O QUE O ROSTO DIZ SOBRE A COR DECLARADA: um modelo probabilístico treinado com fotografias de candidatos às eleições brasileiras (2016–2024)

WHAT A FACE SAYS ABOUT DECLARED COLOR: *a probabilistic model trained on photographs of Brazilian election candidates (2016–2024)*

Jean Pierre Ribeiro dos Santos[1]

RESUMO

No Brasil, a cor ou raça é autodeclarada, e a mesma pessoa pode se declarar de formas diferentes ao longo do tempo. Este estudo verificou quanto uma fotografia do rosto informa sobre essa declaração. Foram usadas as bases públicas do Tribunal Superior Eleitoral: 1.569.137 candidaturas de 2016 a 2024 e 113.491 fotografias coloridas de 2018 a 2024, separadas por pessoa em treino, validação e teste. Entre as 290.724 pessoas que concorreram mais de uma vez, 27,7% mudaram a declaração; dois terços dessas mudanças foram entre branca e parda. Um modelo de visão (CLIP ViT-B/16 ajustado), treinado só com declarações estáveis, atingiu acurácia balanceada de 75,0% (IC 95%: 74,0% a 76,0%) entre branca, parda e preta, com revocação de 65,7% para parda. Os erros entre branca e preta ficaram abaixo de 2%. Em 96,8% das fotos de quem mudou de declaração, a classe prevista estava entre as já usadas pela pessoa. As probabilidades calibradas ordenaram os grupos numa escala contínua, e quem alternou entre duas categorias ficou entre elas. Uma versão compacta de 8,4 MB, que roda no navegador, manteve acurácia de 74,2%. A detecção de autodeclaração indígena teve AUC de 0,90 a 0,92, mas encontrou só cerca de metade das pessoas indígenas. Conclui-se que a fotografia informa sobre a declaração, mas até um limite dado pela própria variação das declarações. O resultado é uma distribuição, não um veredito, e não serve para decidir sobre pessoas, como em bancas de heteroidentificação.

Palavras-chave: Cor ou raça; Autodeclaração; Visão computacional; Aprendizado de máquina; Eleições.

ABSTRACT

In Brazil, color or race is self-declared, and the same person may declare it differently over time. This study measured how much a facial photograph reveals about that declaration. It used public data from the Superior Electoral Court: 1,569,137 candidacies from 2016 to 2024 and 113,491 color photographs from 2018 to 2024, split by person into training, validation and test sets. Among the 290,724 people who ran more than once, 27.7% changed their declaration; two thirds of these changes were between white and brown. A vision model (fine-tuned CLIP ViT-B/16), trained only on stable declarations, reached a balanced accuracy of 75.0% (95% CI: 74.0% to 76.0%) across white, brown and black, with 65.7% recall for brown. Errors between white and black stayed below 2%. For 96.8% of the photographs of people who changed their declaration, the predicted class was one the person had already used. Calibrated probabilities placed the groups on a continuous scale, and people who alternated between two categories fell between them. A compact 8.4 MB version that runs in the browser kept 74.2% balanced accuracy. Detection of Indigenous self-declaration reached an AUC of 0.90 to 0.92 but found only about half of Indigenous people. Photographs carry information about the declaration, but only up to a limit set by the variability of the

1 Pesquisador independente. Contato e demonstração: https://www.jpierreribeiro.com.br/meio-termo/.

declarations themselves. The output is a distribution, not a verdict, and it must not be used to make decisions about people, as in racial verification committees.

Keywords: Color or race; Self-identification; Computer vision; Machine learning; Elections.

1 INTRODUÇÃO

A classificação de cor ou raça usada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) tem cinco categorias (branca, preta, parda, amarela e indígena) e se baseia na autodeclaração. No Censo Demográfico de 2022, 45,3% da população se declarou parda, 43,5% branca e 10,2% preta (IBGE, 2023). A categoria parda é a maior do país e também a mais difícil de delimitar: abrange pessoas de aparência muito variada, e a fronteira com as categorias vizinhas muda conforme a região, a escolaridade e a situação em que a pergunta é feita (Petruccelli; Saboia, 2013; Muniz, 2012).

Essa fronteira tem consequências práticas. As políticas de cotas em universidades e concursos federais combinam a autodeclaração com procedimentos de heteroidentificação, em que uma comissão avalia a aparência do candidato (Brasil, 2012; 2014; 2018a). Os critérios e os procedimentos dessas comissões ainda geram dúvidas e disputas (Dias; Tavares Junior, 2018), e a fronteira da categoria parda é a mais difícil de avaliar. Daí surge uma pergunta que pode ser respondida com dados: quanto da cor declarada está visível no rosto?

As eleições brasileiras oferecem uma base rara para essa pergunta. Desde 2014, o Tribunal Superior Eleitoral (TSE) registra a cor ou raça declarada por cada candidato e publica, junto com os dados da candidatura, a fotografia oficial usada na urna. As fotos seguem um padrão parecido (frontais, com fundo claro), cobrem todas as regiões do país e incluem as mesmas pessoas em eleições diferentes, o que permite observar quem mudou a declaração. Campos e Machado (2015) e Janusz (2021) usaram esses registros para estudar a representação política e a mudança de declaração entre eleições; Bueno e Dunning (2017) recorreram a avaliadores que classificaram as fotografias de urna dos candidatos e compararam essa classificação com a autodeclaração registrada no TSE.

Este trabalho usa as fotografias para outra finalidade: treinar e avaliar um modelo de visão computacional que estima, a partir do rosto, como candidatos com aparência parecida declararam a própria cor. O objetivo não é atribuir uma cor a ninguém. A pergunta é quanto a fotografia informa sobre a declaração, onde esse limite aparece e como apresentar o resultado sem transformá-lo em julgamento. Os objetivos específicos foram:

a) medir a estabilidade da declaração de cor ou raça entre eleições;

b) treinar um classificador com declarações estáveis e avaliá-lo também em pessoas que mudaram de declaração;

c) verificar se as probabilidades do modelo podem ser lidas como frequências, por meio de calibração;

d) produzir uma versão compacta que funcione no aparelho do usuário, sem enviar a foto a servidores.

O artigo está organizado em cinco seções além desta. A seção 2 resume a literatura sobre classificação racial no Brasil e em visão computacional. A seção 3 descreve dados, modelos e métricas, e a seção 4 apresenta os resultados. A seção 5 discute o que o modelo mede e os seus limites, e a seção 6 traz as considerações finais.

2 REFERENCIAL TEÓRICO

2.1 Cor, raça e classificação no Brasil

Os estudos sobre classificação racial no Brasil mostram que a cor não é um atributo fixo da pessoa. Telles (2004) descreveu a sociedade brasileira como orientada pela aparência, e não pela ascendência, com fronteiras fluidas entre as categorias intermediárias. Na Pesquisa das Características Étnico-Raciais da População, o IBGE encontrou dezenas de termos espontâneos para a própria cor, que só em parte coincidem com as cinco categorias oficiais (Petruccelli; Saboia, 2013). Muniz (2012) comparou diferentes formas de medir a cor na mesma amostra nacional e estimou que só cerca de metade dos entrevistados se classificaria de modo consistente em todas elas.

A escolha da medida muda as conclusões. Bailey, Loveman e Muniz (2013) mostraram que as estimativas de desigualdade racial no Brasil variam conforme se use a autodeclaração, a classificação feita pelo entrevistador ou a cor da pele. Para este trabalho, isso significa que um modelo treinado com autodeclarações aprende a relação entre rosto e declaração, e não a cor "verdadeira" da pessoa nem o modo como outras pessoas a classificariam.

2.2 Classificação racial em visão computacional

Em visão computacional, a classificação de "raça" a partir do rosto aparece em conjuntos de dados como o UTKFace (Zhang; Song; Qi, 2017) e o FairFace (Kärkkäinen; Joo, 2021). Nesses conjuntos, o rótulo é atribuído por anotadores, com categorias definidas pelos autores do conjunto. Khan e Fu (2021) mostraram que as mesmas categorias são aplicadas de

modo inconsistente entre conjuntos de dados diferentes e que os classificadores herdam essa inconsistência. Buolamwini e Gebru (2018) documentaram que sistemas comerciais de análise facial erram mais em pessoas de pele mais escura, o que tornou a composição dos dados de treino um tema central na área.

Há também uma crítica de fundo. Stark e Hutson (2022) chamam de "IA fisiognômica" os sistemas que inferem características sociais a partir do rosto e apontam que eles tendem a naturalizar categorias construídas socialmente. Essa crítica orientou duas escolhas deste trabalho: o modelo devolve uma distribuição de probabilidade ligada a uma população de referência explícita (candidatos às eleições), e não uma categoria; e o resultado é descrito como estimativa sobre declarações de terceiros, não como afirmação sobre a pessoa fotografada.

Do ponto de vista técnico, três ideias sustentam o método. Modelos pré-treinados com pares de imagem e texto, como o CLIP (Radford *et al.*, 2021), servem de ponto de partida para tarefas com poucos dados rotulados. Redes neurais modernas tendem a produzir probabilidades mal calibradas, o que se corrige com uma temperatura ajustada em dados separados (Guo *et al.*, 2017). Um modelo grande pode transferir o que aprendeu para um modelo pequeno por destilação, em que o pequeno aprende a reproduzir as probabilidades do grande (Hinton; Vinyals; Dean, 2015).

3 MATERIAL E MÉTODOS

3.1 Dados

Os rótulos vieram dos arquivos de candidaturas do Repositório de Dados Eleitorais do TSE para as eleições de 2016, 2018, 2020, 2022 e 2024 (Tribunal Superior Eleitoral, 2026). Cada linha corresponde a uma candidatura, com a cor ou raça declarada naquela eleição. Como o CPF não é público, a mesma pessoa foi ligada entre eleições pelo nome normalizado (sem acentos, em maiúsculas) junto com a data de nascimento. Esse critério funde homônimos nascidos no mesmo dia e separa quem mudou a grafia do nome. Nenhum dos dois casos foi quantificado. Para cada pessoa, registraram-se todas as declarações feitas nos cinco anos.

As fotografias vieram dos arquivos de fotos de candidatos do TSE, organizados por estado e ano. As de 2016 foram publicadas em escala de cinza e ficaram fora dos modelos, mas os rótulos daquele ano continuaram no histórico de declarações. Para equilibrar as classes sem baixar todo o acervo, sortearam-se pessoas com taxas por cor declarada (branca 6%,

parda 8%, preta 25%, indígena e amarela 100%), com taxa mínima de 30% para quem mudou de declaração.

Em cada foto, o rosto foi detectado com o YuNet (Wu; Peng; Yu, 2023) e aceito quando o escore de detecção era de pelo menos 0,80 e a distância entre os olhos, de pelo menos 12 pixels. Fotos com um segundo rosto maior que 30% do principal foram descartadas. O rosto aceito foi rotacionado para nivelar os olhos e recortado em 160 × 160 pixels, com os olhos a 40% da altura e a distância entre eles igual a 28% do lado. Foram considerados coloridos os recortes com saturação média de pelo menos 12 (escala de 0 a 255) no centro do rosto.

Treino, validação e teste foram separados por pessoa, na proporção de 80%, 10% e 10%, a partir de um *hash* do identificador. A mesma pessoa nunca aparece em dois conjuntos, mesmo que tenha concorrido em vários anos. A categoria amarela, com 3.625 fotos, ficou fora dos modelos. O conjunto final teve 113.491 fotos coloridas com declarações e históricos dentro de branca, parda, preta e indígena.

3.2 Rótulos estáveis e conjunto de pessoas que mudaram de declaração

Os modelos foram treinados apenas com pessoas cuja declaração foi a mesma em todas as eleições observadas. Pessoas que mudaram não foram descartadas: formaram um conjunto de avaliação à parte, em que se verificou se a classe prevista pelo modelo estava entre as declarações que a pessoa já tinha usado. Esse conjunto não tem uma resposta certa única, mas permite verificar se o modelo coloca essas pessoas na região em que a declaração de fato se divide.

3.3 Modelos

O ponto de partida foi o codificador de imagens do CLIP ViT-B/16 (Radford *et al.*, 2021), na implementação da biblioteca timm (Wightman, 2019). Como linha de base, treinou-se uma regressão logística sobre as representações do codificador congelado (sonda linear). O modelo principal, aqui chamado professor, foi o mesmo codificador ajustado por inteiro para três classes (branca, parda e preta) durante seis épocas. O ajuste usou AdamW, com taxa de aprendizado de 10^{-5} no codificador e 10^{-3} na camada de saída, e perda de entropia cruzada com suavização de rótulos de 0,1 e peso por classe proporcional à raiz da frequência inversa. O aumento de dados se limitou a recortes aleatórios entre 60% e 100% da área e ao espelhamento horizontal. Não houve alteração de cor ou brilho, porque a cor da pele é parte

do sinal que se quer medir. Manteve-se o modelo com a maior acurácia balanceada na validação.

Um segundo professor, com o mesmo procedimento, separou a declaração indígena das demais (indígena × outra) durante quatro épocas, com uma foto de pessoa indígena para cada 27 das demais no treino.

As probabilidades de cada professor foram calibradas na validação com uma temperatura T e um viés por classe b, na forma $p = \mathrm{softmax}(z/T + b)$, em que z são as saídas do modelo antes da normalização. O viés corrige a preferência por classe introduzida pelos pesos da perda. As probabilidades de branca, parda e preta valem para as três classes em proporções iguais, que é como elas aparecem no treino. Para comparação, calculou-se também a versão reponderada pelas proporções do Censo de 2022, multiplicando cada probabilidade pela proporção da classe e renormalizando.

Para uso em celulares, os dois professores foram destilados num único modelo pequeno, o aluno: uma MobileNetV3-Large (Howard *et al.*, 2019) com entrada de 160 pixels e cinco saídas, três para branca, parda e preta e duas para outra e indígena. O aluno aprendeu a reproduzir as probabilidades calibradas dos professores, com perda de entropia cruzada sobre alvos suaves, durante 12 épocas, em 90.957 fotos de treino. As fotos de pessoas indígenas foram sobreamostradas quatro vezes. O modelo foi exportado no formato ONNX, com os pesos armazenados em meia precisão (16 bits), num arquivo de 8,4 MB.

3.4 Métricas e testes adicionais

A métrica principal foi a acurácia balanceada (média da revocação das classes), calculada no conjunto de teste com pessoas de declaração estável. Também foram relatadas a revocação e a precisão por classe, a matriz de confusão, o logaritmo negativo da verossimilhança (NLL) e o erro esperado de calibração (ECE), com 15 faixas de probabilidade. Para a classificação indígena, relataram-se a área sob a curva ROC (AUC) e a revocação e a precisão com o limiar de 50%. Para resumir as três probabilidades num número, definiu-se uma escala de 0 a 2, dada por P(parda) + 2 × P(preta), em que 0 corresponde a branca, 1 a parda e 2 a preta.

Fizeram-se três testes adicionais com o aluno. O primeiro aplicou o modelo a fotos de pessoas brancas e negras do UTKFace (Zhang; Song; Qi, 2017), recortadas pelo mesmo procedimento, para verificar o funcionamento fora do padrão de foto do TSE. O segundo simulou mudanças de iluminação em 1.500 fotos de teste (500 por classe), multiplicando os canais de cor no espaço linear: metade, um quarto e o dobro da luz, luz amarelada, luz azulada

e a combinação de metade da luz com luz amarelada. O terceiro mediu quanto a resposta muda quando a foto é espelhada.

3.5 Aspectos éticos

Os dados são públicos e foram obtidos no repositório oficial do TSE. Fotografias de rosto associadas à cor ou raça são, ainda assim, dados pessoais sensíveis segundo a Lei Geral de Proteção de Dados Pessoais (Brasil, 2018b). Por isso, o conjunto de recortes não foi redistribuído, nenhum rosto é exibido neste artigo e todos os resultados são agregados. Testes informais feitos durante o desenvolvimento com fotografias de outras origens, coletadas sem consentimento, foram excluídos de todos os números aqui apresentados. [VERIFICAR com o Comitê de Ética em Pesquisa da instituição se cabe a dispensa prevista na Resolução CNS nº 510/2016, art. 1º, parágrafo único, ou se é preciso submeter o projeto.]

4 RESULTADOS

4.1 A declaração muda entre eleições

A base reúne 1.569.137 candidaturas de 1.178.425 pessoas (Tabela 1). Entre 2016 e 2024, a parcela de candidaturas declaradas brancas caiu de 51,5% para 47,1%, e a de pretas subiu de 8,6% para 11,4%.

Tabela 1 – Candidaturas por eleição e cor ou raça declarada

Eleição	Candidaturas	Branca (%)	Parda (%)	Preta (%)	Indígena (%)	Amarela (%)
2016	498.054	51,5	39,1	8,6	0,3	0,4
2018	29.119	52,4	35,7	10,9	0,5	0,6
2020	551.921	48,6	40,0	10,6	0,4	0,4
2022	29.121	48,4	36,3	14,2	0,7	0,4
2024	460.922	47,1	40,5	11,4	0,6	0,4
Total	1.569.137					

Fonte: elaborado pelo autor (2026).
Nota: dados do Repositório de Dados Eleitorais do TSE. Percentuais sobre as candidaturas de cada eleição.

Das 290.724 pessoas que concorreram em pelo menos duas eleições, 80.584 (27,7%) declararam cores diferentes em algum momento. A mudança se concentrou entre categorias vizinhas: 67,3% dos casos alternaram entre branca e parda e 25,8% entre parda e preta. A alternância entre branca e preta respondeu por 2,6% dos casos. A Tabela 2 mostra as fotos coloridas recortadas, separadas pela estabilidade da declaração.

Tabela 2 – Fotografias coloridas por cor declarada e estabilidade da declaração (2018–2024)

Cor/raça declarada na foto	Declaração sempre igual	Mudou entre eleições	Total
Branca	25.372	12.835	38.207
Parda	25.885	17.471	43.356
Preta	22.097	5.762	27.859
Indígena	3.998	658	4.656
Amarela*	2.339	1.286	3.625
Total	79.691	38.012	117.703

Fonte: elaborado pelo autor (2026).
Nota: as taxas de sorteio por cor tornam essas contagens diferentes da composição das candidaturas. * A categoria amarela ficou fora dos modelos.

4.2 Desempenho entre branca, parda e preta

A sonda linear sobre o CLIP congelado atingiu acurácia balanceada de 73,6%, e o professor ajustado, 75,0% depois da calibração (Tabela 3), com intervalo de confiança de 95% de 74,0% a 76,0%. Os intervalos foram obtidos por *bootstrap* com reamostragem de pessoas, e não de fotos, em 1.000 repetições. A diferença pequena entre os dois e a estabilização da validação a partir da terceira época de ajuste indicam que o limite está mais no rótulo do que no modelo. A classe parda teve a menor revocação (65,7%, IC 95% de 63,8% a 67,6%; Tabela 4). Os erros se concentraram entre classes vizinhas: das 2.543 fotos de pessoas sempre brancas, 17 foram previstas como pretas, e das 2.181 sempre pretas, 39 foram previstas como brancas.

Tabela 3 – Desempenho dos modelos entre branca, parda e preta no conjunto de teste (pessoas de declaração estável, 7.339 fotos)

Modelo	Acurácia balanceada (%)	Revocação parda (%)	Precisão parda (%)	ECE (%)	NLL
Sonda linear (CLIP ViT-B/16 congelado)[1]	73,6	62,1	64,1	–	–
Professor ViT-B/16 ajustado, sem calibrar	74,8	62,3	66,0	2,6	0,601
Professor ViT-B/16 ajustado, calibrado	75,0	65,7	65,6	2,6	0,591
Aluno MobileNetV3 (site), com espelhamento	74,2	65,8	64,2	1,7	0,596

Fonte: elaborado pelo autor (2026).
Nota: [1] execução em GPU registrada no diário de treino e não refeita na verificação local; o conjunto de teste pode diferir um pouco do usado nas demais linhas. ECE = erro esperado de calibração; NLL = logaritmo negativo da verossimilhança.

Tabela 4 – Matriz de confusão do professor calibrado no conjunto de teste (pessoas de declaração estável)

Declarada \ Prevista	Branca	Parda	Preta	Total	Revocação (%)
Branca	2.034	492	17	2.543	80,0
Parda	485	1.718	412	2.615	65,7
Preta	39	410	1.732	2.181	79,4
Precisão (%)	79,5	65,6	80,2		

Fonte: elaborado pelo autor (2026).
Nota: linhas = cor declarada; colunas = classe de maior probabilidade.

No conjunto de teste de pessoas que mudaram de declaração (3.613 fotos), a classe prevista estava entre as declarações já usadas pela pessoa em 96,8% dos casos. Entre pardos de declaração estável, a mediana de P(parda) foi de 59,2%, e só 31,6% das fotos passaram de 70%. O valor máximo foi 92,8%. Para uma pessoa parda típica, então, o modelo distribui uma parte considerável da probabilidade entre as categorias vizinhas.

O desempenho variou entre regiões e entre homens e mulheres (Tabela 5). O Sudeste teve a maior acurácia balanceada (75,0%) e o Nordeste, a menor (67,1%). A diferença está sobretudo na classe branca: a revocação de branca foi de 92,0% no Sul e de 56,3% no Nordeste. Isso combina com a ideia de que a fronteira entre branca e parda muda conforme a região (Telles, 2004), e o modelo, treinado com o país inteiro, aplica uma fronteira média. Entre mulheres, a revocação de parda foi de 57,6%, contra 70,1% entre homens. Como os grupos regionais têm composições de cor muito diferentes, essas comparações descrevem o modelo nesses dados e não devem ser lidas como diferenças entre as populações.

Tabela 5 – Desempenho do professor calibrado por região e gênero no conjunto de teste (pessoas de declaração estável)

Grupo	Fotos	Acurácia balanceada (%) [IC 95%]	Revocação branca (%)	Revocação parda (%)	Revocação preta (%)
Norte	715	70,5 [66,0–74,9]	67,8	73,2	70,4
Nordeste	2.011	67,1 [64,5–69,5]	56,3	68,4	76,7
Centro-Oeste	628	73,5 [70,1–76,9]	79,5	63,9	77,1
Sudeste	2.813	75,0 [73,4–76,7]	78,5	59,7	87,0
Sul	1.172	71,3 [67,2–74,6]	92,0	56,2	65,6
Mulheres	2.681	74,2 [72,5–75,8]	83,6	57,6	81,3
Homens	4.658	75,4 [74,1–76,8]	78,0	70,1	78,2

Fonte: elaborado pelo autor (2026).
Nota: região da UF da candidatura; gênero conforme registro do TSE. IC 95% por *bootstrap* de pessoas (1.000 repetições).

4.3 Uma escala, não três caixas

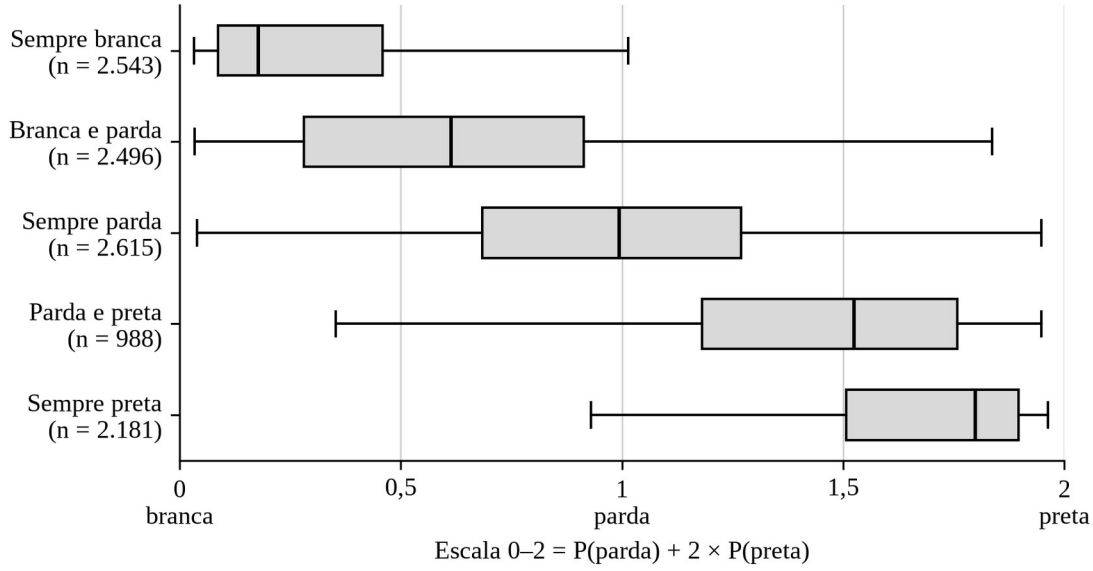
A escala de 0 a 2 ordenou os grupos na sequência esperada (Tabela 6 e Figura 1). Pessoas sempre brancas ficaram em 0,32, sempre pardas em 0,97 e sempre pretas em 1,66. Quem alternou entre branca e parda ficou em 0,63, entre os dois grupos estáveis, e quem alternou entre parda e preta, em 1,45. As distribuições se sobrepõem bastante: a caixa dos pardos estáveis cobre boa parte das caixas dos dois grupos que mudaram.

Tabela 6 – Probabilidades médias e escala de 0 a 2 por histórico de declaração (conjunto de teste)

Declarações da pessoa (2016–2024)	Fotos	P(branca)	P(parda)	P(preta)	Escala 0–2: média (DP)
Sempre branca	2.543	71,9	24,0	4,1	0,32 (0,32)
Branca e parda	2.496	45,3	46,6	8,1	0,63 (0,38)
Sempre parda	2.615	24,1	54,3	21,6	0,97 (0,44)
Parda e preta	988	6,7	41,7	51,6	1,45 (0,35)
Sempre preta	2.181	5,3	24,0	70,7	1,66 (0,34)
Branca e preta	85	20,0	42,1	37,9	1,18 (0,52)
Branca, parda e preta	44	18,2	44,3	37,5	1,19 (0,50)

Fonte: elaborado pelo autor (2026).
Nota: P = probabilidade calibrada do professor; DP = desvio padrão.

Figura 1 – Distribuição da escala de 0 a 2 por histórico de declaração

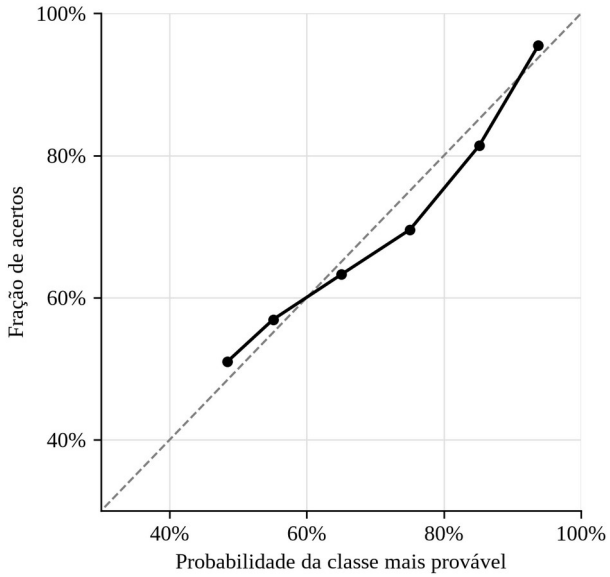


Fonte: elaborado pelo autor (2026).
Nota: caixas = intervalo interquartil; traço interno = mediana; hastes = 1,5 vez o intervalo interquartil. Conjunto de teste.

4.4 Calibração

O professor já saiu do ajuste com probabilidades próximas das frequências observadas (ECE de 2,6%). A calibração pouco mudou o ECE, mas reduziu o NLL de 0,601 para 0,591 e corrigiu a preferência por classe: a revocação de parda passou de 62,3% para 65,7% (Tabela 3). A Figura 2 mostra que, quando o professor dá 60% à classe mais provável, ela está certa em cerca de 60% das fotos, e o mesmo vale nas outras faixas.

Figura 2 – Diagrama de confiabilidade do professor calibrado no conjunto de teste



Fonte: elaborado pelo autor (2026).
Nota: cada ponto reúne as fotos numa faixa de 10 pontos percentuais da maior probabilidade; a linha tracejada indica calibração perfeita.

A escolha da população de referência muda a leitura. Com as proporções do Censo de 2022, em que pretos são cerca de 10% da população, as fotos de pessoas sempre pretas receberam em média 47,5% de probabilidade de preta e 42,9% de parda. Nas proporções do treino, esses valores eram 70,7% e 24,0%. A mesma foto, portanto, tem respostas diferentes conforme a pergunta: entre candidatos em proporções iguais ou entre pessoas na proporção da população.

4.5 Autodeclaração indígena

O conjunto de teste da classificação indígena teve 11.376 fotos, das quais 424 (3,7%) eram de pessoas sempre declaradas indígenas. O professor atingiu AUC de 0,919 (IC 95% de 0,902 a 0,936; Tabela 7). Sem calibração, as probabilidades de indígena estavam muito acima das frequências reais (ECE de 24,6%), efeito da sobrerrepresentação no treino. Depois da calibração, o ECE caiu para 0,5%. Com o limiar de 50%, o professor encontrou 53,8% das pessoas indígenas (IC 95% de 48,7% a 59,0%), e 83,8% das fotos que apontou eram de fato de pessoas indígenas. A probabilidade média foi de 53,2% para pessoas sempre indígenas, de 18,9% para quem alternou entre indígena e outra categoria e de 1,2% a 2,3% para pessoas sempre brancas, pardas ou pretas.

Tabela 7 – Classificação indígena × outra no conjunto de teste

Modelo	AUC	Revocação em P ≥ 50% (%)	Precisão em P ≥ 50% (%)	ECE de P(indígena) (%)
Professor, sem calibrar	0,919	–	–	24,6
Professor, calibrado	0,919	53,8	83,8	0,5
Aluno (site), com espelhamento	0,899	50,2	87,7	0,4

Fonte: elaborado pelo autor (2026).
Nota: P ≥ 50% = foto classificada como indígena quando a probabilidade calibrada é de pelo menos 50%.

Testes informais durante o desenvolvimento sugeriram que o modelo se apoia em pinturas corporais e adornos, como cocares, presentes em parte das fotos oficiais. Essa hipótese não foi quantificada com dados públicos [PENDENTE: medir a revocação com e sem adorno nas fotos de teste do TSE]. Ela seria coerente com a revocação baixa, já que a autodeclaração indígena não corresponde a um traço visual único.

4.6 Modelo compacto e testes adicionais

O aluno, com média sobre a foto original e a espelhada, atingiu acurácia balanceada de 74,2%, 0,8 ponto abaixo do professor, com ECE de 1,7% (Tabela 3). A classe de maior probabilidade coincidiu com a do professor em 90,5% das fotos de teste, e a diferença média na escala de 0 a 2 foi de 0,09. Na classificação indígena, o aluno teve AUC de 0,899, revocação de 50,2% e precisão de 87,7% (Tabela 7).

O espelhamento da foto mudou a maior probabilidade em 2,7 pontos percentuais na mediana e em 9,6 pontos no percentil 90. Por isso, a aplicação de demonstração calcula a média de seis variações de cada foto: três escalas, cada uma com e sem espelhamento.

No UTKFace, a classe mais provável foi branca para 89,8% das 1.399 fotos de pessoas brancas e preta para 89,4% das 1.412 fotos de pessoas negras. As probabilidades médias foram 77,6% e 77,1%, e a probabilidade média de indígena ficou em 1,3% e 2,1%. Os rótulos do UTKFace usam categorias norte-americanas atribuídas por anotadores, e a correspondência entre "negro" e "preta" é aproximada.

Na simulação de iluminação (Tabela 8), a acurácia balanceada ficou entre 70,3% e 72,5%, contra 72,3% nas fotos originais. A mudança média na escala de 0 a 2 foi de 0,04 a 0,08. Mudanças uniformes de luz e de cor, portanto, alteram pouco o resultado. A simulação não reproduz sombras localizadas, contraluz, ruído de câmeras em pouca luz nem filtros de aplicativos.

Tabela 8 – Efeito de mudanças simuladas de iluminação no aluno (1.500 fotos de teste, uma passagem)

Condição simulada	Acurácia balanceada (%)	P(branca) média entre brancos (%)	Mudança média na escala 0–2
Original	72,3	69,2	0,00
Metade da luz (−1 stop)	71,6	69,3	0,05
Um quarto da luz (−2 stops)	70,3	70,4	0,08
Dobro da luz (+1 stop)	71,7	68,4	0,08
Luz quente (amarelada)	71,9	66,6	0,05
Metade da luz e quente	72,5	68,6	0,05
Luz fria (azulada)	71,9	68,0	0,04

Fonte: elaborado pelo autor (2026).
Nota: ganhos aplicados no espaço linear do sRGB; luz quente = (1,15; 1; 0,7) e luz fria = (0,85; 1; 1,25) nos canais R, G e B.

5 DISCUSSÃO

O resultado principal é um limite. Com cerca de 113 mil fotos e um modelo pré-treinado em larga escala, a acurácia balanceada parou perto de 75%, e o ajuste fino ganhou pouco sobre a sonda linear. Esse limite combina com o que a base mostra sobre as próprias declarações: mais de um quarto de quem concorreu mais de uma vez mudou a cor declarada, quase sempre para uma categoria vizinha. Se a mesma pessoa, com o mesmo rosto, declara cores diferentes em eleições diferentes, nenhum classificador acerta todas as declarações a partir da foto. Esse comportamento está de acordo com a literatura sobre a fluidez das categorias intermediárias (Telles, 2004; Muniz, 2012) e mostra que a incerteza não vem só do modelo.

Por isso, o resultado faz mais sentido como distribuição do que como rótulo. A escala de 0 a 2 coloca quem alternou entre duas categorias entre os grupos estáveis correspondentes, e as probabilidades calibradas podem ser lidas como frequências: entre candidatos com rosto parecido, tal proporção se declarou parda. Na aplicação de demonstração, essa frase substitui a pergunta "qual é a minha cor?", para a qual a fotografia não tem resposta.

A comparação com as proporções do Censo mostra que nenhuma distribuição é neutra. A mesma foto de uma pessoa sempre preta recebe 70,7% de preta quando as três classes têm o mesmo peso e 47,5% quando se usa a composição da população brasileira. As duas respostas são corretas para perguntas diferentes. Qualquer uso desses números precisa declarar a população de referência.

O modelo também não mede como as outras pessoas veem quem está na foto. O rótulo é autodeclaração, e o que o modelo aprende é a associação entre aparência e declaração entre candidatos. A heteroidentificação faz outra pergunta, sobre o julgamento de terceiros a partir do fenótipo (Brasil, 2018a). Medir esse julgamento exigiria várias avaliações independentes por foto, como fizeram Bueno e Dunning (2017), o que não foi feito aqui. Mesmo que fosse,

uma acurácia de 75%, com a classe parda em 66%, seria inadequada para decidir sobre pessoas. O modelo não deve ser usado em seleção, verificação de cotas ou qualquer decisão individual.

A classificação indígena é o caso mais delicado. A identidade indígena se define por autodeclaração e reconhecimento da comunidade, e não por aparência. O modelo acerta a maior parte das fotos que aponta como indígenas, mas deixa de reconhecer cerca de metade das pessoas indígenas. Um valor baixo, portanto, não indica que a pessoa não seja indígena.

O trabalho tem limitações. A população é de candidatos, que não representam a população brasileira em idade, escolaridade e região. As fotos oficiais são frontais e bem iluminadas, o que favorece o desempenho no teste e não informa sobre fotos de celular com sombras, contraluz ou filtros. A ligação entre eleições depende de nome e data de nascimento, com erros não medidos. As taxas de sorteio por cor mudam a composição da amostra em relação à base, e as proporções em que as probabilidades são válidas precisam ser declaradas. A categoria amarela ficou fora dos modelos. Por fim, só se usaram fotos coloridas de 2018 a 2024, e a série histórica se limita a cinco eleições.

6 CONSIDERAÇÕES FINAIS

Este estudo usou fotografias e declarações públicas de candidatos às eleições brasileiras para medir quanto o rosto informa sobre a cor ou raça declarada. A fotografia informa bastante sobre os extremos, com menos de 2% de confusão entre branca e preta, e pouco sobre a fronteira da categoria parda, em que o próprio rótulo é instável. O limite de cerca de 75% de acurácia balanceada acompanha a mudança de declaração observada em 27,7% de quem concorreu mais de uma vez.

O produto do trabalho é uma estimativa calibrada de como pessoas com rosto parecido se declararam, com a população de referência explícita, e uma versão compacta que funciona no aparelho do usuário sem enviar a foto. Essa forma de apresentar o resultado, como distribuição, procura evitar que um modelo estatístico seja lido como juízo sobre a identidade de alguém.

Para trabalhos futuros, sugere-se coletar várias classificações independentes por foto, para separar o que é autodeclaração do que é percepção de terceiros, e montar um conjunto de teste com fotos de celular, obtidas com consentimento, para medir o efeito de iluminação e câmera. O código das análises e do site está disponível em https://github.com/jpierreribeiro/meio-termo, e a demonstração, em

https://www.jpierreribeiro.com.br/meio-termo/. Os recortes de rosto e as tabelas de rótulos não são redistribuídos.

REFERÊNCIAS

BAILEY, Stanley R.; LOVEMAN, Mara; MUNIZ, Jeronimo O. Measures of "race" and the analysis of racial inequality in Brazil. **Social Science Research**, [*s. l.*], v. 42, n. 1, p. 106-119, 2013.

BRASIL. Lei nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, DF, 30 ago. 2012.

BRASIL. Lei nº 12.990, de 9 de junho de 2014. Reserva aos negros 20% (vinte por cento) das vagas oferecidas nos concursos públicos para provimento de cargos efetivos e empregos públicos no âmbito da administração pública federal. **Diário Oficial da União**: seção 1, Brasília, DF, 10 jun. 2014.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Portaria Normativa nº 4, de 6 de abril de 2018. Regulamenta o procedimento de heteroidentificação complementar à autodeclaração dos candidatos negros. **Diário Oficial da União**: seção 1, Brasília, DF, 10 abr. 2018a. p. 34.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, DF, 15 ago. 2018b.

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. **Diário Oficial da União**: seção 1, Brasília, DF, 24 maio 2016.

BUENO, Natália S.; DUNNING, Thad. Race, resources, and representation: evidence from Brazilian politicians. **World Politics**, [*s. l.*], v. 69, n. 2, p. 327-365, 2017.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. *In*: CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY, 1., 2018, New York. **Proceedings** [...]. [*S. l.*]: PMLR, 2018. p. 77-91. (Proceedings of Machine Learning Research, v. 81).

CAMPOS, Luiz Augusto; MACHADO, Carlos. A cor dos eleitos: determinantes da sub-representação política dos não brancos no Brasil. **Revista Brasileira de Ciência Política**, Brasília, n. 16, p. 121-151, jan./abr. 2015. DOI: 10.1590/0103-335220151606.

DIAS, Gleidson Renato Martins; TAVARES JUNIOR, Paulo Roberto Faber (org.). **Heteroidentificação e cotas raciais**: dúvidas, metodologias e procedimentos. Canoas: IFRS, 2018.

GUO, Chuan; PLEISS, Geoff; SUN, Yu; WEINBERGER, Kilian Q. On calibration of modern neural networks. *In*: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 34., 2017, Sydney. **Proceedings** [...]. [*S. l.*]: PMLR, 2017. p. 1321-1330. (Proceedings of Machine Learning Research, v. 70).

HINTON, Geoffrey; VINYALS, Oriol; DEAN, Jeff. Distilling the knowledge in a neural network. **arXiv**, [*s. l.*], 2015. arXiv:1503.02531.

HOWARD, Andrew *et al*. Searching for MobileNetV3. *In*: IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2019, Seoul. **Proceedings** [...]. Piscataway: IEEE, 2019. p. 1314-1324.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2022**: identificação étnico-racial da população, por sexo e idade: resultados do universo. Rio de Janeiro: IBGE, 2023.

JANUSZ, Andrew. Electoral incentives and elite racial identification: why Brazilian politicians change their race. **Electoral Studies**, [*s. l.*], v. 72, 102340, 2021. DOI: 10.1016/j.electstud.2021.102340.

KÄRKKÄINEN, Kimmo; JOO, Jungseock. FairFace: face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. *In*: IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION, 2021, Waikoloa. **Proceedings** [...]. Piscataway: IEEE, 2021. p. 1548-1558.

KHAN, Zaid; FU, Yun. One label, one billion faces: usage and consistency of racial categories in computer vision. *In*: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021, [*s. l.*]. **Proceedings** [...]. New York: ACM, 2021. p. 587-597.

MUNIZ, Jerônimo Oliveira. Preto no branco? Mensuração, relevância e concordância classificatória no país da incerteza racial. **Dados**: Revista de Ciências Sociais, Rio de Janeiro, v. 55, n. 1, p. 251-282, 2012. DOI: 10.1590/S0011-52582012000100007.

PETRUCCELLI, José Luis; SABOIA, Ana Lucia (org.). **Características étnico-raciais da população**: classificações e identidades. Rio de Janeiro: IBGE, 2013. (Estudos e Análises: informação demográfica e socioeconômica, n. 2).

RADFORD, Alec *et al*. Learning transferable visual models from natural language supervision. *In*: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 38., 2021, [*s. l.*]. **Proceedings** [...]. [*S. l.*]: PMLR, 2021. p. 8748-8763. (Proceedings of Machine Learning Research, v. 139).

STARK, Luke; HUTSON, Jevan. Physiognomic artificial intelligence. **Fordham Intellectual Property, Media & Entertainment Law Journal**, New York, v. 32, n. 4, p. 922-[VERIFICAR página final no PDF], 2022.

TELLES, Edward. **Race in another America**: the significance of skin color in Brazil. Princeton: Princeton University Press, 2004.

TRIBUNAL SUPERIOR ELEITORAL. **Repositório de Dados Eleitorais**. Brasília, DF: TSE, 2026. Disponível em: https://dadosabertos.tse.jus.br/. Acesso em: 14 nov. 2026.

WIGHTMAN, Ross. **PyTorch Image Models**. [*S. l.*]: GitHub, 2019. Disponível em: https://github.com/huggingface/pytorch-image-models. Acesso em: 15 nov. 2026.

WU, Wei; PENG, Hanyang; YU, Shiqi. YuNet: a tiny millisecond-level face detector. **Machine Intelligence Research**, [*s. l.*], v. 20, n. 5, p. 656-665, 2023.

ZHANG, Zhifei; SONG, Yang; QI, Hairong. Age progression/regression by conditional adversarial autoencoder. *In*: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2017, Honolulu. **Proceedings** [...]. Piscataway: IEEE, 2017. p. 5810-5818.