



II CONGRESSO INTERNACIONAL DE TECNOLOGIAS EMERGENTES

Onde as ideias do presente criam as soluções do futuro

14 e 15 de Novembro de 2025

APLICAÇÃO DE REDES NEURAIIS NA VALIDAÇÃO DE DOCUMENTOS DE IDENTIDADE

Diego Alejandro Bejarano Terceros¹

Marcelo Chamy Machado²

Eduardo Palhares Júnior³

Daniel Fonseca de Souza⁴

Alyson de Jesus dos Santos⁵

¹Especialização em Aprendizado de Máquina - Instituto Federal do Amazonas - IFAM, Campus Zona Leste – Manaus, AM – Brasil, diegoalejandrol32@gmail.com

²Instituto Federal do Amazonas - IFAM, Campus Distrito Industrial – Manaus, AM – Brasil, marcelo.chamy@ifam.edu.br

³ Instituto Federal do Amazonas - IFAM, Campus Distrito Industrial – Manaus, AM – Brasil, eduardo.palharesjr@ifam.edu.br

⁴ Instituto Federal do Amazonas - IFAM, Campus Distrito Industrial – Manaus, AM – Brasil, danielsouza@ifam.edu.br

⁵ Instituto Federal do Amazonas - IFAM, Campus Distrito Industrial – Manaus, AM – Brasil, alyson.santos@ifam.edu.br

Resumo

DOI:10.5281/zenodo.17917770

Este trabalho propõe o uso de redes neurais para otimizar a validação cadastral que envolve documentos de identidade, automatizando tarefas que tradicionalmente dependem de análise manual. O algoritmo desenvolvido realiza três etapas principais comumente presentes em sistemas de validação cadastral: classificação do tipo de documento, extração e validação do texto do documento e verificação facial. Para as tarefas de classificação e verificação facial, foram testados diferentes modelos de redes neurais encontrados no estado da arte, apresentando desempenhos semelhantes. Assim, foram realizadas comparações entre as arquiteturas ResNet-101, DenseNet-169, ViT-B16 e EfficientNet-B3, sendo esta última a escolhida pela melhor relação entre precisão e eficiência. Na verificação facial, foram analisados os modelos FaceNet, FaceNet512 e VGG-Face, destacando-se o FaceNet512 pelo melhor equilíbrio entre acurácia e precisão. A abordagem proposta automatiza o processo de validação, reduzindo significativamente o tempo de análise e filas para conferência manual. Como trabalhos futuros, sugere-se aprimorar o pré-processamento de imagens de forma específica para cada classe de documento, com o objetivo de aumentar a precisão do OCR, além de incorporar uma etapa anterior a classificação, a de detecção de documentos, visando maior robustez em cenários reais. Também se propõe a construção de uma base de dados que contenha imagens de pessoas

segurando seus documentos de identidade, ou documentos simulados com rostos correspondentes, para aprimorar o treinamento e avaliação da etapa de verificação facial.

Palavras-Chaves: Validação documental; Classificação de documentos; Verificação facial; Validação textual; Redes neurais.

1. INTRODUÇÃO.

Com o avanço da tecnologia, a obtenção de dados para cadastros e inscrições, tanto em instituições públicas quanto privadas, passou de um processo predominantemente manual para uma etapa majoritariamente digital. Essa transformação trouxe ganhos significativos de agilidade e eficiência, reduzindo o tempo de processamento e a necessidade de intervenção humana em diversas etapas. No entanto, a validação de informações relevantes, como dados pessoais e documentos de identidade, ainda é realizada, em grande parte, de forma manual, o que prolonga o tempo de análise e pode gerar gargalos operacionais, como destacado por Dias et al. (2019).

Um exemplo disso é a validação de identidade em protocolos de cadastro. Em determinados contextos, como o cadastro de pessoas físicas ou jurídicas em empresas, a verificação documental demanda uma equipe numerosa e um tempo significativo. Com a automação, esse processo poderia ser otimizado por meio de algoritmos que validam os dados inseridos no momento do cadastro, como destacado por Farias (2023). Esses sistemas utilizariam uma margem de confiança para indicar a precisão das informações, tornando o procedimento mais ágil e eficiente.

Diante desse cenário, surge a necessidade de uma solução que automatize a validação de identidade, reduzindo o tempo necessário para essa ação e tornando todo o fluxo mais eficiente. Este trabalho apresenta a aplicação de modelos de *deep learning* para validação de informações de identidade, abrangendo três tarefas principais: classificação de documentos, extração e validação de texto e verificação facial.

Para a classificação de documentos, foram comparados os modelos ResNet-101, EfficientNet-B3, DenseNet-169 e ViT-B16, visando identificar o mais eficiente para a versão final do algoritmo. Na verificação facial, foram avaliados os modelos FaceNet, FaceNet512 e VGG-FACE, considerando a precisão na comparação entre imagens faciais. Já para a extração de texto, foi utilizado o PaddleOCR como ferramenta principal, escolhido por sua robustez e desempenho apresentados.

O algoritmo analisa dados comumente preenchidos em formulários online e realiza: classificação do tipo de documento ou indicação de não reconhecimento; extração e validação de informações textuais em comparação com dados de referência e verificação facial entre a foto presente no documento e uma imagem do rosto fornecida. Ao final, ele retorna a classificação do documento, a validação dos dados e a confirmação de correspondência facial.

Além de propor e avaliar um modelo automatizado de validação de identidade, este trabalho também busca servir como um guia prático para a implementação dessas três tarefas em contextos empresariais ou organizacionais de menor porte. Em muitos casos, essas instituições não dispõem de grandes volumes de dados nem de equipes altamente especializadas para treinar modelos do zero. Assim, este estudo apresenta uma abordagem acessível e aplicável, demonstrando como combinar técnicas já consolidadas de aprendizado profundo para alcançar bons resultados e otimizar fluxos de validação cadastral de forma eficiente e viável.

2. METODOLOGIA.

O projeto desenvolvido faz uso de modelos de *deep learning* para a validação automatizada de documentos, com o objetivo de otimizar o fluxo de trabalho e reduzir a necessidade de intervenção humana. Dessa forma, caracteriza-se como uma pesquisa de natureza aplicada, uma vez que busca soluções práticas para um problema real.

O estudo pode ser classificado como pesquisa quantitativa, pois sua avaliação e desempenho são mensurados estatisticamente, por meio de métricas e gráficos.

A organização metodológica deste trabalho foi organizada em 4 partes principais: base de dados, modelos, métricas de resultado e execução.

2.1 Base de dados.

Como o tema deste trabalho está relacionado a documentos, realizou-se uma busca por bases de dados que disponibilizassem imagens nítidas de documentos contendo rostos de pessoas e arquivos GT (*Ground Truth*) contendo o texto das imagens. Contudo, devido à restrição na divulgação de dados pessoais, não foi possível encontrar um dataset que reunisse todas essas informações simultaneamente. Diante disso, por meio de uma revisão do estado da arte, foram identificadas algumas bases de dados capazes de fornecer ao menos parte dos recursos necessários para o desenvolvimento do projeto.

Farias (2023), em seu trabalho sobre a utilização de OCR em documentos, empregou uma base de dados pública composta por simulações de documentos brasileiros, idênticas aos originais, mas sem rostos de pessoas e contendo apenas dados fictícios e que continham os

arquivos GT. O dataset, denominado BID Dataset (*Brazilian Identity Document Dataset*), apresenta imagens de alta nitidez, suprimindo, assim, a necessidade de uma base de dados adequada às tarefas de classificação, leitura e validação das informações contidas em documentos. A base é composta por oito classes, sendo elas: CPF (frente e verso), RG (frente, verso e aberto) e CNH (frente, verso e aberta) e junto de cada imagem de documento há seu arquivo GT correspondente contendo todo o conteúdo textual da imagem.

Como não foi possível encontrar um dataset que contivesse imagens de documentos com rostos, ou de pessoas segurando suas identidades ou outros documentos com foto, foi utilizado o LFW Dataset (*Labeled Faces in the Wild*) como base de dados para a execução da tarefa de verificação facial. Esse dataset contém diversas classes, representadas pelos nomes das pessoas, e cada classe possui várias imagens dos respectivos rostos, possibilitando, assim, a comparação entre faces de indivíduos iguais e diferentes.

2.2 Modelos.

A escolha dos modelos foi orientada pela necessidade de atender às tarefas específicas deste trabalho, ao mesmo tempo em que se buscou diversificar as arquiteturas utilizadas. Devido à limitação de recursos computacionais e à ausência de bases de dados suficientemente robustas, outro critério adotado foi a utilização de modelos pré-treinados.

2.2.1 Classificação de documentos.

Revisando o estado da arte, foram encontradas algumas arquiteturas que apresentaram bom desempenho na tarefa de classificação de imagens.

A primeira arquitetura escolhida foi a ResNet uma das mais populares e bem-sucedidas atualmente, conhecida por sua facilidade de uso e bom desempenho com imagens estruturadas (PRANAV e SETHI, 2024). A versão utilizada foi a ResNet-101, que não precisa de tanta complexidade computacional como a versão 152, versão com maior número de camadas, porém apresenta resultados semelhantes.

A segunda arquitetura selecionada foi a EfficientNet, que, segundo Pranav e Sethi (2024), oferece um bom equilíbrio entre precisão e eficiência computacional, sendo indicada especialmente para bases de dados menores. Estudos como o de Yang, Zhang e Du (2021) também apontam que a EfficientNet pode superar a ResNet em cenários com conjuntos de dados reduzidos. A versão utilizada foi a EfficientNet-B3.

A terceira arquitetura considerada foi a Vision Transformer (ViT), reconhecida por sua capacidade de capturar relações globais em imagens e pelo destaque crescente nas tarefas de visão computacional (ZHAO et al., 2021). A versão utilizada foi a ViT-B16.

Por fim, foi selecionada a arquitetura DenseNet, cuja principal característica é a reutilização das saídas de camadas anteriores, promovendo eficiência e redução no número de parâmetros (PRANAV e SETHI, 2024). A versão escolhida foi a DenseNet-169.

As versões escolhidas de cada arquitetura são versões que não necessitam de uma complexidade computacional muito grande, mas ao mesmo tempo oferecem resultados semelhantes às versões que exigem poder computacional maior.

Todos esses modelos estão presentes na biblioteca pytorch que foi a biblioteca escolhida como framework para implementação dos modelos de classificação de imagem.

2.2.2 Validação de informações com OCR.

Durante a busca no estado da arte, dois modelos se destacaram por sua recorrência na literatura: o Tesseract OCR e o EasyOCR. Um ponto importante sobre ambos é que possuem suporte para múltiplos idiomas, incluindo o português, conforme indicado por Park e Shin (2023). Além disso, os dois modelos são de fácil utilização, não exigindo configurações complexas para início do uso.

O Tesseract, por padrão, não utiliza GPU, o que pode impactar o desempenho em tarefas mais exigentes. Por outro lado, o EasyOCR oferece suporte ao uso de GPU, o que melhora significativamente sua performance na extração de texto. Além disso, diversos estudos apontam que o EasyOCR apresenta resultados superiores ao Tesseract em diferentes tarefas de OCR, como evidenciado por Park e Shin (2023), Vedhaviyassh et al. (2022) e Beerten (2025). Apesar disso, como evidenciado por Beerten (2025), outros dois modelos apresentaram resultados melhores que os dois primeiros: MMOCR e PaddleOCR. Dos dois, somente o segundo tem adaptação para reconhecimento de caracteres em português e também ao realizar alguns testes, se mostrou mais fácil para configurar o modelo além de ter se mostrado tão rápido quanto os primeiros dois modelos citados mesmo sem uso de GPU. Portanto, o modelo utilizado para OCR nos documentos foi o PaddleOCR.

2.2.3 Verificação facial.

Com o objetivo de agilizar o processo de escolha dos modelos a serem utilizados na etapa de verificação facial, foi realizada uma pesquisa por bibliotecas especializadas nessa tarefa.

A partir dessa busca, foi identificada a biblioteca DeepFace, que disponibiliza, de forma prática e acessível, modelos amplamente referenciados no estado da arte para a criação de *embeddings* faciais.

Foi feita então, uma busca por modelos mais citados no estado da arte e que ao mesmo tempo estivessem disponíveis na biblioteca. Com isso, foram selecionados os modelos VGG-Face e FaceNet512, por apresentarem boa eficiência em verificação facial, como analisado por Tomillo et al. (2024), e também a variante mais leve e antiga do segundo modelo, o FaceNet, que também apresenta bons resultados como comentado por Firmansyah (2023).

2.3 Métricas de resultado.

As métricas são utilizadas para avaliar a qualidade dos resultados obtidos, considerando o contexto de cada tarefa. Dessa forma, cada etapa deste trabalho possui sua própria métrica de avaliação, definida com base, principalmente, nos estudos utilizados como referência.

2.3.1 Classificação de documentos.

Para esta tarefa, assim como proposto por Almeida e Silva (2022), foram utilizadas as métricas acurácia, precisão, recall e F1-score, que são amplamente empregadas na avaliação de modelos de classificação de imagens.

De acordo com Junior *et al.* (2022), a acurácia representa a proporção de resultados corretos em relação ao total de previsões realizadas, a precisão indica a proporção de amostras classificadas como positivas que, de fato, pertencem à classe positiva, ou seja, um modelo com alta precisão apresenta poucos falsos positivos, o recall, por sua vez, avalia a capacidade do modelo de identificar corretamente os casos positivos e, por fim, o F1-score combina a precisão e o recall em uma única medida, atribuindo pesos equilibrados a ambas.

2.3.2 Validação de informações com OCR.

Para avaliar a precisão do OCR aplicado aos documentos, foram adotadas, como referência, as métricas utilizadas por Farias (2023): precisão e taxa de erro. A taxa de erro representa a proporção de dados ausentes em relação ao total de texto real, enquanto a precisão corresponde à proporção de dados corretamente extraídos em relação à quantidade total de informações esperadas.

2.3.3 Verificação facial.

Para avaliar a similaridade entre os rostos presentes em duas imagens, foi utilizada a distância euclidiana, também empregada por Tomillo *et al.* (2024) em sua análise comparativa de algoritmos de verificação facial aplicados a veículos aéreos não tripulados. Essa métrica consiste em um método de cálculo da distância entre dois pontos em um espaço multidimensional, sendo amplamente utilizada na verificação de similaridade facial. Nesse contexto, valores menores indicam maior similaridade entre os rostos comparados.

2.4 Execução.

2.4.1 Ambiente de desenvolvimento.

Para a execução prática do trabalho, foi utilizado o ambiente de desenvolvimento do Jupyter, o Jupyter Notebook.

O código foi inicialmente escrito e testado em um notebook de uso pessoal. Após os testes iniciais, a execução completa do projeto foi realizada em um outro computador, cujas especificações estão no quadro 1.

Quadro 1 - Especificações do computador utilizado para execução do trabalho

SISTEMA OPERACIONAL	LINUX UBUNTU 20.04.1
CPU	AMD RYZEN 7 5800X COM 8 NÚCLEOS, 3,8GHZ
GPU	NVIDIA RTX 3080 COM 10GB DE VRAM

2.4.2 Pré-processamento.

Antes do treinamento dos modelos de classificação, da aplicação dos modelos de geração de *face embeddings* e da leitura de texto das imagens pelo modelo de OCR, foi necessário realizar algumas transformações nas imagens utilizadas nessas etapas, com o objetivo de maximizar a eficiência do processo.

Para as imagens de documentos utilizadas no treinamento dos modelos de classificação, foram aplicadas as etapas de redimensionamento, normalização (utilizando os valores recomendados para cada modelo) e conversão para escala de cinza, por meio da biblioteca OpenCV. Além disso, com o intuito de ampliar a base de dados e tornar o modelo mais robusto a situações em que as imagens não estejam no formato ideal, foi empregada a técnica de *data augmentation*. Para cada imagem original, foram geradas cinco novas variações, com modificações de brilho, rotação, ruído, oclusão e distorção de perspectiva.

No caso das imagens destinadas ao OCR, o pré-processamento consistiu em aumento do DPI, conversão para escala de cinza, remoção de ruído e aumento de contraste, também utilizando a biblioteca OpenCV. Como, para o OCR, os arquivos *ground truth* (GT) são igualmente relevantes, foi realizada uma limpeza dos arquivos, que originalmente continham dados de posição do texto e cabeçalhos. Após esse processo, permaneceram apenas os textos correspondentes às imagens, em letras maiúsculas.

Para as imagens do LFW *Dataset*, o pré-processamento foi executado após a detecção e extração dos rostos dos pares de imagens. Nessa etapa, os rostos foram reposicionados e centralizados, procedimento realizado com o uso da biblioteca MediaPipe, por meio da função FaceMesh. Em seguida, ambas as imagens foram redimensionadas conforme os tamanhos exigidos por cada modelo.

2.4.3 Treinamento e teste dos modelos de classificação.

Para o treinamento dos modelos, o conjunto de dados foi dividido de forma a destinar 80% dos dados para o treino e 20% para os testes. Dessa forma, foram utilizadas 28.785 imagens originais para o treinamento, que, após a aplicação da técnica de *data augmentation*, resultaram em 143.925 amostras, e 5.757 imagens foram reservadas para teste. Definiu-se um tamanho de lote (*batch size*) de 32 para os modelos, com exceção do EfficientNet, que, devido às limitações de alocação de memória na GPU, foi treinado com *batch size* igual a 16. O treinamento foi configurado para 50 épocas, estabelecidas como limite máximo. No entanto, foi implementado um mecanismo de *early stopping* para evitar o *overfitting*. Esse mecanismo foi definido com paciência de três épocas e delta mínimo de 0,01, de modo que, caso o modelo não apresente melhora na função de perda (*loss*) após três épocas consecutivas, o treinamento é interrompido automaticamente.

Com os modelos já treinados, foi desenvolvido um script para realizar as previsões utilizando os dados de teste, cujos resultados são apresentados na seção de Resultados.

2.4.4 Avaliação do modelo de OCR e validação de informações.

Com o objetivo de demonstrar a eficácia do modelo na detecção e extração de texto, foi desenvolvido um script que percorre as pastas contendo as imagens originais e as pré-processadas, realiza o OCR, extrai o texto de cada imagem e o compara com o respectivo arquivo *ground truth* (GT), possibilitando a medição do desempenho do modelo de OCR em imagens com e sem pré-processamento.

Após essa etapa de avaliação, foi criado outro script para realizar o OCR nas imagens, obtendo o conteúdo textual e comparando-o com parâmetros de referência, como nome, CPF e

data de nascimento. Além disso, foi implementada uma verificação adicional, indo além da simples busca por *strings* de CPF e data de nascimento. Essa verificação considera a possibilidade de encontrar os valores apenas pelos números, permitindo substituir traços, pontos ou barras por espaços em branco, ou ainda remover esses caracteres especiais e unir a sequência numérica, de modo a compensar possíveis pequenas falhas na extração de texto pelo modelo.

2.4.5 Geração dos *face embeddings* e melhores métricas para cada modelo.

As imagens das classes foram organizadas em pares formados por pessoas iguais e diferentes, totalizando 16 mil pares sendo metade positivos (mesma pessoa) e metade negativos (pessoas distintas). Em seguida, foi desenvolvido um script responsável por detectar os rostos em ambas as imagens, realizar o pré-processamento e, posteriormente, gerar os *face embeddings* por meio de cada modelo. Para cada par, foi calculada a distância euclidiana entre os vetores gerados e os resultados foram armazenados. Por fim, foi criado um arquivo contendo essas informações para cada modelo, conforme ilustrado na figura 1.

Figura 1 - Parte da lista do modelo FaceNet contendo o caminho das imagens, a informação se o par se trata ou não da mesma pessoa e a distância euclidiana calculada.

```
{
  "img1_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/Tony_Blair/Tony_Blair_0003.jpg",
  "img2_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/Tony_Blair/Tony_Blair_0087.jpg",
  "same_person": true,
  "distance": 13.673136744311616
},
{
  "img1_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/Junichiro_Koizumi/Junichiro_Koizumi_0005.jpg",
  "img2_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/Junichiro_Koizumi/Junichiro_Koizumi_0008.jpg",
  "same_person": true,
  "distance": 10.14124441624863
},
{
  "img1_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/David_McCallum/David_McCallum_0001.jpg",
  "img2_path": "/home/diego/lfw_datasetttt/lfw-deepfunneled/lfw-deepfunneled/Nona_Gaye/Nona_Gaye_0001.jpg",
  "same_person": false,
  "distance": 16.809793748827342
},
}
```

Fonte: autoria própria.

Com as listas criadas contendo os pares de imagens e as respectivas distâncias calculadas para cada modelo, o último passo foi determinar um *threshold* (limiar) a ser utilizado como critério de decisão para classificar se duas imagens pertencem ou não à mesma pessoa.

Para isso, inicialmente foi calculada a média das distâncias entre os pares positivos de cada modelo. A partir desse valor, foram testados limiares próximos a essa média com o objetivo de encontrar aquele que proporcionasse o melhor equilíbrio entre precisão e acurácia. Dessa forma, buscou-se um valor que maximizasse a correta identificação dos pares, porém dando um peso maior para precisão, por ser melhor obter um falso negativo do que falso positivo nesta tarefa.

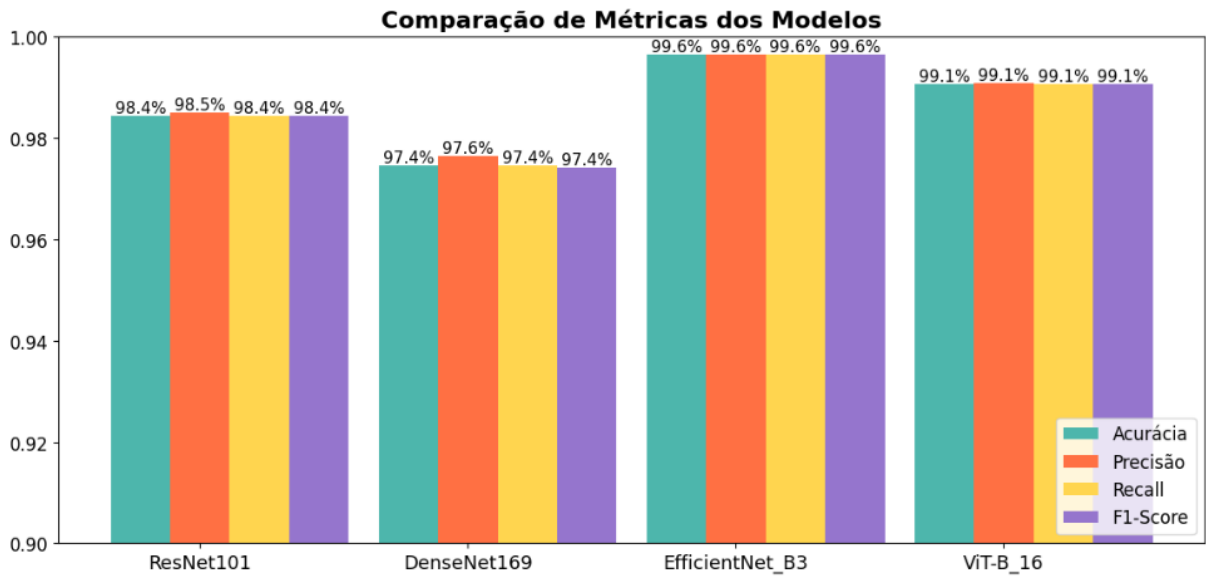
3. RESULTADOS E DISCUSSÕES.

Essa seção apresenta os resultados obtidos a partir da execução dos algoritmos desenvolvidos. Os resultados são detalhados para promover uma compreensão clara dos desempenhos de cada abordagem adotada. A partir da análise desses dados, será possível definir os modelos mais adequados para representar cada tarefa do projeto.

3.1 Classificação de documentos.

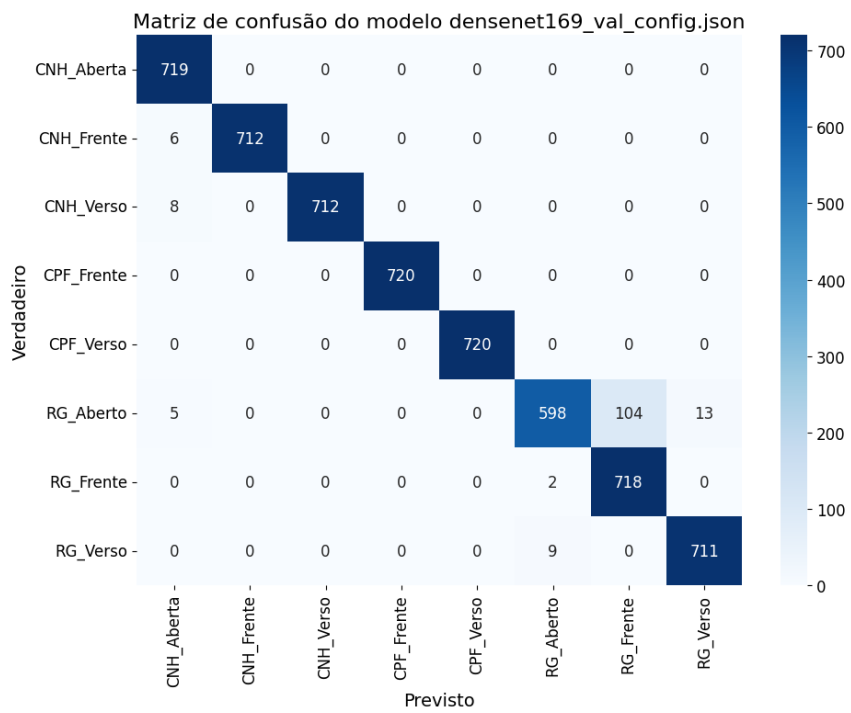
Os modelos de classificação de documentos apresentaram resultados consistentes em todas as arquiteturas avaliadas, com todos os modelos apresentando uma acurácia acima de 96%. Entre os modelos testados, o EfficientNet-B3 destacou-se, alcançando 99,6% de acurácia, demonstrando excelente desempenho e generalização, como é possível ver de forma detalhada na figura 2. Entretanto, ao analisar detalhadamente as figuras 3 e 4, contendo as matrizes de confusão do modelo com pior desempenho e modelo com melhor desempenho, observa-se que os modelos apresentaram uma certa dificuldade na classificação de imagens contendo mais de um documento simultaneamente ou documentos com partes abertas ou parcialmente visíveis, denominadas neste trabalho de classes abertas (como “RG aberto” e “CNH aberta”). Com isso, é possível perceber que quando há mais de um documento na mesma imagem, o modelo só consegue tentar classificar um deles, o que pode limitar o desempenho em cenários mais complexos, como por exemplo tentar classificar vários documentos presentes em uma única imagem. Com isso, para uma implementação mais eficaz em cenários menos controlados, poderá ser necessário extrair individualmente os documentos de cada imagem por meio de um modelo de detecção, e, em seguida, submeter apenas a imagem do documento extraída ao modelo classificador.

Figura 2 - Gráfico onde estão presentes as informações de desempenho de cada modelo.



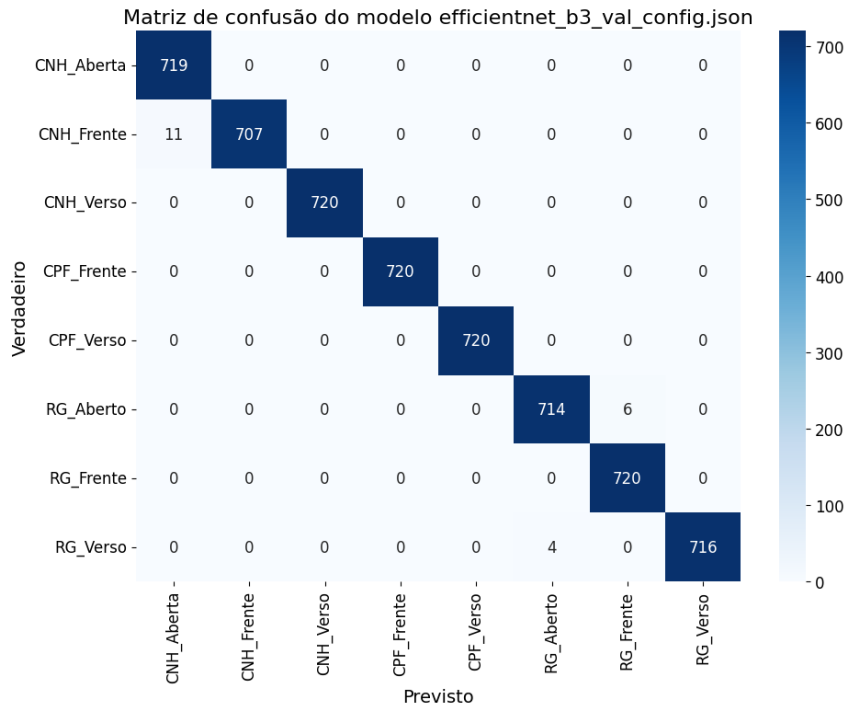
Fonte: autoria própria.

Figura 3 - Matriz de confusão do modelo DenseNet 169.



Fonte: autoria própria.

Figura 4 - Matriz de confusão do modelo EfficientNet b3.



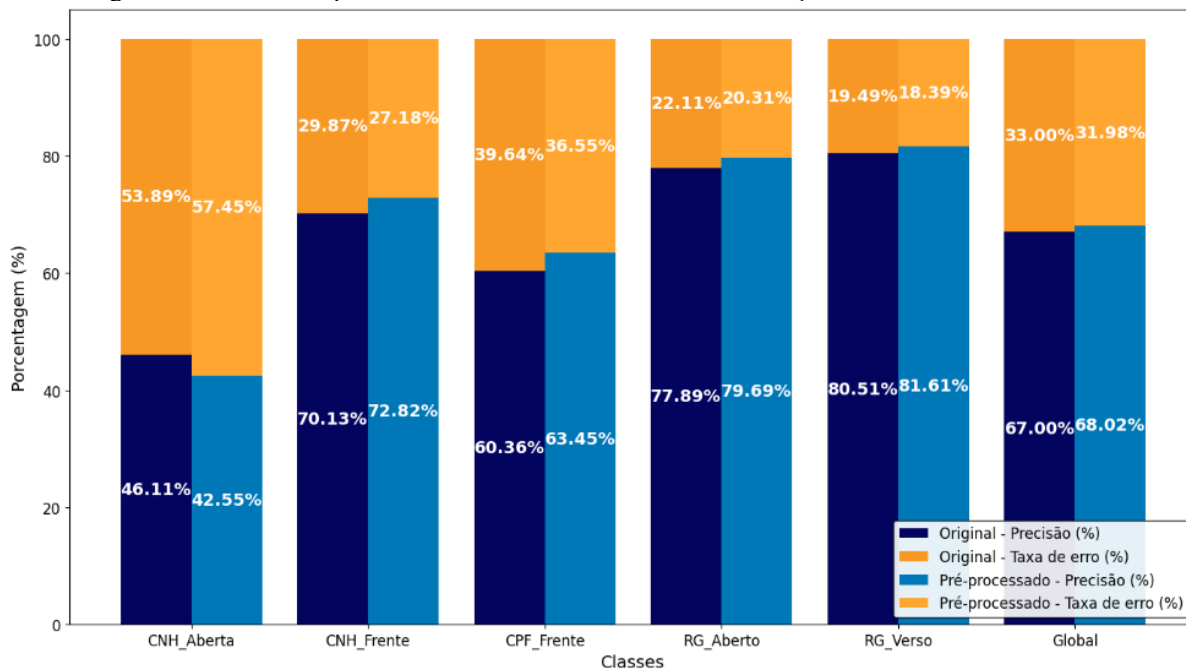
Fonte: autoria própria.

3.2 Validação de informações com OCR.

Com a extração concluída, o texto coletado e armazenado foi comparado com os arquivos GT do *dataset*. O objetivo dessa comparação foi mensurar a taxa de erro e a precisão do modelo, considerando dois cenários distintos: imagens sem qualquer tipo de alteração e imagens submetidas ao pré-processamento definido anteriormente. A Figura 5 mostra a relação

de desempenho do OCR no aspecto classe por classe e geral, mostrando o resultado obtido com as imagens com e sem pré-processamento.

Figura 5 - Gráfico comparativo dos resultados obtidos com a extração de texto com PaddleOCR.



Fonte: autoria própria.

A figura 5 revela que o modelo de OCR apresentou desempenho satisfatório na maioria das classes, com cinco das seis classes avaliadas alcançando precisão acima de 60%, considerando o conteúdo textual total dos documentos. O pré-processamento contribuiu para um leve aumento geral na precisão média, porém a melhoria foi modesta, a maior variação positiva observada foi de 3% em relação à versão sem pré-processamento. Em contrapartida, uma das classes (“CNH aberta”) apresentou até mesmo uma redução de desempenho, obtendo precisão de 46% sem pré-processamento e 42% com. Esses resultados sugerem que a aplicação de um pré-processamento mais específico, com metodologias distintas para cada classe, pode contribuir para o aprimoramento da precisão do OCR.

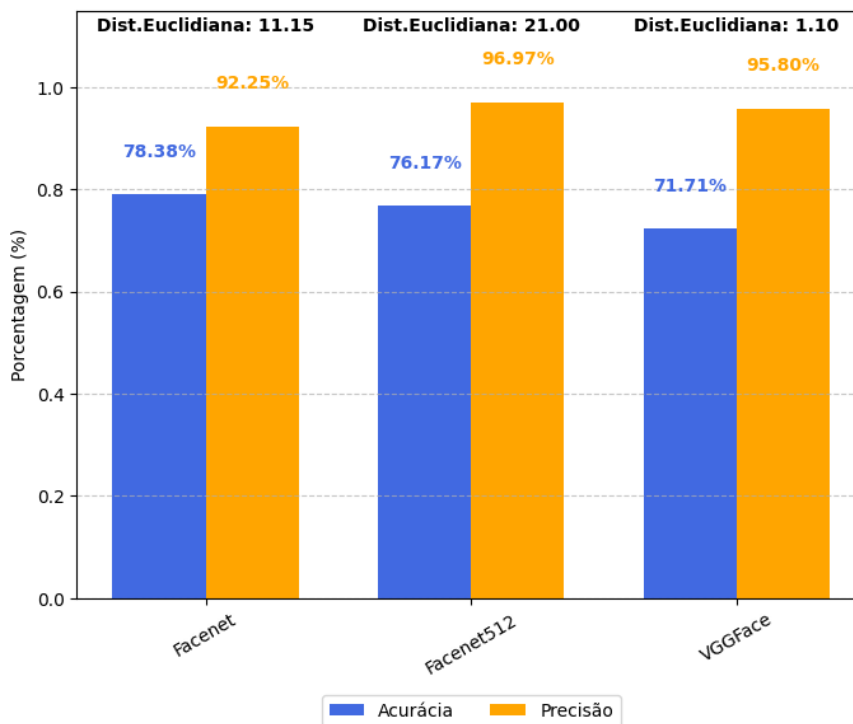
3.3. Verificação facial

Após a realização dos testes para determinar o valor de limiar (*threshold*) mais adequado da distância euclidiana, capaz de separar com maior precisão os pares positivos e negativos, esta subseção apresenta os resultados obtidos.

A Figura 6 ilustra graficamente os desempenhos dos modelos de verificação facial avaliados. No gráfico, são apresentados os modelos FaceNet, FaceNet512 e VGG-Face,

acompanhados das respectivas métricas de acurácia e precisão alcançadas, considerando o limiar de distância euclidiana determinado para cada um.

Figura 6 - Gráfico comparativo de resultados dos modelos de verificação facial.



Fonte: autoria própria.

A Figura 6 mostra que o desempenho dos três modelos foi semelhante, sendo o FaceNet512 o que obteve a maior precisão, enquanto o FaceNet apresentou a maior acurácia. Contudo, o primeiro foi escolhido como modelo representativo para esta tarefa no algoritmo final, uma vez que apresentou maior precisão e acurácia similar à do segundo modelo citado. Essa escolha se justifica pela importância da classificação correta, pois, neste contexto, é preferível minimizar os falsos positivos.

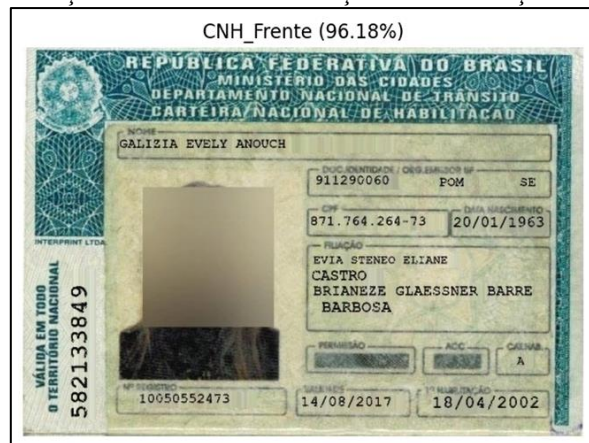
É necessário destacar que, embora a precisão tenha sido alta e a acurácia satisfatória, os testes foram conduzidos com pares de imagens de mesma resolução e sem o uso de imagens extraídas de documentos de identidade. Assim, o experimento configura-se como um cenário ideal, no qual as imagens possuem a mesma resolução e características fornecidas pelo dataset de origem.

Para um cenário mais especializado, é possível aplicar a mesma metodologia adotada neste trabalho, desde que se disponha de uma base de dados mais específica, por exemplo, um conjunto de imagens de documentos contendo rostos, ou de pessoas segurando seus documentos com foto ao lado do próprio rosto.

3.4 Algoritmo final

Uma vez com os modelos definidos, os scripts de cada tarefa foram unificados para que seja feita uma execução demonstrativa com dados pessoais reais, porém por falta de autorização de expor dados de pessoas neste projeto, foi apenas utilizada a imagem de um dos autores. As figuras 7,8,9 e 10 exibem de forma ilustrativa o resultado da execução de cada tarefa e o resultado final.

Figura 7 - Exibição do resultado da execução da classificação de documento.



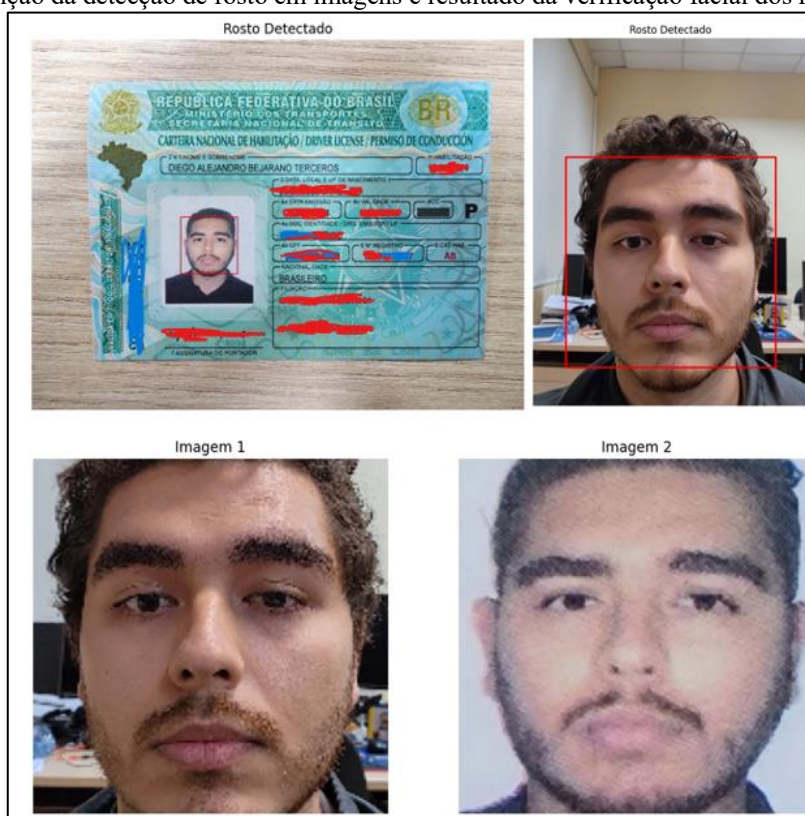
Fonte: autoria própria.

Figura 8 - Exibição do texto detectado na imagem do documento pré-processada junto com o resultado da validação das informações.



Fonte: autoria própria.

Figura 9 - Exibição da detecção de rosto em imagens e resultado da verificação facial dos rostos detectados.



Fonte: autoria própria.

Figura 10 - Resumo dos resultados de cada tarefa

```
CLASSE PREVISTA: CNH FRENTE com 96.18% de confiança  
NOME VÁLIDO: True  
DATA DE NASCIMENTO VÁLIDA: True  
CPF VÁLIDO: True  
MESMA PESSOA: True, COM DISTANCIA EUCLIDIANA DE: 15.085263222288054
```

Fonte: autoria própria.

A utilização deste algoritmo torna o fluxo de validação documental mais eficiente, reduzindo a quantidade de pessoas e o tempo necessários para a realização dessas validações. Na máquina utilizada para este projeto, a verificação de todas as informações levou aproximadamente 15 segundos, podendo ser executada de forma ainda mais rápida em equipamentos com configurações superiores.

4. CONSIDERAÇÕES FINAIS

Este projeto demonstrou que a aplicação de modelos de aprendizado de máquina pode otimizar significativamente o processo de validação documental, reduzindo o tempo necessário para análise de documentos de horas ou dias para poucos segundos (utilizando a configuração de máquina adotada neste trabalho). Foram exploradas abordagens para as três principais etapas

desse processo: classificação de documentos, extração e validação de dados textuais e verificação facial.

Os resultados indicaram que o sistema foi eficaz em cenários ideais, apresentando boa precisão na classificação de documentos, melhoria na extração de texto após o pré-processamento e desempenho satisfatório na verificação facial em condições controladas. No entanto, algumas limitações foram identificadas, tais como: classificação limitada a uma única classe por imagem; variação na eficácia da extração de texto entre diferentes classes de documentos; e desempenho da verificação facial restrito a cenários controlados, com imagens de qualidade similar, sem avaliação em condições mais desafiadoras.

Além de apresentar uma solução funcional, este trabalho também se propõe a servir como um guia prático para empresas, instituições públicas ou organizações de menor porte, que muitas vezes não dispõem de grandes volumes de dados ou recursos computacionais para treinar modelos do zero. Todo o desenvolvimento foi realizado com modelos pré-treinados e bases de dados públicas, o que reforça a viabilidade de replicação e aplicação do método em contextos reais, mesmo com infraestrutura limitada.

Com isso, o projeto pode atuar como um ponto de partida robusto para o desenvolvimento de soluções personalizadas, permitindo que cada uma das três tarefas abordadas, classificação de documentos, extração e validação textual e verificação facial, possa ser aprofundada e especializada de acordo com as necessidades de cada instituição. Essa flexibilidade possibilita que o trabalho seja expandido tanto de forma independente, com foco em uma tarefa específica, quanto de forma integrada, em fluxos automatizados de validação de identidade mais complexos.

Apesar das limitações identificadas, os resultados obtidos confirmam o potencial da abordagem para otimizar a validação documental. Com os aprimoramentos propostos para trabalhos futuros, como a segmentação de múltiplos documentos em uma única imagem, o pré-processamento otimizado para OCR e a ampliação do dataset para verificação facial, espera-se que o sistema se torne mais robusto e aplicável a um número maior de cenários, incluindo aqueles com maior complexidade e variabilidade nas imagens analisadas.

REFERÊNCIAS

ALMEIDA, Leonardo de Borges; SILVA, Marcus Vinicius da. Estudo comparativo de redes neurais convolucionais para a classificação da qualidade de imagens de documentos de identidade. 2022. Disponível em: <<https://bdm.unb.br/handle/10483/31168>>. Acesso em: 4 mar. 2025.

BEERTEN, Toon. *OCR comparison: Tesseract versus EasyOCR vs PaddleOCR vs MMOCR*. Medium. Disponível em: <<https://toon-beerten.medium.com/ocr-comparison-tesseract-versus-easyocr-vs-paddleocr-vs-mmocr-a362d9c79e66>>. Acesso em: 2 mar. 2025.

DIAS, T.; SANO, H.; MEDEIROS, M. F. M. D. **Inovação e tecnologias da comunicação e informação na administração pública**. [s.l.: s.n.], 2019. Disponível em:<https://repositorio.enap.gov.br/bitstream/1/4284/1/8_Livro_Inova%C3%A7%C3%A3o%20e%20tecnologias%20da%20comunica%C3%A7%C3%A3o%20e%20informa%C3%A7%C3%A3o%20na%20administra%C3%A7%C3%A3o%20p%C3%BAblica.pdf>. Acesso em: 24 mar. 2025.

FARIAS, Walisson Nascimento de. **APERFEIÇOANDO O RECONHECIMENTO ÓPTICO DE CARACTERES EM IMAGENS DE DOCUMENTOS PESSOAIS**. [s.l.: s.n.], 2023. Disponível em:< <https://dspace.sti.ufcg.edu.br/handle/riufcg/34848>>. Acesso em: 12 jan. 2025.

FIRMANSYAH, A.; KUSUMASARI, T. F.; ALAM, E. N. *Comparison of Face Recognition Accuracy of ArcFace, Facenet and Facenet512 Models on Deepface Framework*. In: CONFERÊNCIA INTERNACIONAL DE CIÊNCIA DA COMPUTAÇÃO, TECNOLOGIA DA INFORMAÇÃO E ENGENHARIA DE 2023 (ICCOSITE). 16 fev. 2023. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/10127799>>. Acesso em: 5 mar. 2025.

JUNIOR, Guanis de Barros Vilela; LIMA, Bráulio; PEREIRA, Adriano de Almeida; *et al.* MÉTRICAS UTILIZADAS PARA AVALIAR A EFICIÊNCIA DE CLASSIFICADORES EM ALGORITMOS INTELIGENTES. **Revista CPAQV - Centro de Pesquisas Avançadas em Qualidade de Vida**, v. 14, n. v14n2, p. 1–1, 2022. Disponível em: <https://www.researchgate.net/publication/359541310_METRICAS_UTILIZADAS_PARA_AVALIAR_A_EFICIENCIA_DE_CLASSIFICADORES_EM_ALGORITMOS_INTELIGENTES>. Acesso em: 20 nov. 2024.

PARK, Youngki ; SHIN, Youhyun. *Gradual OCR: An Effective OCR Approach Based on Gradual Detection of Texts*. **Mathematics**, v. 11, n. 22, p. 20, 2023. Disponível em: <<https://www.mdpi.com/2227-7390/11/22/4585>>. Acesso em: 24 fev. 2025.

PRANAV, Jeevan ; SETHI, Amit. *Which Backbone to Use: A Resource-efficient Domain*

Specific Comparison for Computer Vision. **arXiv.org**, p. 10, 2024. Disponível em: <<https://arxiv.org/abs/2406.05612v1>>. Acesso em: 5 fev. 2025.

YANG, Yuan; ZHANG, Lin; DU, Mingyu; *et al.* *A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions*. **Computers in Biology and Medicine**, v. 139, p. 27, 2021. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC8461289/pdf/main.pdf>>. Acesso em: 15 fev. 2025.

ZHAO, Peng; LI, Chen; RAHAMAN, Md Mamunur; *et al.* *A Comparative Study of Deep Learning Classification Methods on a Small Environmental Microorganism Image Dataset (EMDS-6): from Convolutional Neural Networks to Visual Transformers*. **arXiv.org**, p. 24, 2021. Disponível em: <<https://arxiv.org/abs/2107.07699>>. Acesso em: 27 fev. 2025.

TOMILLO, J. D.; CALERO, J. M. A.; WANG, Q. *Face Verification Algorithms for UAV Applications: An Empirical Comparative Analysis*. **JOURNAL OF COMMUNICATIONS SOFTWARE AND SYSTEMS**, v. 20, n. 1, mar. 2024. Disponível em: <<https://jcoms.fesb.unist.hr/10.24138/jcomss-2023-0165/>>. Acesso em: 29 mar. 2025.

VEDHAVIYASSH, D.R.; SUDHAN, R.; SARANYA, G.; *et al.* *Comparative Analysis of EasyOCR and TesseractOCR for Automatic License Plate Recognition using Deep Learning Algorithm*. **IEEE Xplore**, 2022. Disponível em: <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10009215>>. Acesso em: 1 mar. 2025.