

# DETECÇÃO DE FISSURAS EM CONCRETO UTILIZANDO REDES NEURAI CONVOLUCIONAIS: UMA AVALIAÇÃO COMPARATIVA COM APRENDIZADO POR TRANSFERÊNCIA

Luis Roberto Molotto (UEPG) E-mail: 23018226@uepg.br  
Lilian Tais de Gouveia (DENGE-UEPG) E-mail: ltgouveia@uepg.br  
Luciano Jose Senger (DEINFO-UEPG) E-mail: ljsenger@uepg.br

**Resumo:** A detecção de patologias, como fissuras, é essencial em sistemas de avaliação de estruturas de concreto, como pontes, túneis e edifícios. Os métodos tradicionais baseiam-se na inspeção visual por especialistas, sendo frequentemente demorados e potencialmente perigosos, a depender da localização das estruturas. Em substituição ao deslocamento e à inspeção visual por profissionais especializados, imagens coletadas por veículos aéreos não tripulados podem ser utilizadas para o levantamento das condições estruturais e, posteriormente, processadas por sistemas automatizados para identificação de patologias. As imagens coletadas apresentam variações de iluminação e textura e estão sujeitas a ruído, impondo desafios aos algoritmos de aprendizado de máquina voltados à construção de modelos para detecção automática de defeitos. Para esse fim, redes neurais convolucionais têm sido consideradas no desenvolvimento de modelos de classificação. Este trabalho investiga o uso de aprendizado de máquina para detecção e classificação de fissuras em estruturas de concreto, por meio da aplicação de redes neurais convolucionais a imagens de concreto. Um conjunto de 3.822 imagens (1.050 com fissuras; 2.772 sem fissuras) foi redimensionado e normalizado. As arquiteturas VGG16, VGG19, MobileNet, MobileNetV2, ResNet152V2 e DenseNet201 foram avaliadas, com ajuste das camadas finais, aplicação de pesos de classe e parada antecipada de treinamento. A avaliação empregou validação cruzada estratificada com cinco partições e as medidas acurácia, precisão, sensibilidade e F1-score, além de matrizes de confusão para comparação do desempenho dos modelos. O principal achado foi que a MobileNet obteve os maiores valores de sensibilidade e F1-score, minimizando falsos negativos e configurando a melhor escolha para triagem conservadora. O modelo MobileNetV2 obteve a maior precisão (0,980), enquanto a MobileNet apresentou os maiores valores de sensibilidade e F1-score (0,976 e 0,974), reduzindo falsos negativos. DenseNet e VGG16 mantiveram desempenho elevado e estável, ao passo que ResNet152V2 e VGG19 apresentaram resultados inferiores. Os resultados indicam que a detecção automatizada de fissuras tem potencial para tornar as inspeções em estruturas de concreto mais ágeis, seguras e padronizadas.

**Palavras-chave:** Detecção de fissuras; Redes neurais convolucionais; Aprendizado profundo; Visão computacional.

## CRACK DETECTION IN CONCRETE USING CONVOLUTIONAL NEURAL NETWORKS: A COMPARATIVE EVALUATION WITH TRANSFER LEARNING

**Abstract:** The detection of pathologies, such as cracks, is essential in condition assessment systems for concrete structures, including bridges, tunnels, and buildings. Traditional methods rely on expert visual inspection, which is often time-consuming and potentially hazardous, depending on the location of the structures. Instead of on-site inspection by specialists, images collected by unmanned aerial vehicles can be used to survey structural conditions and subsequently processed by automated systems for pathology identification. The collected images exhibit variations in illumination and texture and are subject to noise, posing challenges to machine learning algorithms aimed at building models for automatic defect detection. To this end, convolutional neural networks have been considered for the development of classification models. This study investigates the use of machine learning for crack detection and classification in concrete structures through the application of convolutional neural networks to concrete images. A dataset of 3,822 images (1,050 with cracks; 2,772 without) was resized and normalized. The VGG16, VGG19, MobileNet, MobileNetV2, ResNet152V2, and DenseNet201 architectures were evaluated, with fine-tuning of the final layers, application of class weights, and early stopping. Evaluation

employed stratified 5-fold cross-validation and the metrics accuracy, precision, sensitivity, F1-score, and confusion matrices for performance comparison. The main finding was that MobileNet achieved the highest sensitivity and F1-score, minimizing false negatives and representing the best choice for conservative screening. MobileNetV2 obtained the highest precision (0.980), while MobileNet presented the highest sensitivity and F1-score values (0.976 and 0.974), reducing false negatives. DenseNet and VGG16 maintained high and stable performance, whereas ResNet152V2 and VGG19 yielded inferior results. The findings indicate that automated crack detection has the potential to make concrete structure inspections more efficient, safer, and standardized.

**Keywords:** Crack detection; Convolutional neural networks; Deep learning; Computer vision

## 1. Introdução

Na Engenharia Civil, o termo patologia refere-se ao estudo das manifestações anômalas que comprometem o desempenho e a durabilidade das estruturas ao longo do tempo. Em estruturas de concreto, essas manifestações incluem fissuração, deterioração, corrosão das armaduras e desgaste superficial, entre outras, sendo analisadas quanto às suas causas, mecanismos de propagação e impactos estruturais, com o objetivo de subsidiar intervenções corretivas e preventivas. Dentre essas manifestações, as fissuras destacam-se pela elevada frequência de ocorrência e pelo potencial de comprometer a integridade estrutural. Fissuras são descontinuidades superficiais que facilitam a penetração de agentes agressivos, e que podem acelerar processos de degradação. Sua origem está associada a múltiplos fatores, como retração do concreto, carregamentos mecânicos, variações térmicas, movimentações do solo, reações químicas decorrentes da exposição ambiental e falhas de projeto ou execução. A Figura 1 apresenta exemplos de fissuras em concreto, evidenciando variações de iluminação, textura, tonalidade e presença de ruído nas imagens, características que impõem desafios à análise automatizada.

Devido a preocupações relacionadas à segurança estrutural e aos custos de manutenção, estruturas de grande porte, como pontes, túneis e barragens, demandam inspeções periódicas com o objetivo de detectar fissuras e diagnosticar suas causas (ELGHAISH *et al.*, 2021). Nos casos em que a intervenção imediata não é viável, torna-se necessário o monitoramento da evolução da fissuração. A avaliação dessas manifestações é tradicionalmente realizada por meio de inspeções visuais, que requerem a presença de profissionais experientes para análise das condições do concreto. Entretanto, tais métodos são, em geral, demorados, operacionalmente complexos e potencialmente arriscados, sobretudo quando as áreas inspecionadas apresentam difícil acesso, como em pontes e barragens. Diante dessas limitações, veículos aéreos não tripulados (VANTs) têm sido amplamente empregados na inspeção de infraestrutura civil, possibilitando a aquisição de imagens que podem ser utilizadas no treinamento de modelos de aprendizado de máquina (GOPALAKRISHNAN *et al.*, 2018).

A utilização de imagens digitais associadas a técnicas de aprendizado de máquina tem sido considerada como alternativa para a identificação automatizada de defeitos em obras de construção civil (VAZ *et al.*, 2023; CHEN *et al.*, 2021). Dentre as abordagens empregadas em processamento de imagens, as redes neurais convolucionais (CNNs) destacam-se na resolução de problemas de visão computacional, em razão de sua capacidade de extrair automaticamente características relevantes diretamente dos dados de entrada (ELGHAISH *et al.*, 2022; CHEN *et al.*, 2023).

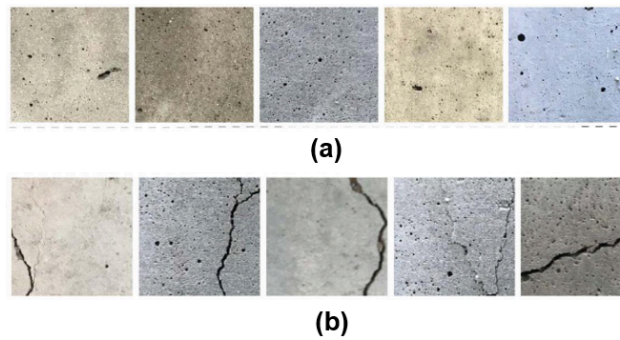


Figura 1 – Exemplos imagens de concreto sem (a) e com (b) fissuras. Fonte: Silva e Lucena (2018)

As CNNs realizam extração hierárquica de características diretamente dos dados de entrada, combinando camadas convolucionais, de agrupamento e totalmente conectadas para mapear padrões visuais em probabilidades de classe (Figura 2). No contexto do aprendizado por transferência, arquiteturas pré-treinadas são adaptadas a novos domínios por meio do ajuste das camadas finais (*fine-tuning*), estratégia amplamente empregada quando o volume de dados disponível é limitado (SIMONYAN; ZISSERMAN, 2015).

Na análise de imagens de superfícies de concreto, redes neurais convolucionais podem identificar padrões característicos associados à presença de fissuras, possibilitando sua detecção automatizada (OLIVEIRA *et al.*, 2021). Estudos anteriores demonstram que essas arquiteturas aprendem representações complexas dos dados visuais, mostrando-se adequadas para tarefas de detecção de fissuras em estruturas de concreto (CRUZ-RAMÍREZ *et al.*, 2020; ALI *et al.*, 2021). Xu *et al.* (2019) evidenciam ainda o potencial das CNNs na identificação automatizada de fissuras em pontes, associando ganhos de eficiência e precisão às inspeções estruturais, além de disponibilizarem uma base unificada de imagens de concreto.

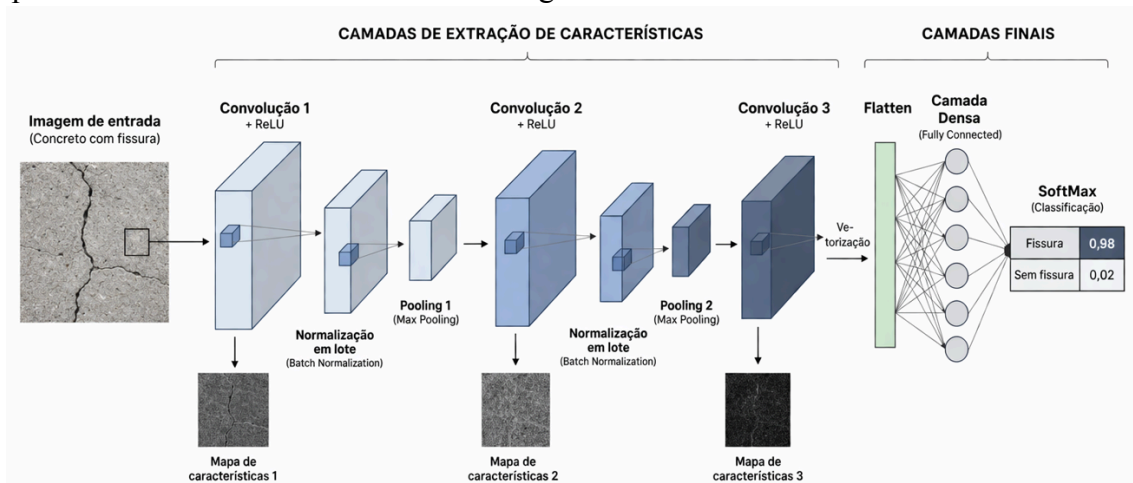


Figura 2 – Arquitetura de rede neural convolucional para classificação de fissuras em concreto.

Apesar dos avanços reportados na literatura quanto ao uso de redes neurais convolucionais para detecção de fissuras, ainda são limitados os estudos que realizam comparações sistemáticas entre múltiplas arquiteturas sob protocolo experimental

padronizado, com validação cruzada e análise estatística formal das diferenças de desempenho. Além disso, poucos trabalhos discutem explicitamente o impacto do equilíbrio entre precisão e sensibilidade sob a perspectiva prática da inspeção estrutural. Nesse contexto, o presente estudo contribui ao (i) avaliar comparativamente diferentes arquiteturas profundas sob condições experimentais uniformes, (ii) empregar análise estatística para identificação de diferenças significativas entre as arquiteturas de rede pré-treinadas e (iii) estabelecer critérios objetivos de seleção orientados às exigências operacionais do processo de inspeção.

O presente trabalho investiga o emprego de redes neurais convolucionais com aprendizado por transferência para identificação de fissuras em imagens de concreto. Foram exploradas arquiteturas pré-treinadas amplamente utilizadas na literatura, incluindo VGG, MobileNet, ResNet e DenseNet, com adaptação das camadas finais para classificação binária. O protocolo experimental compreende etapas de pré-processamento das imagens, tratamento do desbalanceamento de classes, aplicação de mecanismos de regularização e avaliação por meio de validação cruzada e análise estatística.

## 2. Material e Métodos

O protocolo experimental foi desenvolvido na linguagem Python, empregando as bibliotecas TensorFlow (ABADI *et al.*, 2016) e Keras (CHOLLET, 2015) para a modelagem, treinamento e avaliação de redes neurais convolucionais. As operações numéricas vetorizadas foram conduzidas com o auxílio da biblioteca NumPy (HARRIS *et al.*, 2020), enquanto as etapas de particionamento dos dados, validação e cálculo de medidas de avaliação foram realizadas por meio do Scikit-learn (PEDREGOSA *et al.*, 2011). A visualização e análise gráfica dos resultados foram conduzidas utilizando as bibliotecas Matplotlib (HUNTER, 2007) e Seaborn (WASKOM, 2021), permitindo uma interpretação mais clara e consistente do desempenho dos modelos.

As imagens foram obtidas do conjunto público *Concrete Bridge Crack Dataset* (ZOUBIR *et al.*, 2021), composto por duas classes: com fissuras e sem fissuras. O conjunto totalizou 3.822 imagens, sendo 1.050 com fissuras e 2.772 sem fissuras. As imagens foram redimensionadas para  $224 \times 224$  pixels e normalizadas conforme o padrão esperado pelos modelos pré-treinados. A preparação incluiu verificação básica de duplicidades, inspeção de ruído e organização em diretórios compatíveis com os geradores de dados do TensorFlow. Adotou-se a técnica de aprendizado por transferência, com congelamento das camadas convolucionais das arquiteturas pré-treinadas e ajuste das camadas finais para classificação binária. Foram avaliadas as arquiteturas VGG16, VGG19, MobileNet, MobileNetV2, ResNet152V2 e DenseNet201, selecionadas por ampla utilização na literatura e diferentes níveis de complexidade computacional.

O conjunto de dados apresenta desbalanceamento entre as classes, com 2.772 imagens sem fissuras e 1.050 com fissuras, resultando em uma proporção aproximada de 2,6:1. Esse desequilíbrio pode induzir viés nos modelos treinados, favorecendo a classe majoritária e comprometendo a sensibilidade na detecção de fissuras. Para mitigar esse efeito, aplicaram-se pesos de classe inversamente proporcionais à frequência de cada classe durante o treinamento, penalizando proporcionalmente os erros cometidos na classe minoritária. As imagens foram normalizadas com reescalonamento dos valores de pixel para o intervalo  $[0, 1]$ , compatível com o padrão de entrada das arquiteturas pré-treinadas. Estratégias de aumento de dados não foram empregadas, o que constitui uma limitação do estudo, uma vez que a diversificação

artificial das amostras poderia contribuir para maior robustez dos modelos frente a variações de iluminação, textura e ruído presentes nas imagens de campo.

O treinamento foi realizado com tamanho de lote igual a 8, utilizando o otimizador Adam e a função de perda entropia cruzada binária. Para mitigar o desbalanceamento entre classes, aplicaram-se pesos de classe durante o treinamento. Empregou-se a técnica de parada antecipada para prevenir sobreajuste, com monitoramento do desempenho em validação e preservação dos melhores pesos do modelo.

A avaliação foi conduzida por meio de validação cruzada estratificada com cinco partições (*5-fold cross-validation*). Antes da divisão, o conjunto de dados foi embaralhado aleatoriamente, e a estratificação garantiu que a proporção entre as classes fosse mantida em cada partição, preservando a representatividade do desbalanceamento original. Para cada *fold*, o modelo foi treinado nas quatro partições restantes e avaliado na partição retida, sendo calculadas as medidas acurácia, precisão, sensibilidade e F1-score. Adicionalmente, construíram-se matrizes de confusão para análise dos padrões de erro de classificação. Os resultados reportados correspondem à média das cinco execuções. As diferenças entre arquiteturas foram examinadas por meio de análise de variância de um fator e teste de Tukey, adotando diagramas de caixas como síntese gráfica da variabilidade observada (MONTGOMERY, D. C.; RUNGER, 2021). Os experimentos foram executados em computador equipado com unidade de processamento gráfico (GPU), visando reduzir o tempo de treinamento dos modelos.

### 3. Resultados

Os modelos avaliados apresentaram desempenhos distintos conforme a medida de avaliação adotada. O modelo MobileNetV2 (Figura 3) apresentou a maior precisão (0,980), indicando maior confiabilidade na identificação positiva de imagens com fissuras. Entretanto, seu valor de sensibilidade igual a 0,943 foi inferior ao obtido pelo modelo MobileNet, igual a 0,976, sugerindo maior ocorrência de falsos negativos em comparação a esse modelo. O modelo MobileNet (Figura 4) apresentou os maiores valores de sensibilidade e F1-score (0,976 e 0,974, respectivamente), evidenciando melhor desempenho em cenários nos quais a detecção completa das fissuras é prioritária. Observa-se que essa arquitetura apresenta leve redução na precisão em troca de maior capacidade de identificação da classe positiva.

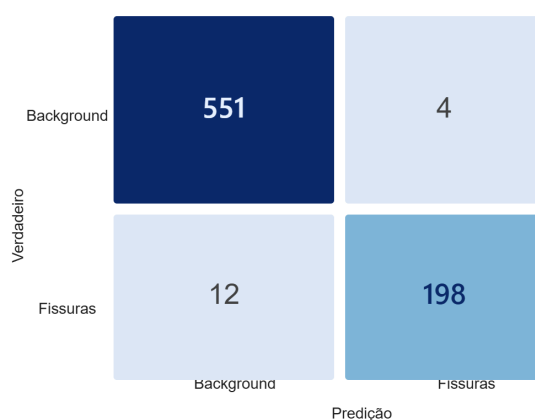


Figura 3 – Matriz de confusão do modelo MobileNetV2.

Os modelos DenseNet e VGG16 (Figuras 5 e 6) mantiveram desempenho elevado em termos de acurácia global, apresentando diferentes equilíbrios entre precisão e

sensibilidade. Por sua vez, ResNet152V2 e VGG19 (Figuras 7 e 8) apresentaram desempenho inferior no conjunto avaliado. As matrizes de confusão permitiram identificar que os erros principais de classificação estão associados a fissuras de baixa evidência visual, como fissuras muito finas ou em regiões com sombra intensa.

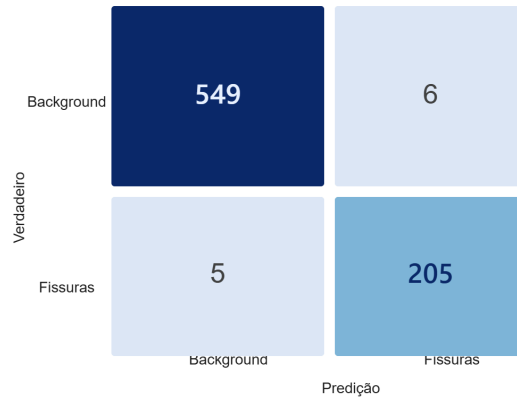


Figura 4 – Matriz de confusão do modelo MobileNet.

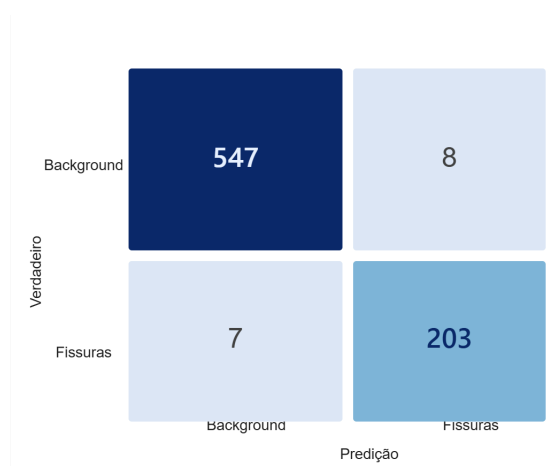


Figura 5 – Matriz de confusão do modelo DenseNet201.

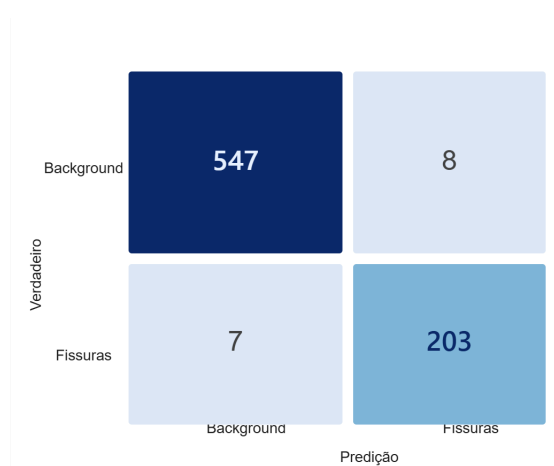


Figura 6 – Matriz de confusão do modelo VGG16.

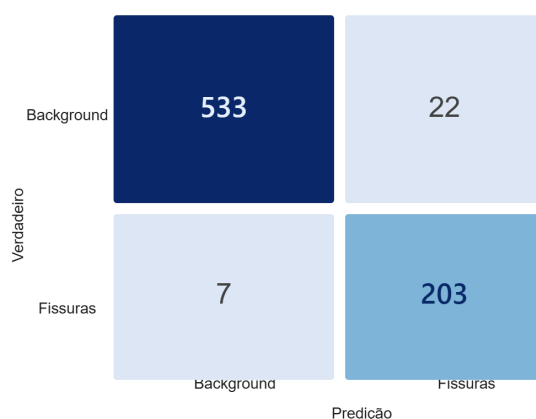


Figura 7 – Matriz de confusão do modelo ResNet152V2.

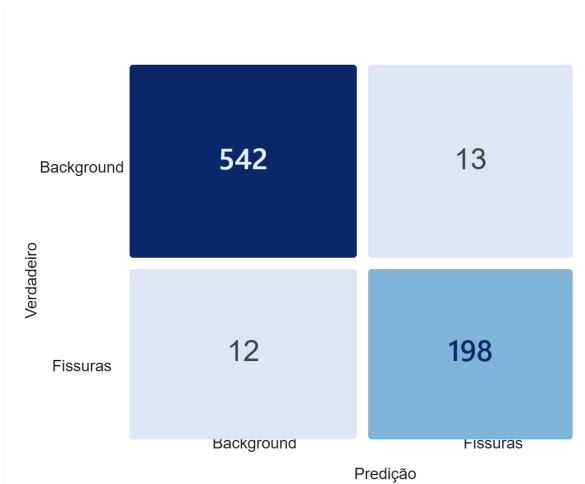


Figura 8 – Matriz de confusão do modelo VGG19.

Aplicou-se análise de variância de um fator para comparação entre os modelos em cada medida de avaliação (Tabela 1). Foram observadas diferenças estatisticamente significativas entre os modelos para acurácia, precisão e F1-score, enquanto para a sensibilidade, não se observou diferença global significativa. O teste de Tukey (Tabela 2) identificou os pares de arquiteturas com diferenças estatisticamente significativas, sendo VGG19 o modelo consistentemente inferior nas comparações de acurácia e precisão. Para sensibilidade e F1-score, a ausência de diferença global significativa na ANOVA dispensou a aplicação do teste post-hoc.

De forma integrada, os resultados quantitativos e a análise estatística indicam que a MobileNet apresenta desempenho mais equilibrado quando a prioridade é reduzir falsos negativos, em virtude dos maiores valores de sensibilidade e F1-score. A MobileNetV2 destacou-se pela maior precisão, sendo mais adequada em cenários que demandam maior seletividade na classificação da classe positiva. DenseNet e VGG16 mantiveram desempenho elevado e estável, configurando alternativas robustas conforme restrições operacionais. Assim, a escolha do modelo deve considerar o equilíbrio entre precisão e sensibilidade de acordo com o contexto de aplicação.

Medida de Avaliação	F	p-valor
Acurácia	5.706072	0,001321
Precisão	4.266361	0,006443
Sensibilidade	1.615873	0,193972
F1 Score	5.603633	0,001470

Tabela 1 – Valores referentes a análise de variância.

Medida	Par de modelos	Diferença	p-valor
Acurácia	DenseNet201 × VGG19	0,2093	0,0080
Acurácia	MobileNet × VGG19	0,2137	0,0066
Acurácia	MobileNetV2 × VGG19	0,1875	0,0206
Precisão	DenseNet201 × VGG19	0,4872	0,0257
Precisão	MobileNet × VGG19	0,5791	0,0057

Tabela 2 — Pares de arquiteturas com diferenças estatisticamente significativas pelo teste de Tukey.

## CONCLUSÃO

Este trabalho apresentou e avaliou, de forma sistemática, o uso de redes neurais convolucionais com aprendizado por transferência para identificação de fissuras em concreto, empregando validação cruzada, medidas de avaliação de desempenho e análise estatística por meio de análise de variância e do teste Tukey. Os resultados evidenciaram que o modelo MobileNet é a alternativa mais indicada quando a prioridade é minimizar falsos negativos, cenário típico de triagem conservadora em inspeções estruturais, enquanto a rede MobileNetV2 apresentou maior precisão, sendo mais adequada quando se exige maior seletividade na rotulagem de imagens com fissuras. DenseNet e VGG16 mantiveram desempenho elevado e estável, configurando alternativas robustas conforme restrições operacionais, como custo computacional, memória disponível e tempo de processamento.

A convergência entre medidas e testes estatísticos reforça a consistência das conclusões e fornece base quantitativa para a escolha do modelo em função do contexto de aplicação. Do ponto de vista prático, a abordagem proposta contribui para a padronização do fluxo de pré-processamento, treinamento e avaliação, oferecendo critérios objetivos para apoio à decisão em inspeções estruturais baseadas em imagens.

O estudo apresenta limitações que devem ser consideradas na interpretação dos resultados. O conjunto de dados utilizado é composto exclusivamente por imagens estáticas de laboratório, sem validação em condições reais de campo, o que pode limitar a generalização dos modelos a cenários com maior variabilidade de iluminação, ângulo de captura e tipo de fissura. O desbalanceamento entre classes, pode ainda influenciar o desempenho em situações com prevalência distinta da observada no *dataset*. A ausência de estratégias de aumento de dados representa uma oportunidade de melhoria, uma vez que a diversificação artificial das amostras poderia contribuir para maior robustez frente

a variações fotométricas e de textura. Do ponto de vista prático, os resultados obtidos indicam que a abordagem proposta é aplicável em sistemas de triagem automatizada para inspeção de estruturas de concreto, com potencial de integração a plataformas embarcadas em VANTs. Investigações futuras devem contemplar a validação em imagens coletadas em campo, a incorporação de técnicas de aumento de dados e a exploração de arquiteturas mais recentes, como EfficientNet e Vision Transformers, visando ampliar a capacidade de generalização dos modelos.

## AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Fundação Araucária e à Universidade Estadual de Ponta Grossa (UEPG) pelo apoio institucional e pela concessão de bolsa de iniciação científica. Agradecem ainda ao Departamento de Informática e ao Laboratório de Computação de Alto Desempenho (LCAD) pelo espaço e pela infraestrutura disponibilizados para a realização deste trabalho.

## REFERÊNCIAS

- ABADI, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. **arXiv preprint arXiv:1603.04467**, 2016.
- ALI, L. et al. Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures. **Sensors**, v. 21, n. 5, p. 1688, 1 mar. 2021.
- CHEN, Lingkun et al. Convolutional neural networks (CNNs)-based multi-category damage detection and recognition of high-speed rail (HSR) reinforced concrete (RC) bridges using test images. **Engineering Structures**, v. 276, p. 115306, 2023.
- CHOLLET, F. Keras: deep learning for humans. 2015. Disponível em: <https://keras.io>. Acesso em: 23 mar. 2026.
- CRUZ-RAMÍREZ, N., ZHU, H.; ANUMBA, C. J. Deep Learning for Automatic Crack Detection in Concrete Structures. In **36th International Symposium on Automation and Robotics in Construction (ISARC 2019)**, 2020.
- ELGHAISH, F. et al. Deep learning for detecting distresses in buildings and pavements: a critical gap analysis. **Construction Innovation**, v. 22, n. 3, p. 554–579, 2021.
- ELGHAISH, Faris et al. Deep learning for detecting distresses in buildings and pavements: a critical gap analysis. **Construction Innovation**, v. 22, n. 3, p. 554–579, 2022.
- GOPALAKRISHNAN, Kasthurirangan et al. Crack damage detection in unmanned aerial vehicle images of civil infrastructure using pre-trained deep learning model. **Int. J. Traffic Transp. Eng.**, v. 8, n. 1, p. 1–14, 2018.
- HARRIS, C. R. et al. Array programming with NumPy. **Nature**, v. 585, p. 357–362, 2020.
- HUNTER, J. D. Matplotlib: a 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.
- MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 7. ed. Rio de Janeiro: LTC, 2021.
- OLIVEIRA, R.; MAURICIO, C.; SANTOS, V.; PERES, F. Detecção de fissuras utilizando redes neurais convolucionais. In: **34th SIBGRAPI Conference on Graphics, Patterns and Images**, 2021. p. 212–217.
- PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

RAWAT, Waseem; WANG, Zenghui. Deep convolutional neural networks for image classification: A comprehensive review. **Neural computation**, v. 29, n. 9, p. 2352-2449, 2017;

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2015.

VAZ, E. S.; GASPARELLO, L. F.; DE GOUVEIA, L. T.; SENGER, L. J. Detecting damage in roads using convolutional neural networks. **Iberoamerican Journal of Applied Computing**, v. 11, n. 1, 2023.

WASKOM, M. **Seaborn: statistical data visualization**. 2021. Disponível em: <https://seaborn.pydata.org>. Acesso em: 23 mar. 2026.

XU, H.; SU, X.; WANG, Y.; CAI, H.; CUI, K.; CHEN, X. Automatic bridge crack detection using a convolutional neural network. **Applied Sciences**, Basel, v. 9, n. 14, p. 2867, 2019.

ZOUBIR, Hajar; RGUIG, Mustapha; ELAROUSSI, Mohammed. Crack recognition automation in concrete bridges using Deep Convolutional Neural Networks. In: **MATEC web of conferences**. EDP Sciences, 2021. p. 03014.