

Modelo de Deep Learning para detecção de fissuras em imagens térmicas

1th Letícia Vieira Gonçalves
UFERSA

Pau dos Ferros, Brasil
leticia.vieira@alunos.ufersa.edu.br

2th Rosana Cibely Batista Rego
DETEC
UFERSA

Pau dos Ferros, Brasil
rosana.rego@ufersa.edu.br

3th Paulo Henrique Araújo Bezerra
DETEC
UFERSA

Pau dos Ferros, Brasil
paulo.bezerra@ufersa.edu.br

Abstract—A detecção automatizada de fissuras em estruturas é crucial para garantir a durabilidade e segurança das edificações, especialmente em regiões sujeitas a variações térmicas. Este estudo propõe um modelo de *Deep Learning* baseado em Redes Neurais Convolucionais (RNCs) para detecção e classificação de fissuras em imagens térmicas de elementos construtivos. Utilizou-se um conjunto de dados balanceados com 128 imagens térmicas, ampliado para 2000 amostras via técnicas de aumento da quantidade de dados. A arquitetura da RNC incluiu camadas convolucionais, *pooling*, *dropout* e a função de ativação *Sigmoide* para classificação binária.

Index Terms—Detecção, fissura, redes neurais convolucionais, infravermelho.

I. INTRODUÇÃO

A termografia infravermelha é uma técnica utilizada para a inspeção de estruturas [1], permitindo a identificação de descontinuidades e defeitos que podem comprometer a durabilidade das edificações. Esse método baseia-se na captura da radiação térmica emitida pelos materiais, possibilitando a detecção de variações de temperatura associadas a anomalias em elementos construtivos.

A preservação da integridade das edificações é fundamental para garantir a segurança e o conforto dos usuários [2]. No entanto, a formação de fissuras pode comprometer a durabilidade e a funcionalidade das construções. Dentre esses fatores, as variações de temperatura desempenham um papel significativo, sobretudo em regiões sujeitas a grandes amplitudes térmicas [3].

Algoritmos de Aprendizado Profundo têm sido utilizados para detectar fissuras na construção civil, conforme [4], e essa abordagem vem se consolidando como uma ferramenta eficiente na inspeção automatizada de estruturas [5]. Nesse sentido, estudos recentes têm explorado e aprimorado essas técnicas, demonstrando seu potencial para aumentar a confiabilidade e a precisão na análise de fissuras [6].

Este estudo propõe um modelo de aprendizado de máquina para a detecção automatizada de fissuras em imagens térmicas de elementos construtivos, com ênfase naquelas originadas por variações térmicas. Essa abordagem apresenta um potencial para aprimorar as inspeções em edificações, viabilizando a identificação precoce de danos, permitindo a otimização dos processos de manutenção e a consequente redução de custos.

Este trabalho está estruturado da seguinte maneira: na seção II, Metodologia, são descritos o conjunto de dados, as técnicas

de aumento de amostras e o pré-processamento aplicado às imagens térmicas, além da arquitetura da rede neural convolucional proposta. Na seção III, Resultados e Discussão, são analisadas métricas como acurácia, perda e matriz de confusão, revelando tanto o potencial das RNCs para detecção de fissuras em imagens térmicas, quanto os desafios práticos desse tipo de aplicação. Por fim, na seção IV, Considerações Finais, destaca-se a necessidade de integrar dados termográficos a outras modalidades espectrais para aprimorar o modelo e a seção V, Referências.

II. METODOLOGIA

A. Conjunto de dados

Este estudo utiliza um conjunto de dados privado para detecção de fissuras em superfícies de elementos construtivos, composto por duas classes: Positivo (imagens térmicas com fissura) e Negativo (imagens térmicas sem fissuras), conforme ilustram as Figuras 1 e 2. O conjunto de treinamento inicial contém 128 imagens, enquanto o conjunto de teste inclui 32 imagens.

Para garantir um treinamento adequado, as classes foram balanceadas utilizando a técnica de *Random Over Sampling* com reposição, replicando aleatoriamente amostras de classes minoritárias e resultou em um aumento efetivo na quantidade de dados. O processo possibilitou que ambas as classes no conjunto de treinamento possuíssem o mesmo número de amostras, totalizando 1000 imagens para cada classe. Diante disso, dividiu-se o conjunto total de dados nos subconjuntos de treino e teste, respectivamente, 80% e 20% dos dados. Essa divisão garante que o modelo seja avaliado em dados não vistos durante o treinamento, permitindo medir sua capacidade de generalização. O método *train_test_split* da biblioteca *sklearn* foi utilizado para dividir os dados, permitindo que o modelo aprendesse padrões nas imagens e fosse testado com dados inéditos. As imagens do conjunto de dados tinham dimensões de 320x240 pixels e estavam na escala RGB.

B. Rede Neural Convolucional

As Redes Neurais Convolucionais (RNCs) emergiram no campo da inteligência artificial, impulsionando avanços significativos em diversas áreas, com destaque para o processamento de imagens. Essa arquitetura permite a extração automática de características complexas a partir de dados

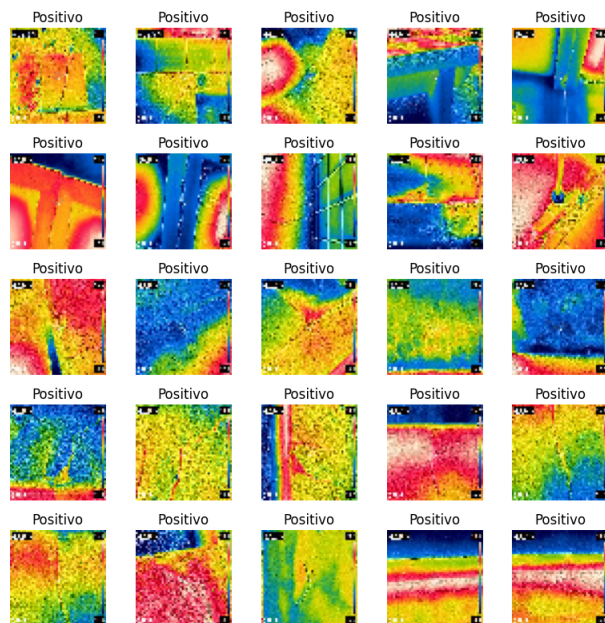


Fig. 1. Imagem com fissura - Positivo.

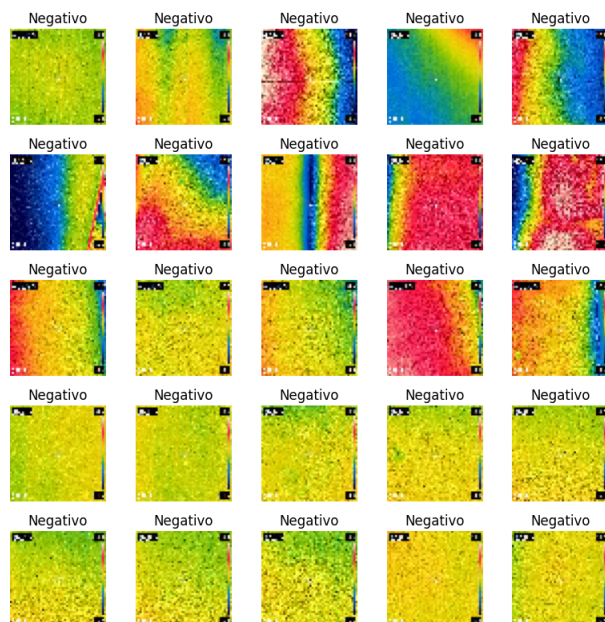


Fig. 2. Imagens sem fissuras - Negativo.

visuais, superando as limitações das abordagens tradicionais [7]. Dessa forma, as RNCs compõem uma das arquiteturas de redes neurais artificiais amplamente aplicadas a diferentes domínios e tipos de dados. Esse tipo de rede é projetado para algoritmos de Aprendizado Profundo, sendo altamente eficaz no reconhecimento de imagens e no processamento de informações visuais em nível de pixel [8].

O funcionamento da estrutura das RNCs se assemelha ao do cérebro humano, no qual os neurônios são organizados de maneira específica para processar estímulos visuais [9]. Por sua vez, as RNCs possuem camadas de neurônios interconectados que operam de forma análoga ao córtex visual, responsável pelo processamento de informações visuais no cérebro [10]. Essa disposição permite que a RNC analise toda a região de uma imagem de maneira eficiente, evitando limitações comuns em redes neurais convencionais. Comparadas a modelos anteriores, as RNCs oferecem um desempenho superior no tratamento de imagens e podem ser aplicadas a outras formas de entrada, como áudio e fala, garantindo maior precisão na extração de padrões [11].

A arquitetura da RNC empregada neste trabalho apresenta duas camadas convolucionais (Conv2D). A primeira camada captura características locais simples, como bordas e texturas, e a segunda camada aprofunda a representação, extraindo padrões abstratos e robustos. Essas camadas aplicam filtros (kernels) à imagem de entrada, realizando convoluções e gerando mapas de características. A função de ativação *ReLU* (*Rectified Linear Unit*) é utilizada após as convoluções, melhorando a eficiência do treinamento. Em seguida, camadas de *Max Pooling* reduzem a dimensão espacial dos mapas de características e o número de parâmetros, e auxiliam na prevenção do *Overfitting*.

O tamanho de *kernel* 3x3 foi utilizado por sua eficiência na captura de padrões locais e por manter equilíbrio entre capacidade de aprendizado e custo computacional. A escolha de 512 e 256 filtros visa aumentar a capacidade de extração de características visuais relevantes. As camadas de *Dropout* foram inseridas com taxa de 20% com o objetivo de reduzir o risco de *Overfitting*, promovendo maior generalização do modelo durante o treinamento.

A definição dos hiperparâmetros foi realizada com base em testes exploratórios e em boas práticas da literatura. Foram ajustadas manualmente quantidades de filtros, tamanho de *kernel* e a taxa de aprendizado do otimizador *Adam* em 0,001.

Os experimentos foram conduzidos em um ambiente local com o sistema operacional *Windows*, utilizando uma *GPU NVIDIA GeForce RTX 3050* com 4 GB de memória dedicada, *driver* versão 566.36 e suporte ao *CUDA 12.7*. A execução foi realizada com suporte ao *driver WDDM* (*Windows Display Driver Model*).

A transição das camadas convolucionais para as densas foi feita pela camada *Fatten*, que "achata" o tensor multidimensional em um vetor unidimensional, essencial para conectar as características espaciais às camadas de classificação. As camadas densas combinam linearmente essas características, seguidas por ativações não-lineares. Na saída, utiliza-se a

função *Sigmoid* para problemas de classificação binária, pois ela transforma a saída em uma probabilidade entre 0 e 1.

III. RESULTADOS E DISCUSSÕES

Como critérios de avaliação foram consideradas a precisão, a perda de treinamento e validação, bem como a matriz de confusão.

A Rede Neural Convolutiva (RNC) atingiu uma acurácia de 0,99 no treinamento e de 1,0 ao final da validação (Figura 3), apresentando também perdas baixas, de 0,03 no treinamento e de 0 na validação (Figura 4). Apesar do desempenho promissor, os testes revelaram que o modelo enfrenta dificuldades em identificar fissuras em algumas imagens específicas, resultando em falsos negativos e falsos positivos, como observado na matriz de confusão (Figura 5) e nos resultados de classificação (Figura 6). Este cenário evidencia a necessidade de investigar outras abordagens para melhorar a capacidade de detecção do modelo.

As métricas do modelo, nos testes realizados, obtiveram uma acurácia de 0,875, precisão de 0,894, recall de 0,875 e F1-score de 0,862. Em outro cenário de avaliação, a acurácia foi de 0,6875, com precisão de 0,749, recall de 0,688 e F1-score de 0,709. Esses resultados reforçam a importância de uma análise mais ampla para identificar limitações específicas do modelo.

No presente estudo, o conjunto de validação foi composto a partir de uma divisão simples dos dados em treino e validação, sem a aplicação de técnicas de validação cruzada. Assim, embora o desempenho nas métricas tenha sido elevado, não é possível assegurar que o modelo mantenha tal desempenho em diferentes subconjuntos de dados. Adicionalmente, a ausência de validação cruzada limita a generalização dos resultados, uma vez que a performance foi avaliada em uma única configuração de divisão dos dados.

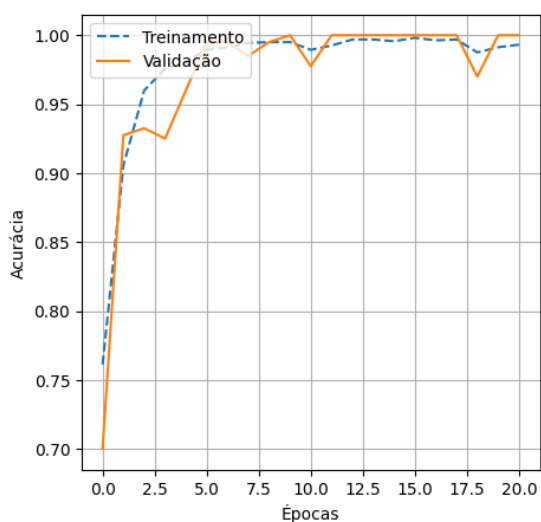


Fig. 3. Acurácia da RNC.

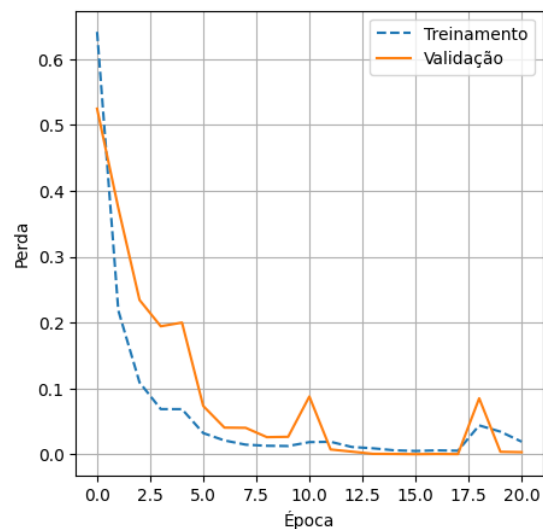


Fig. 4. Perda da RNC

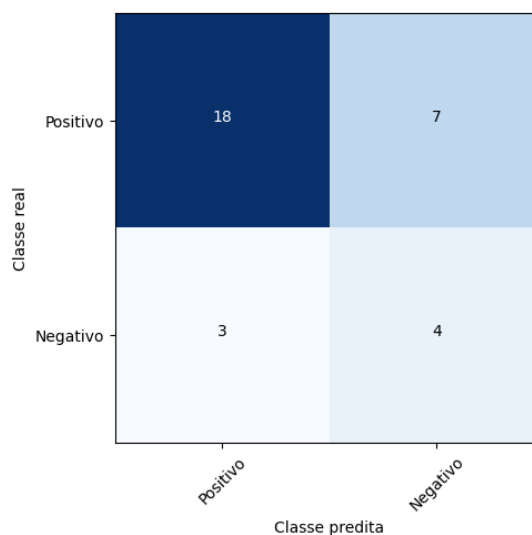


Fig. 5. Matriz de confusão.

IV. CONSIDERAÇÕES FINAIS

O presente estudo demonstrou o potencial das Redes Neurais Convolucionais (RNCs) na classificação de imagens RGB, destacando sua eficácia na extração automática de características complexas. O modelo atingiu alta acurácia e baixa perda na validação, evidenciando sua capacidade de aprender os padrões apresentados. No entanto, a ocorrência de falsos positivos e falsos negativos apontou para desafios relacionados à generalização, além de possíveis limitações no conjunto de dados utilizados.

A análise dos gráficos de desempenho revelou que o modelo atingiu rapidamente altos valores de acurácia e baixos valores

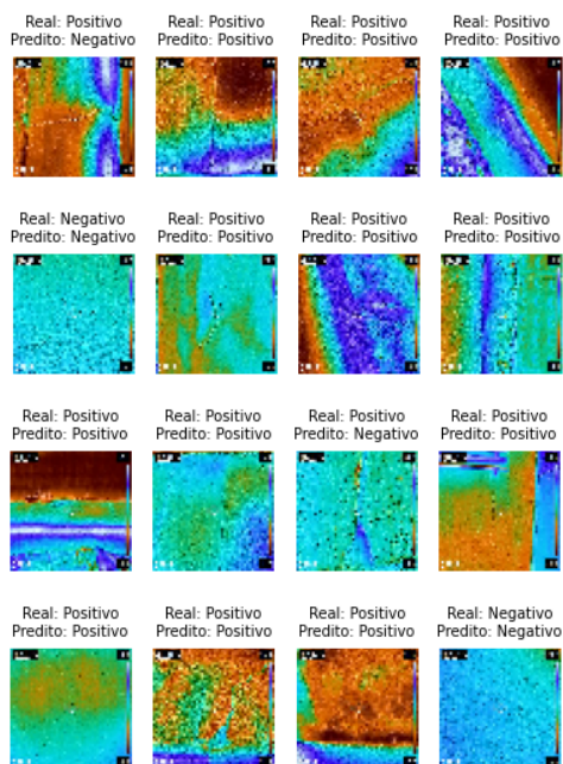


Fig. 6. Resultado da classificação com aplicação do modelo RNC.

de perda, tanto no treinamento quanto na validação, o que pode indicar uma tarefa pouco desafiadora devido à baixa diversidade dos dados. A ausência de técnicas de validação cruzada também limita a avaliação da robustez do modelo, sendo recomendada a adoção de estratégias de validação mais rigorosas em estudos futuros.

Experimentos com imagens térmicas destacaram dificuldades adicionais, como baixa resolução de textura e contraste insuficiente, especialmente em ambientes homogêneos. Esses fatores indicam a necessidade de ajustes na arquitetura da rede e da utilização de técnicas de regularização para melhor lidar com diferentes tipos de imagens.

Em síntese, este trabalho reforça a aplicabilidade das RNCs em visão computacional para a detecção de fissuras em elementos construtivos, sobretudo com a integração de dados térmicos e espectrais. Como sugestões para trabalhos futuros, propõe-se o aumento da diversidade do conjunto de dados, a utilização de diferentes tipos de câmeras e o aprimoramento da arquitetura do modelo, visando a melhoria da capacidade de generalização e a robustez em cenários mais desafiadores.

V. REFERÊNCIAS

- [1] Angela Busheska, Nicholas Sabella, Nara Almeida, e Eudes Rocha, "Machine Learning and Thermography Applied to the Detection and Classification of Cracks in Buildings," *2023 IEEE Conference on Technologies for Sustainability (SusTech)*, pp. 247–251, 2023. doi: 10.1109/SusTech57309.2023.10129614. .
- [2] Sorin Muntean e Ioan Adrian Anton (Eds.), "Coordinating Engineering for Sustainability and Resilience: Proceedings of the 4th International Conference CESARE 2024, Timișoara, Romania: Springer, 2024. doi: 10.1007/978-3-031-57800-7.
- [3] F. Forest, H. Porta, D. Tuia, e O. Fink, "From classification to segmentation with explainable AI: A study on crack detection and growth monitoring," *Automation in Construction*, vol. 165, p. 105497, 2024. doi: 10.1016/j.autcon.2024.105497. .
- [4] R. Ali, J. H. Chuah, M. S. A. Talip, N. Mokhtar, e M. A. Shoaib, "Structural crack detection using deep convolutional neural networks," *Automation in Construction*, vol. 133, p. 103989, 2022. doi: 10.1016/j.autcon.2021.103989.
- [5] Ali Waqas e Mohamad T. Araj, "Machine learning-aided thermography for autonomous heat loss detection in buildings," *Energy Conversion and Management*, vol. 304, p. 118243, 2024. doi: https://doi.org/10.1016/j.enconman.2024.118243.
- [6] S. R. Jambula, S. C. Mathi, S. N. Polu, K. G. O. JM, V. P. Mateti, e S. Yadav, "Enhanced UAV with image-driven concrete crack detection," *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1730–1737, 2023. doi: 10.1109/ICECA58529.2023.10395650.
- [7] S. Awati, *Deep learning for image recognition: Convolutional Neural Networks (RNCs)*. Academic Press, 2023.
- [8] Younes Hamishebahar, Hong Guan, Stephen So e Jun Jo, "A Comprehensive Review of Deep Learning-Based Crack Detection Approaches," *Applied Sciences*, vol. 12, p. 1374, 2022. doi: 10.3390/app12031374.
- [9] Paolo Muratore, Sina Tafazoli, Eugenio Piasini, Alessandro Laio, e Davide Zoccolan, "Prune and distill: similar reformatting of image information along rat visual cortex and deep neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1–13, 2022.
- [10] Jea Kwon, SungJun Lim, Kyungwoo Song, e Chulhee Lee, "Brain-inspired L_p -Convolution benefits large kernels and aligns better with visual cortex," *International Conference on Learning Representations (ICLR)*, pp. 1–15, 2025.
- [11] Jack Lindsey, Samuel Ocko, Surya Ganguli, e Stephane Deny, "A Unified Theory of Early Visual Representations from Retina to Cortex through Anatomically Constrained Deep CNNs," *International Conference on Learning Representations (ICLR)*, pp. 1–10, 2019.