito maior do que se imaginava, pois faltavam modelos que pudessem representar como o cérebro humano interpreta as imagens (MILANO; HONORATO, 2014).

A visão humana é capaz de interpretar objetos de forma extremamente rápida, um processo que ocorre no córtex visual. Inspirados por isso, pesquisadores como os do MIT se dedicam a “ensinar computadores a enxergar como humanos” (MILANO; HONORATO, 2014). Assim, a visão computacional capacita as máquinas a realizarem tarefas inteligentes a partir de dados visuais, aproximando-as da inteligência humana.

2.2 MACHINE LEARNING

Machine Learning (ML) é uma área da Inteligência Artificial (IA) que visa desenvolver algoritmos e modelos capazes de aprender automaticamente a partir de dados, sem intervenção explícita. O principal objetivo é criar sistemas que possam identificar padrões e tomar decisões baseadas nesses padrões, adaptando-se e aprimorando-se com o tempo (GOODFELLOW; BENGIO; COURVILLE, 2016).

Os algoritmos de ML são geralmente classificados em três tipos: supervisionados, não supervisionados e por reforço. No aprendizado supervisionado, o modelo é treinado com dados rotulados, ou seja, exemplos de entradas e saídas conhecidas.

Já o aprendizado não supervisionado busca descobrir padrões ou estruturas ocultas nos dados sem a necessidade de rótulos. Por fim, o aprendizado por reforço envolve a interação de um agente com o ambiente, aprendendo a tomar decisões que maximizem uma recompensa ao longo do tempo (SUTTON; BARTO, 2018).

Para este projeto, foi adotado o aprendizado supervisionado. Essa abordagem é fundamental, pois o objetivo da pesquisa é treinar as arquiteturas de CNNs para classificar o sexo biológico do indivíduo nas imagens faciais. O uso de um dataset composto por dados previamente rotulados é importante para que os modelos aprendam a mapear características visuais.

2.3 DEEP LEARNING

Deep learning (DL), ou aprendizado profundo, é uma subárea do aprendizado de máquina que se baseia na utilização de Redes Neurais Artificiais (RNAs) com múltiplas camadas de processamento. Ao emular a arquitetura e o funcionamento do cérebro biológico, essas redes são projetadas para identificar e aprender representações hierárquicas e complexas de grandes volumes de dados. O que distingue o aprendizado profundo de modelos mais tradicionais é a sua profundidade, definida pela extensão dos Caminhos de Atribuição de Crédito (*CAPs*). Esses caminhos representam as cadeias de ligações causais e modificáveis que conectam ações e seus efeitos, permitindo que a rede determine quais de seus componentes são responsáveis pelo sucesso ou falha de uma tarefa, mesmo em cenários com dependências de longo prazo (SCHMIDHUBER, 2014).

Em vez de depender de uma engenharia de características manual, o DL possibilita que a rede descubra, de maneira autônoma e em cada camada, representações de dados progressivamente mais abstratas. Na camada inicial, a rede pode extrair características de baixo nível, como bordas e texturas. Essas características são então combinadas e transformadas nas camadas subsequentes para formar representações de alto nível, como faces ou objetos inteiros. Esse processo de aprendizado hierárquico permite que o modelo generalize para novos dados com uma precisão notável. A viabilidade prática de treinar redes tão profundas foi impulsionada por avanços em algoritmos e, principalmente, pelo desenvolvimento de *hardware* de processamento paralelo, como as Graphics Processing Unit (*GPUs*) (SCHMIDHUBER, 2014).

2.4 REDES NEURAIS CONVOLUCIONAIS

As CNNs são uma classe de redes neurais profundas amplamente utilizadas em tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação. Elas são compostas por camadas convolucionais que utilizam filtros para extrair características locais das imagens, permitindo que a rede aprenda

padrões hierárquicos de forma eficiente. As CNNs se destacam por sua capacidade de capturar características espaciais das imagens e lidar com variações em escala, rotação e translação (LECUN; BENGIO; HINTON, 2015).

Este trabalho utilizou as CNNs como a principal abordagem para o desenvolvimento do modelo de classificação de imagens, visto que sua estrutura e capacidade de aprender padrões complexos são promissoras para o sucesso do projeto. Desde sua popularização em tarefas de reconhecimento de imagens, as CNNs têm sido aplicadas com sucesso em diversas outras áreas, como processamento de áudio e vídeo, reconhecimento de texto e até mesmo em modelos de redes generativas (GOODFELLOW; BENGIO; COURVILLE, 2016). A principal vantagem das CNNs para este projeto reside em sua capacidade de reduzir significativamente o número de parâmetros necessários para treinamento, compartilhando pesos nos filtros e tornando o processo mais eficiente. No entanto, a aplicação eficaz de CNNs ainda enfrenta desafios, como a eficiência computacional e a necessidade de grandes volumes de dados rotulados para treinar modelos robustos (KRIZHEVSKY; SUTSKEVER; HINTON, 2017).

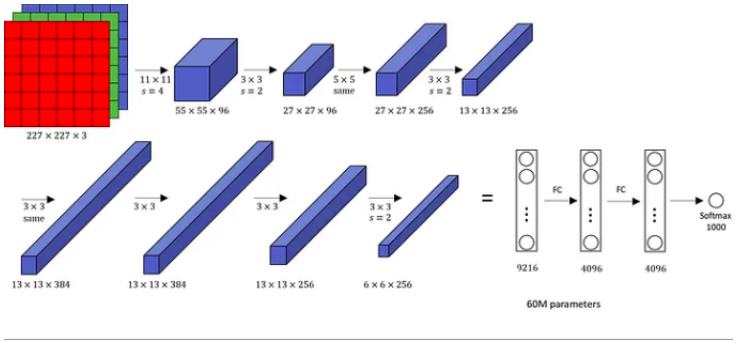
Neste contexto, as CNNs são fundamentais para o sucesso do projeto, permitindo que a rede aprenda a identificar e classificar padrões específicos em imagens de forma eficiente e precisa, atendendo às necessidades do sistema proposto.

2.4.1 AlexNet

A AlexNet surgiu em 2012, criada por Alex Krizhevsky, Ilya Sutskever e Geoffrey Hinton, e ganhou destaque ao vencer a competição anual *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC). Sua vitória foi um marco histórico, pois ela obteve uma taxa de erro de apenas 15,3% no desafio, uma melhora drástica em comparação com os 26,2% do segundo colocado. Esse feito reacendeu o interesse da comunidade científica e tecnológica pelo aprendizado profundo (*deep learning*), que até então era visto com certo ceticismo (BANGAR, 2022).

A AlexNet é uma arquitetura, ilustrado na Figura 1, de CNN pioneira que revolucionou o campo da visão computacional. Sua principal função é a classificação de imagens, ou seja, ela é treinada para reconhecer e categorizar objetos em imagens.

Figura 1 – Arquitetura AlexNet



Fonte: (BANGAR, 2022)

A referida rede é composta por 8 camadas, sendo 5 delas convolucionais e 3 totalmente conectadas ela otimiza o desempenho e evita o sobre-ajuste (*overfitting*) por meio de diversas técnicas. Ela emprega camadas de *max-pooling* para diminuir a dimensionalidade dos dados e camadas de normalização de resposta local para auxiliar na generalização da rede. Para garantir que o modelo não "decore"os dados de treinamento, a AlexNet usa duas estratégias principais: o *Data Augmentation*, que aumenta o conjunto de dados com transformações como recortes e rotações de imagens, e o *Dropout*, que desativa aleatoriamente neurônios durante o treinamento, forçando a rede a aprender características mais robustas e menos dependentes (BANGAR, 2022).

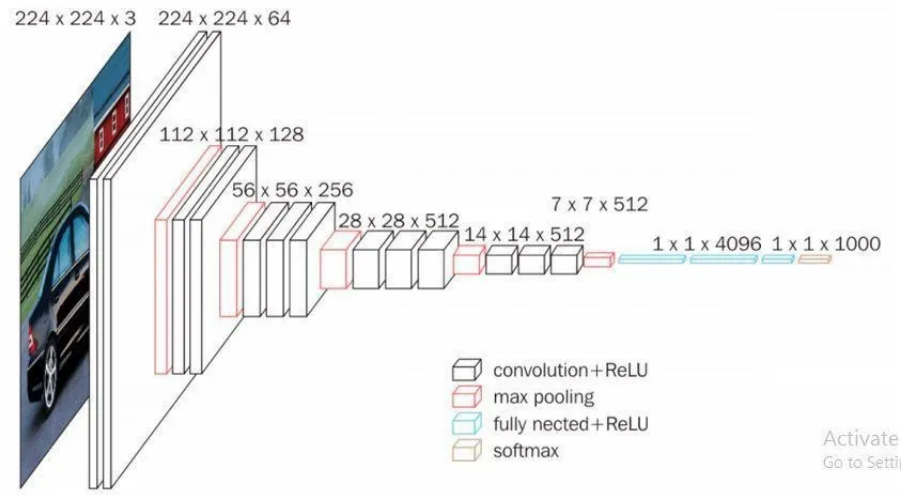
2.4.2 VGG16

A VGG16 é uma arquitetura de CNN nomeada em homenagem ao *Visual Geometry Group* da Universidade de Oxford, sendo frequentemente utilizada em tarefas de visão computacional, como o reconhecimento facial. Sua designação “16” refere-se ao número de camadas do modelo, que são divididas em 13 camadas convolucionais e 3 camadas totalmente conectadas. Na arquitetura, as camadas convolucionais atuam progressivamente na extração de características visuais (como bordas, texturas e padrões), enquanto as três camadas totalmente conectadas, posicionadas ao final, funcionam como o componente de decisão. Estas últimas recebem o conjunto de características processado e o sintetizam para determinar a similaridade entre duas imagens de entrada, que é o objetivo principal no reconhecimento facial. O modelo foi treinado para essa finalidade utilizando o conjunto de dados (ALI *et al.*, 2024).

O modelo utiliza imagens de entrada padronizadas em um tamanho de 224x224 *pixels*. No processo de reconhecimento facial, a VGG16 foca em características essenciais dos rostos, como olhos, nariz e boca, e não em informações de cor. Para isso, as imagens são convertidas para escala de cinza, o que também ajuda a diminuir a complexidade computacional (RUSSAKOVSKY *et al.*, 2015).

A arquitetura VGG16, representada na Figura 2, usa filtros convolucionais de tamanho pequeno (3x3), o que permite uma extração de características mais detalhada e eficiente. As camadas de pooling (geralmente *max pooling*) são usadas para reduzir a dimensionalidade dos mapas de características e o número de parâmetros, o que ajuda a otimizar o desempenho da rede (ALI *et al.*, 2024).

Figura 2 – Arquitetura VGG16



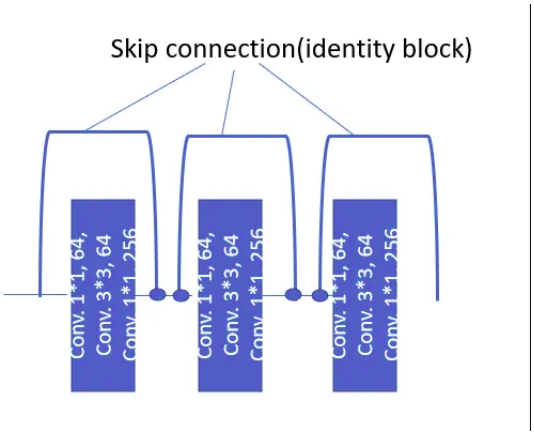
Fonte: (ROHINI, 2021)

2.4.3 ResNet50

A ResNet50 (*Residual Network-50*) é uma arquitetura de CNN profunda, composta por 50 camadas, que se destacou por seu desempenho no conjunto de dados *ImageNet* em 2015. Sua criação foi uma resposta ao desafio do gradiente de desaparecimento, um problema comum em redes neurais profundas que impede que o gradiente se propague eficientemente através das camadas, comprometendo o aprendizado (SHARMA, 2023).

A solução para esse problema foi a introdução das "*skip connections*" ou conexões residuais, mostrado na Figura 3. Essas conexões permitem que o gradiente ignore certas camadas, facilitando o fluxo de informações e permitindo que a rede aprenda uma função residual, representado na Figura 3, o que se mostrou mais eficaz. A ResNet, da qual a ResNet50 faz parte, foi proposta em dezembro de 2015 por pesquisadores da *Microsoft Asia*, e a abordagem foi tão bem-sucedida que uma variante, a ResNet152, venceu a competição de classificação ILSVRC 2015 (SHARMA, 2023).

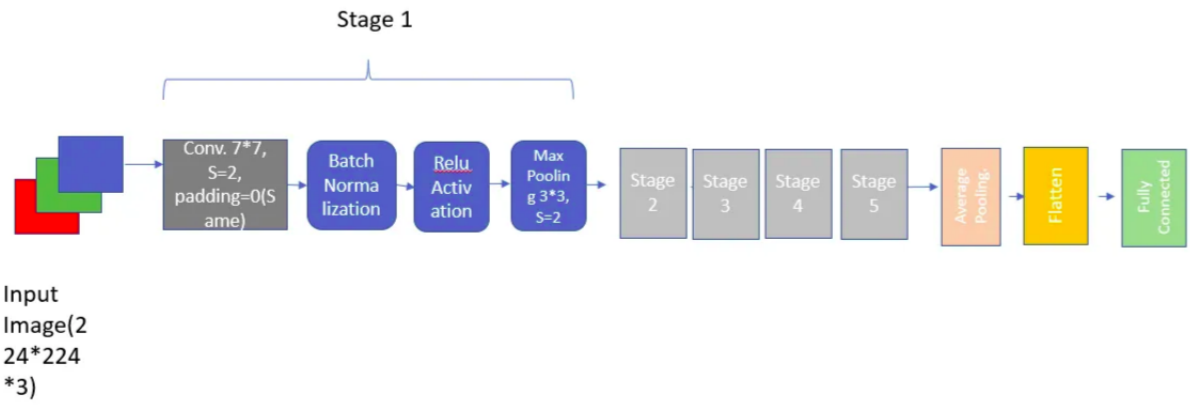
Figura 3 – Conexões Residuais



Fonte: (SHARMA, 2023)

Teoricamente, a ResNet50 se baseia na ideia de que a degradação da precisão em redes mais profundas não é resultado de *overfitting*, mas sim uma limitação fundamental da própria arquitetura. Para combater isso, os blocos de identidade foram criados para garantir que o desempenho não diminua com a adição de mais camadas. Eles permitem que camadas desnecessárias sejam "ignoradas"com um peso de regularização zero, enquanto camadas úteis podem continuar a melhorar o desempenho. A ResNet50, especificamente, aceita uma imagem de 224x224x3 como entrada e utiliza componentes como camadas convolucionais, normalização de lote (*batch normalization*) e ativação *Rectified Linear Unit* (ReLU) para extrair características e processar a imagem, representado na Figura 4 (SHARMA, 2023).

Figura 4 – Tipo de entrada padrão da CNN ResNet50

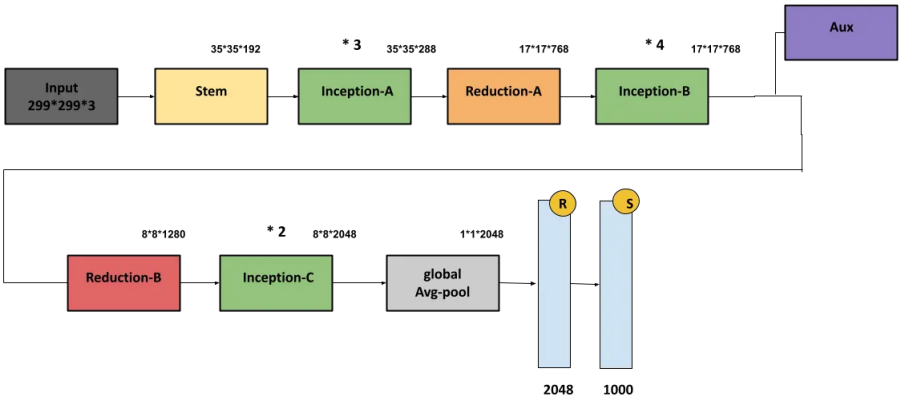


Fonte: (SHARMA, 2023)

2.4.4 InceptionV3

A arquitetura de CNN conhecida como Inception, ou GoogLeNet, emergiu como um marco em 2014 ao vencer a competição ILSVRC. Sua contribuição mais significativa não foi apenas a superação de modelos anteriores, como a AlexNet, com uma taxa de erro de 6,67%, mas sim a introdução de um novo paradigma de *design* de redes. A Inception foi concebida com o objetivo de otimizar o uso de recursos computacionais, garantindo alta precisão com uma notável redução no número de parâmetros e na carga de processamento, tornando-a particularmente eficiente na tarefa de classificação de imagens em larga escala (GHANTIWALA, 2022), arquitetura ilustrada na Figura 5.

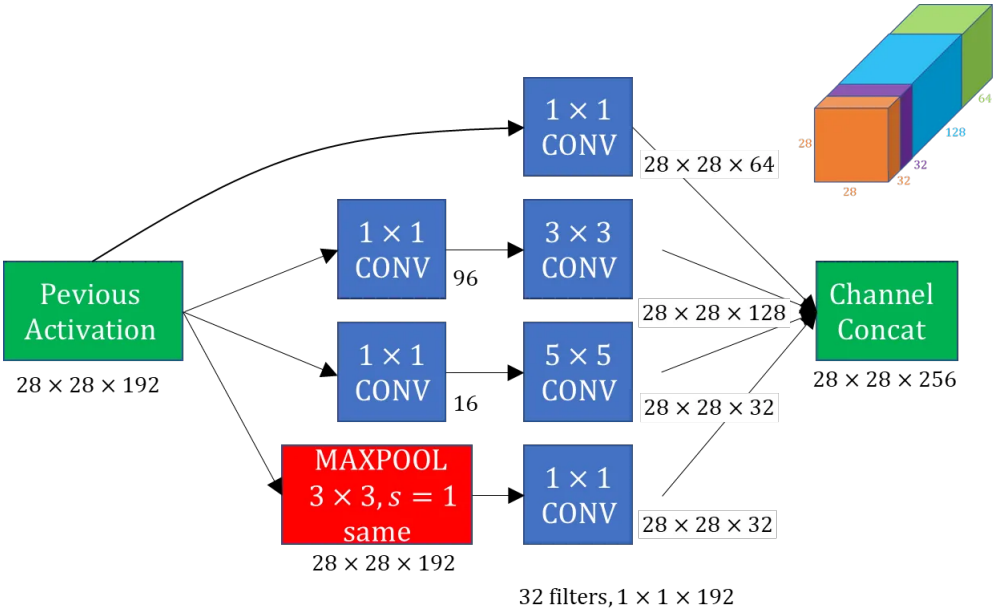
Figura 5 – Arquitetura Inception



Fonte: (BRITAL, 2021)

O cerne da arquitetura Inception reside no seu módulo homônimo, uma estrutura que resolve o desafio de escolher o tamanho ideal do filtro convolucional para uma determinada camada, retratado na Figura 6. Em vez de usar um único tipo de filtro, o módulo Inception emprega múltiplas operações de forma paralela, incluindo convoluções de 1×1, 3×3, 5×5 e operações de *max-pooling*. Os resultados dessas operações são então concatenados, permitindo que a rede extraia características em diferentes escalas e níveis de abstração simultaneamente. Essa abordagem inovadora permitiu um desempenho superior ao capturar informações de forma mais completa e eficiente (GHANTIWALA, 2022).

Figura 6 – Filtro Inception



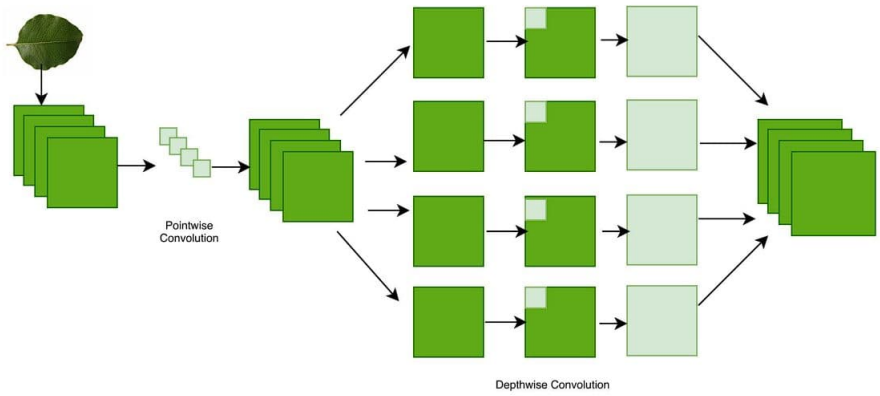
Fonte: (ISBAROV, 2021)

2.4.5 Xception

A arquitetura Xception, cujo nome deriva de "*Extreme Inception*", representa uma evolução significativa do conceito introduzido pela Inception. Proposta em 2017 por François Chollet, essa arquitetura se baseia na hipótese de que as correlações espaciais e as correlações de canal em um mapa de características podem ser modeladas de forma independente. Em vez de usar as convoluções tradicionais, que consideram essas correlações simultaneamente, a Xception introduz um novo tipo de camada, as convoluções separáveis em profundidade (KIM, 2025).

O principal diferencial da Xception é justamente o uso dessas convoluções, que funcionam em duas etapas sequenciais, conforme a Figura 7. Primeiro, uma Convolução de Profundidade (*Depthwise Convolution*) aplica um filtro único a cada canal de entrada de forma independente. Em seguida, uma Convolução de Ponto (*Pointwise Convolution*), que utiliza filtros de 1×1, é aplicada para combinar os canais de saída da etapa anterior. Essa abordagem não apenas reduz drasticamente o número de parâmetros do modelo, mas também aumenta a eficiência computacional, permitindo que a rede seja mais profunda e mais larga sem a sobrecarga de uma convolução padrão (KIM, 2025).

Figura 7 – Convolução de Profundidade e Convolução de Ponto

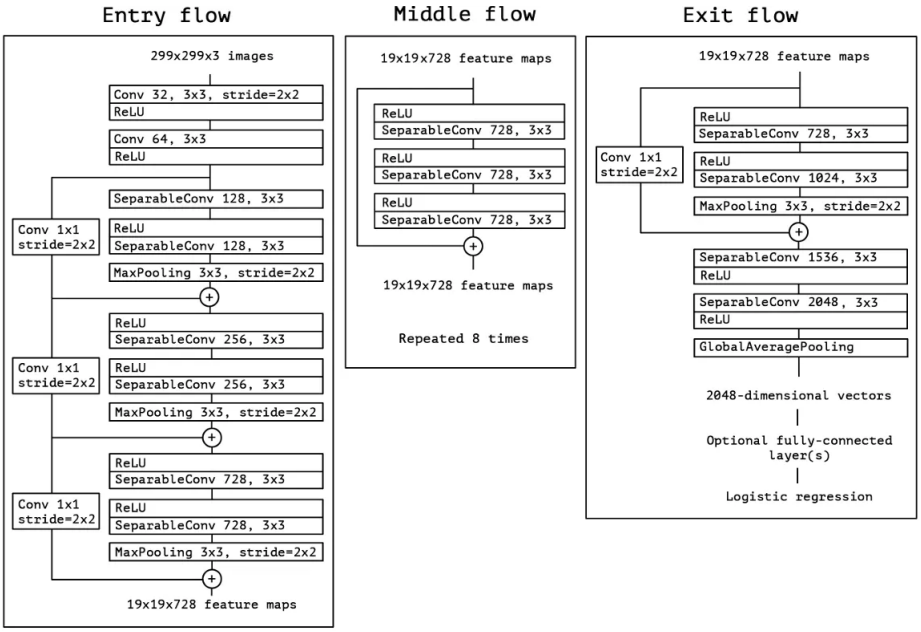


Fonte: (BOASCH, 2024)

A arquitetura da Xception é composta por 36 camadas convolucionais, estruturadas em 14 módulos. O fluxo de dados da rede é dividido em três partes: o *Entry flow*, que processa a entrada; o *Middle flow*, que se repete várias vezes e é o coração da rede onde as convoluções separáveis em profundidade são aplicadas; e o *Entry flow*, que consiste nas camadas finais que levam à classificação. Ao substituir completamente o módulo Inception original por esses blocos de convoluções separáveis, a Xception simplifica a arquitetura e o treinamento, inclusive evitando o uso de classificadores auxiliares (KIM, 2025).

Em resumo, a Xception levou o conceito de modularidade da Inception ao extremo, separando a convolução em suas dimensões espacial e de canal. Essa inovação resultou em uma arquitetura mais leve, mais rápida de treinar e com desempenho superior, validando a ideia de que a modelagem independente de correlações pode ser a chave para o avanço em redes neurais (KIM, 2025), arquitetura Xception representada na Figura 8.

Figura 8 – Arquitetura Xception



Fonte: (HESARAKI, 2023)

3 TRABALHOS RELACIONADOS

A classificação de gênero a partir de imagens faciais é uma tarefa específica dentro do campo da visão computacional que tem recebido atenção crescente com o avanço das técnicas de Aprendizado de Máquina, em especial das CNNs. Esta seção apresenta uma revisão da literatura relevante, abordando os métodos propostos para esta tarefa, os desafios associados, como o viés demográfico, e posicionando a presente pesquisa no contexto atual.

3.1 CLASSIFICAÇÃO DE GÊNERO COM APRENDIZADO DE MÁQUINA

A literatura recente tem explorado diferentes abordagens para a classificação de gênero. Um exemplo notável é o trabalho de (WICAKSONO; SHIDIQ, 2024), que realiza uma comparação direta entre os algoritmos K-Nearest Neighbors (KNN), um método clássico de aprendizado de máquina, e as CNNs para a classificação de gênero baseada em imagens dos olhos. O estudo conclui que as CNNs, devido à sua capacidade de aprender e extrair características complexas hierarquicamente, apresentam um desempenho superior para esta tarefa. Esta conclusão reforça a escolha metodológica desta pesquisa de focar exclusivamente em arquiteturas de CNNs, que representam o estado da arte na extração de características de imagens faciais.

Apesar da eficácia das CNNs, a busca bibliográfica aponta que há uma quantidade limitada de trabalhos que realizam uma análise comparativa extensiva entre diferentes arquiteturas de CNNs para a tarefa específica de classificação de gênero. Muitos estudos propõem e validam uma única arquitetura ou comparam uma CNN com métodos mais tradicionais. Essa lacuna evidencia a necessidade de investigações que avaliem o comportamento e a eficiência de múltiplos modelos de DL consolidados, como os abordados nesta monografia.

3.2 O DESAFIO DO VIÉS DEMOGRÁFICO EM MODELOS DE ANÁLISE FACIAL

Um ponto crítico na literatura de análise facial é a questão do viés. Embora muitos modelos de classificação alcancem alta acurácia, seu desempenho pode não ser uniforme entre diferentes grupos demográficos. Trabalhos como os de (ALI; AL-ARBO, 2025) e (MARVELLOUS, 2025) investigam este problema no contexto mais amplo do reconhecimento facial, mas suas conclusões são diretamente aplicáveis à classificação de gênero.

O trabalho apresentado em (ALI; AL-ARBO, 2025) explora o uso de aprendizado adversário para mitigar o viés demográfico, mostrando que é possível treinar modelos

para que sejam menos sensíveis a atributos como etnia ou idade. Por sua vez, (MARVELLOUS, 2025) demonstra a importância do rebalanceamento e da amostragem do conjunto de dados (dataset) como uma técnica eficaz para garantir a equidade dos resultados. Ambos os estudos ressaltam que a composição do dataset de treinamento é um fator determinante para a generalização e a justiça do modelo final. Essa discussão é fundamental para esta pesquisa, pois a análise de desempenho das cinco arquiteturas de CNNs não deve se limitar à acurácia geral, mas também considerar como cada modelo se comporta diante da diversidade de faces presentes no conjunto de dados.

3.3 POSICIONAMENTO DA PESQUISA E TABELA COMPARATIVA

A análise da literatura revela que os estudos existentes ou focam na comparação de CNNs com métodos mais antigos para a classificação de gênero, ou abordam o problema do viés em um escopo mais geral. A Tabela 1 resume os principais trabalhos analisados e os contrasta com a proposta desta pesquisa.

Tabela 1 – Tabela comparativa dos trabalhos

Autor/Ano	**Modelo(s) usado(s)**	**Dataset**	**Métricas**	**Resultados / Contribuição**
Rizky & Shidio / 2024	KNN e CNN	Dataset de imagens dos olhos (privado)	Acurácia	CNNs apresentam desempenho superior ao KNN para classificação de gênero.
Ali A & Al-Arbo / 2025	Redes Adversariais (Adversarial Learning)	FairFace e outros (públicos)	Taxa de erro e disparidade entre grupos	Demonstra que é possível mitigar o viés demográfico em modelos de análise facial.
Marvellous / 2025	Não especifica uma arquitetura; foca na técnica de reamostragem.	UTKFace (público)	Acurácia e outras métricas de equidade	A reamostragem do dataset é uma técnica eficaz para reduzir o viés do modelo.
ESTE TRABALHO	VGG16, ResNet50, Xception, InceptionV3, AlexNet	**Dataset Piauiense (Original)**	Acurácia, Curva ROC, Recall, F1-Score, Matriz de Confusão	Análise comparativa de 5 arquiteturas de CNN para classificação de gênero em um contexto regional específico.

Fonte: Produzido pelo autor.

Como destacado na tabela, este trabalho se diferencia e contribui para a área de três formas principais:

- Análise Comparativa Direta: Ao contrário de outros estudos, esta pesquisa realiza uma avaliação comparativa do desempenho de cinco arquiteturas de CNNs distintas e amplamente reconhecidas (AlexNet, InceptionV3, VGG16, ResNet50 e Xception) aplicadas estritamente à tarefa de classificação de gênero;

- Foco Específico: O objetivo central é entender como diferentes arquiteturas se comportam e quais características faciais são mais relevantes para cada uma na tarefa de distinguir entre os gêneros masculino e feminino, preenchendo uma lacuna na literatura que carece de comparações diretas;
- Uso de Dataset Original: A utilização de um conjunto de dados com imagens de indivíduos do Piauí confere originalidade à pesquisa e contribui para a análise de desempenho dos modelos em um contexto demográfico específico, que é menos representado nos datasets internacionais amplamente utilizados.

Dessa forma, esta monografia não apenas avalia a eficácia de diferentes modelos, mas também gera conhecimento sobre sua aplicabilidade em um cenário regional, alinhando-se à discussão sobre a importância de dados diversos e representativos.

4 METODOLOGIA

A presente pesquisa adota uma abordagem metodológica baseada no aprendizado de máquina, utilizando a técnica de transferência de aprendizado para a tarefa de classificação de sexo. O trabalho avalia o desempenho de cinco arquiteturas de CNNs pré-treinadas: AlexNet, InceptionV3, ResNet50 VGG16 e Xception. O objetivo é comparar a capacidade de cada modelo para classificar o gênero do indivíduo em duas classes de identidade (feminino e masculino), utilizando o mesmo conjunto de dados e parâmetros de treinamento para uma análise justa e controlada.

A metodologia de treinamento é aplicada a todos os modelos de CNN, respeitando as especificidades de cada uma, conforme mencionado na fundamentação teórica. Isso significa que, embora o conjunto de dados, métricas, épocas e otimizadores sejam os mesmos, as particularidades de cada arquitetura, como o tipo de entrada, foram devidamente ajustadas para garantir a compatibilidade e a execução correta da análise.

A Figura 9 ilustra o fluxograma da metodologia proposta para a classificação de sexo dos indivíduos a partir da face. As etapas desta metodologia são detalhadas nos subtópicos seguintes.

Figura 9 – Fluxograma da metodologia proposta para classificação do sexo

Fonte: Produzido pelo autor.

4.1 AQUISIÇÃO DE IMAGENS

A obtenção de um conjunto de dados robusto e heterogêneo é fundamental para o desenvolvimento de sistemas de visão computacional. Neste estudo, utilizamos um conjunto de dados privado, composto por 33.033 imagens faciais, capturadas da região do queixo até a sobrancelha. A coleta dessas imagens foi realizada mediante a autorização expressa dos indivíduos, em conformidade com a Lei Geral de Proteção de Dados (LGPD), garantindo a conformidade ética e a privacidade dos dados, utilizadas unicamente para esta pesquisa.

As imagens do conjunto, que incluem tanto projeções frontais quanto laterais, apresentam dimensões originais de 1285x1285 *pixels*. O conjunto foi estrategicamente particionado para o desenvolvimento e a validação do modelo, dividindo-se em um subconjunto de treinamento com 23.123 imagens, um subconjunto de teste com 4.955 imagens e, por fim, o conjunto de validação com 4.955 imagens. Essa divisão permite o treinamento confiável do sistema com a divisão 70%, 15% e 15%, repectivamente, e sua posterior avaliação em um conjunto de dados independente.

A rotulagem do conjunto de dados foi efetuada manualmente por meio da cate-

gorização das imagens em pastas. Essa organização foi baseada no sexo biológico, agrupando as imagens de mulheres no diretório 01_feminino e as de homens no diretório 02_masculino. Essa rotulagem é essencial para a compreensão da composição do *dataset* e para eventuais análises de viés de gênero no desempenho do modelo.

4.2 CNNS UTILIZADAS

Para a tarefa de detecção de sexo a partir de recortes faciais, foram utilizadas cinco CNNs pré-treinadas no ImageNet: AlexNet, InceptionV3, ResNet50 VGG16 e Xception. Esses modelos foram selecionados por sua capacidade comprovada na extração de características e no reconhecimento de padrões em imagens. A entrada padrão para todas essas CNNs é uma imagem colorida (RGB), mas cada uma exige um tamanho específico. A VGG16 e a ResNet50 foram projetadas para receber imagens de 224x224 *pixels*, enquanto a Xception e a InceptionV3 requerem imagens de 299x299 *pixels*. Já a AlexNet utiliza imagens de 227x227 *pixels*. Para garantir o correto funcionamento dos modelos, todas as imagens de entrada foram redimensionadas e normalizadas de acordo com as especificações de cada arquitetura.

4.2.1 Imagenet

O ImageNet se estabeleceu como um dos mais importantes recursos na área de visão computacional, servindo como um catalisador fundamental para o avanço do deep learning. Esse vasto e detalhado banco de dados, composto por mais de 14 milhões de imagens anotadas, tornou-se um padrão de referência para treinar redes neurais em tarefas de alta complexidade. Sua influência é inegável, pois forneceu a base para que arquiteturas de deep learning pudessem ser treinadas e avaliadas em larga escala, permitindo que a inovação florescesse (JOCHER, 2023).

No nosso estudo, a influência do ImageNet é bastante importante, pois em vez de treinar as redes do zero, uma abordagem que exigiria um volume massivo de dados e poder computacional e por isso optamos por utilizar o método de *transfer learning* (MUREL; KAVLAKOGLU, 2024). As cinco arquiteturas de CNNs selecionadas foram carregadas com pesos pré-treinados no ImageNet, baseando-se no princípio de que as camadas iniciais de uma CNN aprendem a extrair características genéricas, como bordas e texturas, úteis para diversas tarefas de visão computacional.

Dessa forma, as camadas convolucionais de cada modelo foram congeladas, mantendo os pesos otimizados, enquanto apenas as camadas densas finais, responsáveis pela classificação, foram removidas e substituídas por novas camadas, treinadas especificamente para a nossa tarefa binária de classificação de sexo (masculino/feminino). Essa estratégia permitiu um treinamento mais rápido e eficiente, aproveitando a capacidade de generalização das arquiteturas para alcançar um maior

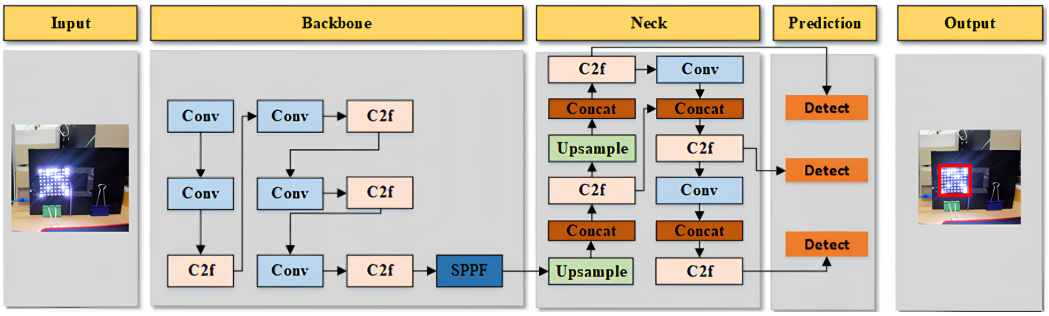
desempenho com o nosso conjunto de dados de imagens faciais.

4.2.2 YOLOv8

O YOLOv8 (*You Only Look Once*) versão 8 é um modelo de detecção de objetos de última geração, lançado pela Ultralytics em 2023. Ele é considerado uma solução de ponta no campo da detecção de objetos , projetado para identificar e localizar com precisão objetos em imagens ou vídeos, como rostos. Sua finalidade é executar essa tarefa mantendo um equilíbrio otimizado entre alta precisão e velocidade em tempo real (YASEEN, 2024).

Para detectar uma face, o YOLOv8 utiliza sua arquitetura de rede neural dividida em três partes, representada na Figura 10. Primeiro, o *Backbone* (Espinha Dorsal) extrai características visuais da imagem em diferentes níveis. Em seguida, o *Neck* (Pescoço) funde e refina essas informações para entender o contexto e a escala , o que é crucial para encontrar rostos de tamanhos variados. Por fim, o *Head* (Cabeça) analisa essas características consolidadas para prever a localização do objeto através de uma caixa delimitadora e classificá-la como "objeto". Ele usa uma abordagem "*anchor-free*" (livre de âncoras), que o ajuda a se adaptar melhor a diferentes formatos de rosto sem caixas de referência predefinidas (YASEEN, 2024).

Figura 10 – Etapas da detecção por meio do YOLOv8



Fonte: (YASEEN, 2024)

Para este trabalho, após o preparo inicial das imagens, citado nas seções acima, o YOLOv8 é aplicado com a finalidade específica de detectar e isolar a região da face. A detecção é focada estritamente na face, compreendendo a área vertical do meio da testa até a metade do queixo, e horizontalmente da metade da bocheca para dentro, excluindo deliberadamente orelhas ou outras partes, representado na Figura 9. O modelo recorta essa região facial identificada, e esse recorte é a etapa indispensável que serve como entrada para todo o processamento subsequente de imagens e o treinamento dos modelos.

4.3 PRÉ-PROCESSAMENTO DAS IMAGENS

O pré-processamento é uma etapa fundamental para uniformizar as características das imagens e otimizar a entrada para as redes neurais convolucionais. Para o nosso estudo de classificação de sexo a partir de imagens faciais, implementamos um fluxo focado em padronização e realce de atributos relevantes.

Primeiramente, para que as imagens se adequem ao formato de entrada das arquiteturas, todas foram padronizadas, conforme mencionado nas sessões anteriores. Após o recorte da face realizado pelo YOLOv8, aplicamos a imagem o preenchimento com zeros (*zero padding*) para transformá-las em um formato quadrado e, em seguida, redimensionamos para as dimensões (*pixels*) de entrada de cada CNN. Essa abordagem garante que não haja distorção das feições faciais, preservando as características essenciais para a classificação.

Para eliminar os efeitos de variações de iluminação e contraste, comuns em fotografias, aplicamos a Técnica de Equalização de Histograma Adaptativa com Limitação de Contraste (CLAHE) (ZAVALA, 2025). Esta técnica divide a imagem em pequenos blocos e equaliza o histograma de cada região, realçando detalhes finos e melhorando a visibilidade das características faciais para os modelos.

Para identificar a arquitetura mais adequada para a tarefa de classificação de sexo, realizamos um estudo comparativo utilizando cinco CNNs de alta performance, todas pré-treinadas no vasto conjunto de dados ImageNet (IMAGENET, 2025). O uso do *transfer learning* permite que os modelos aproveitem o conhecimento adquirido na detecção de características genéricas e o apliquem ao nosso domínio específico. As arquiteturas selecionadas para o estudo são AlexNet, InceptionV3, ResNet50 VGG16 e Xception.

O procedimento de treinamento foi padronizado para todas as arquiteturas, com as camadas convolucionais de cada modelo congeladas para reter os pesos pré-treinados no ImageNet, enquanto as camadas densas finais foram substituídas por novas camadas de classificação para aprenderem carácteristicas especificas do dataset do presente estudo. Essa abordagem permite uma comparação justa entre os modelos, com base em sua capacidade de extrair e utilizar as características faciais para a classificação.

4.3.1 Aumento de Dados

O aumento de dados (*data augmentation*) é uma técnica utilizada no treinamento de modelos de aprendizado de máquina, especialmente em tarefas de visão computacional, para aumentar a quantidade e a diversidade dos dados disponíveis sem a necessidade de coletar novos dados. Isso é feito por meio de transformações que alteram as imagens de treinamento, gerando versões modificadas e mantendo as

informações relevantes, o que pode melhorar a generalização do modelo e evitar o *overfitting* (PEREZ; WANG, 2017).

Para o presente trabalho, essas transformações incluíram deslocamento horizontal e vertical de 10% e 8%, respectivamente, rotação de até 4 graus, espelhamento horizontal e um zoom aleatório de até 2%. Tais transformações foram escolhidas com o objetivo de manter a integridade das características importantes das imagens, enquanto proporcionam ao modelo maior capacidade de generalização. Os valores selecionados para cada parâmetro visam equilibrar a diversidade das imagens sem introduzir distorções excessivas. Adicionalmente, todas as imagens foram submetidas a um processo de normalização, no qual os valores dos *pixels* foram redimensionados para o intervalo entre 0 e 1.

A Figura 11 apresenta a imagem original e a aplicação das transformações. Nela, é possível observar como o aumento de dados foi implementado, gerando variações da mesma imagem por meio de deslocamento, rotação, espelhamento e zoom.

Figura 11 – Aumento de Dados

Fonte: Produzido pelo autor.

Essas transformações permitem visualizar como a imagem original foi modificada, criando novas versões que mantêm suas características essenciais, mas introduzem pequenas alterações. O deslocamento altera ligeiramente a posição da imagem, a rotação ajusta seu ângulo, o espelhamento inverte sua orientação e o *zoom* modifica sua escala.

4.4 MÉTRICAS DE AVALIAÇÃO

Para avaliar o desempenho dos modelos, serão utilizadas diversas métricas. A análise quantitativa do desempenho de cada modelo será apresentada por meio de gráficos e tabelas. A matriz de confusão será usada para visualizar o número de predições corretas e incorretas, mostrando a contagem de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos para cada classe (masculino e feminino). Complementarmente, serão calculados o *F1-score*, a acurácia, curva ROC e o índice *Kappa*.

4.4.1 Matriz de Confusão

Para a avaliação de modelos de classificação, a matriz de confusão organiza as predições em quatro categorias: verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). Matematicamente, a matriz é construída a partir do cruzamento entre os rótulos verdadeiros e as predições do modelo, permitindo visualizar, de forma consolidada, o desempenho em cada classe. Em termos simples, VP representa as amostras corretamente identificadas como positivas, VN as amostras corretamente identificadas como negativas, FP são negativas incorretamente classificadas como positivas e FN são positivas incorretamente classificadas como negativas (MARIANO, 2021), conforme ilustrado na Tabela 2.

Tabela 2 – Matriz de Confusão

		Valor Predito	
		Sim	**Não**
Real	**Sim**	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: Produzido pelo autor.

4.4.2 F1-Score

O F1-score é a média harmônica entre a precisão (precision) e a revocação (recall), buscando equilibrar a parcimônia de falsos positivos com a captura de verdadeiros positivos. Matematicamente, a precisão é VP/(VP+FP) e a revocação é VP/(VP+FN). O F1-score é então calculado como:

$$F1 = 2 \times \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}}$$

Em termos conceituais, o F1-score penaliza tanto falsos positivos quanto falsos negativos, oferecendo uma métrica estável especialmente em cenários com classes desequilibradas, onde a simples acurácia pode mascarar déficits de desempenho em uma das classes (POWERS, 2011).

4.4.3 Curva ROC e AUC

A curva ROC (*Receiver Operating Characteristic*) é uma ferramenta gráfica utilizada para avaliar o desempenho de classificadores binários. Ela ilustra o *trade-off* (a troca) entre a capacidade do modelo de identificar corretamente os casos positivos

e o risco de classificar incorretamente os casos negativos, através da variação de múltiplos limiares de decisão.A curva é construída plotando a Taxa de Verdadeiros Positivos (TPR), também conhecida como Sensibilidade ou Revocação (recall), no eixo Y, em função da Taxa de Falsos Positivos (FPR) no eixo X. As fórmulas para estes eixos são:

$$TPR = \frac{VP}{VP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

O desempenho geral do modelo é frequentemente sumarizado pela métrica AUC (*Area Under the Curve*), que representa a área sob a curva ROC. Um valor de AUC de 1.0 indica um classificador perfeito (capaz de distinguir todas as classes sem erro), enquanto um valor de 0.5 sugere um desempenho não superior ao de uma classificação aleatória (POWERS, 2011).

4.4.4 Acurácia

A acurácia representa a proporção de predições corretas em relação ao total de predições. Matematicamente, é definida como:

$$Acurcia = \frac{VP + VN}{VP + VN + FP + FN}$$

Intuitivamente, descreve a capacidade global do modelo de classificar corretamente, sem distinguir entre classes. Em conjuntos de dados com distribuição de classes muito desigual, a acurácia pode ser enganosa, o que justifica a complementaridade com métricas como F1-score e k (POWERS, 2011).

4.4.5 Índice de Kappa (k)

O índice de Kappa mede a concordância entre as predições do modelo e os rótulos verdadeiros, levando em conta a concordância que poderia ocorrer por acaso. A expressão tradicional é:

$$k = \frac{p_0 - p_e}{1 - p_e},$$

Onde p0 é a precisão observada quantia de predições corretas e pe é a precisão esperada pelo acaso, calculada com base nas margens da matriz de confusão. Em termos práticos, o k fornece uma avaliação mais robusta da concordância, com interpretação comumente apresentada em faixas que vão de piores que o acaso até acordos quase perfeitos. Valores mais próximos de 1 indicam concordância elevada além do que seria esperado ao acaso, enquanto valores próximos de 0 sugerem concordância semelhante à do acaso (TANG *et al.*, 2015).

A tabela Kappa apresentada na Figura 12, ilustra de forma clara como interpretar os diferentes valores do coeficiente, estabelecendo faixas que qualificam o

nível de concordância observado. Para valores inferiores a zero, considera-se que não há nenhuma concordância entre avaliadores; entre 0 e 0,19, a concordância é classificada como pobre; de 0,20 a 0,39, como suave; de 0,40 a 0,59, moderada; de 0,60 a 0,79, substancial; e de 0,80 a 1,00, quase perfeita. Assim, ao associar o valor do Kappa obtido em uma análise ao respectivo intervalo na tabela, é possível determinar o grau de confiabilidade entre os julgamentos dos avaliadores ou entre as previsões do modelo e os rótulos verdadeiros, tornando a interpretação dos resultados mais objetiva e comparável entre diferentes estudos (LANDIS; KOCH, 1977).

Figura 12 – Tabela Kappa

Valores do Coeficiente Kappa	Interpretação
< 0	Nenhuma Concordância
0 < k < 0,19	Concordância Pobre
0,20 < k < 0,39	Concordância Suave
0,40 < k < 0,59	Concordância Moderada
0,60 < k < 0,79	Concordância Substancial
0,80 < k <1,00	Concordância Quase Perfeita

Fonte: (LANDIS; KOCH, 1977).

5 RESULTADOS

Nesta seção, são apresentados os delineamentos experimentais adotados, bem como os critérios de segmentação das imagens submetidas à análise. Os resultados são expostos por meio de representações gráficas, seguidos da discussão das principais adversidades enfrentadas na classificação do gênero facial. Destacam-se, ainda, questões relativas à semelhança facial, ambiguidade de gênero e as dificuldades relacionadas à delimitação precisa entre os gêneros, evidenciando as principais fontes de confusão e suas implicações para o desempenho do estudo.

5.1 DESENHOS DOS EXPERIMENTOS

O objetivo deste experimento consistiu em desenvolver um método capaz de detectar o sexo de um indivíduo a partir de uma porção específica da face, compreendendo desde a metade da testa até a metade do queixo. Essa abordagem foi selecionada com o propósito de eliminar indicadores externos como cabelo, adornos e roupas, que poderiam influenciar os resultados. A pesquisa teve como foco identificar a melhor rede neural para essa tarefa específica. Com o intuito de reduzir vieses raciais e étnicos e aumentar a precisão para a população local, optou-se por treinar o modelo utilizando um conjunto de dados próprio, composto por fotografias de indivíduos piauienses, obtidas mediante consentimento dos participantes, conforme a LGPD.

O projeto foi executado em um notebook com especificações intermediárias e uma placa de vídeo modelo 3050, utilizando-se o *Jupyter Notebook* como ambiente de desenvolvimento. Diversas bibliotecas de ML e IA, tais como o *TensorFlow 2.20.0* e o *Keras 3.11.3*, foram empregadas para o treinamento e a avaliação dos modelos. O conjunto de dados, criado especificamente para esta finalidade, foi cuidadosamente rotulado e categorizado em duas classes: feminino e masculino.

O procedimento iniciou-se com a organização do conjunto de dados, dividindo-se as imagens em duas pastas principais: treino, teste e validação , com proporções de 70%, 15%, 15% respectivamente, essa divisão é adequada para o treinamento de redes neurais, assegurando que o modelo tenha acesso a um volume expressivo de dados para aprendizado, ao mesmo tempo em que a avaliação é realizada sobre amostras representativas e equilibradas (AMAZON WEB SERVICES, 2025).

Após a preparação, as imagens seguiram para o treinamento dos modelos, onde todas as CNNs utilizaram uma configuração padronizada. Esta configuração definiu parâmetros, como o tamanho da imagem e o número de épocas para duas fases distintas: o treinamento superficial (ou *shallow*), que envolve treinar apenas as camadas finais (classificadoras) mantendo a base da rede pré-treinada congelada, e

o treinamento profundo (ou *deep*), que consiste no ajuste fino de mais camadas da rede, ou até mesmo dela por completo, geralmente com uma taxa de aprendizado menor. Além disso, foram estabelecidos fatores de otimização, como a paciência para interrupção precoce e a taxa de abandono (*dropout*) para regularização.

Para a implementação, foram definidas funções específicas tanto para a arquitetura MLP quanto para as etapas de treinamento *shallow* e *deep* mencionadas. O controle do processo foi assegurado pelo uso de pesos de classe (*class_weight*), para lidar com possíveis desbalanceamentos nos dados, e validadores para monitorar o desempenho. Fundamentalmente, a utilização de *callbacks* como o *ModelCheckpoint*, para salvar o melhor modelo encontrado, e o *EarlyStopping*, para interromper o processo caso o desempenho pare de melhorar, foi importante para otimizar os resultados e garantir que o modelo aprendesse de maneira eficiente sem apresentar sobreajuste (*overfitting*).

A justificativa para a utilização desses parâmetros e configurações reside em sua eficácia apontada pela na literatura especializada para tarefas de reconhecimento e detecção de objetos. As estratégias implementadas, como o uso de ajuste fino em duas etapas (*shallow* e *deep*) e a aplicação de *callback*, permitiram um controle preciso sobre o processo de treinamento, visando otimizar o desempenho do modelo, conter o *overfitting* e garantir que o melhor desempenho fosse preservado.

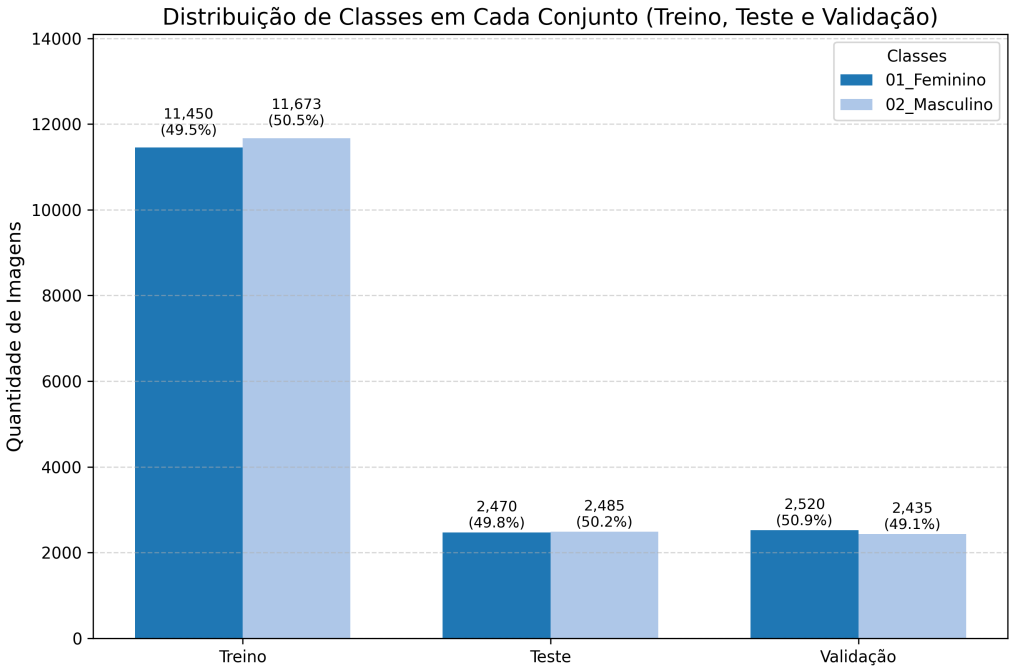
5.1.1 Divisão das imagens

O primeiro passo na análise consiste em compreender a composição do conjunto de dados utilizado na validação dos modelos. A distribuição das imagens é idêntica para todas as arquiteturas avaliadas, garantindo consistência na comparação dos resultados. O conjunto de dados totaliza 33.033 imagens, divididas em três subconjuntos principais: Treino, Teste e Validação.

O conjunto de treino foi balanceado para otimizar o processo de aprendizado e evitar que o modelo se incline para uma classe específica. Conforme ilustrado na Figura 13, a diferença percentual entre as classes “01_feminino” e “02_masculino” é de aproximadamente 4,8 pontos percentuais, o que representa uma distribuição próxima ao equilíbrio ideal (50%/50%). Essa adequação é necessária para garantir que o modelo aprenda de forma equitativa as características visuais de ambas as classes, reduzindo vieses durante o treinamento.

Já os conjuntos de teste e validação mantêm a distribuição natural e não balanceada do conjunto original, simulando de maneira fidedigna o cenário real de aplicação do modelo. Assim, as métricas de desempenho são calculadas sobre essa distribuição preservada, refletindo a eficácia dos modelos em lidar com proporções reais entre as classes, conforme evidenciado na Figura 13.

Figura 13 – Distribuição das Classes Femino e Masculino



Fonte: Produzido pelo autor.

5.2 MÉTRICAS DE DESEMPENHO

Esta seção apresenta a análise detalhada dos resultados de classificação obtidos pelas seguintes CNNs: AlexNet, InceptionV3, VGG16, ResNet50 e Xception. A avaliação será dividida em subseções, uma para cada modelo, onde o desempenho será discutido com foco principal na métrica do Coeficiente Kappa.

Ao final, uma tabela comparativa consolidará as demais métricas de desempenho essenciais para todas as redes, incluindo Acurácia, a Curva ROC com a métrica AUC, coeficiênte Kappa, Matriz de Confusão e *f1-score*. Esta análise visa identificar o melhor modelo para a tarefa de classificação proposta.

5.2.1 Análise dos Resultados

Os resultados obtidos para o coeficiente Kappa nas diferentes arquiteturas são apresentados na Tabela 3 e interpretados à luz das categorias definidas na Tabela 12. A ResNet50 obteve o maior valor de Kappa (0,8957), enquadrando-se na faixa de "concordância quase perfeita" e indicando que suas previsões apresentam alto grau de acordo com os rótulos verdadeiros. A VGG16 apresentou Kappa de 0,7737, classificado como "concordância substancial", enquanto as arquiteturas Xception (0,6120) e AlexNet (0,6071) situam-se no limite inferior dessa mesma faixa, revelando concordância ainda elevada, porém menos precisa. Já o InceptionV3, com Kappa de 0,6680, ocupa uma

posição intermediária dentro da categoria de concordância substancial, evidenciando que, embora todas as arquiteturas alcancem pelo menos um nível satisfatório de concordância, a ResNet50 se destaca como o modelo com desempenho mais consistente segundo a escala de Landis e Koch (1977).

Tabela 3 – Resultados das Métricas de Classificação por Arquitetura

Métrica	**AlexNet**	**ResNet50**	**InceptionV3**	**Xception**	**VGG16**
Kappa	0.6071	0.8957	0.6680	0.6120	0.7737
Acurácia	0.8039	0.8957	0.8358	0.8020	0.8872
ROC-AUC	0.8930	0.9609	0.9042	0.9188	0.9568
F1-score	0.8043	0.8958	0.8355	0.8024	0.8874

Fonte: Produzido pelo autor.

Os resultados das métricas de classificação, apresentados na Tabela 3, evidenciam diferenças significativas entre as arquiteturas. A ResNet50 obteve o melhor desempenho geral, com acurácia e F1-score de 0,8957 e ROC-AUC de 0,9609. A VGG16 apresentou resultados muito próximos, com acurácia de 0,8872, F1-score de 0,8874 e ROC-AUC de 0,9568. O InceptionV3 teve desempenho intermediário, com acurácia de 0,8358 e ROC-AUC de 0,9042, enquanto AlexNet e Xception registraram os menores valores, ambos próximos de 0,80. Ressalta-se que a proximidade de desempenho entre a ResNet50 e a VGG16 se deve às semelhanças estruturais entre essas arquiteturas, que compartilham um nível de complexidade semelhante.

Os resultados apresentados permitem compreender como a complexidade arquitetural das redes impacta diretamente o desempenho obtido nas tarefas de classificação. A ResNet50 destacou-se como a arquitetura de melhor desempenho geral, possivelmente por sua estrutura baseada em blocos residuais. Essa característica permite mitigar o problema da perda de gradiente em redes mais profundas, facilitando o aprendizado de representações mais robustas sem comprometer a capacidade de generalização. Assim, a ResNet50 alcançou um equilíbrio eficiente entre profundidade e estabilidade no treinamento, o que explica seus altos valores de Kappa, acurácia e F1-score.

Por outro lado, a AlexNet apresentou desempenho inferior em comparação às demais arquiteturas. Sua estrutura relativamente simples, com menor número de camadas e capacidade de extração de características, pode ter limitado a generalização dos padrões presentes nos dados, resultando em menor precisão nas classificações. Essa limitação torna-se evidente diante do desempenho mais consistente das redes com maior capacidade de modelagem, como a ResNet50 e a VGG16.

Quanto às arquiteturas mais complexas, como a InceptionV3 e a Xception, seus resultados intermediários sugerem que o potencial de generalização não foi plenamente explorado. Essas redes, por possuírem estruturas mais profundas e sofisticadas, geralmente demandam conjuntos de dados maiores e mais variados para alcançar convergência adequada de seus kernels e pesos. Em contextos com restrições de dados, essa necessidade pode resultar em um ajuste subótimo dos parâmetros, comprometendo a capacidade discriminatória e o desempenho final. Dessa forma, os resultados obtidos reforçam a importância de alinhar a escolha da arquitetura ao tamanho e à complexidade do conjunto de dados disponível, visando maximizar a eficiência do modelo e a qualidade das previsões.

A Tabela 4 apresenta os principais componentes da matriz de confusão para as classes “Feminino” e “Masculino”. Nota-se que a arquitetura ResNet50 obteve os maiores percentuais de verdadeiros positivos (88,5%) e verdadeiros negativos (90,8%), evidenciando acurada distinção entre as classes. VGG16 e InceptionV3 também tiveram desempenho expressivo, com valores elevados nessas métricas. Por outro lado, o modelo Xception mostrou o maior percentual de verdadeiros negativos (93,4%), porém com baixo índice de verdadeiros positivos (69,3%), indicando dificuldade em identificar corretamente a classe “Feminino”. Quanto aos erros, ResNet50, VGG16 e Xception apresentaram os menores valores de falso positivo, enquanto AlexNet e InceptionV3 tiveram índices mais altos. Por fim, Xception se destacou negativamente com elevado percentual de falsos negativos (30,7%), ao contrário de ResNet50 (11,5%) e VGG16 (12,9%), que apresentaram pouca ocorrência desse erro. O conjunto dos dados revela o bom desempenho de ResNet50 e VGG16 na classificação correta das classes, com taxas de erro reduzidas em comparação às demais arquiteturas analisadas.

5.3 DESAFIOS ENCONTRADOS NA CLASSIFICAÇÃO DE GÊNERO FACIAL

Os modelos de classificação de gênero facial demonstraram uma dificuldade notável em casos onde as características faciais se desviam dos estereótipos binários de gênero. A análise da matriz de confusão revelou um conjunto específico de erros que merecem destaque, sendo o principal a confusão na identificação de indivíduos com semelhanças intergênero.

5.3.1 Semelhança Facial e Ambiguidade de Gênero

O cerne do desafio reside na semelhança facial entre gêneros, que é amplificada pela presença ou ausência de elementos de expressão visual que o modelo interpreta como características determinantes. Visto que o treinamento foi realizado com recortes estritos da área facial (incluindo olhos, nariz e boca, e excluindo o cabelo), a performance do classificador é altamente sensível aos padrões visuais contidos nessa

Tabela 4 – Componentes da Matriz de Confusão (Classes: Feminino e Masculino)

Parâmetro	**AlexNet**	**ResNet50**	**InceptionV3**	**Xception**	**VGG16**
Verdadeiro Positivo (VP) (Real: Feminino, Previsto: Feminino)	78.8	88.5	86.5	69.3	87.1
Verdadeiro Negativo (VN) (Real: Masculino, Previsto: Masculino)	82.3	90.8	80.0	93.4	90.6
Falso Positivo (FP) (Real: Masculino, Previsto: Feminino)	17.7	9.2	20.0	6.6	9.4
Falso Negativo (FN) (Real: Feminino, Previsto: Masculino)	21.2	11.5	13.5	30.7	12.9

Fonte: Produzido pelo autor. A classe “Positiva” foi definida como “Feminino” e a “Negativa” como "Masculino". Dados em porcentagem(%).

região. A análise demonstrou que o modelo desenvolveu uma forte, mas simplista, correlação entre certas características e o rótulo de gênero.

Maquiagem como Correlato de Gênero: Uma das principais fontes de erro está ligada à maquiagem facial. No conjunto de dados de treinamento, a maioria das imagens rotuladas como feminino continha features de maquiagem (como delineador, sombra, rímel ou batom). Consequentemente, o modelo aprendeu a tratar a presença de maquiagem como um forte indicador preditivo do gênero feminino. Essa limitação pode ser solucionada com a adição de novas amostras de dados que diminuíssem esse padrão, como a inclusão de mais imagens de mulheres sem maquiagem e, se possível, homens com maquiagem, para que o modelo aprenda a focar em características faciais mais diversas e menos em indicadores superficiais.

5.3.2 A Confusão nos Limites

Homens com Maquiagem: Conforme citado em seções anteriores, as imagens foram rotuladas de acordo com seu sexo biológico. Em decorrência disso, indivíduos rotulados como masculino que utilizavam maquiagem facial em suas imagens foram frequentemente classificados erroneamente como mulheres. A presença dessas características confundiu o modelo, pois elas ativavam o indicador de gênero feminino (maquiagem) aprendido no treinamento, sobrepondo-se às características morfológicas associadas ao rótulo masculino.

Este é um exemplo prático do viés algorítmico (ou discriminação algorítmica),

conforme discutido anteriormente. O erro não é aleatório, mas sistemático, expondo uma falha no conjunto de dados de treinamento. O modelo aprendeu uma correlação espúria (maquiagem = gênero feminino) devido à provável sub-representação de imagens de homens maquiados. A mitigação desse tipo de viés é um desafio na área de "IA Justa" (*Fair AI*). Para evitar essa falha específica, seria necessário intervir no conjunto de treinamento. Estratégias como data *augmentation* focado (coletar ou gerar mais imagens de homens com maquiagem) ou reponderação (*re-weighting*) das amostras durante o treinamento poderiam forçar o modelo a aprender as características morfológicas subjacentes, em vez de depender do atalho da maquiagem.

Mulheres sem Maquiagem: Em menor proporção, a ausência total de maquiagem em faces de indivíduos do grupo 'feminino' gerou alguns erros de classificação. Isso ocorreu quando, além da ausência de maquiagem, a face possuía traços morfológicos que o modelo aprendeu a associar mais fortemente ao grupo 'masculino' [1]. Embora esses casos sejam mínimos, eles reforçam a dependência excessiva do modelo em relação à maquiagem como feature de classificação.

Mulheres com Traços Masculinizados: Além da maquiagem, indivíduos do gênero feminino que possuem características faciais consideradas tradicionalmente "masculinas" ou "grosseiras" [2] foram rotulados incorretamente como homens, sinalizando uma falha em dissociar a morfologia facial dos estereótipos de gênero.

Isso levanta um questionamento fundamental para a área: até que ponto um sistema de visão computacional pode, e eticamente deve, tentar inferir um construto social tão complexo a partir de dados puramente visuais? Um algoritmo treinado para essa tarefa estaria, na prática, classificando a expressão de gênero (a forma como alguém se apresenta) e não a identidade de gênero (quem a pessoa é), correndo o risco de perpetuar estereótipos e realizar classificações incorretas e potencialmente danosas. A tarefa torna-se ainda mais inviável e eticamente questionável se considerarmos a inferência de orientação sexual, que não possui correlatos visuais válidos.

Apesar da elevada acurácia geral alcançada pelos modelos, as falhas observadas nos casos de ambiguidade de gênero apontam para uma sensibilidade residual a certos atributos de apresentação e estilo presentes no dataset, como a maquiagem, iluminação, quantidade de imagens, etc. Isso evidencia uma limitação do modelo e é um indicativo de que, ao atingir alto desempenho, ele incorporou um viés de rotulagem nos poucos casos de fronteira. A dependência dessas features mais variáveis e cosméticas, ainda que minoritária, acaba por comprometer a confiabilidade dos modelos em cenários específicos onde a expressão de gênero se desvia dos padrões majoritários aprendidos.

---

[1] olhos mais afastados, narizes maiores, bocas menores, etc.

[2] mandíbulas proeminentes ou sobrancelhas espessas.

6 CONCLUSÃO

Este trabalho se propôs a realizar uma avaliação comparativa do desempenho de cinco distintas arquiteturas de CNNs, AlexNet, InceptionV3, ResNet50, VGG16 e Xception, na tarefa de classificação de gênero a partir de recortes faciais. O principal objetivo foi identificar a arquitetura mais eficaz, utilizando um conjunto de dados local para mitigar vieses étnicos e garantir maior aplicabilidade à diversidade da população brasileira.

A análise dos resultados demonstrou que todas as arquiteturas pré-treinadas apresentaram um desempenho robusto, validando a eficácia do uso de transfer learning para esta tarefa. No entanto, a comparação direta das métricas revelou diferenças significativas de performance entre os modelos. A arquitetura ResNet50 destacou-se como a mais performática, alcançando os melhores índices em todas as métricas.

Entende-se que, uma das principais contribuições desta pesquisa foi o treinamento dos modelos com um dataset próprio, composto por imagens de indivíduos piauienses, visando a criação de um classificador mais justo e adaptado à realidade local. Contudo, a análise dos erros revelou um desafio importante: os modelos desenvolveram uma forte correlação entre a presença de maquiagem e o gênero feminino. Consequentemente, ocorreram erros de classificação em cenários de ambiguidade, como homens que usavam maquiagem sendo classificados como mulheres e mulheres sem maquiagem com traços faciais mais robustos sendo classificadas como homens. Essa descoberta sublinha que, apesar da alta acurácia, o modelo incorporou um viés baseado em padrões de apresentação visual presentes no dataset, comprometendo sua confiabilidade em casos que fogem dos estereótipos de gênero aprendidos.

Conclui-se, portanto, que a arquitetura ResNet50 é a mais adequada para a tarefa de classificação de gênero a partir de recortes faciais, oferecendo a maior precisão e poder discriminatório entre os modelos avaliados. O estudo reafirma a importância da escolha da arquitetura de CNN e, ao mesmo tempo, destaca a necessidade de trabalhos futuros focados na mitigação de vieses aprendidos, a fim de desenvolver um classificador de sexo mais preciso, ético e generalizável para a diversidade das características humanas.

6.1 TRABALHOS FUTUROS

A presente investigação, ao demonstrar a eficácia da arquitetura ResNet50 na classificação de gênero, estabelece uma base sólida para futuras pesquisas. Embora os resultados sejam promissores, foram identificadas limitações que delineiam direções claras para o aprimoramento e a expansão deste trabalho.

Portanto, para trabalhos futuros, pretende-se adicionar novas imagens ao *dataset* que possam trazer uma representatividade mais heterogênea. Além disso, planeja-se realizar um estudo de aumento de dados (*data augmentation*) focado especificamente na parcela de dados que representa mulheres com traços morfológicos mais associados ao masculino e também sem o uso de maquiagem, bem como na parcela de dados que representa homens que utilizam maquiagem. O objetivo é verificar se essa abordagem proporciona um aprendizado mais generalista para as CNNs, reduzindo a dependência do modelo em características superficiais. Por último, pretende-se testar o desempenho de novas arquiteturas de CNNs e de *Vision Transformers* (ViT) nesta tarefa de classificação.

REFERÊNCIAS

ALI, A.; AL-ARBO, Y. Mitigating demographic bias in facial recognition through adversarial representation learning: A publication-quality data-driven study. **International Research Journal of Innovations in Engineering and Technology**, 2025. 10.47001/IRJIET/2025.906035.

ALI, N. S. *et al.* An effective face detection and recognition model based on improved yolo v3 and vgg 16 networks. **International Journal of Computer, Mechatronics and Electrical Engineering**, 2024. 10.18280/ijcmem.120201.

AMAZON WEB SERVICES. **Dividir os dados em dados de treinamento e de avaliação**. 2025. Disponível em: https://docs.aws.amazon.com/pt_br/machine-learning/latest/dg/splitting-the-data-into-training-and-evaluation-data.html. Acesso em: 29 out. 2025.

BANGAR, S. **Arquitetura AlexNet explicada**. 2022. Medium. Disponível em: https://medium.com/@siddheshb008/alexnet-architecture-explained-b6240c528bd5. Acesso em: 18 set. 2025.

BENABDELKADER, C.; GRIFFIN, P. A local region-based approach to gender classification from face images. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops**, 2005. 10.1109/CVPR.2005.388.

BOASCH, G. **Xception model: Analyzing depthwise separable convolutions**. 2024. Viso.ai. Disponível em: https://viso.ai/deep-learning/xception-model. Acesso em: 22 set. 2025.

BRITAL, A. **Inception V3 CNN architecture explained**. 2021. Medium. Disponível em: https://medium.com/@AnasBrital98/inception-v3-cnn-architecture-explained-691cfb7bba08. Acesso em: 22 set. 2025.

CEIA, E. M. *et al.* **Tecnologia, segurança e direitos: Os usos e riscos do sistema de reconhecimento facil no Brasil**. [S.l.]: Fundação Konrad Adenauer, 2022. ISBN 978-65-89432-29-6.

DURÃES, U. **Reconhecimento facial: erros expõem falta de transparência e viés racista**. 2024. UOL. Disponível em: https://noticias.uol.com.br/cotidiano/ultimas-noticias/2024/04/28/reconhecimento-facial-erros-falta-de-transparencia.htm. Acesso em: 22 jan. 2025.

GHANTIWALA, A. **Understanding architecture of Inception Network and applying it to a real-world dataset**. 2022. Medium. Disponível em: https://gghantiwala.medium.com/understanding-the-architecture-of-the-inception-network-and-applying-it-to-a-real-world-dataset-169874795540. Acesso em: 18 set. 2025.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: The MIT Press, 2016. ISBN 9780262035613.

GUO, G. *et al.* A study on automatic age estimation using a large database. **IEEE**, 2009. 10.1109/ICCV.2009.5459438.

HESARAKI, S. **Xception**. 2023. Medium. Disponível em: https://medium.com/@saba99/xception-cd1adc84290f/. Acesso em: 22 set. 2025.

IMAGENET. **ImageNet**. 2025. Disponível em: https://www.image-net.org/. Acesso em: 20 jan. 2025.

ISBAROV, J. **Going deeper with convolutions: The Inception paper, explained**. 2021. Medium. Disponível em: https://medium.com/aiguys/going-deeper-with-convolutions-the-inception-paper-explained-841a0c661fd3. Acesso em: 22 set. 2025.

JAIN, A. K.; ROSS, A.; NANDAKUMAR, K. **Introduction to biometrics**. [S.l.]: Springer New York, NY, 2011. 10.1007/978-0-387-77326-1. ISBN 978-0-387-77326-1.

JOCHER, G. **ImageNet dataset | Ultralytics docs**. 2023. Disponível em: https://docs.ultralytics.com/pt/datasets/classify/imagenet/. Acesso em: 18 set. 2025.

KIM, D.-K. **Xception: Deep learning's leap beyond Inception**. 2025. Medium. Disponível em: https://medium.com/@kdk199604/xception-deep-learnings-leap-beyond-inception-05a708c205f9. Acesso em: 18 set. 2025.

KODANDARAM, R. *et al.* Face recognition using truncated transform domain feature extraction. **The International Arab Journal of Information Technology (IAJIT)**, 2015. Volume 12, Number 03, pp. 1 - 9,.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. *et al.* (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2017. v. 25. 10.1145/3065386.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159–174, Mar 1977.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature 521**, 2015. 436–444. 10.1038/nature14539.

MALTONI, D. *et al.* **Handbook of fingerprint recognition**. [S.l.]: Springer London, 2009. 10.1007/978-1-84882-254-2. ISBN 978-1-84882-254-2.

MARIANO, D. **Métricas de avaliação em machine learning**. 2021. Disponível em: https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/. Acesso em: 29 out. 2025.

MARVELLOUS, A. Bias mitigation through dataset resampling: A case study on facial recognition systems. **Pattern Analysis and Applications**, 2025. Disponível em: https://www.researchgate.net/publication/392595033. Acesso em: 29 out.2025.

MILANO, D.; HONORATO, L. B. **Visão computacional Palavras-Chaves**. [S.l.]: UNICAMP – Universidade Estadual de Campinas FT – Faculdade de Tecnologia, 2014. Disponível em: https://www.academia.edu/9621896/VIS%C3%83O_COMPUTACIONAL_Palavras_Chave. Acesso em: 27 jan. 2025.

MORENO, J. C. *et al.* Robust periocular recognition by fusing sparse representations of color and geometry information. **Journal of Signal Processing Systems**, 2016. 10.1007/s11265-015-1023-3.

MUREL, J.; KAVLAKOGLU, E. **What is transfer learning?** 2024. IBM. Disponível em: https://www.ibm.com/topics/transfer-learning. Acesso em: 18 set. 2025.

NETO, P. C. *et al.* Compressed models decompress race biases: What quantized models forget for fair face recognition. **INESCTEC**, arXiv:2308.11840, 2023.

OLHAR DIGITAL. **Sistemas de reconhecimento facial podem ter viés racial, dizem pesquisadores**. 2016. Disponível em: https://olhardigital.com.br/2016/07/08/seguranca/sistemas-de-reconhecimento-facial-podem-ter-vies-racial-dizem-pesquisadores/. Acesso em: 14 jan. 2025.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. **arXiv preprint arXiv:1712.04621**, 2017.

POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc, informedness, markedness correlation. **Journal of Machine Learning Technologies**, 2011. 10.48550/arXiv.2010.16061.

RAMALHO, J. P. V. **Reconhecimento facial e a autenticidade: a questão de vieses étnicos**. 2024. SBC Horizontes. ISSN 2175-9235. Disponível em: https://horizontes.sbc.org.br/index.php/2024/11/reconhecimento-facial-e-a-autenticidade-uma-questao-de-vieses-etnico/. Acesso em: 22 jan. 2025.

ROHINI, G. **Everything you need to know about VGG16**. 2021. Medium. Disponível em: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918. Acesso em: 22 set. 2025.

RUSSAKOVSKY, O. *et al.* Imagenet large scale visual recognition challenge. **International Journal of Machine Learning Technology**, 2015. 2:1, pp.37-63. 10.48550/arXiv.2010.16061.

SATO, K.; SHAH, S.; AGGARWAL, J. Partial face recognition using radial basis function networks. In: **Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition**. [S.l.: s.n.], 1998. p. 288–293. 10.1109/AFGR.1998.670963.

SCHMIDHUBER, J. In: **Deep learning in neural networks: An overview**. [S.l.]: Neural Networks, 2014. v. 25, p. 85–117. 10.1016/j.neunet.2014.09.003.

SHARMA, T. **Detailed explanation of Resnet CNN model**. 2023. Medium. Disponível em: https://medium.com/@sharma.tanish096/detailed-explanation-of-residual-network-resnet50-cnn-model-106e0ab9fa9e. Acesso em: 18 set. 2025.

SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. [S.l.]: The MIT Press, 2018. ISBN 9780262039246.

SZELISKI, R. **Computer vision: Algorithms and applications**. [S.l.]: Springer Cham, 2022. 10.1007/978-3-030-34372-9. ISBN 978-3-030-34371-2.

TANG, W. *et al.* Kappa coefficient: A popular measure of rater agreement. **Shanghai Archives of Psychiatry**, 2015. 10.11919/j.issn.1002-0829.21501.

TARAWNEH, A. S. *et al.* Invoice classification using deep features and machine learning techniques. In: **2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)**. [S.l.: s.n.], 2019. p. 855–859. 10.1109/JEEIT.2019.8717504.

WICAKSONO, R. D.; SHIDIQ, G. F. Comparison of knn and cnn algorithms for gender classification based on eye images. **Scientific Journal of Informatics**, 2024. 11(4):915-924. 10.15294/sji.v11i4.13529.

YASEEN, M. What is yolov8: An in-depth exploration of the internal features of the next-generation object detector. **Department of Sciences and Humanities, National University of Computer and Emerging Science**, 2024. 10.48550/arXiv.2408.15857.

ZAVALA, F. **Histogram equalization CLAHE algorithm**. 2025. Medium. Disponível em: https://medium.com/imagecraft/histogram-equalization-clahe-algorithm-8841d402fc76. Acesso em: 22 set. 2025.

ZHAO, W. *et al.* Face recognition: A literature survey. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 35, n. 4, p. 399–458, 2003. ISSN 0360-0300. 10.1145/954339.954342.