

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística e Departamento de Informática
Programa de Especialização em *Data Science* e *Big Data*

Nilton Garcia Sainz

Comparação entre modelos de CNNs para estimação étnico-racial da classe política
brasileira

Curitiba
2025

Nilton Garcia Sainz

Comparação entre modelos de CNNs para estimação étnico-racial da classe política
brasileira

Artigo apresentado ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito para a obtenção do grau de especialista.

Orientador: Prof. Dr. Anderson Ara

Curitiba

2025

Comparação entre modelos de CNNs para estimação étnico-racial da classe política brasileira

Comparison between CNN models for ethnic-racial estimation of the Brazilian political class

Nilton Garcia Sainz¹, Anderson Ara²

¹ Discente do curso de Data Science & Big Data, Departamento de Estatística e Departamento de Informática, Setor de Ciências Exatas, Universidade Federal do Paraná (UFPR)

² Departamento de Estatística (Dest), Setor de Ciências Exatas, Universidade Federal do Paraná (UFPR)

Resumo

A sub-representação de pretos e pardos na política brasileira é um problema histórico, cuja análise é dificultada pela ausência de dados raciais de candidatos antes de 2014. Este artigo explora o uso de Redes Neurais Convolucionais (CNNs) para estimar a etnia de políticos brasileiros a partir de fotos de urna, visando preencher essa lacuna. Comparamos cinco arquiteturas de CNN (VGG-19, ResNet-50, Fair Face ResNet-34, MobileNet-V2 e RGCNN) em cenários de classificação binária (branco/não branco) e ternária (branco/pardo/preto). Os resultados indicam que, no cenário binário, o modelo customizado RGCNN obteve o melhor desempenho (BACC de 0.746), demonstrando a viabilidade da classificação em duas classes. No entanto, a classificação ternária mostrou-se mais desafiadora, com o MobileNet-V2 alcançando o melhor resultado (BACC de 0.636), evidenciando a dificuldade de capturar as nuances do espectro racial brasileiro. O estudo conclui que, embora promissora, a classificação racial automatizada requer um debate ético aprofundado e o desenvolvimento de modelos mais justos e precisos para evitar a perpetuação de vieses.

Palavras-chave: Redes Neurais Convolucionais, Classificação Racial, Ciências Sociais Computacionais, Sub-representação de Raça, Classe Política Brasileira.

Abstract

The underrepresentation of Black and Brown people in Brazilian politics is a historical problem, the analysis of which is hampered by the absence of racial data for candidates prior to 2014. This article explores the use of Convolutional Neural Networks (CNNs) to estimate the ethnicity of Brazilian politicians from their ballot photos, aiming to fill this data gap. We compare five CNN architectures (VGG-19, ResNet-50, Fair Face ResNet-34, MobileNet-V2, and RGCNN) in binary (White/Non-White) and ternary (White/Brown/Black) classification scenarios. The results indicate that in the binary scenario, the custom RGCNN model achieved the best performance (BACC of 0.746), demonstrating the feasibility of two-class classification. However, the ternary classification proved more challenging, with MobileNet-V2 achieving the best result (BACC of 0.636), highlighting the difficulty of capturing the nuances of the Brazilian racial spectrum. The study concludes that, although promising, automated racial classification requires a thorough ethical debate and the development of fairer and more accurate models to avoid perpetuating biases.

Keywords: Convolutional Neural Networks, Racial Classification, Computational Social Science, Underrepresentation of race, Brazilian Political Class.

1 Introdução

A sub-representação histórica de pretos e pardos nos espaços de poder no Brasil é um fenômeno amplamente documentado, refletindo desigualdades estruturais que persistem desde as candidaturas até os resultados eleitorais (Campos & Machado, 2017). Uma barreira significativa para o estudo aprofundado dessa dinâmica reside na ausência de dados étnico-raciais de candidatos em pleitos anteriores a 2014, ano em que o Tribunal Superior Eleitoral (TSE) passou a exigir a autodeclaração racial. Essa lacuna de dados impede a análise longitudinal da participação e do desempenho de minorias raciais na política brasileira,

limitando a compreensão sobre a evolução da representatividade no país.

Nesse contexto, as Ciências Sociais Computacionais (CSS) emergem como um campo promissor, oferecendo novas ferramentas e abordagens para enfrentar problemas sociais complexos por meio de técnicas de mineração de dados, processamento, modelagem e simulação computacional (Cioffi-Revilla, 2014; Lazer et al., 2009). A aplicação de métodos de Ciência de Dados, em particular as Redes Neurais Convolucionais (CNNs), apresenta uma oportunidade para superar, de forma ética e responsável, a carência de dados raciais na Justiça Eleitoral.

Este artigo investiga o potencial de diferentes arquiteturas de CNNs para realizar a estimação étnico-racial de candidatos brasileiros com base em suas fotografias de urna. O objetivo geral é comparar o potencial de modelos como Fair Face ResNet-34, MobileNet-V2, RGCNN, VGG Face-19 e ResNet-50, avaliando seu desempenho em cenários de classificação binária (branco e não branco) e ternária (branco, pardo e preto). Adicionalmente, busca-se discutir as implicações e os limites éticos da construção e validação de modelos de aprendizado de máquina para classificação étnico-racial, considerando a complexidade do construto de raça no Brasil e a sensibilidade de sua aplicação em estudos sobre representação política.

2 Fundamentação Teórica

2.1 Visão computacional e reconhecimento étnico-racial

A classificação de atributos demográficos a partir de imagens faciais, como raça e etnia, consolidou-se como um campo de pesquisa interdisciplinar, impulsionado pelos avanços em Redes Neurais Convolucionais (CNNs). A trajetória dessa área evoluiu de métodos tradicionais, baseados na extração manual de características texturais e espaciais com algoritmos como Local Binary Patterns (LBP) e Principal Component Analysis (PCA), para abordagens de aprendizado profundo que automatizam a extração de características complexas diretamente dos dados (Bishop, 2006).

As CNNs, em particular, revolucionaram a visão computacional por seu desempenho superior em tarefas de classificação. Arquiteturas como a VGG-19 (Simonyan & Zisserman, 2015) demonstraram a eficácia do empilhamento de camadas convolucionais com filtros pequenos (3x3) para aumentar a profundidade e a capacidade de aprendizado. No entanto, o aprofundamento das redes trouxe desafios como o desaparecimento do gradiente. As Redes Residuais (ResNets), como a ResNet-50 (He et al., 2015), superaram essa barreira ao introduzir “conexões de atalho” (*shortcut connections*), que permitem o treinamento de redes muito mais profundas, aprendendo funções residuais em vez de mapeamentos diretos.

Buscando maior eficiência computacional, especialmente para aplicações em dispositivos móveis, surgiram arquiteturas como a MobileNet-V2 (Sandler et al., 2018). Este modelo utiliza convoluções separáveis em profundidade (*depthwise separable convolutions*) e blocos de “gargalo linear” (*linear bottlenecks*) com resíduos invertidos, reduzindo drasticamente o número de parâmetros e o custo computacional sem sacrificar significativamente a performance. Em paralelo, modelos customizados

como a RGCNN (Mikaeel et al., 2024) foram desenvolvidos especificamente para a tarefa de reconhecimento de raça e gênero, buscando um equilíbrio entre leveza (menor número de parâmetros) e alta precisão.

Contudo, a aplicação de tais modelos não está isenta de desafios éticos. A literatura aponta para um viés significativo nos bancos de dados de treinamento, majoritariamente compostos por rostos de pessoas brancas, o que resulta em modelos com desempenho inferior para alguns grupos raciais sub-representados (Karkkainen & Joo, 2019). Essa disparidade pode amplificar desigualdades existentes e levar a conclusões enviesadas em pesquisas demográficas e sociais.

A criação de datasets mais balanceados, como o FairFace, e a adoção de métricas de avaliação focadas em justiça (*fairness*), como a Acurácia Balanceada (*Balanced Accuracy*), são estratégias cruciais para mitigar esses vieses e garantir que os sistemas funcionem de forma equitativa entre diferentes grupos. A complexidade da classificação racial no contexto brasileiro, marcado pela miscigenação, adiciona uma camada de desafio, exigindo que os modelos sejam desenvolvidos e avaliados com especial atenção às suas implicações sociais e culturais.

3 Materiais e Métodos

A metodologia deste estudo envolve a comparação de cinco arquiteturas de Redes Neurais Convolucionais (CNNs) para a tarefa de classificação étnico-racial, aplicadas a dois conjuntos de dados distintos (binário e ternário), derivados das fotografias de urna de candidatos. A seguir, detalham-se os componentes da metodologia empregada.

3.1 Conjunto de Dados

A justificativa para utilizar Minas Gerais como *proxy* do cenário nacional baseia-se na notável similaridade das distribuições étnico-raciais entre o estado e o Brasil. Conforme demonstram os dados de 2016 apresentados no Gráfico 1, as três categorias principais apresentam distribuições praticamente idênticas: candidatos brancos representam 51,4% em MG versus 51,5% nacionalmente (diferença de apenas 0,1 ponto percentual), pardos constituem 37,8% em MG contra 39,1% no Brasil (diferença de 1,3 pontos percentuais). A categoria preta apresenta ligeira divergência: 10,3% em MG versus 8,6% nacionalmente (diferença de 1,7 pontos percentuais). As categorias amarela e indígena, com representação inferior a 1%, mostram variações mínimas que não comprometem a representatividade do recorte estadual.

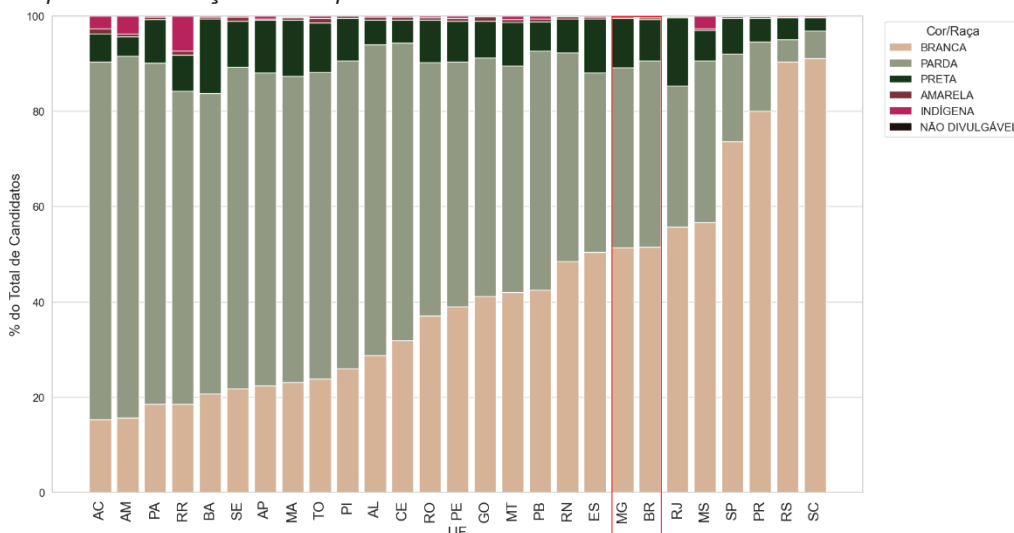
Esta convergência distributiva é particularmente relevante para o estudo, uma vez que MG concentra o segundo maior colégio eleitoral do país e apresenta diversidade geográfica interna significativa, abrangendo desde regiões

metropolitanas até áreas rurais e de mineração. A similaridade das proporções raciais entre MG e Brasil, fornece base metodológica sólida para inferências sobre padrões de classificação racial aplicáveis ao contexto nacional, minimizando vieses de seleção geográfica que poderiam comprometer a validade externa dos achados.

O estudo apresenta limitações importantes que devem ser consideradas na interpretação dos resultados. Primeiro, a restrição geográfica a Minas Gerais limita a generalização dos achados para o contexto nacional, uma vez que a classificação racial varia regionalmente no Brasil devido a diferenças culturais e demográficas. Segundo, a defasagem temporal entre os dados de treinamento (2016) assume estabilidade dos padrões fenotípicos e das dinâmicas de classificação racial, premissa que não se sustenta completamente. Terceiro, o uso de autodeclaração como *ground truth* carrega as limitações inerentes desse método, incluindo possível inconsistência individual e influência de contextos sociais específicos do momento da candidatura.

O conjunto de dados original continha 77.088 imagens de candidatos. A partir deste conjunto, foram criados dois *datasets* para os experimentos: um binário e um ternário.

Gráfico 1: Distribuição das autodeclarações raciais por UF nas eleições municipais de 2016



Fonte: Elaboração própria.

O *dataset* binário foi construído com duas classes: "Branca" e "Não Branca". A classe Não Branca agrupou as categorias autodeclaradas como preta, parda, amarela e indígena. O objetivo desta binarização foi testar a capacidade dos modelos em distinguir a classe majoritária (branca) das minorias raciais agrupadas.

O *dataset* ternário foi elaborado com três classes: branca, parda e preta. Esta abordagem buscou um nível de detalhamento maior, alinhado às categorias

mais representativas do IBGE e do debate sobre sub-representação racial na política brasileira Brasil. Para a construção deste dataset, os experimentos foram divididos em duas decisões binárias hierárquicas: primeiro, um modelo distinguia entre Branco e Não Branco; em seguida, um segundo modelo classificava os classificados como não brancos como pretos ou pardos.

Em todos os experimentos, o conjunto de dados foi dividido em 80% para treinamento, 10% para validação e 10% para teste, utilizando uma semente de reprodutibilidade (*seed* 42) para garantir a consistência dos resultados.

3.2 Arquiteturas de CNNs

Foram selecionadas cinco arquiteturas de CNN para comparação, abrangendo desde modelos clássicos e profundos até arquiteturas mais leves e especializadas, a fim de avaliar diferentes abordagens para a tarefa de classificação racial.

VGG-19 (Face): Utilizou-se uma adaptação do modelo VGG-19 (Simonyan & Zisserman, 2015), pré-treinado no dataset VGG-Face (Parkhi et al., 2015). Este modelo é conhecido por sua profundidade e pelo uso de múltiplos filtros convolucionais de 3x3, que permitem a extração de características complexas. A escolha justifica-se por ser uma arquitetura de base consolidada em tarefas de reconhecimento facial.

ResNet-50 e Fair Face ResNet-34:

Foram incluídos dois modelos da família ResNet (He et al., 2015). O ResNet-50 é uma arquitetura profunda que utiliza "blocos residuais" para otimizar o treinamento e evitar a degradação de performance. O Fair Face ResNet-34 é uma adaptação do ResNet-34, treinado no dataset FairFace (Karkkainen & Joo, 2019), que é balanceado por raça, gênero e idade, buscando

mitigar vieses de classificação.

MobileNet-V2: Esta arquitetura (Sandler et al., 2018) foi escolhida por sua eficiência computacional. Projetada para ambientes com recursos limitados, como dispositivos móveis, utiliza "convoluções *depthwise separable*" e "blocos residuais invertidos" para reduzir o número de parâmetros e o custo de processamento, representando uma abordagem de modelo leve.

RGCNN (Race Gender Convolutional Neural Network): Trata-se de um modelo especializado e de arquitetura

mais leve, proposto por Mikaeel et al. (2024), com foco específico na predição de raça e gênero. Sua inclusão visa comparar o desempenho de um modelo customizado e com atenção a parâmetros éticos frente a arquiteturas de propósito mais geral.

3.3 Parâmetros e configurações

Os modelos foram treinados com configurações padronizadas para garantir a comparabilidade dos resultados. O treinamento foi realizado por 30 épocas para o dataset binário e entre 30 e 40 épocas para o ternário, utilizando um mecanismo de *early stopping* com *patience* de 7 épocas e uma variação mínima de 0.001 (0.002 para o ternário) para evitar sobreajuste.

A função de perda utilizada foi a *CrossEntropyLoss*, e o otimizador foi o *AdamW* com *weight decay* de $1e-4$. Para a regularização e melhoria da generalização, foram aplicadas técnicas de *Data Augmentation*, incluindo *RandomResizedCrop*, *RandomHorizontalFlip*, *RandomRotation* e *ColorJitter*. O agendador de taxa de aprendizado (*scheduler*) foi o *OneCycleLR* com estratégia de cosseno. Os experimentos foram conduzidos em um ambiente com GPU (NVIDIA RTX 4070), permitindo o uso de precisão mista (*mixed precision*) para acelerar o treinamento.

4 Resultados

A avaliação dos modelos nos experimentos binário e ternário foi realizada com foco em métricas que refletem não apenas a acurácia geral, mas também o equilíbrio e a justiça da classificação entre as classes, aspecto fundamental em tarefas de classificação racial. As principais métricas utilizadas foram a Acurácia Balanceada (BACC), a área sob a curva ROC (AUC) e o F1-Macro.

4.1 Experimento binário

No cenário de classificação binária (Branco vs. Não Branco), o modelo RGCNN demonstrou superioridade em todas as métricas principais, como F1-Macro, Acurácia Balanceada (BACC) e Acurácia geral (ver Tabela 1). Este resultado destaca o potencial de arquiteturas customizadas, que incorporam mecanismos para lidar com padrões fenotípicos específicos, superando modelos de propósito geral. O RGCNN foi o único a apresentar valores elevados tanto de BACC quanto de AUC (ver Anexo I), um forte indicativo de que o modelo aprendeu a distinguir efetivamente os limites entre as classes, em vez de apenas acertar estatisticamente devido ao desequilíbrio do dataset. Em contraste, o modelo VGG-19 apresentou o desempenho mais baixo em todos os critérios avaliados.

Tabela 1: Matriz de performance dos modelos binários

Métrica	ResNet-50	ResNet-34	MobileNet-V2	VGG-19	RGCNN
ACC	0.741	0.736	0.732	0.693	0.747
BACC	0.740	0.735	0.732	0.691	0.746
F1-M	0.740	0.735	0.732	0.691	0.746
Sens	0.686	0.706	0.719	0.619	0.710
Spec	0.793	0.764	0.745	0.763	0.781

Fonte: Elaboração própria.

Na análise da matriz de confusão do modelo RGCNN fica reforçado seu desempenho superior. O modelo foi capaz de detectar corretamente a maioria dos candidatos da classe minoritária, acertando 2673 de 3766 instâncias. Este é um ponto crucial, pois erros de classificação sobre grupos minoritários são particularmente sensíveis em contextos sociais e políticos, podendo mascarar desigualdades e reduzir a visibilidade estatística de grupos sub-representados.

A análise mais detalhada das performances contrastantes revela limitações arquiteturais específicas. O baixo desempenho do VGG-19 em todos os critérios pode ser atribuído à sua arquitetura “plana” sem conexões residuais, que sofre do problema de degradação em redes profundas, dificultando a otimização dos gradientes (Simonyan & Zisserman, 2015). Com 21 milhões de parâmetros concentrados principalmente em camadas totalmente conectadas, o VGG-19 demonstra ineficiência computacional (19.6 bilhões de FLOPs) e inadequação para capturar nuances fenotípicas específicas necessárias para classificação racial. Essa ineficiência contrasta com arquiteturas mais modernas que incorporam conexões de atalho e especialização para reconhecimento facial.

Particularmente revelador é o caso da ResNet-50, que apresentou um padrão preocupante de BACC elevado (0.740) combinado com AUC extremamente baixo (0.429) - inferior ao classificador aleatório. Este paradoxo indica um tipo específico de *overfitting* ao desequilíbrio de classes: o modelo, com seus 30 milhões de parâmetros (Karkkainen & Joo, 2019), desenvolveu um viés sistemático em direção à classe majoritária (branca), obtendo alta acurácia balanceada por acertar estatisticamente, mas falhando completamente na capacidade discriminativa real entre classes. Esse fenômeno demonstra que a ResNet-50 memorizou que “a maioria é branca” ao invés de aprender características faciais genuinamente discriminativas, evidenciando como modelos com capacidade excessiva podem desenvolver estratégias estatisticamente eficazes, mas conceitualmente falhas em *datasets* desbalanceados.

4.2 Experimento ternário

Quando a tarefa foi ampliada para três classes (Branco, Pardo e Preto), o desafio de capturar variações raciais mais sutis tornou-se evidente. Nenhum dos modelos

alcançou um desempenho considerado satisfatório para uma aplicação com implicações sociais tão sensíveis. Os modelos *MobileNet-V2* e *FairFace ResNet-34* apresentaram os melhores resultados em termos de Acurácia Balanceada (BACC), com 0.636 e 0.633, respectivamente (ver Tabela 2). O bom desempenho do *FairFace*, um modelo pré-treinado com foco em diversidade facial, sugere que o pré-treinamento em datasets balanceados pode conferir maior aptidão para tarefas de classificação racial com múltiplas classes.

Tabela 2: Matriz de performance dos modelos ternários

Métrica	ResNet-50	Fair Face	MobileNet-V2	VGGFace-19	RGCNN
ACC	0.642	0.663	0.657	0.653	0.651
BACC	0.616	0.633	0.636	0.615	0.627
F1-M	0.581	0.629	0.624	0.617	0.617
Sens	0.616	0.633	0.636	0.615	0.627
Spec	0.817	0.805	0.803	0.797	0.797

Fonte: Elaboração própria.

Apesar do modelo *Fair Face* ter liderado em termos de equilíbrio entre as classes (ver Anexo II), a análise da sua matriz de confusão revelou as dificuldades inerentes à tarefa. Embora o modelo tenha conseguido detectar corretamente a maioria dos candidatos negros, a confusão entre as classes “Parda” e “Branca” foi significativa, refletindo a ambiguidade fenotípica e a complexidade da classificação racial no Brasil. Este resultado evidencia os limites dos modelos atuais em capturar as nuances da autoidentificação e da heteroidentificação racial no país.

5 Discussão

Os resultados obtidos revelam uma distinção clara no desempenho dos modelos de CNN conforme a complexidade da tarefa de classificação racial. No cenário binário, a capacidade de distinguir entre 'brancos' e 'não brancos' se mostrou uma tarefa viável, com o modelo *RGCNN* alcançando um desempenho notável. A superioridade de uma arquitetura customizada e mais leve sobre modelos mais profundos e de propósito geral, como o *ResNet-50*, sugere que a especialização da rede para características fenotípicas específicas de raça e gênero pode ser mais eficaz do que a simples profundidade da rede. O alto desempenho do *RGCNN*, validado tanto pela acurácia balanceada quanto pela AUC, indica que o modelo foi capaz de aprender padrões discriminatórios robustos, minimizando o viés em direção à classe majoritária.

Em contrapartida, o experimento ternário expôs as limitações dos modelos atuais diante da complexidade da classificação racial no Brasil. A queda acentuada no desempenho de todos os modelos, incluindo o

RGCNN, evidencia a dificuldade em capturar as nuances que separam as categorias 'branca', 'parda' e 'preta'. A confusão frequente entre as classes 'parda' e 'branca' é particularmente reveladora, refletindo a sobreposição fenotípica e a fluidez das fronteiras raciais no contexto brasileiro. O fato de os modelos com pré-treinamento em datasets faciais diversos (*FairFace ResNet-34*) e os modelos mais leves e eficientes (*MobileNet-V2*) terem se destacado, ainda que com performance modesta, aponta para a importância de duas frentes: a utilização de dados de treinamento mais representativos da diversidade racial e o desenvolvimento de arquiteturas que possam capturar características sutis de forma eficiente.

A análise das curvas ROC reforça essa interpretação, revelando um padrão no experimento binário: enquanto o *RGCNN* demonstra capacidade discriminativa genuína (AUC=0.819), o *ResNet-50* apresenta um fenômeno preocupante de BACC elevado (0.740) combinado com AUC baixo (0.429) - inferior ao classificador aleatório. Este paradoxo evidencia que o modelo não aprendeu a distinguir efetivamente as classes, mas sim desenvolveu um viés sistemático em direção à classe majoritária, obtendo alta acurácia por meio de classificação estatisticamente enviesada.

No contrário, o experimento ternário exhibe convergência de performance entre todos os modelos (AUCs macro-average entre 0.795-0.811), contrastando com a ampla dispersão observada no binário (0.429-0.819). Esta convergência sugere que todos os modelos atingem um 'teto de complexidade' fundamental, o que representa não apenas uma limitação arquitetural, mas possivelmente um limite inerente à própria abordagem de classificação visual para distinções étnico-raciais sutis.

Os resultados encontram eco nas especificidades da construção social da raça no Brasil, particularmente na fluidez categorial que caracteriza o sistema racial brasileiro. A literatura sociológica documenta que 22,9% dos brasileiros “mudam de raça” entre pesquisas consecutivas, e 40% dos pardos optam por se declarar brancos quando eliminada a categoria intermediária (Senkevics, 2017).

Esta fluidez temporal e contextual das categorias raciais explica, em parte, a dificuldade dos modelos em estabelecer fronteiras classificatórias consistentes no experimento ternário. A alta confusão entre “pardo” e “branco” observada nos resultados não representa apenas limitação técnica, mas espelha a realidade de que os pardos constituem grupo heterogêneo que inclui tanto “pardos-brancos” (fenotipicamente próximos da brancura) quanto “pardos-negros” (identificados com a negritude) (Telles, 2012; Silva & Leão, 2012). A ambiguidade racial brasileira permite maior fluidez classificatória, mas mantém hierarquias cromáticas (Telles, 2012). Os modelos CNN, operando exclusivamente por critério de “marca” (aparência fenotípica), enfrentam contradição

estrutural com a definição oficial do IBGE para “pardo”, baseada em “origem” (ancestralidade multirracial).

A superioridade da classificação binária pode refletir que a divisão branco/não-branco captura mais adequadamente as hierarquias raciais operantes na sociedade brasileira, validando empiricamente argumentos sociológicos sobre a existência de uma hierarquização fundamental que persiste apesar das categorias intermediárias (Telles, 2012). Os modelos CNN mostram-se mais alinhados com as bancas de heteroidentificação das políticas de cotas - que também operam exclusivamente por critérios fenotípicos, ignorando ancestralidade - do que com a classificação oficial do IBGE.

Este achado sugere tensão entre autenticidade cultural, ou seja, a manutenção da fluidez categorial brasileira e operacionalidade política (necessidade de categorias mais rígidas para implementação de políticas públicas). Para fins de automação em larga escala, pode ser necessário “endurecer” as classificações raciais, aproximando o Brasil de sistemas mais rígidos, com implicações importantes para a especificidade histórica do modelo brasileiro de relações raciais.

Esses achados levantam importantes questões sobre os limites éticos e práticos da classificação racial automatizada. Embora a abordagem binária mostre potencial para a geração de dados em larga escala, a dificuldade na classificação ternária reforça que a raça é um construto social complexo, que não pode ser reduzido a meros padrões visuais sem uma perda significativa de informação e a introdução de vieses.

Os achados também sugerem que abordagens hierárquicas (classificação binária seguida de refinamento) podem ser mais promissoras do que classificação multiclasse direta, explorando os pontos fortes demonstrados no cenário binário enquanto reconhecem as limitações fundamentais da distinção visual entre categorias raciais próximas.

A ambiguidade na classificação de pardos, por exemplo, pode levar à invisibilização deste grupo ou à sua assimilação inadequada a outras categorias, com sérias implicações para estudos sobre desigualdade e representação política. Portanto, a aplicação dessas tecnologias deve ser feita com cautela, transparência e uma profunda reflexão sobre seus pressupostos e potenciais impactos sociais.

6 Considerações finais

Este estudo investigou a viabilidade de utilizar Redes Neurais Convolucionais (CNNs) para a estimação étnico-racial de políticos brasileiros, visando suprir a lacuna de dados anterior a 2014. Esta representa a primeira comparação sistemática de CNNs para

classificação étnico-racial política brasileira, estabelecendo um baseline metodológico replicável.

A comparação entre cinco arquiteturas de CNN em cenários de classificação binária e ternária demonstrou que, embora a tecnologia apresente potencial, sua aplicação é complexa e repleta de desafios. No cenário binário, o modelo customizado RGCNN alcançou performance superior (BACC=0.746, AUC=0.819), indicando que a classificação entre 'brancos' e 'não brancos' é uma tarefa factível e com potencial para gerar dados em larga escala para análises longitudinais da representatividade política.

Contudo, a transição para uma classificação ternária ('branco', 'pardo' e 'preto') revelou uma queda significativa no desempenho de todos os modelos (BACC≈0.636), evidenciando a dificuldade em capturar as sutilezas do espectro racial brasileiro. A confusão entre as classes, especialmente entre 'pardos' e 'brancos', sublinha os limites da classificação baseada unicamente em fenótipos e reforça a natureza da raça como um construto social. Os resultados sugerem que modelos pré-treinados em datasets balanceados, como o *FairFace*, oferecem um caminho promissor, mas que ainda requer aprimoramento.

O estudo apresenta limitações importantes que devem ser consideradas na interpretação dos resultados. Os dados utilizados se restringem ao estado de Minas Gerais, e a generalização dos achados para o contexto nacional brasileiro requer validação adicional em outros estados com diferentes perfis demográficos. Ademais, o uso de dados de 2016 para inferir sobre o período 2004-2012 assume estabilidade temporal dos padrões fenotípicos que pode não se sustentar completamente.

Conclui-se que, embora as CNNs sejam ferramentas poderosas, sua utilização para classificação racial deve ser acompanhada de transparência sobre limitações, validação com especialistas em ciências sociais e engajamento com comunidades afetadas. O estudo contribui ao estabelecer baseline metodológico para classificação racial automatizada em contextos políticos, destacando tanto as possibilidades quanto as limitações da abordagem. Futuros trabalhos devem se concentrar no desenvolvimento de modelos mais justos, na construção de bancos de dados mais representativos da diversidade brasileira e na exploração de abordagens que combinem a análise visual com outras fontes de informação, sempre com o objetivo de promover a equidade e a visibilidade de grupos historicamente sub-representados.

Referências

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

Campos, Luiz Augusto & Machado, Carlos. O que afasta pretos e pardos da representação política? Uma análise

a partir das eleições legislativas de 2014. *Rev. Sociol. Polit.* 25 (61), 2017.

Cioffi-Revilla, C. (2014). *Introduction to computational social science: Principles and applications*. Springer. <https://doi.org/10.1007/978-1-4471-5661-1>

Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2. ed. Sebastopol: O'Reilly Media, 2019.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Xplore*.

Karkkainen, K. & Joo, J. (2021) FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age, *IEEE Xplore*.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>

Mikaeel, V., Turan, B., & Abdulrazzaq, M. (2024). Facial Race and Gender Recognition Based on Convolutional Neural Network Models. *Gaziosmanpasa Journal of Scientific Research (GBAD)*, 13(3), 1–18910. Disponível em: <http://dergipark.gov.tr/gbad9>.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In *British Machine Vision Conference*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks, *IEEE Xplore*.

Senkevics, Adriano Souza. De Brancos para Negros? Uma análise longitudinal da reclassificação racial no Enem 2010-2014. In: 38ª Reunião Nacional da ANPEd, 2017, São Luís. *Anais da 38ª Reunião Nacional da ANPEd*.

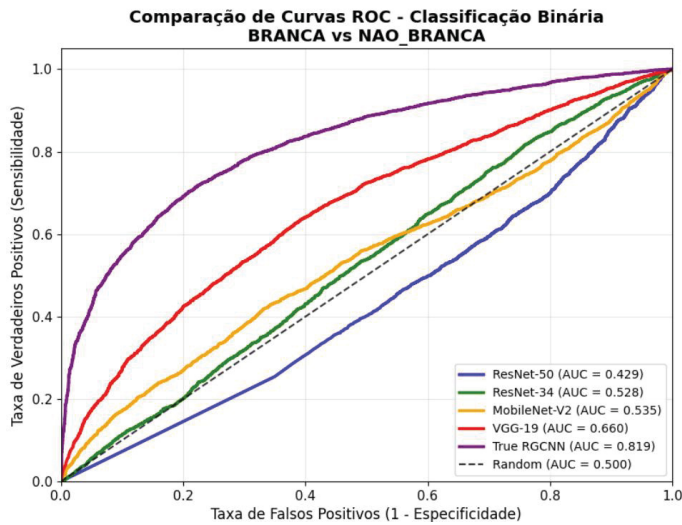
Silva, Graziella Moraes & Leão, Luciana T. de Souza. O paradoxo da mistura: identidades, desigualdades e percepção de discriminação entre brasileiros pardos. *Rev. bras. Ci. Soc.*, v. 27, n. 80, 2012.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition, *ICLR-2015 conference paper*.

Telles, Edward E. *O Significado da Raça na Sociedade Brasileira*. Princeton University Press, 2012.

ANEXOS

Anexo I – Comparação de curvas ROC da classificação binária



Anexo II – Comparação Micro e Macro da classificação ternária

