BOCA
BOLETIM DE CONJUNTURA

VERIFICAÇÃO FACIAL COM ARCFACE E FACENET EM FOTOGRAFIAS COM UMA E MÚLTIPLAS FACES: ESTUDO EXPLORATÓRIO COM AUDITORIA DA FACE SELECIONADA

FACIAL VERIFICATION USING ARCFACE AND FACENET ON PHOTOGRAPHS CONTAINING SINGLE OR MULTIPLE FACES: AN EXPLORATORY STUDY WITH AN AUDIT OF THE SELECTED FACE

VERIFICACIÓN FACIAL CON ARCFACE Y FACENET EN FOTOGRAFÍAS CON UNO O MÚLTIPLES ROSTROS: ESTUDIO EXPLORATORIO CON AUDITORÍA DEL ROSTRO SELECCIONADO

https://doi.org/10.56238/bocav25n84-045

Data de submissão: 25/10/2026 **Data de publicação:** 25/11/2026

Arthur de Morais Sarmanho
Graduando em Sistemas de Informação
Instituição: Centro Universitário Santa Terezinha (CEST)
E-mail: sarmanhoarthur7@gmail.com

Byanca Joaquiny Mendes Sousa
Graduanda em Sistemas de Informação
Instituição: Centro Universitário Santa Terezinha (CEST)
E-mail: byancamendessousa93@gmail.com

Dadilton Bastos Melo
Orientador

RESUMO

Este estudo analisa ArcFace e FaceNet em um protocolo pareado de verificação facial com fotografias de uma ou múltiplas faces. O conjunto experimental reúne 40 imagens públicas de quatro identidades, com duas referências e oito testes por identidade, totalizando 512 decisões que reutilizam as mesmas imagens e pessoas. O RetinaFace foi usado como detector comum, e os embeddings foram comparados por distância cosseno. Com os limiares padrão do DeepFace 0.0.101 — 0,68 para ArcFace e 0,40 para FaceNet —, o ArcFace apresentou acurácia de 92,58%, FMR de 3,65% e FNMR de 18,75%; no FaceNet, os valores foram 89,84%, 1,04% e 37,50%. A auditoria visual realizada pelos autores cobriu todas as 128 comparações genuínas e os nove falsos positivos. Entre as genuínas, a face selecionada foi classificada como distratora em 4 de 64 casos do ArcFace (6,25%) e em 2 de 64 do FaceNet (3,13%). Nos nove falsos positivos, a auditoria não registrou nova seleção classificada como distratora segundo o critério adotado. Duas seleções distratoras do ArcFace foram aceitas e apareceram como verdadeiros positivos na rotulagem por fotografia. A análise de sensibilidade mostrou ainda que a ordenação dos modelos depende do limiar: com FaceNet em 0,44, a acurácia chegou a 94,53%, resultado exploratório que não foi adotado como novo ponto operacional por ter sido obtido no mesmo conjunto de teste.

Palavras-chave: Reconhecimento Facial. Verificação Biométrica. ArcFace. FaceNet. Avaliação de Desempenho.

ABSTRACT
This study analyzes ArcFace and FaceNet within a paired facial verification protocol using photographs containing one or multiple faces. The experimental dataset comprises 40 public images representing four identities, with two reference images and eight test images per identity, totaling 512 decisions that reuse the same images and individuals. RetinaFace was employed as a common detector, and embeddings were compared using cosine distance. Applying the default thresholds from DeepFace 0.0.101—0.68 for ArcFace and 0.40 for FaceNet—ArcFace yielded an accuracy of 92.58%, an FMR of 3.65%, and an FNMR of 18.75%; for FaceNet, these values were 89.84%, 1.04%, and 37.50%, respectively. A visual audit conducted by the authors covered all 128 genuine comparisons and the nine false positives. Among the genuine comparisons, the selected face was classified as a "distractor" in 4 out of 64 cases for ArcFace (6.25%) and in 2 out of 64 cases for FaceNet (3.13%). Regarding the nine false positives, the audit did not identify any additional selections classified as distractors according to the established criteria. Two "distractor" selections in the ArcFace results were accepted and appeared as true positives in the photograph-based labeling. Sensitivity analysis further demonstrated that the relative ranking of the models depends on the threshold; with FaceNet set at 0.44, accuracy reached 94.53%—an exploratory result that was not adopted as a new operational point, as it was derived from the same test dataset.

Keywords: Facial Recognition. Biometric Verification. ArcFace. FaceNet. Performance Evaluation.

RESUMEN
Este estudio analiza ArcFace y FaceNet en un protocolo emparejado de verificación facial con fotografías de uno o varios rostros. El conjunto experimental reúne 40 imágenes públicas de cuatro identidades, con dos referencias y ocho pruebas por identidad, lo que suma un total de 512 decisiones que reutilizan las mismas imágenes y personas. Se utilizó RetinaFace como detector común, y los embeddings se compararon mediante la distancia coseno. Con los umbrales predeterminados de DeepFace 0.0.101 —0,68 para ArcFace y 0,40 para FaceNet—, ArcFace presentó una precisión del 92,58 %, una FMR del 3,65 % y una FNMR del 18,75 %; en el caso de FaceNet, los valores fueron 89,84 %, 1,04 % y 37,50 %, respectivamente. La auditoría visual realizada por los autores abarcó las 128 comparaciones genuinas y los nueve falsos positivos. Entre las comparaciones genuinas, el rostro seleccionado fue clasificado como «distractor» en 4 de los 64 casos de ArcFace (6,25 %) y en 2 de los 64 casos de FaceNet (3,13 %). En cuanto a los nueve falsos positivos, la auditoría no registró ninguna selección adicional clasificada como «distractora» según el criterio adoptado. Dos selecciones consideradas distractoras en ArcFace fueron aceptadas y aparecieron como verdaderos positivos en el etiquetado por fotografía. El análisis de sensibilidad reveló, además, que la clasificación de los modelos depende del umbral: con FaceNet ajustado a 0,44, la precisión alcanzó el 94,53 %; no obstante, este resultado exploratorio no se adoptó como nuevo punto operativo, dado que se obtuvo utilizando el mismo conjunto de pruebas.

Palabras clave: Reconocimiento Facial. Verificación Biométrica. ArcFace. FaceNet. Evaluación del Rendimiento.

1 INTRODUÇÃO

Reconhecimento facial é uma modalidade biométrica baseada em características faciais para verificar ou identificar indivíduos. Na verificação 1:1, uma amostra é comparada com a identidade declarada; na identificação 1:N, a busca ocorre entre múltiplas identidades cadastradas (JAIN; ROSS; PRABHAKAR, 2004). Este artigo avalia exclusivamente um protocolo pareado de verificação facial.

Abordagens contemporâneas de reconhecimento facial utilizam redes profundas para gerar embeddings, vetores nos quais a distância entre representações pode apoiar a comparação de identidades. O FaceNet aprende esse espaço por meio de Triplet Loss (SCHROFF; KALENICHENKO; PHILBIN, 2015), enquanto o ArcFace emprega margem angular aditiva para aumentar a separação entre classes (DENG et al., 2019).

O desempenho, entretanto, depende também das condições das imagens e do pipeline empregado. Pose, iluminação, resolução, oclusões e qualidade de aquisição podem dificultar a comparação automática (ZHAO et al., 2003). Em fotografias não controladas, outro fator relevante é a presença de várias pessoas: uma única imagem pode fornecer múltiplas regiões candidatas ao reconhecimento.

Em imagens multiface, o problema não se limita à qualidade de cada rosto. Quando o sistema compara uma referência com todas as faces detectadas e seleciona a menor distância, a decisão pode ser fundamentada por uma pessoa diferente da identidade-alvo da fotografia. Por isso, além da decisão binária de aceitação ou rejeição, é necessário observar qual região facial efetivamente produziu a menor distância.

A avaliação de verificação facial também deve distinguir falsas correspondências e falsas rejeições. A False Match Rate (FMR) mede a proporção de comparações impostoras incorretamente aceitas, enquanto a False Non-Match Rate (FNMR) mede a proporção de comparações genuínas incorretamente rejeitadas (JAIN; ROSS; PRABHAKAR, 2004). Essas métricas complementam a acurácia, mas sua interpretação depende do protocolo e do limiar adotado.

Este trabalho compara ArcFace e FaceNet em um protocolo de verificação facial pareada com fotografias de uma e múltiplas faces, utilizando o RetinaFace como detector comum. As imagens públicas formam um conjunto exploratório, selecionado de modo não probabilístico, com condições não controladas e presença de distratores em parte dos testes.

A questão de pesquisa é: como ArcFace e FaceNet se comportam, nos limiares adotados, em verificação facial pareada e de que forma a seleção da menor distância em imagens multiface afeta a

validade das decisões? O objetivo geral é comparar os reconhecedores sob as mesmas condições de detecção e verificar a consistência entre a face selecionada e a identidade-alvo. Especificamente, busca-se mensurar os erros do protocolo, comparar imagens com uma e múltiplas faces, auditar as decisões genuínas e os falsos positivos e quantificar, nesse subconjunto auditado, as seleções de faces diferentes da pessoa-alvo, distinguindo as aceitas das rejeitadas.

A principal contribuição deste trabalho é observar não apenas a decisão final do sistema, mas também qual face realmente produziu a menor distância nas fotografias com várias pessoas. Assim, a comparação entre ArcFace e FaceNet é acompanhada por uma auditoria da região selecionada. O objetivo não é afirmar que um modelo seja melhor em qualquer situação, mas mostrar o comportamento dos dois reconhecedores neste conjunto específico e como a regra da menor distância pode mudar a interpretação de uma decisão. Além desta introdução, a Seção 2 apresenta o referencial teórico e os trabalhos relacionados, a Seção 3 descreve a metodologia, a Seção 4 apresenta e discute os resultados e a Seção 5 reúne as conclusões e limitações.

2 REFERENCIAL TEÓRICO

2.1 RECONHECIMENTO FACIAL E VERIFICAÇÃO

Na biometria facial, a verificação compara uma amostra com uma identidade declarada em relação 1:1, enquanto a identificação pesquisa uma amostra entre múltiplas identidades (JAIN; ROSS; PRABHAKAR, 2004). O presente estudo utiliza apenas verificação pareada.

Um sistema de verificação facial normalmente passa por etapas de detecção, alinhamento, geração de embeddings e comparação. O uso de redes profundas e alinhamento explícito foi consolidado por trabalhos como DeepFace (TAIGMAN et al., 2014), enquanto revisões posteriores mostram que arquitetura, função de perda, base de dados, protocolo e condições de aquisição também influenciam o resultado (KORTLI et al., 2020; WANG; DENG, 2021). Na verificação, o detector localiza as faces, o reconhecedor gera suas representações vetoriais e uma medida de distância ou similaridade é usada na decisão. FMR e FNMR permitem separar aceitações impostoras de rejeições genuínas, mas devem ser interpretadas de acordo com o ground truth e o ponto operacional escolhido (JAIN; ROSS; PRABHAKAR, 2004).

2.2 FACENET

O FaceNet aprende um espaço de embeddings no qual a distância entre vetores representa similaridade facial. Seu treinamento utiliza Triplet Loss, aproximando representações da mesma identidade e afastando identidades diferentes por uma margem. A proposta está centrada na representação facial; a métrica e o limiar usados para converter distâncias em decisões dependem do protocolo de aplicação (SCHROFF; KALENICHENKO; PHILBIN, 2015).

2.3 ARCFACE

O ArcFace utiliza Additive Angular Margin Loss para ampliar a separação angular entre identidades e produzir embeddings discriminativos. Embora essa estratégia organize o espaço de características de forma distinta do FaceNet, o desempenho observado em uma aplicação também depende do detector, das imagens, da métrica e do limiar adotado; por isso, diferenças experimentais não devem ser atribuídas apenas à função de perda (DENG et al., 2019).

2.4 RETINAFACE

O RetinaFace é um detector facial de estágio único que estima caixas delimitadoras e landmarks úteis ao alinhamento em condições não controladas (DENG et al., 2020). Neste estudo, ele foi adotado como detector comum para ArcFace e FaceNet, reduzindo a influência de diferenças entre detectores e permitindo que a comparação se concentre nos reconhecedores.

2.5 TRABALHOS RELACIONADOS

Os trabalhos relacionados abrangem tanto imagens com várias pessoas quanto comparações de reconhecedores. Manyam et al. (2011) estudaram reconhecimento conjunto em fotografias de grupo, e Huang et al. (2022) avaliaram verificação em entradas multiface. Firmansyah, Kusumasari e Alam (2023) compararam ArcFace, FaceNet e FaceNet512 no DeepFace, enquanto Serengil e Özpınar (2024) analisaram combinações de modelos, detectores, métricas e alinhamento.

Macarulla Rodriguez et al. (2024) avaliaram ArcFace, FaceNet e QMagFace em imagens e vídeos de vigilância, enquanto Sowan, Ibrahim e Keshlaf (2026) compararam ArcFace e FaceNet no DeepFace em 1.000 pares derivados do LFW. Os valores desses estudos não são diretamente comparáveis aos deste artigo porque diferem em base, escala, tarefa e regra de decisão.

O Quadro 1 sintetiza essas diferenças. O presente experimento se concentra na divergência entre o rótulo da fotografia e a face efetivamente selecionada pela regra da menor distância em verificação pareada, aspecto distinto dos protocolos anteriores.

Quadro 1 – Síntese dos trabalhos relacionados

Estudo	Dados/cenário	Protocolo	Métricas/avaliação	Achado principal	Diferença para este estudo
Manyam et al. (2011)	Fotografias de grupo; dois conjuntos	Reconhecimento conjunto de múltiplas faces	Acurácia de reconhecimento conjunto	Explora dependências entre faces da mesma fotografia	Não audita a regra de menor distância em verificação pareada
Huang et al. (2022)	500 imagens de 25 indivíduos; grupos com 1 a 4 faces	Verificação multiface com MTCNN e MDCNN	Acurácia; tempo no cenário multiface	Avalia verificação em entradas com múltiplas faces	Usa sistema e regra de seleção próprios, não a auditoria do mínimo empregada aqui
Firmansyah, Kusumasari e Alam (2023)	Conjunto de faces indonésias	DeepFace: ArcFace, FaceNet e FaceNet512	Acurácia	Mostra diferenças de desempenho entre modelos	Comparação de modelos sem foco na seleção de face em fotografia multiface
Serengil e Özpınar (2024)	Benchmark de pipelines	Combina modelos, detectores, métricas e alinhamento	Acurácia por combinações de módulos	Mostra influência do pipeline completo	Foco em combinações de módulos, não na divergência rótulo da foto × face selecionada
Macarulla Rodriguez et al. (2024)	Imagens e vídeos de vigilância	ArcFace, FaceNet e QMagFace com medidas de qualidade	Cllr (custo log-likelihood ratio)	Destaca efeitos de pose, qualidade e seleção de quadros	Protocolo de vigilância distinto do pareamento multiface deste estudo
Sowan, Ibrahim e Keshlaf (2026)	1.000 pares derivados do LFW	DeepFace: ArcFace × FaceNet	Acurácia; taxa de erro; tempo de verificação	Compara acurácia e tempo de verificação	Verificação por pares, sem seleção entre múltiplas faces na mesma fotografia

Fonte: Elaborado pelos autores com base em Manyam et al. (2011), Huang et al. (2022), Firmansyah, Kusumasari e Alam (2023), Serengil e Özpınar (2024), Macarulla Rodriguez et al. (2024) e Sowan, Ibrahim e Keshlaf (2026).

3 METODOLOGIA

Este é um estudo aplicado e quantitativo, com comparação exploratória entre ArcFace e FaceNet sob o mesmo protocolo de verificação facial pareada. O principal fator de comparação foi o

modelo de reconhecimento. A quantidade de faces da fotografia foi usada apenas para separar os resultados em imagens com uma face e imagens multiface; o número de pessoas não foi controlado experimentalmente. Por isso, essa divisão é descritiva e não permite atribuir causalidade à quantidade de faces. O protocolo também não representa uma tarefa operacional de identificação 1:N.

3.1 CARACTERIZAÇÃO E CONJUNTO EXPERIMENTAL

O conjunto experimental reuniu 40 fotografias públicas do Wikimedia Commons, distribuídas entre Barack Obama, Neymar Jr., Scarlett Johansson e Taylor Swift. Foram utilizadas duas referências e oito testes por identidade, totalizando oito referências e 32 testes. A seleção foi intencional e não probabilística; não havia critérios adicionais pré-especificados de pose, resolução, iluminação ou número de pessoas. As oito referências continham uma face detectada; entre os testes, dez imagens tinham uma face e 22 eram multiface. Para rastreabilidade das fotografias, foi mantido um manifesto das imagens com nome do arquivo, autoria, fonte, licença e URL original quando registrada. A organização entre imagens de referência e de teste e a contagem de faces detectadas pelo RetinaFace estão documentadas nos arquivos do experimento, incluindo contagem_faces_retinaface.csv. O código-fonte, o manifesto das imagens e os arquivos de resultados estão disponíveis no repositório público do projeto no GitHub: https://github.com/ArthurSarmanho/arcface-facenet-verification.

As 512 decisões produzidas correspondem a combinações repetidas das mesmas quatro identidades e 40 fotografias, não a 512 indivíduos independentes. Por essa razão, os resultados são tratados de forma descritiva e não são utilizados para inferência populacional.

3.2 AMBIENTE E PROCESSAMENTO DAS IMAGENS

O experimento foi implementado em Python com DeepFace 0.0.101, TensorFlow 2.21.0, OpenCV 4.11.0.86 e retina-face 0.0.18, conforme o arquivo requirements.txt. A execução ocorreu em um notebook com Windows 11 Home, processador Intel Core i5 de 13ª geração, 8 GB de memória RAM DDR5 e GPU NVIDIA GeForce RTX 3050 com 6 GB de memória dedicada. O RetinaFace foi usado como detector comum. No app.py, a detecção mantém enforce_detection e alinhamento habilitados, sem limiar de confiança ou filtro adicional de tamanho de face definido explicitamente. Depois da detecção, os embeddings são gerados com o detector em modo skip e normalização base. Na DeepFace 0.0.101, o modelo Facenet chamado pelo projeto é a variante de 128 dimensões, enquanto o ArcFace gera embeddings de 512 dimensões. Nessa versão da biblioteca, o código está

configurado para baixar facenet_weights.h5 e arcface_weights.h5 da release v1.0 do repositório serengil/deepface_models. Os arquivos físicos desses pesos não estão entre os artefatos preservados do projeto; por isso, seus hashes não puderam ser confirmados. Os resultados intermediários foram reutilizados para evitar processamento repetido. A organização do pipeline em detecção, alinhamento, representação e verificação é compatível com a arquitetura modular descrita por Serengil e Özpınar (2020, 2024), enquanto as configurações específicas desta execução estão registradas no código do projeto.

3.3 PROTOCOLO DE COMPARAÇÃO

Cada uma das oito referências foi comparada com as 32 imagens de teste, totalizando 256 decisões por modelo e 512 no conjunto dos dois reconhecedores. Para cada referência, oito testes pertenciam à mesma identidade rotulada e 24 pertenciam às outras três identidades. Assim, por modelo, foram formados $8 \times 8 = 64$ pares genuínos e $8 \times 24 = 192$ pares impostores na rotulagem original das fotografias.

A comparação entre embeddings utilizou distância cosseno. O protocolo multiface possui duas etapas distintas: primeiro, o RetinaFace detecta as regiões faciais e o sistema calcula as distâncias entre as faces da referência e todas as faces do teste; em seguida, a região associada à menor distância é selecionada como candidata. Somente depois dessa seleção a distância mínima é comparada ao limiar do modelo para produzir a decisão de aceitação ou rejeição. Foram utilizados os limiares padrão da DeepFace 0.0.101 para distância cosseno, definidos na configuração da biblioteca: 0,68 para ArcFace e 0,40 para Facenet. Esses valores não foram calibrados a partir do conjunto experimental. Distâncias menores ou iguais ao limiar foram aceitas como correspondência. Em imagens auditadas visualmente, a identidade da face selecionada foi analisada separadamente da decisão binária.

3.4 GROUND TRUTH E AUDITORIA DAS FACES SELECIONADAS

A classificação original usa como referência a identidade associada à fotografia de teste. Em imagens com várias pessoas, porém, esse rótulo não garante que a menor distância tenha sido produzida pela pessoa-alvo. Para verificar esse ponto, os autores fizeram uma auditoria visual de todas as 128 comparações genuínas e dos nove falsos positivos, totalizando 137 decisões. As coordenadas registradas no CSV foram usadas para localizar a região selecionada. Nas comparações genuínas, a face foi classificada como distratora quando a região escolhida correspondia visualmente a uma pessoa

diferente da identidade rotulada na fotografia. Nos pares impostores, a inspeção considerou a identidade rotulada na foto de teste e também verificou se havia evidência visual de seleção da identidade usada como referência.

Não foi produzido ground truth completo para todas as pessoas secundárias detectadas nas imagens. As 375 comparações impostoras rejeitadas não receberam anotação no nível da face, pois a auditoria foi concentrada em todas as comparações genuínas e em todos os falsos positivos. Como as 128 comparações genuínas foram inspecionadas, é possível calcular a frequência de seleção de distratores nesse grupo; esses resultados são apresentados na Seção 4.3. As métricas gerais, por sua vez, continuam baseadas no rótulo da fotografia.

3.5 MÉTRICAS E ANÁLISE DOS RESULTADOS

No protocolo baseado no rótulo das fotografias, verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN) foram definidos pela relação entre o rótulo da imagem e a decisão do modelo. A FMR foi calculada por FP/(FP+VN), a FNMR por FN/(FN+VP) e a acurácia por (VP+VN)/(VP+VN+FP+FN). Em cada modelo, a FMR usa 192 comparações impostoras e a FNMR usa 64 genuínas. Como essas métricas partem do rótulo da fotografia, elas descrevem o protocolo empregado e não substituem taxas biométricas obtidas com ground truth completo no nível de cada face (JAIN; ROSS; PRABHAKAR, 2004). As mesmas imagens e identidades são reutilizadas em vários pares; por isso, as 512 decisões não foram tratadas como observações independentes para cálculo de intervalos de confiança ou inferência populacional.

Os resultados também foram separados conforme a presença de uma ou múltiplas faces. A interpretação considera que os limiares dos dois modelos são pontos operacionais distintos e, portanto, diferenças de FMR e FNMR observadas nesses valores não representam comparação em níveis equivalentes de FMR nem demonstram superioridade universal. Em avaliações biométricas, o desempenho depende do limiar e do compromisso entre diferentes tipos de erro (JAIN; ROSS; PRABHAKAR, 2004).

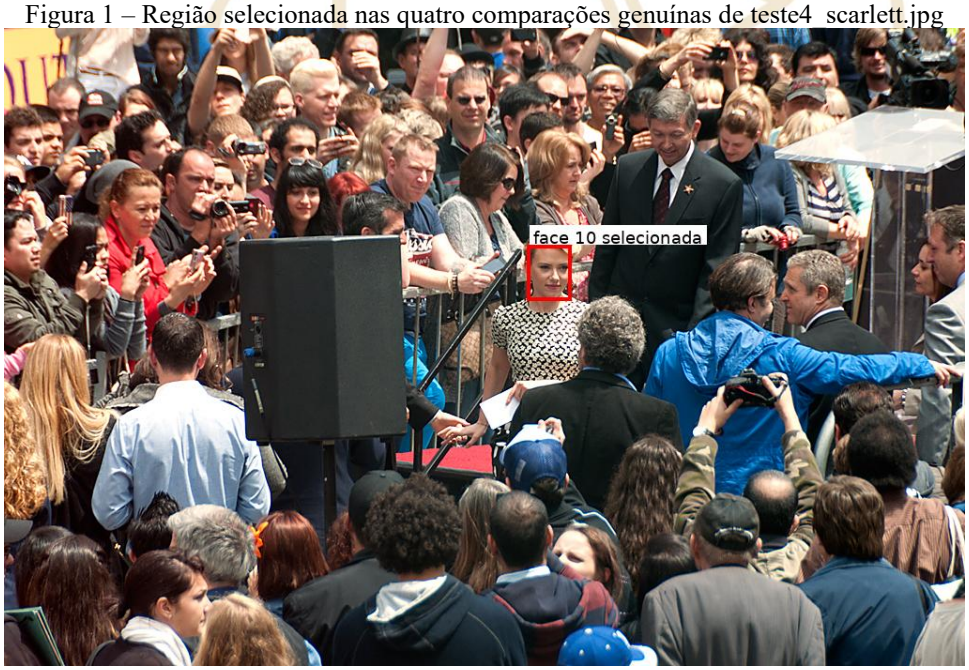
4 RESULTADOS E DISCUSSÃO

4.1 VISÃO GERAL DOS RESULTADOS

Foram realizadas 512 comparações no total, sendo 256 para cada modelo. Em cada modelo, 64 comparações eram genuínas e 192 impostoras. Todas as comparações do experimento foram processadas, sem registros de falha de detecção ou N/A.

Das 32 imagens de teste, dez continham uma única face e 22 apresentavam múltiplas faces. O caso com maior número de detecções foi teste4_scarlett.jpg, com 41 faces. Nas quatro comparações genuínas por rótulo associadas a essa fotografia — duas referências em cada modelo — a menor distância veio da face de índice 10, classificada pelos autores como a pessoa-alvo, localizada em (x=560, y=234, w=41, h=53), e as quatro decisões foram de aceitação. Como não há identidade anotada para as outras 40 faces, elas não foram identificadas individualmente.

Figura 1 – Região selecionada nas quatro comparações genuínas de teste4_scarlett.jpg



Fonte: Wikimedia Commons, arquivo do conjunto experimental; autor mbtrama, licença CC BY 2.0. Caixa e índice elaborados pelos autores a partir do CSV (2026).

Nota: a caixa destaca a face-alvo de índice 10, bbox (560, 234, 41, 53), que também foi a face selecionada pela menor distância nas quatro comparações genuínas associadas a teste4_scarlett.jpg (duas referências × dois modelos); as quatro decisões foram aceitas.

Tabela 1 – Desempenho geral dos modelos na rotulagem por imagem

Modelo	VP	VN	FP	FN	FMR (%)	FNMR (%)	Acurácia (%)
ArcFace	52	185	7	12	3,65	18,75	92,58
FaceNet	40	190	2	24	1,04	37,50	89,84

Fonte: Elaborado pelos autores (2026).
Nota: métricas calculadas segundo a rotulagem original das fotografias; por modelo, FMR usa 192 pares impostores e FNMR 64 pares genuínos. A auditoria da face selecionada é apresentada na Seção 4.3.

Nos limiares usados no experimento, o ArcFace teve maior acurácia e menor FNMR, enquanto o FaceNet apresentou menos falsas correspondências. Esses números, porém, são calculados a partir do rótulo da fotografia e não garantem que a menor distância tenha vindo da pessoa-alvo. Por isso, a Seção 4.3 analisa também a face que efetivamente sustentou a decisão. A ordem observada entre os modelos não se repete em todos os trabalhos anteriores. Firmansyah, Kusumasari e Alam (2023), também utilizando DeepFace, relataram acurácia de 92,1% para FaceNet e 87,8% para ArcFace em um conjunto de faces indonésias. Já Sowan, Ibrahim e Keshlaf (2026), em 1.000 pares derivados do LFW, obtiveram 92,4% para ArcFace e 83,6% para FaceNet. As diferenças entre esses resultados e os deste trabalho mostram que dataset, protocolo e limiar influenciam a comparação, ponto que também aparece na análise de sensibilidade da Seção 4.3.

4.2 COMPARAÇÃO ENTRE IMAGENS COM UMA FACE E MÚLTIPLAS FACES

As dez imagens com uma face originaram 80 comparações por modelo, enquanto as 22 imagens com múltiplas faces originaram 176.

Tabela 2 – Desempenho por cenário na rotulagem por imagem

Modelo	Cenário	N	VP	VN	FP	FN	FMR (%)	FNMR (%)	Acurácia (%)
ArcFace	1 face	80	14	56	4	6	6,67	30,00	87,50
ArcFace	Multiface	176	38	129	3	6	2,27	13,64	94,89
FaceNet	1 face	80	14	59	1	6	1,67	30,00	91,25
FaceNet	Multiface	176	26	131	1	18	0,76	40,91	89,20

Fonte: Elaborado pelos autores (2026).
Nota: métricas calculadas segundo a rotulagem original das fotografias. Por modelo, o cenário de uma face contém 20 pares genuínos e 60 impostores; o cenário multiface contém 44 genuínos e 132 impostores. A auditoria da face selecionada é apresentada na Seção 4.3.

A Tabela 2 mostra comportamentos distintos entre os cenários: no ArcFace, a acurácia foi maior no grupo multiface, enquanto no FaceNet foi ligeiramente maior no grupo com uma face. Esses valores são descritivos e devem ser interpretados em conjunto com FMR, FNMR e as características das imagens.

Os resultados da Tabela 2 não mostram que mais faces tenham melhorado ou piorado o reconhecimento de forma geral. A regra do protocolo escolhe a menor distância entre todas as faces detectadas, então um número maior de candidatos pode, em princípio, gerar uma distância mínima menor. Neste conjunto, porém, isso não apareceu como aumento da FMR: no grupo multiface, ela foi de 2,27% contra 6,67% no ArcFace e de 0,76% contra 1,67% no FaceNet. Como há poucos falsos positivos por cenário e não houve controle de fatores como número de distratores, pose, qualidade e iluminação, esses valores devem ser tratados apenas de forma descritiva.

4.3 ANÁLISE DOS ERROS E AUDITORIA DAS FACES SELECIONADAS

No protocolo original, o ArcFace registrou sete falsos positivos e o FaceNet dois. Alguns erros ficaram próximos aos limiares adotados, como uma distância de 0,678873 no ArcFace para o limite de 0,68 e uma distância de 0,393642 no FaceNet para o limite de 0,40. Esses exemplos mostram que decisões individuais podem mudar com pequenas alterações do ponto de corte. Como os limiares não foram calibrados em uma partição independente deste conjunto, eles são mantidos como configuração do protocolo e não são ajustados a posteriori para maximizar o resultado.

A inspeção visual realizada pelos autores cobriu todas as 128 comparações genuínas e os nove falsos positivos. As seis seleções classificadas como distratoras apareceram nas comparações genuínas, em três fotografias multiface: 4 de 64 casos do ArcFace (6,25%) e 2 de 64 do FaceNet (3,13%). Considerando somente aceitações fundamentadas na face-alvo, o ArcFace aceitou 50 das 64 comparações genuínas (78,13%) e o FaceNet 40 de 64 (62,50%). Nos nove falsos positivos, a auditoria não registrou nova seleção classificada como distratora segundo o critério adotado; esses casos permaneceram classificados como falsos positivos na rotulagem por fotografia. Em duas comparações do ArcFace com teste6_neymar.jpg, foram selecionadas as faces de índices 3 e 2, com distâncias 0,624913 e 0,587971, e ambas foram aceitas pelo limiar de 0,68. Em teste3_taylor.jpg, o ArcFace selecionou as faces 3 e 5, com distâncias 0,973367 e 0,852948, e ambas foram rejeitadas. Em teste3_obama.jpg, o FaceNet selecionou a face 1 em duas comparações, com distâncias 0,872373 e

0,868448, também rejeitadas. A Tabela 3 resume a cobertura da auditoria por modelo, e a Tabela 4 reúne as coordenadas e decisões dos seis casos.

Tabela 3 – Síntese da auditoria visual das comparações genuínas e falsos positivos

Modelo	Decisões auditadas	Seleções de distratores	Aceitas	Rejeitadas
ArcFace	71	4	2	2
FaceNet	66	2	0	2

Fonte: Elaborado pelos autores a partir do subconjunto auditado (2026).

Tabela 4 – Evidências das seleções de distratores no subconjunto auditado

Imagem	Modelo	Ref.	Rótulo	Face / bbox	Dist.	Limiar	Decisão	Auditoria
teste6_neymar.jpg	ArcFace	ref_neymar.jpg	Neymar Jr.	3 / (195,157,63,86)	0,624913	0,68	Aceita (VP)	Distrator
teste6_neymar.jpg	ArcFace	ref2_neymar.jpg	Neymar Jr.	2 / (10,193,57,73)	0,587971	0,68	Aceita (VP)	Distrator
teste3_taylor.jpg	ArcFace	ref_taylor.jpg	Taylor Swift	3 / (613,161,38,52)	0,973367	0,68	Rejeita (FN)	Distrator
teste3_taylor.jpg	ArcFace	ref2_taylor.jpg	Taylor Swift	5 / (240,216,28,37)	0,852948	0,68	Rejeita (FN)	Distrator
teste3_obama.jpg	FaceNet	ref_obama.jpg	Barack Obama	1 / (336,159,247,380)	0,872373	0,40	Rejeita (FN)	Distrator
teste3_obama.jpg	FaceNet	ref2_obama.jpg	Barack Obama	1 / (336,159,247,380)	0,868448	0,40	Rejeita (FN)	Distrator

Fonte: Elaborado pelos autores a partir do CSV de resultados e da auditoria visual (2026).
Nota: "Rótulo" indica a identidade associada à fotografia; "Auditoria" registra a classificação visual da região selecionada.

Figura 2 – Regiões selecionadas nos casos classificados como distratores na auditoria visual



Fonte: Wikimedia Commons: Agencia de Noticias ANDES, CC BY-SA 2.0 (teste6_neymar.jpg); Paolo Villanueva, CC BY 2.0 (teste3_taylor.jpg); Daniel Schwen, CC BY-SA 4.0 (teste3_obama.jpg). Caixas e índices elaborados pelos autores a partir das coordenadas registradas no CSV (2026).

Os seis casos não têm o mesmo efeito sobre as métricas. Nos dois casos aceitos, o resultado aparece como verdadeiro positivo pela rotulagem da fotografia, embora a decisão tenha sido sustentada por outra pessoa; no nível da face selecionada, são falsas correspondências que o rótulo da imagem não revela. Nos quatro casos rejeitados, a face distratora teve a menor distância e, mesmo assim, permaneceu acima do limiar. Se a face-alvo estava entre as detecções, sua distância não foi menor que a do distrator. Nesses casos, a seleção da face muda a interpretação do processo, mas não transforma a rejeição em aceitação.

Também foi feita uma análise exploratória de sensibilidade aos limiares. Foram testados valores em passos de 0,02, numa faixa de ±0,04 em torno dos limiares originais, apenas para observar como FMR, FNMR e acurácia mudam com o ponto de corte. Esses valores não foram usados para escolher um novo limiar e não substituem uma calibração em conjunto independente nem uma comparação em FMR equivalente (Tabela 5).

Tabela 5 – Análise exploratória de sensibilidade aos limiares

Modelo	Limiar	FMR (%)	FNMR (%)	Acurácia (%)
ArcFace	0,64	2,60	20,31	92,97
ArcFace	0,66	2,60	18,75	93,36
ArcFace	0,68	3,65	18,75	92,58
ArcFace	0,70	4,69	14,06	92,97
ArcFace	0,72	7,29	12,50	91,41
FaceNet	0,36	0,00	53,12	86,72
FaceNet	0,38	0,52	45,31	88,28
FaceNet	0,40	1,04	37,50	89,84
FaceNet	0,42	1,56	26,56	92,19
FaceNet	0,44	1,56	17,19	94,53

Fonte: Reanálise exploratória dos autores a partir das distâncias registradas no CSV final (2026).

A Tabela 5 deixa claro que a comparação muda quando o limiar é alterado. No ArcFace, a acurácia ficou entre 91,41% e 93,36% na faixa analisada. No FaceNet, passar de 0,40 para 0,44 reduziu a FNMR de 37,50% para 17,19%, aumentou a FMR de 1,04% para 1,56% e elevou a acurácia para 94,53%, acima do maior valor obtido pelo ArcFace nessa análise. Em 0,44, o FaceNet também apresentou FMR e FNMR menores que as do ArcFace em seu limiar padrão de 0,68. Isso não significa que 0,44 seja um novo limiar recomendado, pois o valor foi observado no mesmo conjunto de teste. O resultado mostra apenas que a ordem vista na Tabela 1 depende do ponto operacional.

5 CONCLUSÃO

O resultado mais importante apareceu na auditoria da face realmente selecionada. Foram inspecionadas todas as 128 comparações genuínas e os nove falsos positivos. Entre as genuínas, o ArcFace selecionou distratores em 4 de 64 casos (6,25%) e o FaceNet em 2 de 64 (3,13%). Em duas dessas ocorrências, o ArcFace aceitou uma região distratora em decisões que apareciam como verdadeiros positivos pela rotulagem da fotografia. Isso mostra que uma métrica calculada apenas pelo rótulo da imagem pode esconder uma falsa correspondência no nível da face usada na decisão.

Nos limiares padrão usados no experimento, o ArcFace teve maior acurácia e menor FNMR, enquanto o FaceNet teve menor FMR. A análise de sensibilidade mostrou que essa ordem pode mudar: com o FaceNet em 0,44, a acurácia chegou a 94,53%, acima dos valores do ArcFace na faixa analisada. Como esse ponto foi observado no próprio conjunto de teste, ele não é proposto como novo limiar. A conclusão é mais limitada: a comparação entre os modelos depende do ponto operacional e não sustenta uma superioridade geral de um reconhecedor.

Há três grupos principais de limitações. Na validade interna, a classificação entre face-alvo e distrator dependeu da inspeção visual dos autores; não houve ground truth independente para todas as faces nem medida de concordância entre avaliadores. Na validade externa, o conjunto é pequeno, com quatro identidades públicas e 40 fotografias escolhidas de forma não probabilística e reutilizadas em vários pares. Por isso, as 512 decisões não devem ser tratadas como observações independentes nem generalizadas para outras populações. Quanto à reprodutibilidade, o código, os CSVs e o manifesto foram preservados, mas os arquivos físicos dos pesos e seus hashes não fazem parte dos artefatos do projeto, e 24 das 40 imagens ainda não têm a URL original localizada no manifesto. Como as quatro identidades são figuras públicas amplamente fotografadas, não foi possível descartar sobreposição de identidades com dados de treinamento dos reconhecedores; também não há evidência de que as fotografias específicas deste conjunto tenham sido usadas no treinamento. Como o estudo realiza tratamento de dados biométricos vinculados a pessoas identificáveis, esse tipo de dado é classificado pela Lei Geral de Proteção de Dados como dado pessoal sensível (BRASIL, 2018). As imagens foram obtidas de fontes públicas, com autoria e licença registradas no manifesto, e foram utilizadas apenas para finalidade acadêmica.

Como continuidade, estudos com mais identidades e bases padronizadas permitiriam avaliar se o comportamento observado se repete em outros conjuntos. Também seria útil controlar a quantidade de distratores por imagem e realizar anotação completa no nível da face. Comparações em pontos

operacionais equivalentes, por curvas ROC/DET ou FNMR em FMR fixada, ajudariam a separar melhor o efeito do modelo do efeito do limiar. O manifesto das imagens, o código-fonte e os arquivos de resultados estão disponíveis no repositório público do projeto no GitHub: https://github.com/ArthurSarmanho/arcface-facenet-verification.

REFERÊNCIAS

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 18 nov. 2026.

DENG, Jiankang; GUO, Jia; XUE, Niannan; ZAFEIRIOU, Stefanos. ArcFace: additive angular margin loss for deep face recognition. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2019, Long Beach, CA. Proceedings [...]. [S. l.]: IEEE/CVF, 2019. p. 4690-4699. Disponível em: https://openaccess.thecvf.com/content_CVPR_2019/html/Deng_ArcFace_Additive_Angular_Margin_Loss_for_Deep_Face_Recognition_CVPR_2019_paper.html. Acesso em: 13 nov. 2026.

DENG, Jiankang; GUO, Jia; VERVERAS, Evangelos; KOTSIA, Irene; ZAFEIRIOU, Stefanos. RetinaFace: single-shot multi-level face localisation in the wild. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2020, Seattle, WA. Proceedings [...]. [S. l.]: IEEE/CVF, 2020. p. 5203-5212. Disponível em: https://openaccess.thecvf.com/content_CVPR_2020/html/Deng_RetinaFace_Single-Shot_Multi-Level_Face_Localisation_in_the_Wild_CVPR_2020_paper.html. Acesso em: 13 nov. 2026.

FIRMANSYAH, Andrian; KUSUMASARI, Tien Fabrianti; ALAM, Ekky Novriza. Comparison of Face Recognition Accuracy of ArcFace, Facenet and Facenet512 Models on Deepface Framework. In: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE, INFORMATION TECHNOLOGY AND ENGINEERING (ICCoSITE), 2023, Jakarta, Indonesia. Proceedings [...]. [S. l.]: IEEE, 2023. p. 535-539. DOI: 10.1109/ICCoSITE57641.2023.10127799. Disponível em: https://doi.org/10.1109/ICCoSITE57641.2023.10127799. Acesso em: 17 nov. 2026.

HUANG, Xixian; ZENG, Xiongjun; WU, Qingxiang; LU, Yu; HUANG, Xi; ZHENG, Hua. Face Verification Based on Deep Learning for Person Tracking in Hazardous Goods Factories. Processes, v. 10, n. 2, art. 380, 2022. DOI: 10.3390/pr10020380. Disponível em: https://doi.org/10.3390/pr10020380. Acesso em: 18 nov. 2026.

JAIN, Anil K.; ROSS, Arun; PRABHAKAR, Salil. An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, v. 14, n. 1, p. 4-20, 2004. DOI: 10.1109/TCSVT.2003.818349. Disponível em: https://doi.org/10.1109/TCSVT.2003.818349. Acesso em: 18 nov. 2026.

KORTLI, Yassin; JRIDI, Maher; AL FALOU, Ayman; ATRI, Mohamed. Face Recognition Systems: A Survey. Sensors, v. 20, n. 2, art. 342, 2020. DOI: 10.3390/s20020342. Disponível em: https://doi.org/10.3390/s20020342. Acesso em: 18 nov. 2026.

MACARULLA RODRIGUEZ, Andrea; GERADTS, Zeno; WORRING, Marcel; UNZUETA, Luis. Improved likelihood ratios for face recognition in surveillance video by multimodal feature pairing. Forensic Science International: Synergy, v. 8, art. 100458, 2024. DOI: 10.1016/j.fsisyn.2024.100458. Disponível em: https://doi.org/10.1016/j.fsisyn.2024.100458. Acesso em: 14 nov. 2026.

MANYAM, Ohil K.; KUMAR, Neeraj; BELHUMEUR, Peter N.; KRIEGMAN, David J. Two faces are better than one: face recognition in group photographs. In: INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS (IJCB), 2011, Washington, DC. Proceedings [...]. [S. l.]: IEEE, 2011. p. 1-8. DOI: 10.1109/IJCB.2011.6117516. Disponível em: https://doi.org/10.1109/IJCB.2011.6117516. Acesso em: 12 nov. 2026.

SCHROFF, Florian; KALENICHENKO, Dmitry; PHILBIN, James. FaceNet: a unified embedding for face recognition and clustering. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2015, Boston, MA. Proceedings [...]. [S. l.]: IEEE, 2015. p. 815-823. Disponível em: https://openaccess.thecvf.com/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html. Acesso em: 13 nov. 2026.

SERENGIL, Sefik Ilkin; ÖZPINAR, Alper. LightFace: a hybrid deep face recognition framework. In: INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS CONFERENCE (ASYU), 2020, Istanbul, Turkey. Proceedings [...]. [S. l.]: IEEE, 2020. p. 23-27. DOI: 10.1109/ASYU50717.2020.9259802. Disponível em: https://doi.org/10.1109/ASYU50717.2020.9259802. Acesso em: 15 nov. 2026.

SERENGIL, Sefik Ilkin; ÖZPINAR, Alper. A benchmark of facial recognition pipelines and co-usability performances of modules. Bilişim Teknolojileri Dergisi, v. 17, n. 2, p. 95-107, 2024. DOI: 10.17671/gazibtd.1399077. Disponível em: https://doi.org/10.17671/gazibtd.1399077. Acesso em: 16 nov. 2026.

SOWAN, Essra; IBRAHIM, Juma; KESHLAF, Ayad Ali. Facial Recognition Accuracy and Verification time Comparison Between ArcFace and Facenet Models on DeepFace Framework. In: IEEE 5TH INTERNATIONAL MAGHREB MEETING OF THE CONFERENCE ON SCIENCES AND TECHNIQUES OF AUTOMATIC CONTROL AND COMPUTER ENGINEERING (MI-STA), 2026, Sebha, Libya. Proceedings [...]. [S. l.]: IEEE, 2026. p. 460-465. DOI: 10.1109/MI-STA68962.2026.11511304. Disponível em: https://doi.org/10.1109/MI-STA68962.2026.11511304. Acesso em: 18 nov. 2026.

TAIGMAN, Yaniv; YANG, Ming; RANZATO, Marc'Aurelio; WOLF, Lior. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2014, Columbus, OH. Proceedings [...]. [S. l.]: IEEE, 2014. p. 1701-1708. Disponível em: https://openaccess.thecvf.com/content_cvpr_2014/html/Taigman_DeepFace_Closing_the_2014_CVPR_paper.html. Acesso em: 16 nov. 2026.

WANG, Mei; DENG, Weihong. Deep face recognition: A survey. Neurocomputing, v. 429, p. 215-244, 2021. DOI: 10.1016/j.neucom.2020.10.081. Disponível em: https://doi.org/10.1016/j.neucom.2020.10.081. Acesso em: 18 nov. 2026.

ZHAO, Wenyi; CHELLAPPA, Rama; PHILLIPS, P. Jonathon; ROSENFELD, Azriel. Face recognition: a literature survey. ACM Computing Surveys, v. 35, n. 4, p. 399-458, 2003. DOI:

10.1145/954339.954342. Disponível em: https://doi.org/10.1145/954339.954342. Acesso em: 17 nov. 2026.