UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Estudo sobre a identificação dos cadastros de cooperados de uma Instituição Financeira Cooperativa real que possuam foto com indício de fraude

Edson Rodrigues Lisboa Júnior

Monografia - MBA em Inteligência Artificial e Big Data

ICMC USP SÃO CARLOS

Edson Rodrigues Lisboa Júnior

Estudo sobre a identificação dos cadastros de cooperados de uma Instituição Financeira Cooperativa real que possuam foto com indício de fraude

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. PhD. Jean Roberto Ponciano

Versão original

São Carlos

2025

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m

Lisboa Júnior, Edson R.
Estudo sobre a identificação dos cadastros de cooperados de uma Instituição Financeira Cooperativa real que possuam foto com indício de fraude / Edson Rodrigues Lisboa Júnior ; orientador Jean Roberto Ponciano. – São Carlos, 2025.
73 p. : il. (algumas color.) ; 30 cm.

Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2025.

1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Dissertação. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. Ponciano, Jean Roberto, orient. II. Título.

Edson Rodrigues Lisboa Júnior

Study on the identification of records of real Cooperative Financial Institution member's who have a photo with evidence of fraud

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. PhD. Jean Roberto Ponciano

Original version

São Carlos

2025

Este trabalho é dedicado aos alunos da USP, como uma contribuição das Bibliotecas do Campus USP de São Carlos para o desenvolvimento e disseminação da pesquisa científica da Universidade.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder sabedoria, força e perseverança para trilhar esta jornada acadêmica e chegar até a conclusão deste trabalho.

À minha amada esposa Fabiana Lisboa e aos meus filhos Matheus, Gabriel e Lucas, expresso minha mais profunda gratidão pelo apoio incondicional e pelo incentivo constante que foram fundamentais para que eu pudesse me dedicar a esta formação. Vocês são minha maior motivação.

Ao Sicoob, instituição cooperativa financeira onde tenho o privilégio de trabalhar, registro meu sincero reconhecimento pelo patrocínio desta formação.

Aos colegas de trabalho Rodolfo Ayala, Rafael Pantoja e Isaac Pessoa, meu especial agradecimento pelo apoio técnico, pelos debates enriquecedores sobre o tema e pelo esclarecimento de dúvidas que surgiram ao longo do processo. A troca de conhecimentos e experiências foi fundamental para o aprimoramento deste trabalho.

Aos colegas Vilaça e Dênio, companheiros de trabalho no Sicoob e também de MBA, agradeço pela parceria e apoio mútuo durante toda esta jornada de formação.

Por fim, expresso minha sincera gratidão ao Prof. PhD Jean Ponciano, meu orientador, pela dedicação, paciência e valiosas contribuições que foram essenciais para o desenvolvimento e aprimoramento deste trabalho.

"O temor a Deus é o princípio da sabedoria."
Provérbios 9:10

RESUMO

LISBOA JR, E. R. **Estudo sobre a identificação dos cadastros de cooperados de uma Instituição Financeira Cooperativa real que possuam foto com indício de fraude**. 2025. 73 p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

A crescente digitalização dos serviços financeiros e o aumento das fraudes digitais têm intensificado a necessidade de sistemas robustos de verificação de identidade em instituições financeiras. Este trabalho estuda a aplicação de técnicas avançadas de processamento de imagens e inteligência artificial para identificar cadastros de cooperados que apresentem indícios de fraude por meio da detecção de duplicidade de fotos faciais. O estudo foi conduzido em duas fases: uma avaliação experimental controlada utilizando um conjunto de dados público de 442 imagens de faces de celebridades e uma aplicação prática em uma base real de 32.192 cadastros de uma cooperativa financeira do Sistema Sicoob. Foram avaliadas 22 técnicas diferentes, organizadas em duas categorias: baseadas nas características da imagem (MSE, SSIM, SIFT, ORB, PCA-SIFT, intersecção de histogramas) e baseadas na estrutura da imagem utilizando Perceptual Hashing (Average Hash, Perception Hash, Difference Hash, Wavelet Hash) ou utilizando Redes Neurais Convolucionais considerando 11 modelos pré-treinados (VGG16, VGG19, ResNet50, ResNet101, ResNet152, InceptionV3, Xception, EfficientNetB7, MobileNet, DenseNet201 e ConvNeXtBase), além de Large Language Models multimodais. A metodologia de avaliação baseou-se em métricas de acurácia, recall e taxa de falsos positivos, utilizando análise de dominância de Pareto para seleção das técnicas mais eficazes. Os resultados experimentais demonstraram que as técnicas baseadas em CNNs, especificamente MobileNet (99,94% de acurácia e 98,64% de recall) e ConvNeXtBase (98,87% de acurácia e 100% de recall), apresentaram o melhor desempenho na detecção de similaridades. A aplicação na base real identificou 662 casos de verdadeiros positivos, validando a eficácia das técnicas selecionadas em um ambiente operacional real. O estudo contribui para o aprimoramento da segurança cibernética em instituições financeiras cooperativas, oferecendo uma metodologia de baixo custo, escalável e eficiente para detecção de fraudes em cadastros de cooperados.

Palavras-chave: Biometria Facial. Detecção de Fraude. Perceptual Hashing. Redes Neurais Convolucionais. Processamento de Imagens. Cooperativismo Financeiro. Segurança Cibernética.

ABSTRACT

LISBOA JR, E. R. **Study on the identification of records of real Cooperative Financial Institution member's who have a photo with evidence of fraud**. 2025. 73 p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

The growing digitalization of financial services and the increase in digital fraud have intensified the need for robust identity verification systems in financial institutions. This work studies the application of advanced image processing and artificial intelligence techniques to identify member records that present evidence of fraud through the detection of facial photo duplicity. The study was conducted in two phases: a controlled experimental evaluation using a public dataset of 442 celebrity face images and a practical application on a real database of 32,192 records from a Sicoob System financial cooperative. Twenty-two different techniques were evaluated, organized into two categories: based on image characteristics (MSE, SSIM, SIFT, ORB, PCA-SIFT, histogram intersection); and based on image structure using Perceptual Hashing (Average Hash, Perception Hash, Difference Hash, Wavelet Hash) or using Convolutional Neural Networks considering 11 pre-trained models (VGG16, VGG19, ResNet50, ResNet101, ResNet152, InceptionV3, Xception, EfficientNetB7, MobileNet, DenseNet201, and ConvNeXtBase), as well as multimodal Large Language Models. The evaluation methodology was based on accuracy, recall, and false positive rate metrics, using Pareto dominance analysis to select the most effective techniques. Experimental results demonstrated that CNN-based techniques, specifically MobileNet (99.94% accuracy and 98.64% recall) and ConvNeXtBase (98.87% accuracy and 100% recall), showed the best performance in similarity detection. The application on the real database identified 662 true positive cases, validating the effectiveness of the selected techniques in a real operational environment. The study contributes to improving cybersecurity in cooperative financial institutions, offering a low-cost, scalable, and efficient methodology for fraud detection in member records.

Keywords: Facial Biometrics. Fraud Detection. Perceptual Hashing. Convolutional Neural Networks. Image Processing. Financial Cooperativism. Cybersecurity.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

USP	Universidade de São Paulo
USPSC	Campus USP de São Carlos
aHash	Average Hash
BCB	Banco Central do Brasil
FEBRABAN	Federação Brasileira de Bancos
SICOOB	Sistema de Cooperativas de Crédito do Brasil
DTVM	Distribuidora de Títulos e Valores Mobiliários
FGCoop	Fundo Garantidor do Cooperativismo de Crédito
MSE	Mean Squared Error (Erro Quadrático Médio)
OCB	Organização das Cooperativas Brasileiras
ORB	Oriented FAST and Rotated BRIEF
SIFT	Scale-Invariant Feature Transform
SSIM	Structural Similarity Index (Índice de Similaridade Estrutural)
PCA	Principal Components Analysis (Análise de Componentes Principais)
PCA-SIFT	Principal Components Analysis - Scale-Invariant Feature Transform
pHash	Perception Hash
dHash	Difference Hash
wHash	Wavelet Hash
DWT	Transformação Wavelet Discreta
PSNR	Peak Signal-to-Noise Ratio
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
PCNN	Pulse Coupled Neural Network
CNN	Convolutional Neural Network (Rede Neural Convolucional)

DCT	Transformada Discreta de Cosseno
FLANN	Fast Library for Approximate Nearest Neighbors
BRIEF	Binary Robust Independent Elementary Features
FAST	Features from Accelerated Segment Test
LLM	Large Language Model
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
FN	Falso Negativo
FP	Falso Positivo
PEM	Pessoas Expostas à Mídia
PEP	Pessoas Expostas Politicamente

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

No cenário atual da transformação digital do setor financeiro, a segurança cibernética passa a ser um pilar fundamental para a proteção dos ativos da instituição financeira e dos seus clientes, que em uma cooperativa financeira são conhecidos como cooperados. Com a consolidação dos canais digitais como a principal forma de relacionamento do cooperado com a sua cooperativa no uso dos produtos e serviços disponíveis, o desafio constante, assim como nas demais instituições do mercado financeiro, é a validação da identidade de quem está realizando a transação financeira pelo canal digital. Além disso, as cooperativas financeiras, por sua natureza de relacionamento próximo com os cooperados, têm uma responsabilidade adicional na proteção de seus membros contra fraudes.

Nesse contexto, a biometria facial representa uma evolução significativa nos sistemas de autenticação e validação de identidade, oferecendo uma combinação de segurança e conveniência. Entretanto, considerando os custos envolvidos para utilização das bases de biometrias faciais do governo e até mesmo de bureau (empresas privadas com bases próprias), uma das estratégias para a eficiência econômica é a utilização das fotos dos clientes existentes em bases cadastrais próprias. Para isso, torna-se essencial a identificação prévia de cadastros com fotos potencialmente fraudulentas que podem comprometer os sistemas de autenticação baseados em biometria facial, principalmente, considerando que, segundo dados da Federação Brasileira de Bancos (FEBRABAN), as tentativas de fraudes digitais contra instituições financeiras aumentaram 165% desde 2020, com prejuízos estimados em R$ 2,5 bilhões (FEBRABAN, 2021; Arimathea *et al.*, 2022).

O objetivo geral deste trabalho é identificar e avaliar métodos/técnicas eficazes para identificar cadastros de cooperados que apresentem indícios de fraude considerando a existência de mais de um cadastro com fotos de registro de uma mesma pessoa, fortalecendo assim a segurança cibernética da instituição e protegendo tanto os ativos financeiros quanto a confiança dos cooperados no sistema.

1.2 Justificativa e Motivação

A escolha deste tema justifica-se pela crescente sofisticação das técnicas de fraude digital, incluindo o uso de manipulação de cadastros e fotos/imagens que podem burlar os sistemas tradicionais de verificação de identidade. Uma das estratégias utilizadas pelos fraudadores é o aliciamento de funcionários das instituições financeiras que possuam credenciais privilegiadas para alteração da foto do cliente pela foto do fraudador em diversos cadastros. Isso evidencia a urgência em desenvolver mecanismos mais robustos de

detecção e prevenção desse tipo de ataque considerando os cadastros ativos já alterados, ainda que processos seguros já tenham sido implantados para controlar e evitar novas alterações dessa natureza.

Para isso, será necessário a aplicação de técnicas avançadas de processamento de imagens e inteligência artificial, buscando a criação de um sistema de detecção que possa ser integrado aos processos existentes de verificação de identidade, fortalecendo assim a segurança cibernética da instituição e protegendo tanto os ativos financeiros quanto a confiança dos cooperados no sistema.

2 REFERENCIAL TEÓRICO

2.1 Cooperativismo Financeiro

O cooperativismo financeiro representa um modelo econômico alternativo ao sistema bancário tradicional, fundamentado nos princípios de autogestão, democracia e participação igualitária. Conforme Meinen e Port (2014, p. 48), "o cooperativismo de crédito baseia-se na associação autônoma de pessoas que se unem voluntariamente para satisfazer aspirações e necessidades econômicas, sociais e culturais comuns por meio de uma empresa de propriedade coletiva e democraticamente gerida". Este modelo tem se destacado no Brasil pela sua capacidade de promover inclusão financeira e desenvolvimento local, especialmente em regiões menos atendidas pelos bancos comerciais e tem avançado cada vez mais para as médias e grandes cidades do país principalmente pela forte atuação nos canais digitais.

Nas cooperativas financeiras, os cooperados encontram os principais serviços disponíveis nos bancos tradicionais, como conta-corrente, investimentos, empréstimos, financiamentos, câmbio, seguro, consórcio, cartão de débito/crédito e demais meios de pagamento e recebimento. Os cooperados têm poder igual de voto independentemente da sua cota de participação no capital social da cooperativa.

O resultado positivo da cooperativa é conhecido como sobra e é distribuído entre os cooperados proporcionalmente às operações que cada associado realiza com a cooperativa. Assim, os ganhos voltam para a comunidade dos cooperados. No entanto, assim como partilha das sobras, o cooperado está sujeito a participar do rateio de eventuais perdas também na proporção dos serviços usufruídos.

As cooperativas de crédito são autorizadas e supervisionadas pelo Banco Central e seus depósitos têm a proteção do Fundo Garantidor do Cooperativismo de Crédito (FGCoop). Esse fundo garante os depósitos e os créditos mantidos nas cooperativas singulares de crédito e nos bancos cooperativos em caso de intervenção ou liquidação extrajudicial dessas instituições. Atualmente, o valor limite dessa proteção é o mesmo em vigor para os depositantes dos bancos.

No contexto nacional, o cooperativismo financeiro ganhou força a partir da Resolução nº 3.106/2003 do Conselho Monetário Nacional, que permitiu a livre admissão de associados às cooperativas de crédito (Jacques; Gonçalves, 2016). Segundo dados da Organização das Cooperativas Brasileiras (OCB), o setor cresceu significativamente na última década, evidenciando sua relevância como alternativa viável e sustentável para o acesso a serviços financeiros (OCB, 2023). O cooperativismo não visa lucros, pois os direitos e deveres de todos são iguais e a adesão é livre e voluntária

De acordo com Búrigo (2010, p. 87), "as cooperativas de crédito diferenciam-se das

demais instituições financeiras não apenas por sua natureza jurídica, mas principalmente pela filosofia que norteia seu funcionamento, centrada no atendimento às necessidades financeiras do quadro social". Esta filosofia traduz-se em taxas mais justas, atendimento personalizado, distribuição dos resultados da instituição entre os cooperados e desenvolvimento das comunidades onde estão inseridas.

2.2 Sistema de Cooperativas de Crédito do Brasil (Sicoob)

O Sistema de Cooperativas de Crédito do Brasil (Sicoob) figura entre os principais sistemas cooperativos financeiros do país, constituindo-se como um agente transformador na oferta de produtos e serviços financeiros. Segundo Soares e Sobrinho (2008, p. 124), "o Sicoob representa um modelo institucional que combina autonomia das cooperativas singulares com os ganhos de escala provenientes da atuação sistêmica".

O Sicoob opera sob uma estrutura organizacional composta por cooperativas singulares, cooperativas centrais e uma confederação, além de contar com um banco cooperativo próprio, o Banco Sicoob, que amplia sua capacidade de atuação no mercado financeiro (SICOOB, 2023) Adicionalmente, o Sicoob possui uma Seguradora, uma Administradora de Consórcio, uma Distribuidora de Títulos e Valores Mobiliários (DTVM), uma Fundação de Previdência, uma processadora de Pagamentos e Recebimentos e um Instituto. Esta estrutura permite que o sistema ofereça aos seus cooperados uma ampla gama de produtos e serviços financeiros competitivos, comparáveis aos disponibilizados pelos bancos tradicionais, e atue fortemente nas comunidades em seu propósito de "promover a justiça financeira e prosperidade".

Conforme destacado por Pinheiro (2008, p. 7), "o crescimento e a consolidação do Sicoob no cenário nacional refletem a maturidade e a eficiência do cooperativismo financeiro brasileiro". Este crescimento, porém, traz novos desafios, especialmente no que tange à segurança das operações e à proteção dos dados dos cooperados, tornando crucial o investimento em tecnologias de segurança cibernética.

2.3 Segurança Cibernética em Instituições Financeiras

A Segurança Cibernética pode ser definida como o conjunto de tecnologias, processos e práticas projetados para proteger redes, dispositivos, programas e dados contra-ataques, danos ou acessos não autorizados.

Na perspectiva das instituições financeiras, o que naturalmente inclui o Sicoob como sistema cooperativo financeiro, a segurança cibernética assume um papel ainda mais crítico, pois envolve a proteção de ativos financeiros e dados sensíveis de clientes em um cenário de ameaças cada vez mais sofisticadas. No passado, visto como uma função meramente operacional, a perspectiva da segurança cibernética dentro das instituições tem assumido

papéis estratégicos essenciais interligados à reputação, perdas financeiras e consequências legais (Akhtar *et al.*, 2021; Darem *et al.*, 2023). Este entendimento fortalece a visão de que, no contexto contemporâneo de transformação digital acelerada, investimentos em segurança cibernética representam não apenas uma medida preventiva contra prejuízos financeiros e reputacionais, mas um diferencial competitivo fundamental para a sustentabilidade das organizações financeiras.

A segurança cibernética tornou-se um componente crítico na estratégia operacional das instituições financeiras, especialmente com a aceleração da digitalização dos serviços bancários. O Banco Central do Brasil, por meio da Resolução nº 4.658/2018, estabeleceu a obrigatoriedade da implementação de uma política de segurança cibernética para todas as instituições financeiras, reconhecendo a criticidade desse tema (BCB, 2018). Esta regulamentação reflete a preocupação crescente com os ataques cibernéticos, que se tornaram mais sofisticados e frequentes.

A sensibilidade dos dados, legislações/regulamentações a serem cumpridas e o elevado valor financeiro envolvido nas transações gerenciadas tornam único e desafiador o papel da segurança cibernética nas instituições financeiras (Darem *et al.*, 2023). Esta realidade é ainda mais relevante para as cooperativas financeiras, cujo modelo de negócio baseia-se na confiança mútua entre a instituição e seus cooperados.

Sendo assim, este cenário evidencia também para o cooperativismo a necessidade de investimentos contínuos em tecnologias de segurança, como a biometria facial, para mitigar riscos e proteger tanto a instituição financeira cooperativa quanto seus cooperados.

2.4 Biometria Facial

A biometria facial é uma tecnologia promissora no contexto da segurança cibernética, oferecendo um método de autenticação mais seguro e conveniente em comparação aos métodos tradicionais. Segundo Jain, Ross e Nandakumar (2011, p. 18), "a biometria facial utiliza características únicas da face humana para estabelecer a identidade de um indivíduo, oferecendo um nível adicional de segurança que é difícil de ser fraudado".

No setor financeiro, a implementação de sistemas de reconhecimento facial tem se mostrado eficaz na prevenção de fraudes e na melhoria da experiência do usuário, o que está completamente alinhado com a visão do Sicoob em oferecer a melhor experiência financeira aos seus cooperados. Conforme destacam Silva e Santos (2023, p. 125), "a adoção de tecnologias biométricas representa um avanço significativo na segurança e na usabilidade dos serviços bancários, reduzindo substancialmente os riscos de fraude".

Conforme destacado por Liu e Silverman (2018, p. 56), "a biometria facial representa um avanço significativo na segurança das transações financeiras digitais, pois vincula a operação diretamente à identidade física do usuário". No entanto, apesar de

seus benefícios, os sistemas de biometria facial enfrentam desafios, como a detecção de tentativas de *spoofing* usando fotografias, vídeos ou máscaras. Além desses desafios mais tradicionalmente conhecidos e, portanto, mais comumente combatidos pelas inúmeras soluções de mercado, os fraudadores têm investido e trabalhado para alterar as fotos dos cadastros das instituições financeiras utilizadas em seus sistemas de biometria facial. Este aspecto é particularmente relevante no contexto deste trabalho, que busca identificar cadastros com fotos potencialmente fraudulentas.

2.5 Métodos/Técnicas da Literatura

O estudo proposto por esse trabalho abrange a aplicação e avaliação de diferentes técnicas para detecção de similaridade e duplicação de imagens.

2.5.1 Mean Squared Error

Uma das mais importantes e difundidas técnicas no contexto de processamento de imagens é a métrica do Erro Quadrático Médio (MSE - Mean Squared Error) (Wang; Bovik, 2009; Sara; Akter; Uddin, 2019). Ao longo das últimas décadas, o MSE tem sido amplamente aplicado na resolução de problemas relativos ao processamento de imagens/sinais, como por exemplo, redução de ruído, correspondência, classificação, detecção e reconhecimento de objetos (Cha, 2007).

Por definição, o MSE é o somatório do quadrado das diferenças de cada ponto/pixel da imagem original e a imagem comparada, dividido pela multiplicação da dimensionalidade da imagem, conforme descrito na equação 2.1:

$$f_{mse}(X,Y) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(X(i,j) - Y(i,j))^2 \tag{2.1}$$

onde m representa o comprimento da imagem, n representa a largura da imagem, $X(i,j)$ descreve um ponto da imagem original (X) e $Y(i,j)$ descreve um ponto da imagem comparada (Y). Assim, quanto menor o valor resultante da Equação 2.1, mais a imagem comparada se aproxima da imagem original.

Na perspectiva de reconhecimento de similaridade entre imagens, comumente utiliza-se um limiar (*threshold*), Δ_{mse}, para nortear a classificação de similaridade ou não entre duas imagens. Caso $f_{mse}(X,Y) \leq \Delta_{mse}$, conclui-se que foi detectada similaridade significativa entre as imagens X e Y; caso contrário, é considerado que não foi identificada similaridade.

O MSE se destaca pela sua simplicidade e facilidade de implementação. Como premissa, essa técnica requer que as imagens avaliadas apresentem as mesmas dimensões e espaço de cores. Seu processamento par a par apresenta baixo custo computacional por

par de imagens comparadas. Por outro lado, substancial tempo de processamento pode ser necessário para a comparação par a par de grandes volumes de imagens. Além disto, o MSE não é capaz de captar distorções estruturais nas imagens, sendo também, sensíveis a pequenas mudanças, como por exemplo, deslocamento, rotações e ruídos.

2.5.2 Structural Similarity Index

Proposta por Wang *et al.* (2004), a métrica de Índice de Similaridade Estrutural (SSIM - Structural Similarity Index) foi apresentada como uma forma alternativa às técnicas que visam mensurar erros entre duas imagens usando propriedades conhecidas do sistema visual humano, como por exemplo, a técnica MSE. A técnica SSIM parte do pressuposto de que a percepção visual humana é capaz de extrair informações estruturais de uma imagem. Assim, essa medida busca mensurar a distorção estrutural para quantificar a similaridade de duas imagens compondo a avaliação de luminância, contraste e coeficiente de correlação, conforme descrito na Equação 2.2 abaixo:

$$f_{ssim}(X, Y) = l(X, Y) \cdot c(X, Y) \cdot s(X, Y) \tag{2.2}$$

onde X representa a imagem original, Y representa a imagem comparada, $l(X, Y)$ descreve a função que compara a luminância entre as imagens, $c(X, Y)$ representa a função que compara o contraste entre as imagens e a função $s(X, Y)$ representa o cálculo do coeficiente de correlação de estrutura das imagens. Para mais detalhes sobre o cálculo da métrica SSIM e as respectivas definições das funções que a compõem, vide (Wang *et al.*, 2004; Wang; Bovik, 2009).

A resultante da métrica SSIM é limitada ao seguinte intervalo: $-1 < f_{ssim}(X, Y) \leq 1$, onde $f_{ssim}(X, Y) = 1$, se e somente se, $X = Y$. É importante destacar que esse índice é dito simétrico, $f_{ssim}(X, Y) = f_{ssim}(Y, X)$, ou seja, duas imagens comparadas resultam no mesmo valor de índice independentemente de sua ordenação. Portanto, se $f_{ssim}(X, Y)$ for maior ou igual a um limiar (*threshold*), $f_{ssim}(X, Y) \geq \Delta_{ssim}$, conclui-se que foi detectada similaridade significativa entre as imagens X e Y; caso contrário, é considerado que não foi identificada similaridade.

Similarmente à técnica MSE, o cálculo do índice SSIM é realizado par a par sendo premissa a mesma dimensionalidade para as imagens comparadas. A métrica SSIM destaca-se também pela sua simplicidade, facilidade de implementação e disponibilização em bibliotecas/frameworks de diversas linguagens de programação. Apesar do baixo custo computacional para o processamento de um par de imagens, a sua aplicação em larga escala pode requerer tempos de processamento proibitivos em aplicações do mundo real.

Por fim, embora o SSIM vise extrair e avaliar mudanças estruturantes entre imagens comparadas, essa técnica pode não capturar bem diferenças perceptuais globais, como, por

exemplo, objetos repetidos. Além do mais, pode ser sensível também a pequenas mudanças como deslocamentos e rotações.

2.5.3 SIFT - Scale-Invariant Feature Transform

O Scale-Invariant Feature Transform (SIFT) é um relevante algoritmo de visão computacional para reconhecimento de objetos inicialmente apresentado em 1999 por (Lowe, 1999). Posteriormente, o próprio autor, David G. Lowe, desenvolveu uma versão aprofundada de seu trabalho anteriormente apresentado implementando uma série de melhorias de estabilidade e invariância (Lowe, 2004). Um dos principais destaques da técnica SIFT é sua capacidade de extração de características independentes da escala, translação e rotação da imagem. Segundo Lowe (1999), sua extração pode ser até mesmo parcialmente invariante a mudanças de iluminação.

De uma forma geral, essa técnica funciona em quatro estágios. No primeiro estágio, são identificados, por meio de diferenças Gaussianas, potenciais pontos de interesse (também denominados de ponto-chave, *keypoints*) que são invariantes a escala e orientação. No segundo estágio, a partir do conjunto de candidatos, pontos de interesse identificados no estágio anterior, esses pontos são selecionados baseados em suas métricas de estabilidade. Em seguida, são atribuídos aos pontos de interesse selecionados uma ou mais orientações com base nas direções do gradiente da imagem local. Por fim, no quarto estágio, descritores dos pontos de interesse são criados com base em sua vizinhança. Detalhes e definições matemáticas sobre estas etapas podem ser consultados em (Lowe, 2004).

No contexto de detecção de similaridade entre duas imagens, uma etapa adicional é necessária para identificação de correspondência de pontos de interesse das imagens. Esses pontos de interesse correspondentes são identificados por meio de seus vizinhos mais próximos. Desta forma, uma integração entre o SIFT e a biblioteca Fast Library for Approximate Nearest Neighbors (FLANN) é uma importante estratégia para encontrar vizinhos aproximados. Isto porque o FLANN cria uma estrutura de índice que permite encontrar vizinhos aproximados de forma eficiente.

O SIFT é popularmente conhecido por sua robustez a mudanças moderadas de iluminação, ruído e oclusões, bem como, sua característica de invariância a escala e rotação. No entanto, essa técnica ainda pode ser sensível a transformações severas da imagem. Complementarmente, destaca-se o custo computacional elevado para cada par de imagens comparadas e a possibilidade de geração de pontos de interesse irrelevantes.

2.5.4 ORB - Oriented FAST and Rotated BRIEF

Proposto como uma alternativa computacionalmente mais eficiente que a técnica SIFT, o método Oriented FAST and Rotated BRIEF (ORB) se apresenta como um descritor binário invariante à rotação e resistente ao ruído (Rublee *et al.*, 2011). O ORB foi

proposto como uma combinação do método de detecção de pontos de interesse (*keypoints*), FAST (Rosten; Drummond, 2006), e o descritor BRIEF (Calonder *et al.*, 2010).

Essa técnica funciona, primeiramente, analisando a intensidade dos pixels, em um formato circular ao redor de um pixel candidato, para identificação de candidatos a pontos de interesse. Um pixel é sinalizado como ponto de interesse, caso um segmento contíguo de pixels no círculo seja significativamente mais claro ou mais escuro que o pixel central. Em seguida, o ORB realiza um refinamento dos pontos de interesse sinalizados na etapa anterior por meio da aplicação do algoritmo de canto Harris. A orientação de cada ponto de interesse é computada, sendo essa orientação repassada ao descritor BRIEF. Para correspondência de pontos de interesse, essa técnica geralmente aplica a distância de Hamming (Whitelaw, 1978) para comparar os descritores binários. Vide Rublee *et al.* (2011), para maiores detalhes sobre o funcionamento do ORB.

A técnica ORB tem por principal característica ser uma opção mais eficiente que o algoritmo SIFT, mantendo os atributos de invariância à escala e à rotação de seu par. Da mesma maneira, essa técnica pode apresentar sensibilidade a transformações severas das imagens avaliadas e geração de muitos pontos de interesse irrelevantes. No processo de detecção de similaridade entre imagens, o processamento par a par de imagens tende a apresentar baixo custo computacional, contudo, sua aplicação em larga escala pode não ser recomendável.

2.5.5 PCA-SIFT - Principal Components Analysis - Scale-Invariant Feature Transform

Proposto também com uma versão aprimorada da técnica SIFT (Lowe, 1999), o algoritmo Principal Components Analysis - Scale-Invariant Feature (PCA-SIFT) foi apresentado em 2004 por Ke e Sukthankar (2004). Os resultados iniciais apresentados por seus autores descrevem o PCA-SIFT como uma opção ao seu predecessor mais eficiente e precisa.

Um dos maiores desafios dos métodos de processamento e análise de imagens é aprimorar o processo de identificação de características distintivas. De maneira geral, os descritores exigem alta dimensionalidade para representar adequadamente os objetos/pontos de interesse, exigindo maior consumo de memória e processamento, como por exemplo, o algoritmo SIFT. O PCA-SIFT se propõe a lidar com esse desafio aplicando análise de componentes principais (PCA) visando reduzir a dimensionalidade, ao mesmo tempo que preserva suas informações descritivas.

Comparativamente, o estudo apresentado por Ke e Sukthankar (2004) descreve o PCA-SIFT com similar robustez à invariância de escala e rotação. Assim como seu predecessor, SIFT, essa técnica pode apresentar sensibilidade a transformações mais severas nas imagens. O custo computacionalmente mais eficiente é outra característica relevante do PCA-SIFT. Todavia, sua aplicação em larga escala pode ainda ser considerada inviável.

2.5.6 Intersecção de Histogramas

Amplamente estudada em técnicas de recuperação de imagens e reconhecimento de objetos, a cor é considerada um atributo relevante de uma imagem (Lee; Xin; Westland, 2005). Nesta perspectiva, estudos descrevem a aplicação eficiente da intersecção de histogramas no contexto de banco de dados de busca de imagens (Niblack *et al.*, 1993; Bach *et al.*, 1996) e indexação de imagens coloridas (Finlayson; Chatterjee; Funt, 1996).

Em linhas gerais, essa técnica mensura o indicador de semelhança entre duas imagens por meio da intersecção dos histogramas de cores de suas respectivas imagens. Dado um espaço discretizado de d cores, esse processo pode ser descrito em três etapas. Primeiro, computa-se o histograma de cada imagem, considerando a frequência de cores. Na segunda etapa é realizada a comparação do nível de intersecção entre os histogramas gerados no passo anterior. A última etapa avalia o indicador gerado na etapa anterior.

De acordo com Lee, Xin e Westland (2005), a intersecção de histogramas, dado duas imagens X e Y, é descrito pela seguinte equação:

$$f_{Hist}(X,Y) = f'_{Hist}(X) \cap f'_{Hist}(Y) = \sum_{i=1}^{d} min(f'_{Hist_i}(X), f'_{Hist_i}(Y)) \tag{2.3}$$

onde d representa o número de partições ou discretização das cores. Quanto maior o valor de $f_{Hist}(X,Y)$, mais semelhantes são consideradas as duas imagens. Portanto, indica-se que duas imagens, X e Y, apresentam significativa similaridade, caso $f_{Hist}(X,Y)$ seja maior ou igual a um limiar (*threshold*), Δ_{Hist}. Em contrapartida, é considerado que não foi identificada similaridade entre as imagens avaliadas.

A simplicidade da técnica de intersecção de histogramas contrasta com a não consideração de características importantes estruturais e espaciais de uma imagem, como objetos, formas e texturas. Tal perspectiva pode gerar falsos positivos em imagens que apresentem a mesma distribuição de cores, mas que descrevam conteúdos distintos. Por fim, seu processamento par a par é considerado de baixo custo, entretanto, a sua aplicação em larga escala pode ser impraticável.

2.5.7 Hash Perceptivo

Em um mundo onde os conteúdos digitais têm sido amplamente difundidos, o surgimento do Hash Perceptivo (Perceptual Hashing) contrasta com a necessidade de detecção de violação de direitos autorais e perícia digital destes materiais (Samanta; Jain, 2021; Hao *et al.*, 2021). Uma vez que pequenas variações no conteúdo de arquivos multimídia possam resultar em valores discrepantes de saída gerados por tradicionais funções Hash, como MD5 (Rivest, 1992), SHA-256 e SHA-512 (Tilborg; Jajodia, 2011), a técnica Hash Perceptivo foi alternativamente proposta.

Os valores de hash gerados por essas técnicas são como uma assinatura que representa o conteúdo digital. Deste modo, técnicas de Hash Perceptivo têm como objetivo a geração de valores de hash que mantenham ao máximo preservadas suas assinaturas, mesmo quando são realizadas alterações que não modifiquem o conteúdo da imagem. Segundo Samanta e Jain (2021), algoritmos de Hash Perceptivo usualmente são divididos em quatro etapas: pré-processamento; extração de características; quantização e geração de hash; e avaliação de similaridade.

Conforme apresentado na Equação 2.4, a avaliação de similaridade entre duas imagens, X e Y, é representada pela função $f_{Hash}(X, Y)$, sendo esta função o resultado da distância de Hamming (Whitelaw, 1978) entre os valores de hash gerados para cada uma das imagens avaliadas.

$$f_{Hash}(X, Y) =\mid f'_{Hash}(X) - f'_{Hash}(Y) \mid \tag{2.4}$$

Sendo assim, é considerado que duas imagens apresentam significativa similaridade, caso $f_{Hash}(X, Y)$ seja menor que o limiar (*threshold*) definido por Δ_{Hash}, ou seja, $f_{Hash}(X, Y) < \Delta_{Hash}$. Por outro lado, $f_{Hash}(X, Y) \geq \Delta_{Hash}$ é considerado que não há similaridade entre as imagens. Caso $X = Y$, $f_{Hash}(X, Y) = 0$.

Ao longo das últimas décadas, diversos estudos relacionados às técnicas de Hash Perceptivo têm sido conduzidos e apresentados na literatura. Na Tabela 1 apresentamos algumas dessas abordagens consideradas neste estudo. Os dados apresentados na tabela descrevem as principais informações destas abordagens de Hash Perceptivo estudadas, como nomenclatura, características e referências bibliográficas.

Tabela 1 – Algoritmos de Hash Perceptivo.

	Sigla	Características	Referência
Average Hash	aHash	Baseado no valor médio dos pixels.	(Yang; Gu; Niu, 2006)
Difference Hash	dHash	Baseado na mudança do gradiente de cor entre pixels adjacentes.	(Ibrahin; Khalifa; Ahmed, 2020)
Perceptual Hash	pHash	Baseado na Transformada Discreta de Cosseno (DCT).	(Zauner; Steinebach; Hermann, 2011)
Wavelet Hash	wHash	Baseado na Transformação Wavelet Discreta (DWT).	(Huang; Troia; Stamp, 2018)

De maneira geral, as técnicas de Hash Perceptivo têm apresentado promissores resultados em suas mais diversas aplicabilidades. Ainda assim, essa técnica pode apresentar sensibilidade a transformações mais severas das imagens, tais como rotações, cortes, grandes alterações de brilho e contraste. Adicionalmente, essas técnicas tendem a apresentar um custo computacional moderado para avaliação de um par de imagens.

2.5.8 LLM - Large Language Model

Nos últimos anos, Large Language Models (LLMs) têm despertado significativo interesse acadêmico e dos mais diversos setores industriais/econômicos (Chang *et al.*, 2024; Raiaan *et al.*, 2024; Liu *et al.*, 2025). Treinados sob enormes volumes de dados, os LLMs são considerados uma classe de modelos de aprendizado profundo de máquina capazes de resolver as mais diversas tarefas.

Comumente mencionado como uma inteligência artificial generativa, LLMs podem ser instruídos a produzir textos, códigos, reconhecer e gerar imagens, vídeos e outros tipos de conteúdos. Por meio de *prompts* (conjunto de instruções em linguagem natural) e contextos fornecidos, os LLMs suportam que usuários solicitem respostas contextualmente pertinentes e coerentes. Complementarmente, LLMs capazes de lidar com o processamento de texto, imagem, áudio e vídeo são conhecidos como modelos multimodais.

Ainda que a avaliação de imagens para detecção de similaridade não seja uma das maiores valências dos LLMs, esses modelos multimodais também suportam este tipo de tarefa. Este processo pode ser conduzido enviando aos LLMs o par de imagens concatenadas e o *prompt* com as instruções para avaliação das imagens, conforme ilustrado na Figura 1.

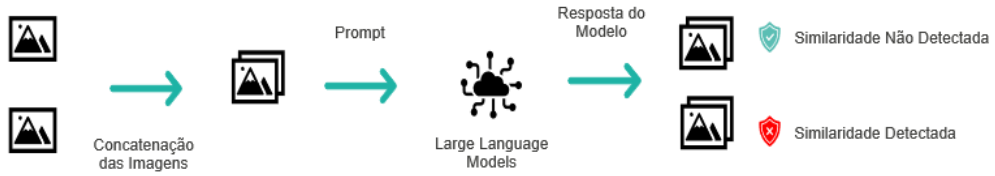

Figura 1 – Processo de Avaliação de Similaridade de Imagens Utilizando LLMs.

Fonte: Elaborado pelo autor (2025)

Dado o grande número de LLMs e plataformas que fornecem acesso aos mais variados modelos, o uso deste tipo de abordagem para detecção de similaridade de imagens é uma fácil e ágil estratégia. Naturalmente, o *prompt* fornecido pode exercer grande influência na resposta obtida. Por outro lado, este tipo de estratégia pode demandar alto custo computacional, principalmente quando aplicado em larga escala.

2.5.9 Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNN - Convolutional Neural Network) são também reconhecidas abordagens de aprendizado profundo de máquinas extensamente difundidas na literatura (Li *et al.*, 2022; Gu *et al.*, 2018). O sucesso das CNNs pode ser explicado por seus relevantes resultados e aplicabilidade nas mais diversas áreas, como por exemplo,

medicina (Yamashita *et al.*, 2018), detecção de objetos (Dhillon; Verma, 2020), agricultura (Kamilaris; Prenafeta-Boldú, 2018), dentre outras áreas.

Em linhas gerais, as CNNs funcionam utilizando operações matemáticas de convolução para identificação de padrões locais e características. O funcionamento das CNNs favorece a sua aplicação em tarefas de classificação e reconhecimento de objetos. A arquitetura deste tipo de rede neural artificial é usualmente composta por três tipos principais de camadas: camada convolucional, camada de agrupamento (*pooling*) e camada completamente conectada. A título de ilustração, as primeiras camadas de uma CNN concentram-se no reconhecimento de características mais simples (cores, bordas, dentre outras). À medida que o processamento avança para as demandas camadas, são reconhecidos elementos ou formas mais elaborados. Ao final, identifica-se o objeto procurado.

No que se refere a aplicação de CNNs para detecção de similaridade entre duas imagens, X e Y, esta tarefa por ser descrita conforme a formulação a seguir:

$$f_{cnn}(X, Y) = 1 - cosseno(\vec{x}, \vec{y}) \tag{2.5}$$

onde a função $cosseno(\vec{x}, \vec{y})$ descreve o cálculo da distância cosseno (Estrada, 2024) entre os vetores, $\vec{x}$ e $\vec{y}$; estes vetores, $\vec{x}$ e $\vec{y}$, representam as características extraídas pelas CNNs das imagens, X e Y, respectivamente. Assim, os vetores $\vec{x}$ e $\vec{y}$ podem ser formalmente definidos segundo as Equações 2.6 e 2.7 descritas abaixo:

$$\vec{x} = f^{'}_{cnn}(X) \tag{2.6}$$

$$\vec{y} = f^{'}_{cnn}(Y) \tag{2.7}$$

Deste modo, quando as imagens de entrada, X e Y, são idênticas, o resultado de $f_{cnn}(X, Y)$ é igual a 1. Portanto, para a detecção de imagens similares, é comum a aplicação de um limiar (*threshold*) definido por Δ_{cnn}. Para os casos em que $f_{cnn}(X, Y) \geq \Delta_{cnn}$, as imagens comparadas são classificadas como similares. Caso contrário, é dito que não há similaridade significativa entre as imagens avaliadas.

Diversos modelos de CNNs pré-treinados podem ser encontrados em bibliotecas e frameworks. Neste estudo, consideramos a avaliação de um conjunto desses relevantes modelos pré-treinados. Particularmente, cada modelo possui uma quantidade de parâmetros, treinamento e custo computacional distintos que podem influenciar diretamente nos resultados obtidos em cada modelo. Assim, a Tabela 2 apresenta brevemente algumas informações dos modelos considerados neste trabalho.

Tabela 2 – Detalhes dos Modelos Pré-treinados de Redes Neurais Convolucionais.

Modelo	Tamanho	Parâmetros	Referência
ConvNeXtBase	338 MB	88,5 M	(Liu *et al.*, 2022)
DenseNet201	80 MB	20,2 M	(Huang *et al.*, 2018)
EfficientNetB7	256 MB	66,7 M	(Tan; Le, 2020)
InceptionV3	92 MB	23,9 M	(Szegedy *et al.*, 2015)
MobileNet	16 MB	4,3 M	(Howard *et al.*, 2017)
ResNet50	98 MB	25,6 M	(He *et al.*, 2015)
ResNet101	171 MB	44,7 M	(He *et al.*, 2015)
ResNet152	232 MB	60,4 M	(He *et al.*, 2015)
VGG16	528 MB	138,4 M	(Simonyan; Zisserman, 2015)
VGG19	549 MB	143,7 M	(Simonyan; Zisserman, 2015)
Xception	88 MB	22,9 M	(Chollet, 2017)

As CNNs apresentam boa capacidade de generalização e robustez a variações de cor, escala e posição (Fawzi; Moosavi-Dezfooli; Frossard, 2017; Li *et al.*, 2022; Zhao *et al.*, 2024). Modelos pré-treinados e disponíveis podem ser muito úteis economizando extenso tempo de treinamento. Todavia, as CNNs podem demandar um custo computacional alto e diferenças finas entre imagens podem não ser detectáveis por esta técnica (Khamaiseh *et al.*, 2022; Rangel *et al.*, 2024).

3 TRABALHOS RELACIONADOS

A aplicação de técnicas avançadas de inteligência artificial e aprendizado de máquina tem potencializado a eficácia dos sistemas de biometria facial. Trabalhos relacionados têm demonstrado que os algoritmos *deep learning*, particularmente as redes neurais convolucionais (CNNs), revolucionaram o campo do reconhecimento facial, alcançando relevantes níveis de precisão (Yadav *et al.*, 2019; Dakhil; Abdulazeez, 2024). Estas tecnologias oferecem ferramentas poderosas para a detecção de anomalias em imagens faciais, contribuindo significativamente para a identificação de possíveis fraudes.

Dentro desse contexto, o artigo *Biometric Authentication Systems in Banking: A Technical Evaluation of Security Measures*, (Syed *et al.*, 2024), analisa a implementação de sistemas biométricos de autenticação no setor bancário, destacando sua relevância crescente como alternativa mais segura aos métodos tradicionais como senhas e PINs. Segundo Syed *et al.* (2024), esses sistemas empregam características físicas ou comportamentais únicas — como impressões digitais, reconhecimento facial, voz e padrões de digitação — para verificar identidades e proteger transações financeiras. O estudo avalia tecnicamente os processos de autenticação biométrica, enfatizando medidas de segurança como criptografia avançada para proteger dados biométricos e detecção de vivacidade para prevenir ataques de falsificação. Os autores argumentam que a autenticação biométrica oferece maior precisão (com taxas de erro tão baixas quanto 0,0001%) e conveniência para os usuários, eliminando a necessidade de memorizar múltiplas senhas. Contudo, o estudo também aponta desafios significativos, incluindo preocupações com privacidade, variações na precisão devido a fatores ambientais e altos custos de implementação para instituições menores. A pesquisa conclui que, apesar desses desafios, os sistemas biométricos demonstram resultados promissores em termos de segurança bancária, oferecendo proteção robusta contra fraudes e furto de identidade, embora sua implementação bem-sucedida exija infraestrutura tecnológica adequada e atenção cuidadosa às preocupações de privacidade dos clientes.

Ainda considerando o contexto bancário/financeiro, o artigo *Biometrics for banking: best practices and barriers to adoption*, de Alan Goode, publicado na revista *Biometric Technology Today* em 2018, analisa a crescente adoção de tecnologias biométricas no setor bancário. Conforme destaca Goode (2018, p. 5), o ambiente bancário evoluiu rapidamente nos últimos anos, impulsionado por eventos digitais significativos como a "Revolução do *Open Banking*", que forçou os bancos a abrirem suas infraestruturas digitais para terceiros. Apesar de não estar citado no artigo, por ter sido evento ocorrido após a sua publicação, a pandemia causada pelo COVID-19 acelerou o processo de transformação digital das instituições financeiras. Nesse contexto, a biometria estabeleceu-se como uma ferramenta fundamental para a verificação de identidade, autenticação e gestão de fraudes. Entre

as tendências que impulsionam essa adoção, destacam-se: o crescimento da autenticação biométrica multimodal em dispositivos móveis, o surgimento de cartões bancários biométricos, a adoção de plataformas biométricas únicas para diversos canais bancários, e a integração com sistemas nacionais de identificação. O autor também discute barreiras importantes para a adoção, como "falhas na detecção de falsificações e prova de vida" (Goode, 2018), preocupações com a privacidade e o gerenciamento do ciclo de vida dos sistemas biométricos.

No que tange à identificação de imagens com indício de fraude por duplicidade, o artigo de Morra e Lamberti (2019) discute a detecção não supervisionada de imagens quase duplicadas, uma técnica essencial para aplicações como análise de mídia social, recuperação de imagens em grande escala e forense digital. Os autores analisam a eficácia de descritores baseados em redes neurais convolucionais profundas em comparação com métodos tradicionais, enfatizando a importância da especificidade, dada a taxa crescente de alarmes falsos em grandes bases de dados. Entre os achados, destaca-se que a adaptação de redes convolucionais oferece um desempenho superior à utilização de características pré-extraídas, mas as diferenças se tornam menores quando são requeridos altos níveis de especificidade. O estudo propõe ainda uma metodologia baseada em análise ROC e mineração de negativos difíceis para caracterizar a especificidade dos descritores. Essa abordagem se mostra relevante para avaliar a viabilidade da detecção não supervisionada em contextos reais, como prevenção de fraudes e rastreamento de conteúdo online. Contudo, um ponto de debate reside na generalização dos resultados para outros conjuntos de dados, uma vez que a dificuldade da detecção pode variar conforme as transformações aplicadas às imagens quase duplicadas.

Adicionalmente, o artigo de Thyagharajan e Kalaiarasi (2020) realiza uma revisão sobre técnicas de detecção de imagens quase duplicadas utilizando visão computacional, enfatizando a importância desse processo para melhorar a precisão dos motores de busca e evitar redundância em bases de dados digitais. A pesquisa destaca que imagens quase duplicadas podem surgir de modificações em imagens originais, como alterações na iluminação, rotação, recorte e compressão. Os autores discutem diferentes métodos de extração de características, classificando-os em baseados em pontos-chave, pixels e áreas, além de explorar abordagens como CNN e hashing sensível à localização. O estudo aponta desafios como a necessidade de métodos mais eficientes e escaláveis para grandes conjuntos de dados. No entanto, uma limitação do artigo é que, embora forneça uma visão abrangente das técnicas existentes, ele poderia incluir mais experimentos comparativos para avaliar o desempenho dos métodos revisados em diferentes cenários práticos. Dessa forma, a pesquisa contribui significativamente para o entendimento da detecção de imagens quase duplicadas, mas abre espaço para investigações futuras que aprimorem a eficiência e precisão das abordagens discutidas.

Não obstante à ligação dos artigos citados acima ao tema proposto nesse estudo, há uma lacuna na literatura quanto à identificação específica de cadastros com fotos possivelmente fraudulentas dentro de uma instituição financeira bancária ou cooperativa real, considerando desafios como a duplicidade de imagens. Diante disso, justifica-se o desenvolvimento do presente trabalho, o qual investigará a aplicação de métodos avançados de inteligência artificial e visão computacional na detecção de indícios de fraude em imagens de uma base cadastral já estabelecida e utilizada em processos de segurança biométrica. Esse estudo contribuirá não apenas para o aprimoramento dos sistemas de verificação de identidade, mas também para a mitigação de riscos operacionais e a conformidade com regulamentações de segurança no setor financeiro, incluindo as cooperativas.

4 METODOLOGIA

A metodologia adotada neste estudo seguiu uma abordagem experimental comparativa, dividida em duas fases principais: um estudo experimental controlado utilizando técnicas avançadas de processamento de imagens e inteligência artifical em um conjunto de dados público e uma aplicação prática das técnicas que se destacaram na primeira fase em um conjunto de dados reais de uma cooperativa financeira.

4.1 Técnicas utilizadas

O estudo proposto por esse trabalho abrangeu a aplicação e avaliação de 22 técnicas diferentes de detecção de similaridade e duplicação de imagens, organizadas em duas categorias principais.

As técnicas baseadas nas características da imagem englobaram métricas tradicionais como Mean Squared Error (MSE), Structural Similarity Index (SSIM), além de métodos de extração de características como Scale-Invariant Feature Transform (SIFT), Oriented FAST and Rotated BRIEF (ORB), PCA-SIFT e intersecção de histogramas.

A categoria de técnicas baseadas na estrutura da imagem incluiu métodos de Perceptual Hashing (Average Hash, Perception Hash, Difference Hash e Wavelet Hash), Large Language Models multimodais e Convolutional Neural Network com 11 modelos pré-treinados (VGG16, VGG19, ResNet50, ResNet101, ResNet152, InceptionV3, Xception, EfficientNetB7, MobileNet, DenseNet201 e ConvNeXtBase).

O detalhamento de cada uma dessas técnicas pode ser consultado na seção 2.5.

4.2 Formas de Avaliação

A avaliação das técnicas foi baseada na matriz de confusão, que é uma tabela usada para avaliar o desempenho de técnicas e modelos de processamento de dados, comparando previsões com os valores reais. A matriz detalha acertos e erros da técnica, mostrando onde ela "se confunde". Essa ferramenta permite uma análise aprofundada do desempenho da técnica.

A matriz é composta pelas métricas a seguir, que são o resultado da comparação entre os valores reais e as previsões da técnica.

Matriz de Confusão

- **VP**: Verdadeiro positivo
- **VN**: Verdadeiro negativo

- **FP**: Falso positivo
- **FN**: Falso negativo

A seleção adequada de métricas para avaliação de técnicas em problemas de detecção de imagens duplicadas ou quase duplicadas representa um aspecto crítico para o sucesso de sistemas de segurança, principalmente, em instituições financeiras. A escolha das métricas deve considerar as características específicas do problema, incluindo o desequilíbrio dos dados e o impacto diferenciado dos tipos de erro na operação do negócio.

No contexto do estudo desse trabalho que está relacionado à detecção de cadastros com fotos fraudulentas em uma instituição financeira cooperativa, é fundamental compreender as implicações econômicas e operacionais dos diferentes tipos de erro. Em problemas de detecção de fraude, o custo associado aos falsos negativos (não detectar uma fraude real) é substancialmente superior ao custo dos falsos positivos (sinalizar incorretamente um caso legítimo como fraudulento).

Esta característica é particularmente relevante no presente estudo, onde:

1. **Verdadeiro positivo (Fraudes detectadas):** Representam cadastros fraudulentos do sistema, podendo ser utilizados para transações financeiras ilegítimas, resultando em prejuízos diretos para a cooperativa e seus cooperados, além de potencial comprometimento da reputação institucional.

2. **Verdadeiro negativo (Cadastros sem duplicidade):** Representam cadastros corretos do sistema, considerando que não existem outros cadastros com a mesma foto facial, garantindo assim segurança ao processo de Biometria Facial para realização de transações financeiras.

3. **Falsos Positivos (Casos legítimos sinalizados):** Embora indesejáveis, resultam apenas em revisões manuais adicionais dos cadastros, sem impacto financeiro direto significativo.

4. **Falsos Negativos (Fraudes não detectadas):** Representam cadastros fraudulentos que permanecerão ativos no sistema, podendo ser utilizados para transações financeiras ilegítimas, resultando em prejuízos diretos para a cooperativa e seus cooperados, além de potencial comprometimento da reputação institucional.

O desequilíbrio sifnificativo, com predominância de casos de não-similaridade em relação aos casos de similaridade entre imagens faciais, é característico de problemas reais de detecção de fraude em bases cadastrais, onde a ocorrência de casos fraudulentos é naturalmente inferior aos casos legítimos.

Além da matriz de confusão fornecer uma visão clara dos resultados da técnica, mostrando as classificações corretas e incorretas, a partir das métricas da matriz é possível calcular outras métricas importantes para avaliação da técnica como acurácia, recall, precisão, taxa de falso positivo e F1 Score, conforme detalhamento a seguir.

Métricas:

Acurácia

Fornece uma visão geral do desempenho do classificador. Apesar das limitações em conjuntos desequilibrados, esta métrica foi mantida para permitir comparação com estudos similares e fornecer uma perspectiva complementar do desempenho global.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \tag{4.1}$$

Recall

Representa a capacidade do sistema em identificar corretamente os casos positivos (fraudulentos). Esta métrica foi priorizada devido ao alto custo associado aos falsos negativos no contexto das instituições financeiras. Em aplicações de segurança financeira, maximizar o Recall é fundamental para garantir que o maior número possível de casos fraudulentos seja detectado, mesmo que isso resulte em um aumento dos falsos positivos.

$$Recall = \frac{VP}{VP + FN} \tag{4.2}$$

Precision

Embora a precisão seja uma métrica valiosa em muitos contextos de classificação, sua relevância é reduzida no presente estudo devido ao desequilíbrio dos dados. Em conjuntos altamente desequilibrados, a Precisão pode apresentar valores artificialmente inflados, não refletindo adequadamente a capacidade real de detecção de fraudes.

$$Precision = \frac{VP}{VP + FP} \tag{4.3}$$

Taxa de Falso Positivo

Quantifica a proporção de casos legítimos incorretamente classificados como fraudulentos. Embora menos crítica que os falsos negativos, esta métrica é importante para avaliar a eficiência operacional do sistema.

$$\text{Taxa de Falso Positivo} = \frac{FP}{FP + TN} \tag{4.4}$$

F1 *Score*

O F1-*Score*, representando a média harmônica entre *Precision* e *Recall*, constitui uma métrica equilibrada que combina ambos os aspectos. Esta métrica é particularmente valiosa quando se busca um equilíbrio entre a identificação de casos positivos e a minimização de falsos positivos.

Apesar de não ter sido priorizada na seleção final devido ao contexto específico do problema, o F1-*Score* permanece como uma métrica complementar importante para uma avaliação abrangente dos modelos, especialmente quando comparado com benchmarks da literatura científica.

$$\text{F1 } Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.5}$$

Outras métricas:

Análise de Dominância de Pareto

Para a seleção final das técnicas mais adequadas, foi aplicada uma análise de dominância de Pareto considerando simultaneamente duas métricas selecionadas. Uma solução A domina uma solução B se A é pelo menos tão boa quanto B em todos os objetivos e estritamente melhor em pelo menos um objetivo.

No contexto multiobjetivo deste estudo, a fronteira de Pareto identifica as técnicas que apresentam os melhores compromissos entre Acurácia e *Recall*, conforme será demonstrado graficamente na análise de resultados no estudo das técnicas, considerando que eventual variação da Taxa de Falsos Positivos embora indesejável, resulta apenas em revisões manuais adicionais dos cadastros, sem exposição adicional às fraudes.

A aplicação desta análise permitirá a identificação objetiva das técnicas com melhor desempenho global, evitando a necessidade de atribuição arbitrária de pesos às diferentes métricas.

Tempo de Execução

Além das métricas de desempenho, foi considerado o tempo de execução como critério complementar, reconhecendo que a viabilidade prática das técnicas em ambiente produtivo depende tanto da precisão quanto da eficiência computacional.

A análise conjunta de desempenho e tempo de execução permite a identificação de soluções que oferecem o melhor compromisso entre eficácia na detecção de fraudes e viabilidade operacional para aplicação em larga escala.

5 AVALIAÇÃO EXPERIMENTAL

Na primeira fase do trabalho, que consistiu em um estudo experimental controlado, foi utilizado um conjunto de dados público "*Celebrity Face Image*" do Kaggle (Garhpati, 2022), composto por 221 fotografias de celebridades (151 homens e 70 mulheres), caracterizado por diversidade de cenas, iluminação, contraste e nitidez.

Para ampliar a base de testes e criar cenários específicos por meio de transformações controladas que permitissem a avaliação do resultado apresentado após a aplicação de cada técnica da literatura bem como do arranjo de técnicas com o objetivo de obter ganhos de performance, foram geradas artificialmente 221 novas imagens, incluindo imagens duplicadas de integral (34 imagens), com variações de brilho (75 imagens), com inversão horizontal (38 imagens) e com rotações (74 imagens), gerando ao final um dataset experimental com 442 imagens para análise das técnicas.

A Figura 2 ilustra a composição e características originais desse conjunto de dados público bem como as transformações controladas realizadas para geração das novas imagens e composição do *DataSet* final utilizado na avaliação experimental.

Figura 2 – *DataSet* Experimental.

Fonte: Elaborado pelo autor (2025)

Para facilitar e garantir a avaliação dos resultados da utilização das técnicas na base experimental, foi estabelecido um padrão de nomenclatura para cada imagem, seguindo a máscara `photo_9999_999.jpg`, conforme pode ser visualizado na Figura 3, onde:

- **photo**: Prefixo constante utilizado para todas as imagens.
- **9999**: Identificador único da imagem original da base pública Kaggle.
- **999**: Identificador da transformação realizada, sendo:

- `000`: Imagem original.
- `001`: Redução de brilho.
- `002`: Aumento de brilho.
- `003`: Inversão horizontal.
- `004`: Rotação para a esquerda.
- `005`: Rotação para a direita.

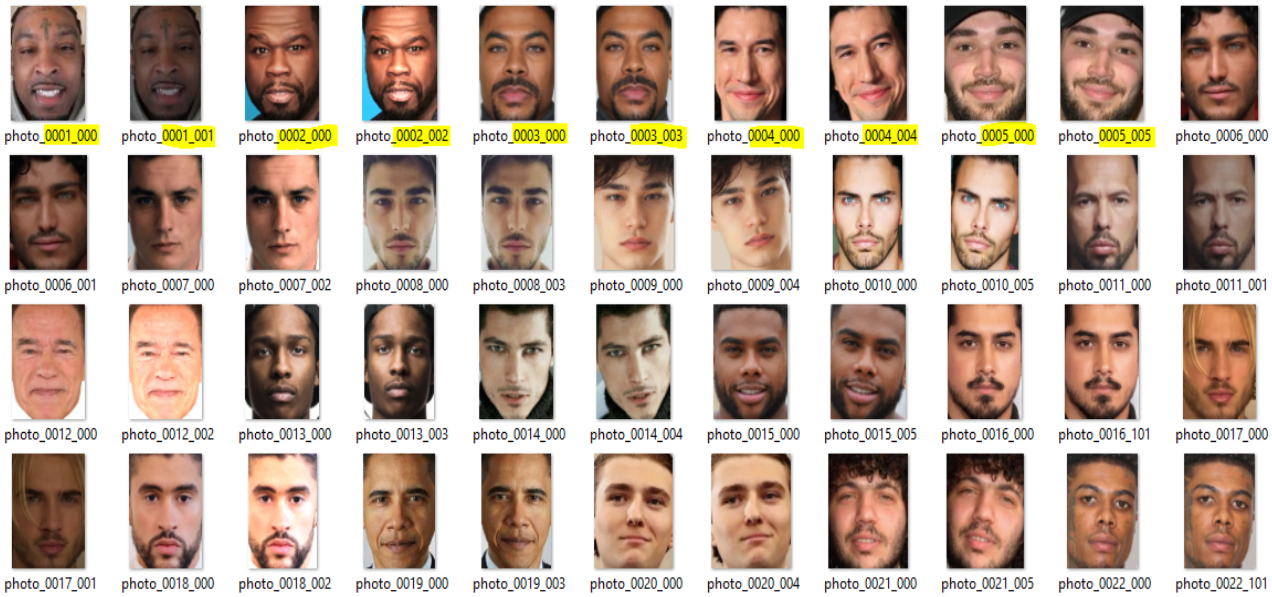

Figura 3 – Exemplos de transformações controladas no *DataSet* Experimental.

Fonte: Elaborado pelo autor (2025)

Com base nesse conjunto de imagens mencionadas, experimentos prévios foram conduzidos para configuração e seleção do parâmetro de limiar de similaridade dos métodos discutidos na Seção 2.5. Para seleção das configurações mais adequadas, foram consideradas as métricas de Acurácia (Equação 4.1), *Recall* (Equação 4.2) e Taxa de Falso Positivo (Equação 4.4). Assim, os dados apresentados na Tabela 3 descrevem os métodos, parâmetros avaliados, as seções do texto onde os parâmetros foram discutidos e os valores selecionados.

Em relação ao estudo computacional utilizando LLMs (previamente discutido na Seção 2.5.8), detalhes do *prompt* considerado nos experimentos são apresentados na Figura 4. Além disto, este estudo foi conduzido utilizando o modelo multimodal *Meta Llama 3.2 Vision*[1,2].

[1] Disponível em github.com/meta-llama/llama-models. Acessado em 18 set. 2025.
[2] Disponível em ollama.com/library/llama3.2-vision. Acessado em 18 set. 2025.

Tabela 3 – Descrição dos parâmetros utilizados para técnicas avaliadas.

Método	Parâmetro	Definição	Valor
MSE	Δ_{mse}	Seção 2.5.1	2.750
SSIM	Δ_{ssim}	Seção 2.5.2	0,4
Histograma	Δ_{Hist}	Seção 2.5.6	10
H. Perceptivo - aHash	Δ_{Hash}	Seção 2.5.7	15
H. Perceptivo - dHash	Δ_{Hash}	Seção 2.5.7	15
H. Perceptivo - pHash	Δ_{Hash}	Seção 2.5.7	20
H. Perceptivo - wHash	Δ_{Hash}	Seção 2.5.7	15
CNN - ConvNeXtBase	Δ_{cnn}	Seção 2.5.9	0,8
CNN - DenseNet201	Δ_{cnn}	Seção 2.5.9	0,9
CNN - EfficientNetB7	Δ_{cnn}	Seção 2.5.9	0,9
CNN - InceptionV3	Δ_{cnn}	Seção 2.5.9	0,9
CNN - MobileNet	Δ_{cnn}	Seção 2.5.9	0,9
CNN - ResNet50	Δ_{cnn}	Seção 2.5.9	0,9
CNN - ResNet101	Δ_{cnn}	Seção 2.5.9	0,9
CNN - ResNet152	Δ_{cnn}	Seção 2.5.9	0,9
CNN - VGG16	Δ_{cnn}	Seção 2.5.9	0,9
CNN - VGG19	Δ_{cnn}	Seção 2.5.9	0,9
CNN - Xception	Δ_{cnn}	Seção 2.5.9	0,9

```
Prompt: Existem duas pessoas na imagem, elas possuem
as mesmas características? Parecem ser a mesma pessoa?
Responda apenas **SIM** ou **NÃO**, no idioma português.
```

Figura 4 – Descrição do *prompt* utilizado para verificação de similaridade entre duas imagens. Como apresentado na Figura 1, as duas imagens são concatenadas em uma para a análise do modelo.

As técnicas foram implementadas utilizando Python e bibliotecas especializadas como OpenCV, TensorFlow, Scikit-image, ImageHash e Weaviate, garantindo reprodutibilidade e robustez nos experimentos.

5.1 Detalhamento do experimento utilizando uma das técnicas

Nesta subseção, apresenta-se, como exemplo, a avaliação experimental realizada com a técnica de Redes Neurais Convolucionais (CNN) utilizando o modelo pré-treinado VGG16, aplicado ao conjunto de dados de imagens de faces de celebridades. O objetivo foi avaliar a capacidade do modelo em identificar similaridades entre imagens, simulando a detecção de fotos com indícios de fraude em cadastros de cooperados que é o estudo proposto por esse trabalho a ser aplicado posteriormente em uma base real. A implementação utilizou a biblioteca TensorFlow para carregar o modelo VGG16, extrair características das imagens

e calcular a similaridade cosseno entre vetores de features.

O fluxo experimental aplicado à técnica CNN-VGG16 seguiu uma metodologia estruturada em cinco etapas sequenciais, conforme apresentado na Figura 5.

Inicialmente, foi realizado o carregamento do modelo VGG16 pré-treinado na base ImageNet, removendo-se as camadas fully connected para focar exclusivamente na capacidade de extração de características visuais da rede. Em seguida, as imagens do conjunto de dados foram submetidas a um processo de pré-processamento padronizado, incluindo redimensionamento para as dimensões de entrada requeridas pelo modelo (224×224 pixels) e normalização dos valores dos pixels. A terceira etapa consistiu na extração de features através da camada *block5_pool*, que produz representações vetoriais de alta dimensionalidade (25.088 características) capturando padrões complexos das imagens faciais. Para quantificar a similaridade entre pares de imagens, foi aplicada a métrica de similaridade cosseno entre os vetores de características extraídos. Na sequência, os resultados obtidos foram comparados com o *ground truth* estabelecido com base na identificação das celebridades, permitindo classificar as predições como verdadeiros ou falsos positivos. Por fim, foram computadas métricas de desempenho abrangentes, incluindo acurácia, precisão, recall, taxa de falsos positivos, F1-score e tempos de execução, proporcionando uma avaliação quantitativa completa da eficácia da técnica.

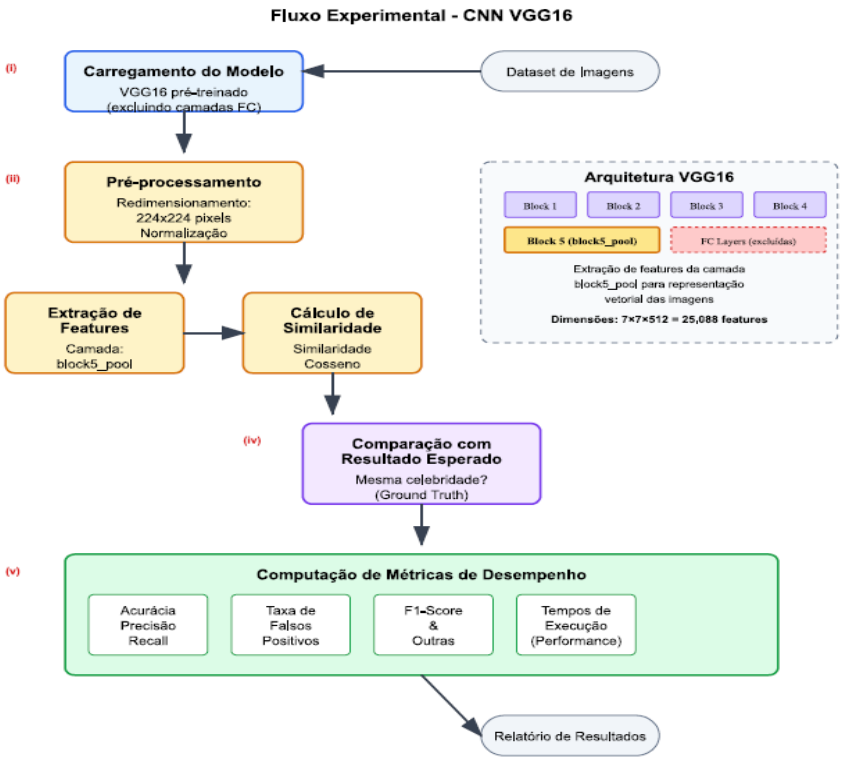

Figura 5 – Fluxo experimental aplicado na técnica CNN-VGG16 para detecção de similaridade entre imagens faciais.

Fonte: Elaborado pelo autor (2025)

O relatório final de resultados da técnica CNN-VGG16 utilizada como exemplo nessa subseção, conforme apresentado na Figura 6, consolida para cada configuração (threshold) o conjunto de métricas que poderão ser utilizadas na comparação de desempenho entre todas as técnicas que foram estudadas nessa avaliação experimental bem como indica a configuração com o melhor desempenho considerando os melhores resultados nas métricas Acurácia, F1 Score e Recall.

CNN - VGG16													
Threshold	VP	VN	FN	FP	Acurácia (%)	Precisão (%)	Recall (%)	F1 Score	Taxa de Falso Positivo (%)	Tempo Médio * (1 Par de Imagens)	Desvio Padrão * (1 Par de Imagens)	Maior Tempo * (1 Par de Imagens)	Tempo Total (Todos os pares para validação de 1 imagem)
0,5	221	479	0	48141	1,43	0,46	100	0,91	99,01	1,00	0,07	1,33	221,62
0,6	221	2551	0	46069	5,68	0,48	100	0,95	94,75	1,02	0,08	1,37	225,18
0,7	221	11014	0	37606	23,00	0,58	100	1,16	77,35	1,00	0,08	1,37	220,82
0,8	221	30259	0	18361	62,41	1,19	100	2,35	37,76	1,06	0,07	1,28	233,57
0,9***	218	47285	3	1335	97,26	14,04	98,64	24,58	2,75	1,03	0,09	1,63	228,34
* Tempo medido em segundos *** Configuração selecionada													

Figura 6 – Relatório final de resultados após aplicação da técnica CNN-VGG16 no conjunto de dados experimental.

Fonte: Elaborado pelo autor (2025)

5.2 Resultado da aplicação das técnicas estudadas

Após a aplicação de todas as técnicas no conjunto de dados experimental com a geração do relatório final de cada uma delas e a definição do melhor valor de *threshold* conforme detalhado na Tabela 3, foi iniciado o trabalho de consolidação, análise comparativa e seleção das técnicas com os melhores resultados/desempenho.

5.2.1 Resultado da aplicação de todas as técnicas estudadas baseadas em Redes Neurais

Esse trabalho foi feito em etapas considerando o agrupamento de técnicas baseadas na estrutura da imagem e selecionando as melhores técnicas com o uso de Redes Neurais e em seguida as técnicas com a utilização de Perceptual Hashing. Para conclusão da análise, as técnicas que se destacaram nesses 2 agrupamentos foram analisadas em conjunto com as demais técnicas baseadas nas características da imagem para seleção final dessa avaliação experimental.

Sendo assim, primeiramente, considerando que foram estudados 11 (onze) modelos pré-treinados de CNN, conforme ilustrado pela Tabela 2, foram consolidados em um mesmo relatório o melhor cenário para cada CNN aplicada, conforme apresentado na Figura 7.

Resultado Comparativo - Modelos CNN													
Método	VP	VN	FN	FP	Acurácia (%)	Precisão (%)	Recall (%)	F1 Score	Taxa de Falso Positivo (%)	Tempo Médio * (1 Par de Imagens)	Desvio Padrão * (1 Par de Imagens)	Maior Tempo * (1 Par de Imagens)	Tempo Total (Todos os pares para validação de 1 imagem)
VGG16	218	47285	3	1335	97,26	14,04	98,64	24,58	2,75	1,03	0,09	1,63	228,34
VGG19	214	48038	7	582	98,79	26,88	96,83	42,08	1,20	1,38	0,28	3,54	304,32
ResNet50	215	48232	6	388	99,19	35,66	97,29	52,18	0,80	0,39	0,11	2,04	86,99
ResNet101	211	47545	10	1075	97,78	16,41	95,48	28,00	2,21	0,62	0,28	4,92	137,53
ResNet152	213	48288	8	332	99,30	39,08	96,38	55,61	0,68	0,81	0,41	7,15	179,41
InceptionV3	216	48366	5	254	99,47	45,96	97,74	62,52	0,52	0,51	0,16	3,01	113,31
Xception	211	48524	10	96	99,78	68,73	95,48	79,92	0,20	0,66	0,16	2,34	145,42
EfficientNet	221	47729	0	891	98,18	19,87	100,00	33,16	1,83	3,04	0,56	11,97	670,91
MobileNet	**218**	**48595**	**3**	**25**	**99,94**	**89,71**	**98,64**	**93,97**	**0,05**	**0,27**	**0,08**	**1,50**	**59,25**
DenseNet	217	47873	4	747	98,46	22,51	98,19	36,62	1,54	0,69	0,39	6,92	151,98
ConvNextBase	**221**	**48070**	**0**	**550**	**98,87**	**28,66**	**100,00**	**44,56**	**1,13**	**3,54**	**0,65**	**9,56**	**782,65**
* Tempo medido em segundos													

Figura 7 – Relatório consolidado de resultados após aplicação de todas as técnicas CNNs no conjunto de dados experimental.

Fonte: Elaborado pelo autor (2025)

Os resultados da avaliação experimental, considerando em um primeiro momento apenas os modelos pré-treinados de Redes Neurais Convolucionais (CNNs), revelam um trade-off claro entre as métricas de acurácia e recall no conjunto de dados experimental de faces de celebridades. As melhores taxas de acurácia foram obtidas pelos modelos MobileNet (99,942%) e Xception (99,782%), respectivamente, destacando sua eficácia em classificações precisas com baixas taxas de falsos positivos (0,051% e 0,197%). Por outro lado, as melhores taxas de recall foram alcançadas pelos modelos EfficientNetB7 (100,00%) e ConvNeXtBase (100,00%), priorizando a detecção completa de similaridades, embora com taxas de falsos positivos ligeiramente mais elevadas (1,832% e 1,131%).

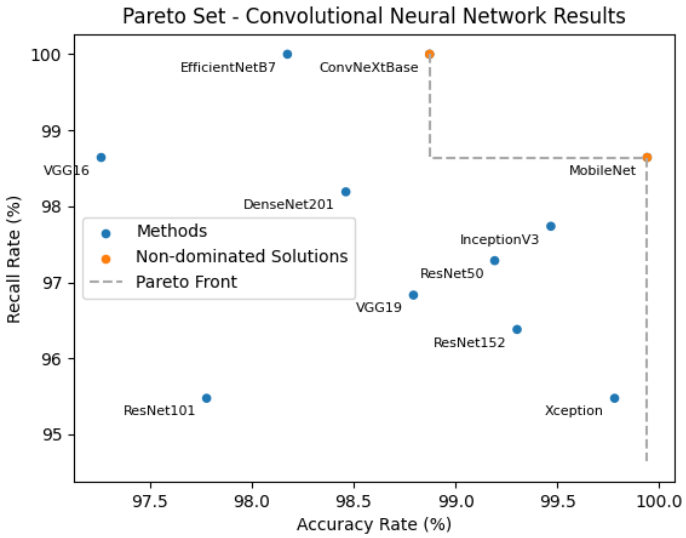

Figura 8 – Análise de dominância de Pareto para as técnicas avaliadas, considerando Acurácia vs. Recall. As soluções não-dominadas formam a fronteira de Pareto, representando os melhores compromissos entre os objetivos conflitantes.

Fonte: Elaborado pelo autor (2025)

A análise de dominância de Pareto, conforme ilustrado na Figura 8, indica que a fronteira de soluções não-dominadas é composta pelos modelos ConvNeXtBase e MobileNet, equilibrando desempenho em acurácia e recall de forma otimizada.

Os modelos apresentaram tempos médios de processamento distintos por par de imagens, com MobileNet (0,2681 segundos) e ResNet50 (0,3936 segundos) se destacando como os mais rápidos, enquanto ConvNeXtBase (3,5414 segundos) e EfficientNetB7 (3,0358 segundos) exibem os piores tempos médios, refletindo o custo computacional associado a arquiteturas mais complexas. Esses achados orientam a seleção de técnicas para detecção de fraudes em fotos de cadastros cooperativos, priorizando equilíbrio entre precisão e eficiência.

5.2.2 Resultado da aplicação de todas as técnicas estudadas baseadas em Perceptual Hashing

Seguindo o mesmo processo detalhado para o contexto das CNNs e considerando as técnicas baseadas em Perceptual Hashing, foram consolidados em um mesmo relatório o melhor cenário para cada técnica aplicada, conforme apresentado na Figura 9.

Perceptual Hashing													
Método	VP	VN	FN	FP	Acurácia (%)	Precisão (%)	Recall (%)	F1 Score	Taxa de Falso Positivo (%)	Tempo Médio * (1 Par de Imagens)	Desvio Padrão * (1 Par de Imagens)	Maior Tempo * (1 Par de Imagens)	Tempo Total (Todos os pares para validação de 1 imagem)
A. Hash	183	43761	38	4859	89,97	3,63	82,81	6,95	9,99	0,01	0,01	0,09	1,13
P. Hash	145	46648	76	1972	95,81	6,85	65,61	12,40	4,06	0,01	0,01	0,33	1,15
D. Hash	148	45454	73	3166	93,37	4,47	66,97	8,37	6,51	0,00	0,01	0,05	1,08
W. Hash	187	43614	34	5006	89,68	3,60	84,62	6,91	10,30	0,02	0,03	0,50	4,93
* Tempo medido em segundos													

Figura 9 – Relatório consolidado de resultados após aplicação de todas as técnicas baseadas em Perceptual Hashing no conjunto de dados experimental.

Fonte: Elaborado pelo autor (2025)

Os resultados do estudo comparativo das técnicas de Perceptual Hashing revelam um *trade-off* claro entre as métricas de acurácia e *recall*, destacando que nenhuma abordagem otimiza ambas simultaneamente no conjunto de dados experimental analisado. As técnicas de Perception Hashing e Difference Hashing apresentaram as melhores taxas de acurácia, respectivamente, e também demonstraram maior precisão na identificação de similaridades entre imagens, ainda que com valores muito baixos. Por outro lado, as técnicas de Wavelet Hashing e Average Hashing obtiveram as maiores taxas de recall, priorizando a captura de verdadeiros positivos mesmo em cenários de variação.

A análise da fronteira de Pareto, conforme ilustrado na Figura 10, confirma que as soluções avaliadas são não dominantes entre si, implicando que a escolha depende dos

requisitos específicos da aplicação, sem uma opção universalmente superior.

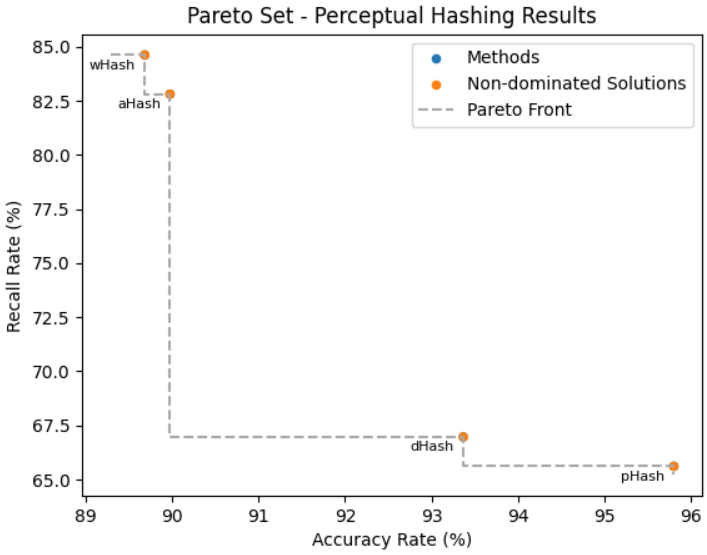

Figura 10 – Análise de dominância de Pareto para as técnicas avaliadas, considerando Acurácia vs. Recall. As soluções não-dominadas formam a fronteira de Pareto, representando os melhores compromissos entre os objetivos conflitantes.

Fonte: Elaborado pelo autor (2025)

Ademais, as técnicas exibem tempos médios de processamento semelhantes, com exceção da Wavelet Hashing, que registrou um tempo médio aproximadamente três vezes superior às demais, sugerindo considerações de eficiência computacional para implementações em larga escala.

5.2.3 Resultado da aplicação de todas as técnicas da Literatura que foram estudadas

Seguindo o mesmo processo e considerando também as técnicas baseadas na característica da imagem, foram consolidados em um mesmo relatório os melhores cenários para cada técnica aplicada, incluindo as técnicas selecionadas dos agrupamentos anteriores, conforme apresentado na Figura 11.

Resultado Comparativo - Literatura													
Método	VP	VN	FN	FP	Acurácia (%)	Precisão (%)	Recall (%)	F1 Score	Taxa de Falso Positivo (%)	Tempo Médio * (1 Par de Imagens)	Desvio Padrão * (1 Par de Imagens)	Maior Tempo * (1 Par de Imagens)	Tempo Total (Todos os pares para validação de 1 imagem)
MSE	169	43772	52	4848	89,97	3,37	76,47	6,45	9,97	0,01	0,01	0,16	1,26
SSIM	168	42812	53	5808	88,00	2,81	76,02	5,42	11,95	0,01	0,02	0,29	2,70
SIFT	204	47546	17	1074	97,77	15,96	92,31	27,22	2,21	0,10	0,08	0,90	21,17
ORB	204	44386	17	4234	91,30	4,60	92,31	8,76	8,71	0,01	0,01	0,20	3,14
PCA-SIFT	182	48620	39	0	99,92	100,00	82,35	90,32	0,00	0,09	0,07	0,92	20,58
Histograma	133	45143	88	3477	92,70	3,68	60,18	6,94	7,15	0,00	0,01	0,06	1,04
A. Hash	183	43761	38	4859	89,97	3,63	82,81	6,95	9,99	0,01	0,01	0,09	1,13
P. Hash	145	46648	76	1972	95,81	6,85	65,61	12,40	4,06	0,01	0,01	0,33	1,15
D. Hash	148	45454	73	3166	93,37	4,47	66,97	8,37	6,51	0,00	0,01	0,05	1,08
W. Hash	187	43614	34	5006	89,68	3,60	84,62	6,91	10,30	0,02	0,03	0,50	4,93
LLM	99	46520	122	2100	95,45	4,50	44,80	8,18	4,32	1,63	0,43	18,00	359,79
CNN - MobileNet	**218**	**48595**	**3**	**25**	**99,94**	**89,71**	**98,64**	**93,97**	**0,05**	**0,27**	**0,08**	**1,50**	**59,25**
CNN - ConvNextBase	**221**	**48070**	**0**	**550**	**98,87**	**28,66**	**100,00**	**44,56**	**1,13**	**3,54**	**0,65**	**9,56**	**782,65**
* Tempo medido em segundos													

Figura 11 – Relatório consolidado de resultados após aplicação de todas as técnicas estudadas no conjunto de dados experimental.

Fonte: Elaborado pelo autor (2025)

A análise comparativa das técnicas avaliadas no conjunto de dados experimental revelou que todas apresentaram taxas de acurácia superiores a 85%, demonstrando uma capacidade geral robusta na classificação de similaridades entre imagens. No entanto, as taxas de recall inferiores a 77% em algumas abordagens (MSE, SSIM, Intersecção de Histogramas, pHash, dHash e LLM) destacam o desafio imposto pelo desbalanceamento entre casos de imagens não similares e similares, o que pode comprometer a detecção de instâncias positivas em cenários reais. Dentre as técnicas, as CNNs baseadas em MobileNet e ConvNeXt, juntamente com SIFT e PCA-SIFT, sobressaíram nas métricas de acurácia e recall, posicionando-se como opções mais eficazes para tarefas de identificação de fraudes em fotos.

Além disso, as soluções utilizando modelos pré-treinados de CNN, especificamente MobileNet e ConvNeXt, foram identificadas como Pareto não-dominadas ao serem comparadas com as demais, conforme ilustrado nas Figuras 12 e 13, indicando um equilíbrio otimizado entre desempenho e eficiência.

Por fim, observou-se variabilidade nos tempos médios de processamento, com destaque para os valores mais elevados nas técnicas baseadas em LLM e ConvNeXt (CNN), sugerindo a necessidade de otimizações computacionais para aplicações em larga escala.

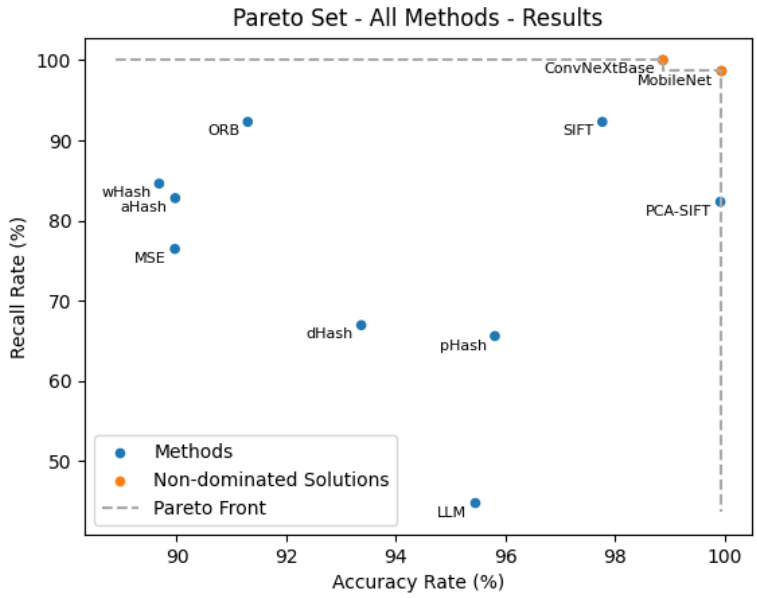

Figura 12 – Análise de dominância de Pareto para as técnicas avaliadas, considerando Acurácia vs. Recall. As soluções não-dominadas formam a fronteira de Pareto, representando os melhores compromissos entre os objetivos conflitantes.

Fonte: Elaborado pelo autor (2025)

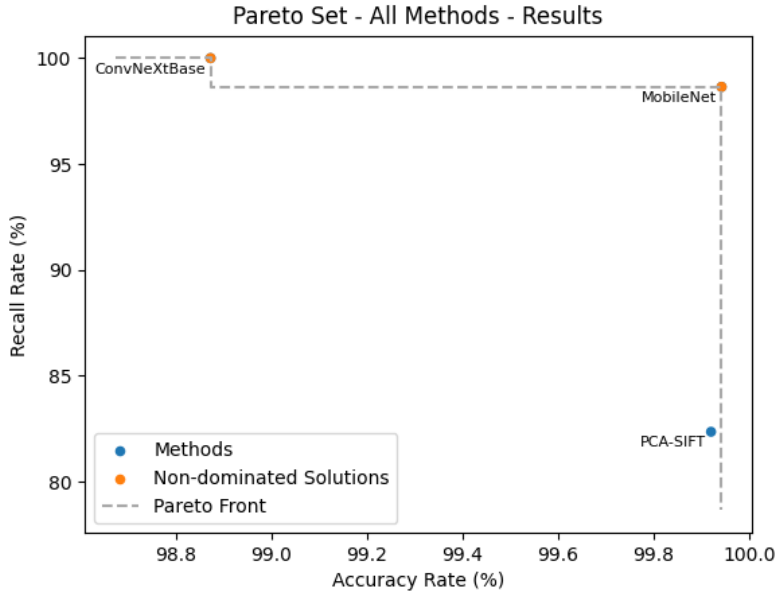

Figura 13 – Análise de dominância de Pareto (com ZOOM) para as técnicas avaliadas, considerando Acurácia vs. Recall. As soluções não-dominadas formam a fronteira de Pareto, representando os melhores compromissos entre os objetivos conflitantes.

Fonte: Elaborado pelo autor (2025)

6 AVALIAÇÃO DO RESULTADO DAS TÉCNICAS SELECIONADAS EM UM CASO REAL

Seguindo a metodologia de estudo proposta e anteriormente discutida na Seção 4, a segunda fase deste trabalho consistiu na realização de um estudo de caso real utilizando as melhores técnicas avaliadas durante a avaliação experimental realizada na primeira fase. Para este estudo de caso foi utilizado um conjunto de dados reais pertencentes à uma Instituição Financeira Cooperativa real com o objetivo de identificar cadastros de cooperados cuja foto apresente indícios de fraude considerando aspectos de similaridade com fotos de outro(s) cadastro(s) da própria base analisada.

6.1 Seleção da Instituição e Período de Análise

A aplicação das técnicas selecionadas na fase experimental foi realizada em uma base real de dados de uma cooperativa singular do estado de Minas Gerais pertencente ao Sistema de Cooperativas de Crédito do Brasil (Sicoob). A escolha desta instituição justifica-se pela representatividade do Sicoob no cenário nacional e mineiro do cooperativismo financeiro e pela necessidade prática de validar as técnicas estudadas em um ambiente operacional real, conforme destacado na contextualização deste trabalho.

A extração da base de dados da instituição cooperativa selecionada foi realizada considerando a data base de Julho/2025. Visando garantir o sigilo, bem como a contemporaneidade das informações relevantes para aplicação das técnicas selecionadas, o processo institucional do Sicoob de descaracterização de dados de produção também foi realizado ao conjunto de dados reais.

6.2 Processo de composição do Dataset

O processo de composição do dataset seguiu uma metodologia estruturada de filtragem progressiva, conforme ilustrado na Figura 14. Ao todo, neste processo de composição, cinco diferentes filtros foram aplicados: filtragem de cadastros apenas de Pessoa Física, filtragem por faixa etária, seleção de cadastros com disponibilidade de fotografia e exclusão de cadastros com selo ouro.

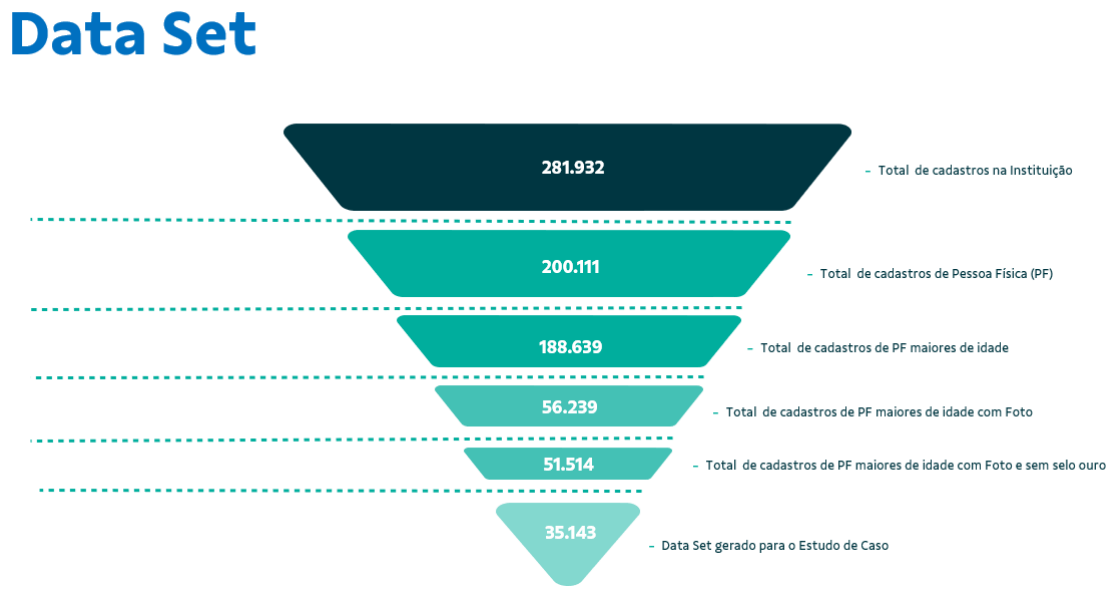

Figura 14 – Funil para geração do DataSet a ser utilizado a partir de uma base cadastral de uma cooperativa real.

Fonte: Elaborado pelo autor (2025)

Este processo pode ser compreendido através das seguintes etapas:

6.2.1 Base Inicial e Primeiro Filtro

O processo iniciou-se com a base completa da cooperativa, contendo 281.932 cadastros totais. A primeira etapa de filtragem consistiu na exclusão de cadastros de Pessoa Jurídica, resultando em 200.111 cadastros de Pessoa Física. Esta decisão metodológica fundamenta-se no fato de que o objetivo do estudo foca na identificação de indícios de fraude em fotografias de pessoas físicas.

6.2.2 Filtro por Faixa Etária

Na segunda etapa, foram excluídos os cadastros de menores de idade, reduzindo a base para 188.639 registros. Esta filtragem mitiga o apontamento de falsos positivos, considerando que a regra de negócio da instituição permitiu durante algum tempo a utilização da foto do responsável legal pelo menor em seu cadastro para autorizações de segurança por meio do processo de reconhecimento facial.

6.2.3 Filtro por Disponibilidade de Fotografia

A terceira etapa consistiu na seleção apenas dos cadastros que possuem foto registrada, resultando em 56.239 registros. Esta filtragem é fundamental para o estudo,

uma vez que as técnicas desenvolvidas dependem da existência de imagens para análise de possível duplicidade e indícios de fraude.

6.2.4 Exclusão de cadastros com Selo Ouro

Na quarta etapa, foram excluídos os cadastros com selo ouro, que são registros já submetidos a processos rigorosos de verificação e validação de identidade em bases governamentais ou de bureaus. Após esta filtragem, obteve-se 51.514 cadastros. A exclusão destes registros justifica-se pelo fato de já terem passado por verificação adicional de autenticidade, reduzindo significativamente a probabilidade de indícios de fraude.

6.2.5 Exclusões específicas para Conformidade

A etapa final envolveu a exclusão de cadastros específicos submetidos ao processo de documentoscopia, que é uma análise técnica especializada de documentos, o que reduz substancialmente a possibilidade de fraude.

Após todas as filtragens, o dataset final para aplicação das técnicas selecionadas foi composto por **35.143 cadastros**.

6.3 Análise qualitativa preliminar do Dataset

Antes da aplicação das técnicas de detecção de duplicidade, foi realizada uma análise qualitativa preliminar do dataset para identificar possíveis inconsistências que poderiam impactar nos resultados das técnicas selecionadas. A Tabela 4 apresenta os resultados desta análise.

Tabela 4 – Análise qualitativa preliminar do dataset real

Categoria de Inconsistência	**Quantidade**	**Porcentagem (%)**
Documentos em Geral	1.335	3,798%
Baixa Qualidade	573	1,630%
Imagens Corrompidas	367	1,044%
Imagens Rotacionadas	332	0,944%
Documentos de Identificação	163	0,463%
Lugares/Objetos	112	0,318%
Múltiplas Faces	43	0,122%
Menores de Idade	26	0,073%
Total de Inadequações	**2.951**	**8,397%**
Total de Cadastros	**35.143**	**100%**
DataSet Final	**32.192**	**91,603%**

6.3.1 Interpretação dos Resultados da Análise Qualitativa

A análise qualitativa, realizada para esse estudo de forma manual, revelou que 8,397% dos cadastros apresentam algum tipo de inadequação que pode impactar na eficácia das técnicas de detecção de duplicidade, categorizadas da seguinte forma em ordem decrescente de representatividade em relação ao DataSet gerado para aplicação das técnicas:

- **Documentos em Geral**: Indica cadastros onde foram armazenadas imagens de documentos ao invés de fotos faciais;
- **Baixa Qualidade**: Imagens com resolução, nitidez ou condições de iluminação inadequadas que podem comprometer a análise;
- **Imagens Corrompidas**: Arquivos de imagem danificados ou ilegíveis;
- **Imagens Rotacionadas**: Fotografias com orientação inadequada que podem afetar o desempenho dos algoritmos de reconhecimento.
- **Documentos de Identificação**: Imagens de Carteiras de Identidade e Carteira Nacional de Habilitação.
- **Lugares/Objetos**: Imagens de lugares e objetos e não de faces.
- **Múltiplas faces**: Imagens com mais de uma face humana.
- **Menores de Idade**: Apesar do cadastro indicar pela data de nascimento que o registro é de um adulto, a foto é de uma criança.

O dataset final com 32.192 registros representa 11,41% da base de cadastros da instituição selecionada e apresenta volume adequado para validação das técnicas em ambiente real, sendo representativo em relação às condições operacionais de uma cooperativa financeira de médio porte.

6.4 Resultado da aplicação dos métodos selecionados na base real

A aplicação das técnicas selecionadas na fase experimental - MobileNet (CNN) e ConvNeXtBase (CNN) - no conjunto de dados real composto por 32.192 cadastros de cooperados validou a eficácia das abordagens em um ambiente operacional autêntico.

Conforme apresentado na Figura 15, os resultados demonstraram capacidade significativa de detecção de possíveis casos de fraude por duplicidade de fotos faciais. A técnica MobileNet (CNN), que havia se destacado na avaliação experimental com 99,94% de acurácia e 98,64% de recall, identificou 662 casos de verdadeiros positivos na base

real, representando aproximadamente 2,05% dos cadastros analisados. O tempo médio de processamento por par de imagens foi de 0,26 segundos, mantendo-se próximo ao desempenho observado na fase experimental e confirmando sua viabilidade para aplicação em larga escala. A técnica ConvNeXtBase (CNN), que apresentou 98,87% de acurácia e 100% de recall na avaliação experimental, detectou 686 casos de verdadeiros positivos, correspondendo a 2,13% da base analisada. Apesar do tempo médio de processamento mais elevado (2,66 segundos por par), esta técnica demonstrou maior sensibilidade na detecção de similaridades, corroborando os resultados experimentais que indicaram seu recall superior.

Método	Situação	Qtde Imagens Comparadas	Verdadeiro Positivo	Falso Positivo	Tempo Par a Par (em segundos)	Tempo Total – Proc. de 1 Imagem (em segundos)
CNN – MobileNet	Concluído	518.146.336	662 casos	56 casos	0,2641 (0,0696)	8.504,35 (≈ 141 minutos)
CNN – ConvNeXt	Concluído	518.146.336	686 casos	100.175 casos	2,665 (0,2706)	85.790,82 (≈ 1.429 minutos)

Figura 15 – Resultado da aplicação dos métodos selecionados na base real.

Fonte: Elaborado pelo autor (2025)

Os resultados obtidos na base real demonstram uma convergência entre as duas técnicas avaliadas. A sobreposição de 662 casos identificados por ambas as técnicas fortalece a confiabilidade das detecções realizadas. Somando-se os 24 casos exclusivos pelo ConvNeXtBase, totaliza-se 686 cadastros únicos com indícios de fraude por duplicidade de fotos. Estes resultados corroboram a aplicabilidade prática das técnicas selecionadas e demonstram seu potencial para integração aos processos de segurança cibernética da cooperativa. A identificação de aproximadamente 2% dos cadastros com possíveis indícios de fraude na instituição financeira selecionada confirma a relevância operacional do estudo desenvolvido.

A diferença observada entre casos verdadeiros positivos e falsos negativos dos modelos MobileNet e ConvNeXtBase evidencia o impacto da seleção do limiar de similaridade selecionado. Limiares mais relaxados, conforme utilizado no modelo ConvNeXtBase, podem ser capazes de detectar mais casos fraudulentos. Em contrapartida, um número expressivo de casos de falsos positivos pode ser identificado, isto porque CNNs podem apresentar dificuldades em detectar diferenças finas entre imagens. Adicionalmente, esta complementaridade das técnicas sugere que uma abordagem combinada pode maximizar a capacidade de detecção, reduzindo a probabilidade de falsos negativos.

Neste cenário de ambiente operacional autêntico, considerando o tempo médio de processamento de 0,26 (MobileNet) e 2,66 (ConvNeXtBase) segundos por par de imagens, um cadastro para ser validado por completo em toda a base necessitou de aproximadamente 141 e 1.429 minutos, utilizando os modelos respectivamente mencionados. A avaliação

completa da base, ou seja, a avaliação de 518.146.336 pares de imagens, demandaria aproximadamente 38.000 e 383.000 horas de processamento sequencial para cada um dos modelos avaliados. Sendo assim, foi realizado o processamento de forma não uniformemente distribuída.

Notoriamente, o processamento sequencial de volumosos bancos de imagens exige tempos de processamento indesejáveis ou proibitivos. Desta forma, a Figura 16 apresenta uma projeção de tempo total de processamento ao se adotar estratégias computacionais de processamento distribuído, para cada uma das técnicas avaliadas. Portanto, observa-se, nessa figura, um comportamento decrescente do tempo total à medida que o número de máquinas é acrescido.

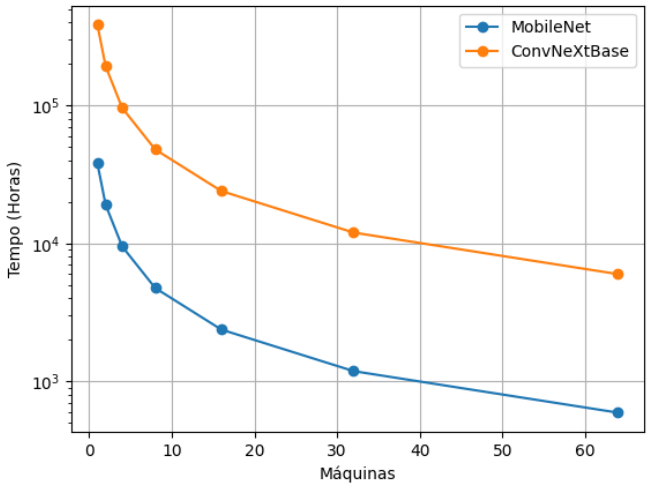

Figura 16 – Estimativa de tempo (em horas) para processamento completo da base.

Fonte: Elaborado pelo autor (2025)

7 CONCLUSÕES

Potencializado pelo atual cenário de transformação digital cotidianamente vivenciado, fraudes digitais assolam os mais diversos setores econômicos, envoltos neste relevante fenômeno global. A sensibilidade dos dados e o relevante valor financeiro que envolvem as transações gerenciadas por instituições financeiras, incluindo os sistemas cooperativos financeiros, têm destacado o fundamental e estratégico papel desempenhado pela segurança cibernética nessas instituições. Preservar e prevenir os sistemas contra fraudes digitais cada vez mais sofisticadas, evitando prejuízos financeiros e reputacionais, bem como consequências legais, são algumas das importantes atividades realizadas por equipes de segurança cibernética.

Inserido neste contexto, este trabalho teve como objetivo geral identificar e avaliar métodos/técnicas eficazes para detectar cadastros de cooperados que apresentem indícios de fraude por meio da identificação de duplicidade de fotos faciais em instituições financeiras cooperativas. Este estudo desenvolveu-se em duas fases complementares que avaliaram e validaram a eficiência e eficácia de diversas técnicas propostas e encontradas na literatura.

Na primeira fase do estudo proposto, foi conduzida uma avaliação experimental controlada utilizando um conjunto de dados público de 442 imagens de faces de celebridades, onde foram testadas 22 técnicas diferentes organizadas em duas categorias: baseadas nas características da imagem (MSE, SSIM, SIFT, ORB, PCA-SIFT, intersecção de histogramas) e baseadas na estrutura da imagem seja utilizando Perceptual Hashing (Average Hash, Perception Hash, Difference Hash, Wavelet Hash) utilizando Redes Neurais Convolucionais (CNNs) com 11 modelos pré-treinados, além de Large Language Models multimodais.

A metodologia de avaliação proposta baseou-se em métricas de acurácia, *recall* e taxa de falsos positivos, utilizando análise de dominância de Pareto para seleção das técnicas mais eficazes. Os resultados experimentais demonstraram que as técnicas baseadas em CNN apresentaram promissor desempenho na detecção de similaridades. Especificamente, o modelo de CNN pré-treinada, MobileNet, alcançou 99,94% de acurácia e 98,64% de *recall*, enquanto o modelo ConvNeXtBase obteve 98,87% de acurácia e 100% de recall, sendo ambas identificadas como soluções não-dominadas na análise de Pareto. Estas técnicas foram selecionadas para aplicação na segunda fase do estudo devido ao equilíbrio otimizado entre desempenho e viabilidade computacional.

Na segunda fase, as técnicas selecionadas foram aplicadas em uma base real de 32.192 cadastros de uma cooperativa financeira do Sistema Sicoob após processo de filtragem que resultou em um *DataSet* final de 32.192 registros. Nessa fase foram identificados

686 cadastros únicos com indícios de fraude por duplicidade de fotos, representando aproximadamente 2% da base analisada. A convergência dos resultados entre as duas técnicas, com sobreposição de 662 casos, fortalece a confiabilidade das detecções e valida a eficácia das abordagens em ambiente operacional real.

Cientes dos promissores resultados e limitações das técnicas avaliadas, este estudo demonstra aplicabilidade em diversos contextos do sistema financeiro cooperativo, incluindo: fortalecimento de sistemas de autenticação biométrica facial, auditoria preventiva de bases cadastrais, suporte a processos de conformidade e compliance, otimização de custos operacionais na validação de identidade e aprimoramento de políticas de segurança cibernética. Adicionalmente, este estudo contribui significativamente para o fortalecimento da segurança cibernética no cooperativismo financeiro, oferecendo uma metodologia de baixo custo, escalável e eficiente para detecção preventiva de fraudes.

7.1 Desafios Identificados

Embora promissores e significativos resultados tenham sido alcançados neste estudo, relevantes desafios foram identificados. Qualidade da base de imagens, desequilíbrio amostral (casos fraudulentos e não fraudulentos), impacto da seleção dos parâmetros de limiar de similaridade das técnicas e variabilidade do custo computacional dos métodos evidenciam a complexidade da aplicação prática das técnicas avaliadas.

Em relação à qualidade do conjunto de dados reais, uma análise qualitativa preliminar revelou que 8,39% dos cadastros apresentavam algum tipo de inadequação, incluindo documentos em geral (3,79%), imagens de baixa qualidade (1,63%), imagens corrompidas (1,04%) e imagens rotacionadas (0,94%). Esses dados destacam que a qualidade final do *DataSet* a ser trabalhado pode ter implicações diretas para a aplicação das técnicas selecionadas na fase experimental. Sendo assim, o processo para uso dos métodos/técnicas deve ser capaz de gerenciar a qualidade das imagens, seja através de:

- Pré-processamento adequado para melhoria da qualidade das imagens;
- Filtros específicos para exclusão automática de imagens inadequadas;
- Ajuste de parâmetros das técnicas para maior robustez a variações de qualidade.

Além disso, em relação ao conjunto de imagens identificadas com qualidade inadequada, considerando que o processo permitiu o cadastramento antes do seu aperfeiçoamento, surge a oportunidade de marcação desses cadastros para exigir a atualização dos mesmos em relação ao registro facial do cooperado.

O desequilíbrio natural dos dados, característico de problemas reais de detecção de fraude, representa outro desafio importante. A predominância de casos legítimos

em relação aos fraudulentos exige cuidadosa calibração das técnicas para maximizar a detecção de verdadeiros positivos sem comprometer excessivamente a especificidade do sistema. Alternativamente, este estudo conduzido é um potencial colaborador para o desenvolvimento e manutenção de uma base de dados equilibrada, representativa e própria para trabalhos futuros.

O custo computacional diferenciado entre as técnicas também se apresenta como desafio operacional. A variabilidade do tempo médio de processamento por par de imagens evidenciou o impacto direto na viabilidade de implementação em larga escala das técnicas avaliadas. Por consequência, exigindo planejamento adequado de recursos computacionais.

7.2 Recomendações para Trabalhos Futuros

Os resultados alcançados e os desafios identificados neste estudo representam um importante progresso para o aprimoramento das estratégias de segurança cibernética no sistema financeiro cooperativista. Desta forma, surgem possibilidades de trabalhos futuros que podem ampliar e aprofundar as contribuições deste estudo.

A primeira possibilidade consiste no desenvolvimento e avaliação de técnicas especializadas em qualificação automática de imagens faciais em cadastros que possam ser integradas ao processo estudado nesse trabalho. Este desenvolvimento deve incluir algoritmos para detecção de imagens com baixa qualidade, com rotações inadequadas, identificação de objetos e documentos em lugar de faces, reconhecimento de múltiplas faces, e detecção de cadastros de menores de idade. Adicionalmente, sugere-se a implementação de sistemas para identificação de Pessoas Expostas à Mídia (PEM) e Pessoas Expostas Politicamente (PEP) que podem apresentar desafios específicos para sistemas de biometria facial devido à ampla disponibilidade de suas imagens, além da possibilidade de validação da imagem da face com os demais dados cadastrais.

A segunda recomendação envolve o estudo e a elaboração de uma métrica quantitativa de risco para avaliar o nível de exposição que uma base de imagens representa para os negócios da instituição. Esta métrica deve integrar resultados das técnicas de qualificação de imagens com os indicadores de detecção de similaridade, fornecendo um score consolidado que possa orientar decisões da gestão executiva sobre investimentos em segurança, políticas de validação de cadastros e estratégias de mitigação de riscos cibernéticos.

Outra direção promissora contempla a investigação de técnicas híbridas que combinem abordagens complementares por meio de arranjos de métodos para maximizar a detecção de cadastros com indícios de fraudes, minimizar falsos negativos e otimizar o tempo para validação de grandes conjuntos de imagens. Tais abordagens, como o uso de hashing perceptivo (por exemplo, Average Hashing combinado com descritores SIFT ou

ORB) e embeddings vetoriais armazenados em bancos de dados vetoriais, permitem uma análise eficiente de grandes volumes de imagens sem depender exclusivamente de soluções proprietárias de alto custo, como as que utilizam bases biométricas governamentais ou de bureaus privados. Adicionalmente, a combinação de técnicas personalizadas pode oferecer maior robustez contra variações em imagens, como alterações de iluminação, rotação ou compressão, que são comuns em tentativas de fraude.

Além disso, há um importante espaço para exploração de técnicas de aprendizado federado (Zhu; Zhang; Jin, 2021; Guo *et al.*, 2024) que permitam compartilhamento de conhecimento entre cooperativas financeiras sem comprometer a privacidade dos dados.

Por fim, recomenda-se a expansão do estudo para análise de outras modalidades biométricas e a investigação da aplicabilidade das técnicas desenvolvidas em diferentes setores que enfrentem desafios similares de validação de identidade, incluindo empresas de telecomunicações, plataformas digitais e órgãos governamentais que mantêm bases cadastrais extensas.

Com isso, conclui-se que esse trabalho estabelece uma base sólida para o avanço da segurança cibernética no cooperativismo financeiro e demonstra o potencial das técnicas de inteligência artificial para enfrentar os desafios crescentes da fraude digital, contribuindo para a construção de um sistema financeiro mais seguro e resiliente.

REFERÊNCIAS

AKHTAR, S. *et al.* Cyber security solutions for businesses in financial services: Challenges, opportunities, and the way forward. **International Journal of Business Intelligence Research**, IGI Global, USA, v. 12, n. 1, p. 82–97, jan. 2021.

ARIMATHEA, B. *et al.* **Golpes Bancários Disparam no País e Devem Gerar Prejuízos de pelo Menos R$ 2,5 Bilhões Neste Ano**. Jornal Estadão, 2022. Disponível em: https://www.estadao.com.br/economia/golpes-bancarios-geram-prejuizos-no-pais/. Acesso em: 18 set. 2025.

BACH, J. R. *et al.* Virage image search engine: an open framework for image management. *In*: **Electronic imaging**. [*S.l.: s.n.*], 1996.

BCB. **Resolução nº 4.658, de 26 de abril de 2018. Dispõe sobre a política de segurança cibernética e sobre os requisitos para a contratação de serviços de processamento e armazenamento de dados e de computação em nuvem a serem observados pelas instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil.** Banco Central do Brasil (BCB), 2018. Disponível em: https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50581/Res_4658_v1_O.pdf. Acesso em: 3 março. 2025.

BÚRIGO, F. L. Finanças e solidariedade: cooperativismo de crédito rural solidário no brasil. **Estudos Sociedade e Agricultura**, v. 18, p. 489–509, 2010.

CALONDER, M. *et al.* Brief: Binary robust independent elementary features. *In*: **Computer Vision – ECCV 2010**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 778–792.

CHA, S.-H. Comprehensive survey on distance/similarity measures between probability density functions. **International Journal of Mathematical Models and Methods in Applied Sciences**, v. 1, 01 2007.

CHANG, Y. *et al.* A survey on evaluation of large language models. **ACM Transactions on Interactive Intelligent Systems**, Association for Computing Machinery, New York, NY, USA, v. 15, n. 3, mar. 2024.

CHOLLET, F. **Xception: Deep Learning with Depthwise Separable Convolutions**. 2017. Disponível em: https://arxiv.org/abs/1610.02357.

DAKHIL, N.; ABDULAZEEZ, A. Face recognition based on deep learning: A comprehensive review. **Indonesian Journal of Computer Science**, v. 13, 06 2024.

DAREM, A. A. *et al.* Cyber threats classifications and countermeasures in banking and financial sector. **IEEE Access**, v. 11, p. 125138–125158, 2023.

DHILLON, A.; VERMA, G. K. Convolutional neural network: a review of models, methodologies and applications to object detection. **Progress in Artificial Intelligence**, Springer, v. 9, n. 2, p. 85–112, 2020.

ESTRADA, E. Communicability cosine distance: similarity and symmetry in graphs/networks. **Computational and Applied Mathematics**, Springer, v. 43, n. 1, p. 49, 2024.

FAWZI, A.; MOOSAVI-DEZFOOLI, S.-M.; FROSSARD, P. The robustness of deep networks: A geometrical perspective. **IEEE Signal Processing Magazine**, v. 34, n. 6, p. 50–62, 2017.

FEBRABAN, F. B. d. B. **Crescem Golpes Envolvendo Manipulação de Vítimas para Roubo de Informações Pessoais**. FEBRABAN, 2021. Disponível em: https://portal.febraban.org.br/noticia/3704/pt-br/. Acesso em: 18 set. 2025.

FINLAYSON, G.; CHATTERJEE, S.; FUNT, B. Colour-texture indexing. *In*: **IEE Colloquium on Intelligent Image Databases**. [*S.l.: s.n.*], 1996. p. 12/1–12/6.

GARHPATI, V. **Celebrity Face Image Dataset**. Kaggle, 2022. Disponível em: https://www.kaggle.com/datasets/vishesh1412/celebrity-face-image-dataset.

GOODE, A. Biometrics for banking: best practices and barriers to adoption. **Biometric Technology Today**, v. 2018, n. 10, p. 5–7, 2018.

GU, J. *et al.* Recent advances in convolutional neural networks. **Pattern Recognition**, v. 77, p. 354–377, 2018.

GUO, W. *et al.* A comprehensive survey of federated transfer learning: challenges, methods and applications. *In*: **Frontiers of Computer Science**. [*S.l.: s.n.*], 2024.

HAO, Q. *et al.* It's not what it looks like: Manipulating perceptual hashing based applications. *In*: **Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security**. New York, NY, USA: Association for Computing Machinery, 2021. (CCS '21), p. 69–85.

HE, K. *et al.* **Deep Residual Learning for Image Recognition**. 2015. Disponível em: https://arxiv.org/abs/1512.03385.

HOWARD, A. G. *et al.* **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**. 2017. Disponível em: https://arxiv.org/abs/1704.04861.

HUANG, G. *et al.* **Densely Connected Convolutional Networks**. 2018. Disponível em: https://arxiv.org/abs/1608.06993.

HUANG, W.-C.; TROIA, F. D.; STAMP, M. Robust hashing for image-based malware classification. *In*: **International Workshop on Behavioral Analysis for System Security**. [*S.l.: s.n.*], 2018. p. 451–459.

IBRAHIN, A. S. B.; KHALIFA, O. O.; AHMED, D. E. M. Plagiarism detection of images. *In*: **2020 IEEE Student Conference on Research and Development (SCOReD)**. [*S.l.: s.n.*], 2020. p. 183–188.

JACQUES, E.; GONçALVES, F. Cooperativas de crédito no brasil: evolução e impacto sobre a renda dos municípios brasileiros. **Economia e Sociedade**, v. 25, p. 489–509, 2016.

JAIN, A.; ROSS, A.; NANDAKUMAR, K. **Introduction to Biometrics**. New York: Springer US, 2011.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. A review of the use of convolutional neural networks in agriculture. **The Journal of Agricultural Science**, Cambridge University Press, v. 156, n. 3, p. 312–322, 2018.

KE, Y.; SUKTHANKAR, R. Pca-sift: a more distinctive representation for local image descriptors. *In*: **Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.** [*S.l.: s.n.*], 2004. v. 2, p. II–II.

KHAMAISEH, S. Y. *et al.* Adversarial deep learning: A survey on adversarial attacks and defense mechanisms on image classification. **IEEE Access**, v. 10, p. 102266–102291, 2022.

LEE, S.; XIN, J.; WESTLAND, S. Evaluation of image similarity by histogram intersection. **Color Research & Application**, v. 30, p. 265 – 274, 08 2005.

LI, Z. *et al.* A survey of convolutional neural networks: Analysis, applications, and prospects. **IEEE Transactions on Neural Networks and Learning Systems**, v. 33, n. 12, p. 6999–7019, 2022.

LIU, F. *et al.* Application of large language models in medicine. **Nature Reviews Bioengineering**, Nature Publishing Group UK London, p. 1–20, 2025.

LIU, S.; SILVERMAN, M. A practical guide to biometric security technology. **IEEE Security & Privacy**, v. 16, n. 1, p. 50–62, 2018.

LIU, Z. *et al.* **A ConvNet for the 2020s**. 2022. Disponível em: https://arxiv.org/abs/2201.03545.

LOWE, D. Object recognition from local scale-invariant features. *In*: **Proceedings of the Seventh IEEE International Conference on Computer Vision**. [*S.l.: s.n.*], 1999. v. 2, p. 1150–1157.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, Kluwer Academic Publishers, USA, v. 60, n. 2, p. 91–110, 2004.

MEINEN, E.; PORT, M. **O cooperativismo de crédito ontem, hoje e amanhã**. Brasília: Confebras, 2014.

MORRA, L.; LAMBERTI, F. Benchmarking unsupervised near-duplicate image detection. **Expert Systems with Applications**, v. 135, p. 313–326, 2019.

NIBLACK, W. *et al.* The qbic project: Querying images by content, using color, texture, and shape. **SPIE Conference on Storage and Retrieval for Image and Video Databases**, v. 1908, p. 173–187, 01 1993.

OCB. **Organização das Cooperativas Brasileiras - Anuário do Cooperativismo Brasileiro 2023**. Brasília: Sistema OCB, 2023.

PINHEIRO, M. A. H. **Cooperativas de crédito: história da evolução normativa no Brasil**. 6. ed. Brasília: Banco Central do Brasil (BCB), 2008.

RAIAAN, M. A. K. *et al.* A review on large language models: Architectures, applications, taxonomies, open issues and challenges. **IEEE access**, IEEE, v. 12, p. 26839–26874, 2024.

RANGEL, G. *et al.* A survey on convolutional neural networks and their performance limitations in image recognition tasks. **Journal of sensors**, Wiley Online Library, v. 2024, n. 1, p. 2797320, 2024.

RIVEST, R. **RFC1321: The MD5 Message-Digest Algorithm**. USA: RFC Editor, 1992.

ROSTEN, E.; DRUMMOND, T. Machine learning for high-speed corner detection. *In*: **Computer Vision – ECCV 2006**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 430–443.

RUBLEE, E. *et al.* Orb: An efficient alternative to sift or surf. *In*: **2011 International Conference on Computer Vision**. [*S.l.: s.n.*], 2011. p. 2564–2571.

SAMANTA, P.; JAIN, S. Analysis of perceptual hashing algorithms in image manipulation detection. **Procedia Computer Science**, v. 185, p. 203–212, 2021. Big Data, IoT, and AI for a Smarter Future.

SARA, U.; AKTER, M.; UDDIN, M. S. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. **Journal of Computer and Communications**, v. 07, p. 8–18, 01 2019.

SICOOB. **Relatório Anual e de Sustentabilidade 2023**. Brasília: Sicoob, 2023.

SILVA, R. P.; SANTOS, M. A. Biometric authentication in financial services: Facial recognition technologies and fraud prevention. **International Journal of Banking Innovation**, v. 45, p. 112–129, 2023.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. 2015. Disponível em: https://arxiv.org/abs/1409.1556.

SOARES, M. M.; SOBRINHO, A. D. M. **Microfinanças: o papel do Banco Central do Brasil e a importância do cooperativismo de crédito**. 2. ed. Brasília: Banco Central do Brasil (BCB), 2008.

SYED, W. K. *et al.* Biometric authentication systems in banking: A technical evaluation of security measures. *In*: **2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)**. Gwalior, India: [*S.l.: s.n.*], 2024. p. 1331–1336.

SZEGEDY, C. *et al.* **Rethinking the Inception Architecture for Computer Vision**. 2015. Disponível em: https://arxiv.org/abs/1512.00567.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2020. Disponível em: https://arxiv.org/abs/1905.11946.

THYAGHARAJAN, K. K.; KALAIARASI, G. A review on near-duplicate detection of images using computer vision techniques. **Archives of Computational Methods in Engineering**, v. 28, p. 897–916, 01 2020.

TILBORG, H.; JAJODIA, S. **Encyclopedia of Cryptography and Security, 2nd Ed**. [*S.l.: s.n.*]: Springer, 2011.

WANG, Z. *et al.* Image quality assessment: from error visibility to structural similarity. **IEEE Transactions on Image Processing**, v. 13, n. 4, p. 600–612, 2004.

WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. **IEEE Signal Processing Magazine**, v. 26, n. 1, p. 98–117, 2009.

WHITELAW, T. **An Introduction to Abstract Algebra**. [*S.l.: s.n.*]: Blackie, 1978.

YADAV, K. *et al.* Real time face recognition based on convolution neural network. *In*: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND SOFTWARE ENGINEERING. **Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019**. [*S.l.: s.n.*]: SSRN, 2019.

YAMASHITA, R. *et al.* Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, Springer, v. 9, n. 4, p. 611–629, 2018.

YANG, B.; GU, F.; NIU, X. Block mean value based image perceptual hashing. *In*: **2006 International Conference on Intelligent Information Hiding and Multimedia**. [*S.l.: s.n.*], 2006. p. 167–172.

ZAUNER, C.; STEINEBACH, M.; HERMANN, E. Rihamark: perceptual image hash benchmarking. *In*: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Media Watermarking, Security, and Forensics III**. [*S.l.: s.n.*]: SPIE, 2011. v. 7880, p. 78800X.

ZHAO, X. *et al.* A review of convolutional neural networks in computer vision. **Artificial Intelligence Review**, v. 57, p. 57–99, 03 2024.

ZHU, H.; ZHANG, H.; JIN, Y. From federated learning to federated neural architecture search: a survey. *In*: **Complex & Intelligent Systems**. [*S.l.: s.n.*], 2021. p. 639–657.