

# StoneScan: Classificação Automatizada de Rochas Ornamentais por Redes Neurais Convolucionais

Rafael Vargas Mesquita dos Santos<sup>1</sup>, Sofia Pereira Bachetti Sartorio<sup>1</sup>,  
Isabella Fornazier Silotti<sup>1</sup>, Alessandro Mion Batista<sup>1</sup>

<sup>1</sup>Instituto Federal do Espírito Santo (IFES)

rafaelv@ifes.edu.br, sofiasartrio@gmail.com

isabella@granimaster.com.br, alessandrombta@gmail.com

**Abstract.** *This work proposes a computer vision system for automatic classification of ornamental stones from smartphone images. Using transfer learning with ResNet18 and stratified 5-fold cross-validation, the model identifies five classes. Data augmentation and grid search yielded mean accuracy of 98.62% ( $\pm 0.83\%$ ) and Cohen's kappa of 0.9826, surpassing a linear probe baseline. Results demonstrate the feasibility of CNNs to discriminate visual patterns in geosciences and support a mobile application for in-field identification.*

**Resumo.** *Este trabalho propõe um sistema de visão computacional para classificação automática de rochas ornamentais a partir de imagens de smartphones. Com transfer learning (ResNet18) e validação cruzada estratificada (5 folds), o modelo identifica cinco classes. Aumento de dados e busca em grade resultaram em acurácia média de 98,62% ( $\pm 0,83\%$ ) e Kappa de 0,9826, superando um baseline linear. Os resultados, no conjunto avaliado, evidenciam a viabilidade de CNNs na discriminação de padrões visuais e sua integração em um aplicativo móvel, indicando potencial de uso em campo.*

## 1. Introdução

O avanço das tecnologias de Inteligência Artificial (IA), em particular o *Deep Learning*, tem permitido sua aplicação em setores com processos manuais e intensivos em conhecimento, como a indústria de rochas ornamentais [Cheng et al. 2022]. O reconhecimento de padrões visuais, antes dependente da experiência humana, pode ser automatizado por algoritmos de Visão Computacional, impulsionando ferramentas para inspeção e classificação visual em diversos domínios industriais.

A identificação de rochas ornamentais envolve a análise de texturas, cores, veios e inclusões, elementos que determinam aplicação estética e funcional, bem como valor comercial. Para profissionais não especializados, identificar por fotografias é complexo e propenso a erros; a ausência de ferramentas acessíveis pode levar a escolhas inadequadas de materiais e retrabalho em projetos.

A ruptura do paradigma de análise visual manual ocorreu com a consolidação das Redes Neurais Convolucionais (CNNs), que permitem o aprendizado de características diretamente a partir dos dados [LeCun et al. 2015]. A extração automática de atributos reduz a necessidade de engenharia manual de características, processo tradicionalmente

dependente de conhecimento especializado [Soares et al. 2025]. A evolução das arquiteturas de *Deep Learning* viabilizou também a execução desses modelos em dispositivos móveis, relevante para inspeções em campo.

O objetivo deste trabalho é desenvolver e validar um modelo de classificação de rochas ornamentais baseado em *transfer learning*, avaliando sua capacidade de discriminar cinco classes de materiais com características visuais similares. Adicionalmente, propõe-se a integração deste modelo em uma aplicação móvel que permita a identificação de rochas a partir de fotografias capturadas em campo.

## 2. Fundamentação Teórica

A digitalização tem se tornado estratégica no setor de construção e mineração [Kumar et al. 2023]. O segmento de rochas ornamentais possui importância econômica na construção civil e no design de interiores. O Brasil ocupa posição de destaque, figurando entre os principais produtores e exportadores de granitos, mármore e quartzitos.

Apesar da maturidade industrial, o setor enfrenta desafios práticos de identificação. As rochas ornamentais apresentam variabilidade visual dentro de uma mesma classificação comercial e diferenças sutis entre materiais distintos, dependendo da mineralogia, tamanho e distribuição dos grãos, e da presença de estruturas secundárias. As dificuldades incluem dependência de condições de iluminação, avaliação subjetiva de textura e interpretação de características microgeológicas. A nomenclatura comercial frequentemente diverge da classificação geológica, gerando confusões e especificações incorretas de materiais.

O uso de IA para reconhecimento de padrões em materiais naturais evoluiu de soluções baseadas em processamento de imagens tradicionais, como histogramas e análise de cores, para abordagens automatizadas baseadas em aprendizado profundo [Liu et al. 2019]. Métodos tradicionais como SIFT (*Scale-Invariant Feature Transform*) e HOG (*Histogram of Oriented Gradients*), embora eficazes para objetos com geometria rígida, apresentam limitações ao lidar com a complexidade das superfícies naturais. A utilização de *transfer learning* consolidou-se como técnica eficaz para problemas com conjuntos de dados limitados [Garg and Dhiman 2021], permitindo aproveitar conhecimento aprendido em grandes bases de dados genéricas como o ImageNet.

A arquitetura ResNet (*Residual Network*) introduziu as conexões residuais (*skip connections*), que permitem o treinamento de redes profundas ao mitigar o problema do desvanecimento do gradiente [He et al. 2016]. A ResNet18, com suas 18 camadas, oferece equilíbrio entre capacidade de representação e custo computacional [Silva 2021]. Essa característica, aliada à eficiência no uso de memória, torna a arquitetura adequada para execução em tempo real em dispositivos com recursos limitados, como smartphones e tablets comumente utilizados em inspeções de campo.

A técnica de aumento de dados (*data augmentation*) consiste em aplicar transformações nas imagens de treino, como rotações, espelhamentos, recortes e ajustes de luminosidade, gerando variações que forcem a rede a aprender características invariantes às condições de captura [Shorten and Khoshgoftaar 2019]. Essa abordagem reduz o risco de sobreajuste (*overfitting*) e melhora a capacidade de generalização do modelo em cenários reais [Srivastava et al. 2014]. A combinação de *transfer learning* com aumento de dados tem se mostrado eficaz em domínios com escassez de dados rotulados.

### 3. Metodologia

A metodologia adotada segue um fluxo de desenvolvimento iterativo e multidisciplinar (Figura 1), abrangendo desde a aquisição dos dados até o treinamento e a avaliação do modelo. A equipe é composta por Ext-F (Técnico em Mineração), Ext-E (Sistemas de Informação) e parceira empresarial do ramo de comercialização de chapas.

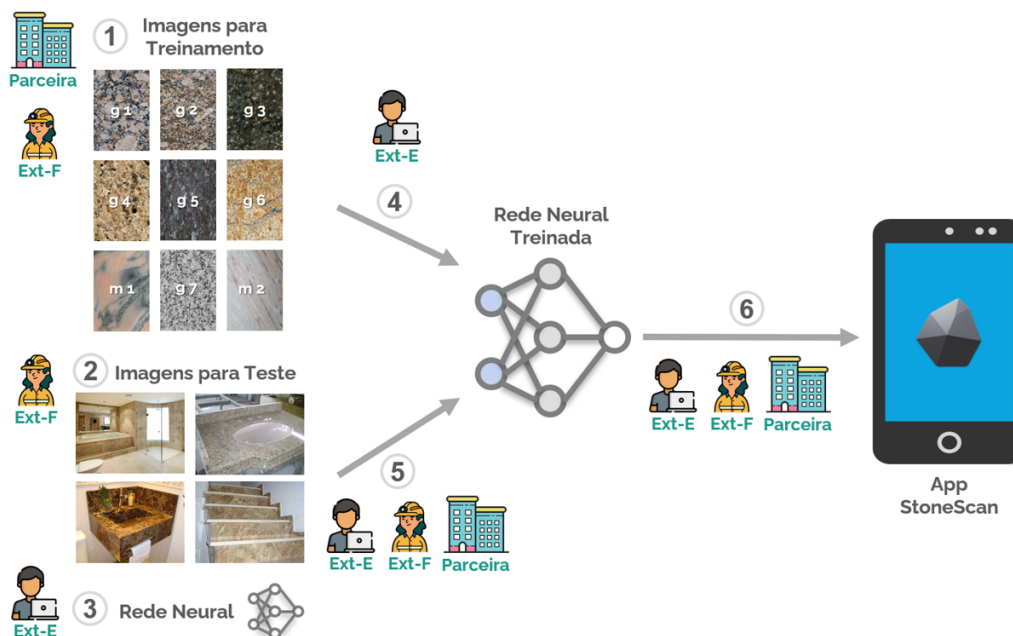


Figura 1. Fluxo metodológico do sistema proposto.

O fluxo compreende seis atividades: (1) coleta de imagens para treinamento, liderada por Ext-F com apoio da parceira; (2) coleta de um conjunto separado de imagens para teste; (3) escolha da arquitetura de rede neural por Ext-E; (4) treinamento do modelo com ajuste de hiperparâmetros; (5) validação colaborativa entre todos os envolvidos; e (6) implementação do aplicativo StoneScan, com contribuições de Ext-F e da parceira nas versões beta.

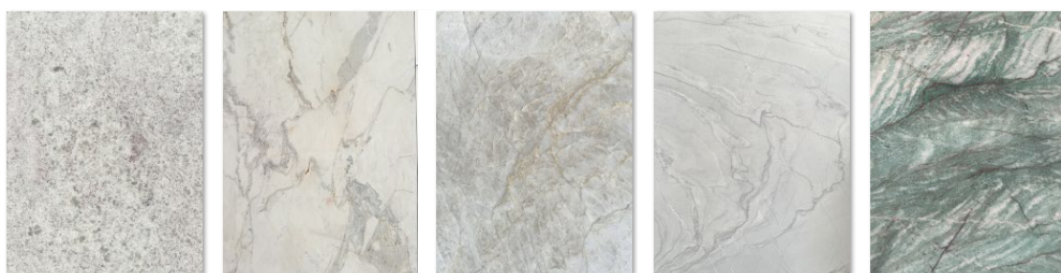
#### 3.1. Conjunto de Dados

Foram capturadas aproximadamente 150 imagens para cada uma das cinco classes: Granito Branco Itaúnas, Mármore Matarazzo, Quartzito Perla, Quartzito Wakanda e Quartzito Verde Gaya, em diferentes distâncias, ângulos e condições de iluminação, utilizando dispositivos móveis.

O contexto industrial contempla ~20 classes de rochas ornamentais; a seleção inicial de cinco classes deve-se à disponibilidade dos materiais durante o processo produtivo.

O protocolo incluiu capturas em diferentes distâncias (detalhes de textura e visão geral), ângulos (incluindo inclinações para registrar brilho) e iluminação (natural e artificial), visando reduzir viés e ampliar a robustez do modelo.

A Figura 2 apresenta as classes utilizadas. Apesar de pertencerem a categorias geológicas distintas, alguns materiais apresentam similaridades visuais que tornam a classificação desafiadora para observadores não especializados.



**Figura 2. Classes de rochas: (a) Granito Branco Itaúnas, (b) Mármore Matarazzo, (c) Quartzito Perla, (d) Quartzito Wakanda, (e) Quartzito Verde Gaya.**

### 3.2. Arquitetura e Treinamento

Foi adotada a arquitetura ResNet18 com pesos pré-treinados no ImageNet, uma base de dados com mais de 14 milhões de imagens em 1.000 categorias. As camadas convolucionais foram congeladas para funcionar como extrator de características, aproveitando o conhecimento prévio da rede. A camada final foi substituída por camadas *dropout* (com taxas de 50% e 30%) e camadas totalmente conectadas para classificação das cinco classes. Esta estratégia reduz o número de parâmetros treináveis para aproximadamente 68.000, diminuindo o risco de sobreajuste e acelerando o treinamento.

Como *baseline*, implementou-se uma sonda linear (*Linear Probe*): uma única camada linear sobre as *features* da ResNet18 congelada, sem aumento de dados nem otimização de hiperparâmetros, medindo a qualidade intrínseca das representações pré-treinadas.

Para avaliação do modelo, foi empregada validação cruzada estratificada com  $k = 5$  partições. O pré-processamento incluiu técnicas de aumento de dados aplicadas apenas durante o treinamento: recortes aleatórios redimensionados, inversões horizontais e verticais, rotações de até 30°, ajustes de brilho, contraste e saturação, e desfoque gaussiano. Ressalta-se que a avaliação foi realizada sobre as imagens originais, sem transformações de aumento, garantindo que as métricas reflitam o desempenho real do modelo.

A seleção de hiperparâmetros foi realizada por busca em grade (*Grid Search*), avaliando combinações de taxa de aprendizado (0.0001, 0.0005, 0.001), tamanho de lote (8, 16, 32) e otimizador (Adam, SGD, AdamW), totalizando 27 combinações. Cada combinação foi submetida à validação cruzada completa, resultando em 135 execuções de treinamento independentes. A Tabela 1 apresenta a configuração selecionada.

## 4. Resultados

Os repositórios contendo os dados e scripts de treinamento<sup>1</sup> e o aplicativo desenvolvido<sup>2</sup> estão publicamente disponíveis.

A validação cruzada obteve acurácia de 98,62% ( $\pm 0,83\%$ ). A Tabela 2 apresenta precisão, revocação e *F1-score* por classe. Destaca-se que o Granito Branco Itaúnas obteve 100% de revocação (todas as amostras corretamente identificadas), enquanto o Quartzito Verde Gaya apresentou as métricas mais equilibradas.

<sup>1</sup><https://github.com/ravarmes/StoneScan-Python>

<sup>2</sup><https://github.com/ravarmes/StoneScan-ReactNative>

**Tabela 1. Parâmetros do treinamento.**

Parâmetro	Valor
Arquitetura	ResNet18 (pré-treinado)
Taxa de aprendizado	0.001
Tamanho do lote	16
Otimizador	Adam
Função de perda	Cross-Entropy (pesos de classe)
Validação cruzada	$k = 5$ folds
Early stopping	7 épocas
Tamanho da imagem	$224 \times 224$ pixels

**Tabela 2. Métricas por classe.**

Classe	Precisão	Revocação	F1	Suporte
Granito Branco Itaúnas	0,98	1,00	0,99	160
Mármore Matarazzo	0,99	0,97	0,98	156
Quartzito Perla	0,97	0,99	0,98	138
Quartzito Wakanda	0,99	0,98	0,99	134
Quartzito Verde Gaya	1,00	1,00	1,00	207
<b>Média Ponderada</b>	0,99	0,99	0,99	795

O coeficiente Kappa de Cohen foi de 0,9826, indicando concordância quase perfeita entre as predições do modelo e as classes verdadeiras. Este valor supera o limiar de 0,81 tipicamente considerado como concordância forte na literatura.

#### 4.1. Análise de Convergência

A Figura 3 apresenta as curvas de aprendizado. A análise revela convergência consistente e paralelismo entre treino e validação, indicando ausência de sobreajuste significativo.

A Figura 4 apresenta a matriz de confusão agregada, evidenciando que a maioria das amostras foi classificada corretamente.

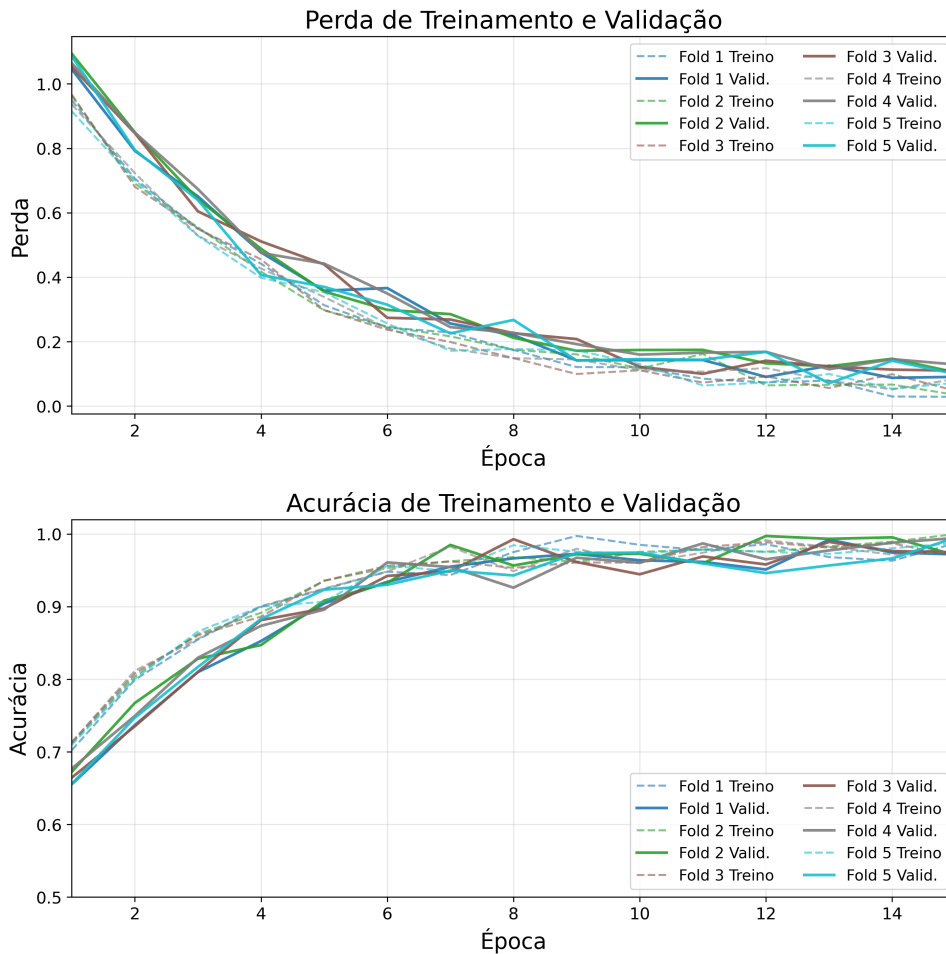
#### 4.2. Discussão

Para contextualizar o desempenho obtido, comparou-se o modelo proposto com o *baseline* por sonda linear, conforme a Tabela 3.

**Tabela 3. Comparação: modelo proposto vs. *baseline* (Linear Probe).**

Métrica	Baseline	Proposto	$\Delta$
Acurácia	$97,36\% \pm 0,62\%$	$98,62\% \pm 0,83\%$	+1,26 p.p.
Kappa de Cohen	0,9667	0,9826	+0,0159
F1 (méd. pond.)	0,9735	0,9861	+0,0126

O *baseline* atinge 97,36% de acurácia, indicando que as *features* do ImageNet são adequadas para este domínio. Contudo, o modelo proposto supera o *baseline* em todas



**Figura 3. Curvas de aprendizado: (superior) perda; (inferior) acurácia. Linhas tracejadas representam treinamento e linhas sólidas representam validação.**

as métricas, evidenciando que a combinação de MLP com *dropout*, aumento de dados e seleção de hiperparâmetros agrega ganho mensurável.

A acurácia obtida pode ser justificada por diversos fatores metodológicos e características do domínio: (1) a validação cruzada estratificada garante resultados não dependentes de uma divisão específica dos dados; (2) múltiplas técnicas de regularização foram empregadas, incluindo *dropout*, *early stopping* e aumento de dados; (3) o congelamento das camadas convolucionais reduziu o número de parâmetros treináveis para aproximadamente 68.000; (4) rochas ornamentais apresentam padrões visuais distintos (granitos com grãos cristalinos, mármore com veios característicos, quartzitos com padrões cromáticos únicos) que facilitam a discriminação por CNNs.

Contudo, limitações são reconhecidas: o desempenho está associado à qualidade do conjunto de dados, e o modelo pode requerer retreinamento para novos materiais.

## 5. Aplicação Prática

O aplicativo móvel integra o modelo a uma interface amigável para identificação de rochas diretamente em campo. A Figura 5 apresenta a interface.

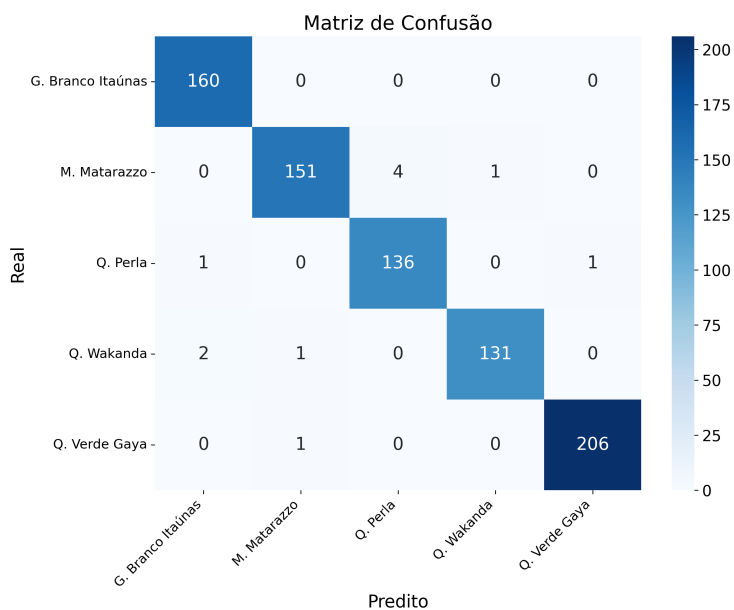


Figura 4. Matriz de confusão agregada dos 5 folds.

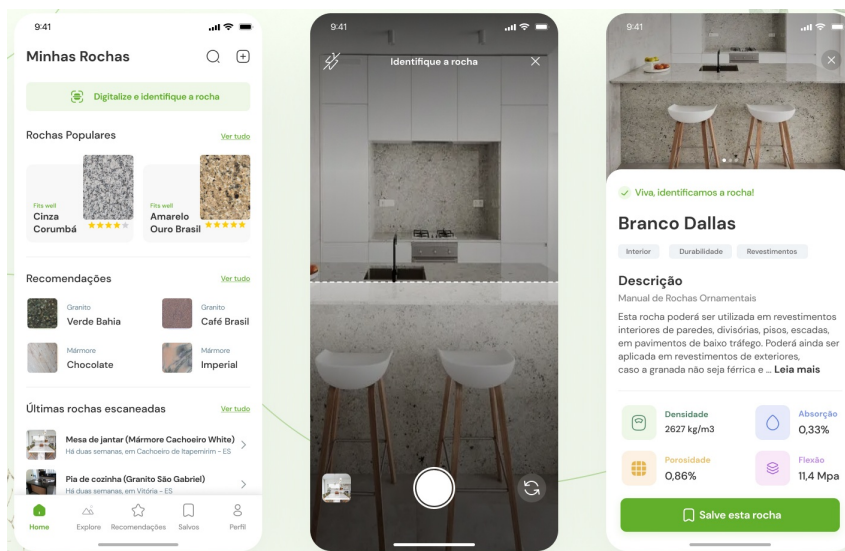


Figura 5. Interface do aplicativo: (a) tela inicial com histórico de escaneamentos, (b) interface de captura de imagem, (c) tela de resultado exibindo a identificação da rocha.

O fluxo de uso consiste em: (1) captura de imagem via câmera ou seleção da galeria; (2) pré-processamento e inferência pelo modelo embarcado; e (3) exibição imediata do resultado. A arquitetura da ResNet18 mostrou-se adequada para execução local, garantindo agilidade e viabilizando o uso em áreas remotas sem conexão com servidores externos.

## 6. Considerações Finais

A contribuição deste trabalho reside na validação de que arquiteturas de *Deep Learning*, ajustadas via *transfer learning*, podem discriminar padrões visuais complexos em materi-

ais geológicos. Os resultados obtidos (acurácia de 98,62% e Kappa de 0,9826), superiores ao *baseline* por sonda linear (97,36%), confirmam que a abordagem proposta agrega valor mensurável mesmo em um cenário onde as representações pré-treinadas já se mostram eficazes.

Trabalhos futuros incluem expansão do número de classes, coleta de dados em condições variadas e a comparação do método proposto com outros *baselines* e arquiteturas no estado da arte.

**Agradecimentos** - Os autores agradecem ao Instituto Federal do Espírito Santo (Ifes) e à Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (Fapes) pelo apoio financeiro, por meio do EDITAL FAPES N° 09/2024 – Extensão Tecnológica, que viabilizou a realização deste trabalho. Agradecem também à empresa GRANIMASTER – Mármore e Granito pela disponibilização das imagens utilizadas na pesquisa.

## Referências

- Cheng, L., Zhang, H., and Wang, Q. (2022). Deep learning applications in material science: a review. *Journal of Materials Research and Technology*, 21:1332–1350.
- Garg, S. and Dhiman, G. (2021). A comprehensive review of transfer learning for image classification. *Artificial Intelligence Review*, 54(1):555–605.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- Kumar, S., Jain, P., and Singh, V. (2023). Digital transformation in the construction industry: A review of ai applications. In *Proceedings of the 2023 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 1201–1207. IEEE.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. (2019). Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, 128(2):261–318.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48.
- Silva, C. F. d. (2021). Redes residuais: Sobre quantidade de camadas e variância intraclasse utilizando aprendizagem não supervisionada. Dissertação de mestrado, Programa de Pós-Graduação em Informática, Universidade Federal da Paraíba (UFPB/CI), João Pessoa, PB.
- Soares, P., Ensina, L., and Foleis, J. (2025). A hybrid approach combining cnn and ensemble algorithms for dermoscopic image classification. In *Anais do XXII Encontro Nacional de Inteligência Artificial e Computacional*, pages 1527–1538, Porto Alegre, RS, Brasil. SBC.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.