

AVALIAÇÃO DO DESEMPENHO DE REDES NEURAIS CONVOLUCIONAIS PRÉ-TREINADAS PARA CLASSIFICAÇÃO DE RACHADURAS DE CONCRETO

Lucas Santos Candido¹, Leonardo Melo De Medeiros², Alexandre Cunha Machado³

¹Discente de graduação em Engenharia Civil - IFAL; ²Professor do curso de Engenharia Civil - IFAL; ³Professor do curso de Engenharia Civil- IFAL. E-mail autor correspondente: lsc7@aluno.ifal.edu.br

Área de conhecimento/Subárea: Área 09 - Multidisciplinar.
ODS vinculado(s): ODS09, ODS11.

INTRODUÇÃO

A manifestação patológica de trincas, fissuras e rachaduras em estruturas de concreto atuam como indicadores de possíveis falhas estruturais, demonstrando que a resistência do material não está sendo o suficiente para resistir aos esforços atuantes sob o mesmo. Estas aberturas, caso não tratadas, servem como porta de entrada de agentes agressivos ao interior da estrutura possibilitando o surgimento de novas degradações, como a infiltração de água e corrosão de armaduras (Silva e Cabral, 2014). A eventual negligência da manutenção prévia destes danos favorece o colapso parcial ou total da estrutura, segundo o IBAPE-SP (2015), 66% dos acidentes prediais ocorridos são relacionados a uma precária ou falta de manutenção. Portanto, o monitoramento de construções deve ser realizado sistematicamente e periodicamente, para garantir a segurança dos usuários e transeuntes do respectivo local.

O método tradicional para detecção de aberturas no concreto baseia-se na observação e inspeção visual de profissionais técnicos, no entanto, Su e Wang (2020) explica que estes métodos são trabalhosos, demorados e comumente apresentam perigos ocultos a segurança dos inspetores. Diante disto, diversas tecnologias computacionais como processamento de imagens, visão computacional e aprendizado de máquina são aplicadas com intenção de fornecer uma metodologia para identificação de rachaduras eficiente e segura. Em especial destaca-se a rede neural convolucional (CNN), habilitada com a alta capacidade de analisar fotos e vídeos, sendo eficaz em identificar padrões e aprender características associadas as imagens, conseguindo classificar os mais diversos tipos de objetos.

Diante desse cenário, estudos como os dos pesquisadores Arafin et al. (2022), Hussain et al. (2024) e Zadeh et al. (2023) objetiva a avaliação do desempenho de CNN pré-treinada no banco de dados da ImageNet para detecção de rachaduras de concreto. Estas pesquisas visam analisar se os respectivos modelos empregues demonstram potencial para o desenvolvimento de métodos inovadores e eficazes de inspeção estrutural. Por conseguinte, a presente pesquisa tem finalidade de analisar e comparar quatro modelos comuns da literatura com transferência de aprendizado para identificar a arquitetura mais adequada para a tarefa de detecção da manifestação patológica de aberturas em estruturas de concreto, promovendo avanços significativos no campo da engenharia civil e da manutenção preditiva.

MATERIAL E MÉTODOS

O vigente projeto fundamenta-se nos conceitos metodológicos definidos por Gil (2017), atuando como uma pesquisa exploratória em seu objetivo de estudo, com o intuito de aprofundar o conhecimento na área de aprendizado profundo para classificação de imagens. Dito isto, a pesquisa classifica-se como de natureza aplicada, pois tem como propósito dirigido à solução do problema de pesquisa, havendo abordagem quantitativa para a análise dos resultados.

Dados Experimentais

O banco de dados empregado neste experimento origina-se de imagens coletadas aleatoriamente de três pesquisas: Özgenel (2019), Dorafshan et al. (2018) e Xu et al. (2019). Contendo 25.000 imagens de diversas superfícies de concreto como edificações, pontes, decks, pátios e paredes, as amostras dividem-se em dois subconjuntos para classificação: o primeiro refere-se a superfícies sem ocorrência de rachaduras, enquanto o segundo contempla cenários positivos para a manifestação patológica, conforme as amostras demonstradas na Figura 1.

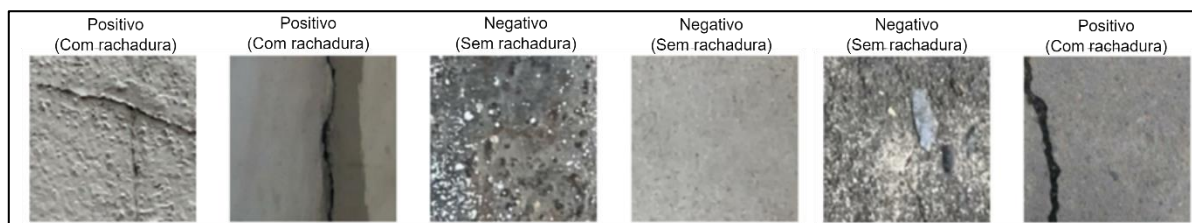


Figura 1. Amostras do banco de dados selecionadas aleatoriamente. IFAL, 2024.

As imagens originais são constituídas em pixels da escala de RGB no intervalo numérico entre [1, 255], esses valores requerem alto custo computacional para o treinamento dos modelos. Portanto os valores são normalizados para [0, 1], consequentemente os algoritmos terão melhor performance computacional ao executar operações matemáticas com as amostras. Em seguida as imagens são redimensionadas para (224, 224, 3), os dados são divididos aleatoriamente em 3 subconjuntos, 17.500 (70%) para o treinamento, 5.000 (20%) para validação e 2.500 (10%) para teste, mantendo a proporção de 50% de cada categoria.

Construção dos Modelos

Os avanços do aprendizado profundo proporcionado pelas CNN impulsionaram a solução de tarefas associadas ao reconhecimento de imagens. A partir de uma quantidade de dados de treinamento suficiente, as CNN podem aprender as características representativas completamente baseadas nessas amostras e aplicá-las a problemas específicos. Porém salienta-se que a eficiência de uma arquitetura de rede neural limita-se a quantidade ou pela baixa qualidade dos dados. A transferência de aprendizado (TL) surge como uma técnica poderosa neste âmbito, pois possibilita o aproveitamento do conhecimento prévio de modelos existentes com intuito de reutilizá-los em outros domínios. Por conseguinte, a TL possibilita contornar a problemática da escassez de dados de treinamento, acrescentando-se do fato que os recursos genéricos influenciam positivamente na classificação de rachaduras de concreto.

A abordagem de TL utilizada refere-se à extração de recursos conforme ilustrado na Figura 2, por intermédio da seleção de modelos pré-treinados no ImageNet, além do ajuste de novas camadas classificadoras de acordo com o número de saídas e o treinamento do modelo no novo conjunto de dados.

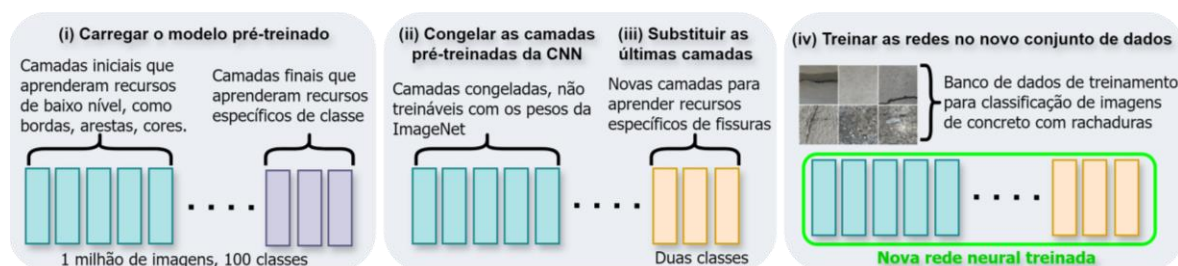


Figura 2. Infográfico da metodologia para extração de recursos com transfer learning. IFAL, 2024.

A seleção das CNN baseou-se na busca extensiva de modelos apresentados na literatura científica atual que performaram em diversas atividades de visão computacional e classificação de imagens, sendo contempladas neste presente trabalho: VGG16 (Simonyan e Zisserman, 2015), MobileNetV2 (Sandler et al, 2018), InceptionV3 (Szegedy et al., 2016), Xception (Chollet, 2017).

- A. **VGG16.** Esta arquitetura tem como característica a sua simplicidade e eficiência, composta por 13 camadas convolucionais empilhadas seguidas de 3 camadas totalmente conectadas. A profundidade da rede facilita o aprendizado da hierarquia complexa das características da imagem, porém apresenta menor eficiência em termos de uso de memória decorrente da grande quantidade de parâmetros do modelo.
- B. **MobileNetV2.** Tendo como fundamento a sua utilização em dispositivos móveis e embarcados, a arquitetura apresenta boa performance enquanto minimiza a quantidade de parâmetros e operações matemáticas o possível. O modelo introduz o uso de estrutura residual invertida em conjunto com camadas de convoluções separáveis para aumentar o desempenho e manter o baixo custo computacional.
- C. **InceptionV3.** O coração desta arquitetura são os módulos Inception, os quais combinam diversas convoluções com diferentes tamanhos de filtros executados paralelamente, isto permite capturar diversas características da imagem em múltiplas escalas. Ao mesmo tempo em que emprega a suavização de rótulos para diminuir a tendência e confiança excessiva da rede em suas previsões, evitando o fenômeno de overfitting.
- D. **Xception.** Surgiu como uma versão estendida da arquitetura Inception, substituindo os tradicionais módulos Inception por camadas de convolução separável em profundidade. Estas camadas utilizam menos parâmetros e requerem custo computacional menor em comparação as camadas regulares de convolução, enquanto mantem alto desempenho em classificação de imagens.

Aprendizado Supervisionado

O procedimento de treinamento das arquiteturas sucede através do aprendizado supervisionado, cada amostra é comparada com sua classe verdadeira e com a predição do modelo, os resultados serão analisados por um supervisor externo da rede, direcionando o reajuste eficiente dos pesos. As configurações dos hiperparâmetros do procedimento ocorreram de forma empírica objetivando sua otimização, estabelecendo 20 épocas de treinamento, com lotes de 32 amostras, uma taxa de aprendizado inicial de 1×10^{-3} , usando função de perda de entropia cruzada e otimização estocástica eficiente por intermédio de ADAM. Ainda convém mencionar que adotou-se a redução da taxa de aprendizado pelo fator 0,1 caso o erro no conjunto de validação não diminuir por três épocas consecutivas.

Medidas de Desempenho

A avaliação do desempenho dos diferentes modelos aplicados a solução desta tarefa, ocorre por meio de quatro métricas. Acurácia refere-se à proporção de classificações corretas para a quantidade total de amostras. Precisão analisa a exatidão da quantidade de imagens com rachaduras que foram previstas corretamente. Sensibilidade calcula a porcentagem de imagens previstas corretamente para manifestação patológica em relação ao número total de amostras de fissuras. Pontuação F1 representa a média harmônica entre precisão e sensibilidade.

RESULTADOS E DISCUSSÃO

Os experimentos nesta pesquisa foram conduzidos no ambiente computacional do Google Colaboratory, versão Colab Pro, em conjunto com a linguagem de programação Python. As

redes neurais artificiais foram implementadas através das versões dos frameworks Tensorflow e Keras disponibilizadas pelo próprio notebook na nuvem.

A performance dos 4 modelos pré-treinados validou-se a partir do uso do mesmo banco de dados para detecção de rachaduras de concreto, a Figura 3 ilustra o histórico do aprendizado supervisionado ao longo de 20 épocas. Optamos por monitorar ao final de cada época o cálculo das previsões corretas das amostras de validação, salvando os pesos do modelo correspondente a época que obteve o melhor resultado, disposto na Tabela 1. Observa-se que VGG16 conquistou o melhor desempenho entre as redes durante o treinamento, com a maior acurácia e menor perda, ao mesmo tempo em que obteve pior resultado total no processo de validação. O InceptionV3 efetuou péssimo exercício de treino, enquanto performou apenas 0,56% mais adiante da acurácia de validação do VGG16, consagrando-se como a segunda pior rede no geral. As redes MobileNetV2 e Xception tiveram desempenhos semelhantes, demonstrando robustez a solução da tarefa, entretanto MobileNetV2 alcança a acurácia ligeiramente melhor nas amostras de validação, sugerindo ser a escolha sólida para detecção de rachaduras de concreto.

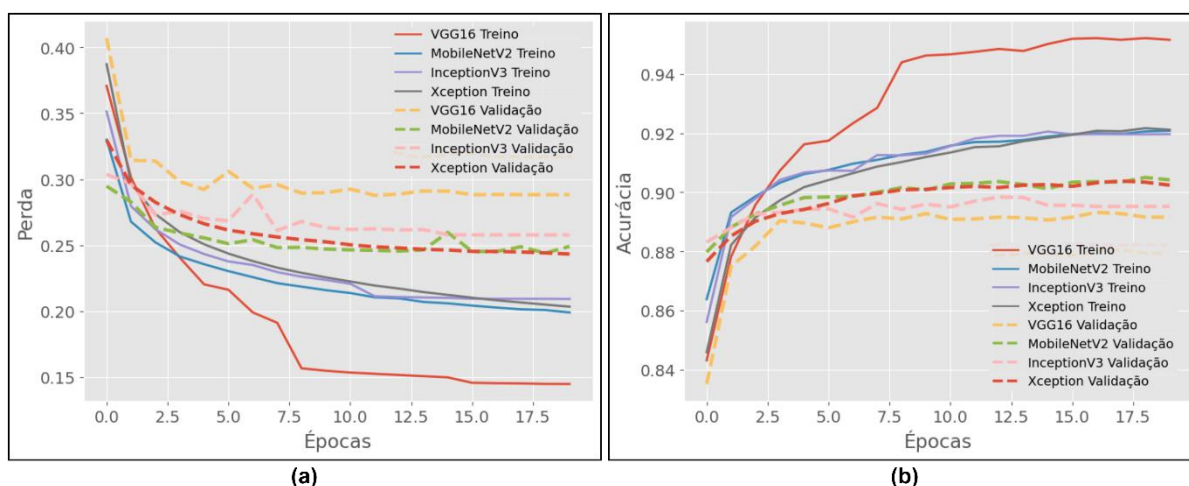


Figura 3. Perda (a) e acurácia (b) durante o procedimento de treinamento e validação. IFAL, 2024.

Tabela 1. Comparação dos modelos salvos com os melhores pesos e resultados. IFAL, 2024.

Modelo Pré-treinado	Melhor Época	Acurácia		Perda		Parâmetros (Milhões)	Tamanho (MB)
		Treino	Validação	Treino	Validação		
VGG16	17	95,21%	89,32%	14,54%	28,84%	14.764	56,32
MobileNetV2	19	92,05%	90,50%	20,00%	24,39%	02.260	08,62
InceptionV3	13	91,90%	89,84%	21,07%	26,15%	21.806	84,19
Xception	18	92,06%	90,38%	20,66%	24,48%	20.861	79,58

Embora o aprendizado supervisionado quantifique a performance de cada modelo durante o treinamento, salienta-se a necessidade de analisar a eficiência das arquiteturas sob imagens nunca visualizadas. Portanto, empregamos o subconjunto de teste com a finalidade de simular um ambiente experimental, estas amostras foram analisadas individualmente pelas CNN e suas classificações expostas na Figura 4. Calculamos as métricas de desempenho de acordo com os valores da matriz de confusão, como disposto na Tabela 2, visualizamos que VGG16 obteve a pior acurácia ao mesmo tempo que tem a segunda melhor precisão; Xception e InceptionV3 alcançam valores semelhantes, com diferença de 1,16% de precisão favorável a Xception; MobileNetV2 se consagra com a melhor acurácia e precisão dentre os modelos, porém apenas apresenta sensibilidade superior a VGG16.

	(a) VGG16	(b) MobilenetV2	(c) InceptionV3	(d) Xception
Verdadeiro				
Negativo	1164	1187	1137	1154
Positivo	168	150	131	148
	86	63	113	86
	1082	1100	1119	1082
	Negativo	Negativo	Negativo	Negativo
	Previsto	Previsto	Previsto	Previsto
	Positivo	Positivo	Positivo	Positivo
	Previsto	Previsto	Previsto	Previsto

Figura 4. Matriz de confusão dos modelos pré-treinados no subconjunto de teste. IFAL, 2024.

Tabela 2. Desempenho dos modelos no subconjunto de teste. IFAL, 2024.

Arquitetura	Acurácia	Precisão	Sensibilidade	Pontuação F1
VGG16	89,84%	92,64%	86,56%	89,50%
MobileNetV2	91,48%	94,58%	88,00%	91,17%
InceptionV3	90,24%	90,83%	89,52%	90,17%
Xception	90,24%	91,99%	88,16%	90,03%

A tabela 3 apresenta uma comparação dos resultados obtidos em pesquisas semelhantes com o desempenho do melhor modelo do presente estudo. Embora as métricas dos modelos de Hussain et al. (2024) e Zadeh et al. (2023) tenham alcançado acurácia maior em seus dados de teste, salienta-se que o dataset utilizado em ambos os trabalhos contém baixa generalização de imagens. Isto é, as arquiteturas foram treinadas e testadas apenas para classificação de trincas em superfícies de concreto limitadas, enquanto alimentamos a MobileNetV2 em três bancos de dados distintos e Arafin et al. (2022) emprega amostras de locais distintos. Os valores mostram que MobileNetV2 obtém acurácia semelhante ao modelo de Arafin et al. (2022), à medida que estabelece uma precisão superior de 11,58%. Indicando a capacidade do modelo em convergir a resultados robustos, em especial a classificação correta de superfícies de concreto com trincas.

Tabela 3. Comparação do desempenho dos modelos de trabalhos relacionados. IFAL, 2024.

Trabalho	Acurácia	Precisão	Aplicação
Hussain et al. (2024)	99,85%	-	Detecção rachaduras de concreto em superfícies de edifícios.
Zadeh et al. (2023)	98,90%	98,90%	
Arafin et al. (2022)	91,00%	83,00%	Detecção rachaduras de concreto em superfícies de construções variadas.
Melhor modelo deste estudo (MobileNetV2)	91,48%	94,58%	

CONCLUSÕES

O uso de tecnologias avançadas de visão computacional, como as CNN pré-treinadas aliadas ao uso de transferência de aprendizado, demonstrou ser uma abordagem promissora para a inspeção e detecção da manifestação patológica de rachaduras em estruturas de concreto. Esta metodologia oferece uma alternativa segura em comparação aos métodos tradicionais de inspeção e monitoramento predial, contribuindo para a segurança de profissionais técnicos e eficiência laboral.

Embora o MobileNetV2 tenha apresentado alta aplicabilidade a solução desta tarefa, assim como outros modelos, ainda se mantém algumas limitações relacionadas ao escopo deste trabalho, como o aumento da generalização dos algoritmos para diferentes tipos de condições das rachaduras. Assim como a possível expansão do banco de dados para a inclusão de novas categorias de manifestações patológicas de concreto, tendo como foco aumentar a variedade de danos em estruturais que possam ser detectados pelas máquinas preditivas.

AGRADECIMENTOS

Este estudo se manifesta como continuação de dois projetos do Programa Institucional de Bolsa de Iniciação Científica do Instituto Federal de Alagoas (IFAL). O primeiro de código PVE73-2021, sob concessão de bolsa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o segundo de código PVE390-2022 com financiamento do IFAL.

REFERÊNCIAS

ARAFIN, Palisa et al. Performance comparison of multiple convolutional neural networks for concrete defects classification. **Sensors**, v. 22, n. 22, p. 8714, 2022.

CHOLLET, F. Xception: Deep Learning with Depthwise Separable Convolutions. **arXiv (Cornell University)**, 7 out. 2016.

Construction and Building Materials, v. 186, p. 1031–1045, out. 2018a.

DORAFSHAN, S.; THOMAS, R. J.; MAGUIRE, M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete.

GIL, Antonio Carlos. Como elaborar projetos de pesquisa. Editora Atlas SA, 2017.

HUSSAIN, Adil et al. A Novel SCD11 CNN Model Performance Evaluation with Inception V3, VGG16 and ResNet50 Using Surface Crack Dataset. Em: **2024 2nd International Conference on Unmanned Vehicle Systems-Oman (UVS)**. IEEE, 2024. p. 1-7.

INSTITUTO BRASILEIRO DE AVALIAÇÕES E PERÍCIAS DE ENGENHARIA DE SÃO PAULO. **Inspeção Predial: a Saúde dos Edifícios**. 2. ed. [s.l.] IBAPE-SP, 2015.

ÖZGENEL, Ç. F. Concrete Crack Images for Classification. **Mendeley**. v. 2, 23 jul. 2019.

SANDLER, M. et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. **arXiv (Cornell University)**, 13 jan. 2018.

SILVA, Luiza Kilvia da; CABRAL, Antonio Eduardo Bezerra. **Levantamento de manifestações patológicas em estruturas de concreto armado no estado do Ceará**. 2014.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **Computer Science**, 2014.

SU, C.; WANG, W. Concrete Cracks Detection Using Convolutional NeuralNetwork Based on Transfer Learning. **Mathematical Problems in Engineering**, v. 2020, p. 1–10. 2020.

SZEGEDY, C. et al. **Rethinking the Inception Architecture for Computer Vision**. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

XU, H. et al. Automatic Bridge Crack Detection Using a Convolutional Neural Network. **Applied Sciences**, v. 9, n. 14, p. 2867, 18 jul. 2019.

ZADEH, Sara Shomal et al. Concrete Surface Crack Detection with Convolutional-based Deep Learning Models. **International Journal of Novel Research in Civil Structural and Earth Sciences**, v. 10, n. 3, p. 25-35, nov. 2023.