

Classificação de patologias em estruturas usando redes neurais convolucionais: diferenciação em trincas, fissuras e rachaduras

Luís Gustavo Werle Tozevich¹, Giovani Rubert Librelotto², Pedro Luís Tozevich³

¹Curso de Ciência da Computação – Universidade Federal de Santa Maria (UFSM)
Santa Maria - RS – Brasil

²Centro de Tecnologia– Universidade Federal de Santa Maria (UFSM)
Santa Maria - RS – Brasil

³Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI)
Santa Rosa - RS – Brasil

lgtozevich@inf.ufsm.br, librelotto@inf.ufsm.br, pedrotz@bol.com.br

Abstract. *This study proposes a convolutional neural networks (CNNs) based approach for classifying cracks, fissures, and fractures through image analysis. The methodology encompasses preprocessing, data balancing, and employs the ResNet50 architecture with pooling, dropout, and regularization layers. Advanced data augmentation techniques are applied to overcome image scarcity. The model achieves approximately 96% accuracy, highlighting its effectiveness. However, areas for improvement are identified, such as continuous expansion of the realistic dataset. In summary, this study provides valuable insights into structural inspection using CNNs, with practical implications for infrastructure security and maintenance.*

Resumo. *Este estudo propõe uma abordagem baseada em redes neurais convolucionais (CNNs) para a classificação de trincas, fissuras e rachaduras por análise de imagens. A metodologia inclui pré-processamento, equilíbrio de dados e usa a arquitetura ResNet50 com camadas de pooling, dropout e regularização. Transformações avançadas de aumento de dados são aplicadas para superar a falta de imagens. O modelo atinge cerca de 96% de precisão, evidenciando sua eficácia. No entanto, oportunidades de aprimoramento são identificadas, como a expansão contínua do conjunto de dados. Em suma, este estudo oferece novas visões para a inspeção estrutural por meio de CNNs, com implicações práticas para a segurança e manutenção de infraestruturas.*

1. Introdução

A detecção e avaliação de defeitos em estruturas desempenham um papel crucial na garantia da segurança, confiabilidade e longevidade de várias estruturas. A detecção precoce e precisa de defeitos é fundamental para prevenir falhas catastróficas e custos substanciais de manutenção. Nas últimas décadas, avanços significativos em visão computacional e aprendizado de máquina abriram novas possibilidades para melhorar os métodos tradicionais de inspeção de estruturas.

O presente estudo tem como foco a aplicação de redes neurais convolucionais (CNNs) na detecção de defeitos em estruturas, com base em um conjunto de imagens coletadas na Universidade Federal de Santa Maria (UFSM). O objetivo é a análise de patologias em edifícios públicos, com ênfase na identificação da quantidade e estágio dessas manifestações patológicas.

Ao empregar as imagens obtidas na UFSM como conjunto de dados de treinamento, este estudo almeja não apenas o desenvolvimento de um modelo de classificação de patologias em estruturas, mas também a produção de informações relevantes sobre os locais de maior urgência para reparos em edifícios. A análise destas patologias e a identificação de áreas prioritárias para intervenção têm o potencial de contribuir de forma significativa para a manutenção preventiva e eficiente gestão de recursos em instalações públicas. É válido ressaltar que os resultados deste estudo possuem aplicabilidade que transcende o contexto da UFSM, podendo ser extrapoladas para diversas instituições públicas e privadas. A capacidade de identificar a diferença entre fissuras, trincas e rachaduras em diferentes contextos, beneficiará a segurança e a durabilidade das estruturas.

Portanto, busca-se contribuir para o avanço da pesquisa e da aplicação prática na detecção de manifestações patológicas, desenvolvendo e avaliando um modelo de classificação de patologias em estruturas com CNNs. Espera-se fornecer uma base sólida para futuros estudos e iniciativas voltadas para aprimorar a segurança e eficiência nas inspeções de edificações.

2. Revisão bibliográfica

Esta seção tem como objetivo analisar as abordagens atuais na detecção de patologias em estruturas, com foco nas aplicações de redes neurais convolucionais. Através de uma análise da literatura existente, pretende-se identificar as principais realizações e desafios, fornecendo ideias sobre as implicações práticas dessas abordagens para a segurança e manutenção de estruturas.

2.1. Tipos de manifestações patológicas em edificações

Quando se pensa em patologias, é comum associar a definição ao campo da medicina, referindo-se às condições ou doenças anormais que afetam diversos aspectos da vida humana [Prabhu 2023]. No entanto, patologias também podem ser observadas em outras áreas, como a construção civil. Dessa forma, assim como os seres humanos necessitam de avaliação e cuidado quando apresentam uma patologia, as edificações também precisam de atenção, para garantir a sua integridade estrutural.

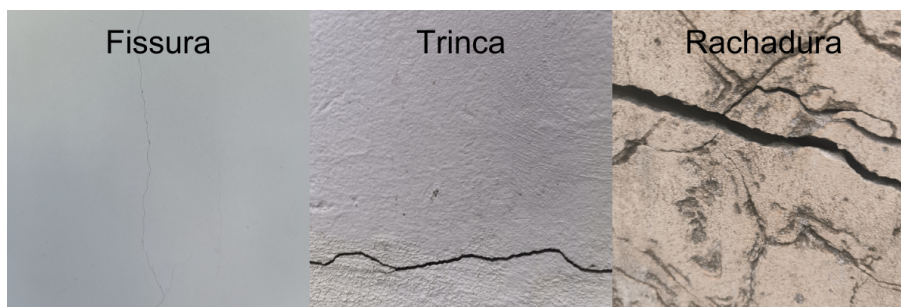
As patologias em construções abrangem uma variedade de questões que podem impactar na segurança de um edifício [Harris 2001]. Segundo Neumann (2017), as manifestações patológicas referem-se a alterações estruturais e/ou funcionais originadas por disfunções no organismo construtivo, ou seja, abrangem todos os elementos que contribuem para a degradação do material ou de suas propriedades físicas e estruturais. Conforme Dias et al. (2020), umidade, carbonatação, deslocamento de revestimento, fissuras, trincas e rachaduras são alguns exemplos de manifestações patológicas. No entanto, neste estudo focaremos apenas nos conceitos relacionados a fissuras, trincas e rachaduras.

De acordo com Oliveira (2012), as fissuras, são comumente identificadas em elementos como alvenarias, vigas, pilares, lajes, pisos, entre outros. Conforme Oliveira

(2012), tais manifestações são geralmente ocasionadas pelas tensões nos materiais, ou seja, quando os materiais são submetidos a esforços que ultrapassam sua resistência, ocorre a falha, resultando em aberturas que, dependendo de sua espessura, são classificadas como fissuras, trincas ou rachaduras.

As fissuras, segundo Corsini (2010), resultam da ruptura do material após alguma ação mecânica ou físico-química. A principal característica que as distingue das demais é que possuem uma abertura de no máximo 0,5 mm, não apresentando grandes perigos à edificação se não evoluírem para graus mais avançados. As trincas são uma evolução das fissuras, e conforme Corsini (2010), elas apresentam aberturas mais profundas e largas do que as fissuras, possuindo abertura maior que 0,5mm e menor que 1,0mm. Elas representam um estágio mais avançado de deformação nos materiais e podem ocorrer em resposta a esforços significativos. Já as rachaduras, segundo Dias et al. (2020), possuem espessura maior que 3,0mm, são maiores e mais acentuadas que as fissuras e trincas. Elas são o último estágio dessas manifestações patológicas, conforme Dias et al. (2020), as rachaduras evidenciam problemas graves na edificação e a necessidade de avaliação e reparos. A diferença entre as fissuras, trincas e rachaduras pode ser visualizada na figura 1.

Figura 1. Diferença entre fissuras, trincas e rachaduras



2.2. Métodos tradicionais de inspeção estrutural e suas limitações

A detecção de manifestações patológicas em edificações demanda a realização de uma vistoria detalhada e estrategicamente planejada, visando à determinação precisa do estado estrutural [Mazer 2012]. Consequentemente, não é possível obter resultados satisfatórios, realizando-se inspeções apenas quando as patologias alcançarem um alto grau de desenvolvimento. Conforme argumentado por Gomide et al. (2020), as inspeções nas edificações devem ser realizadas de modo contínuo. Essa abordagem é necessária para manter a segurança e a conservação da edificação, visando a identificação dos problemas e a estruturação de um plano para reparos [Gomide et al. 2020].

A condução de uma avaliação do estado de uma edificação, como descrito por Mazer (2012), desempenha um papel crucial na determinação dos processos para reabilitação da edificação. Carvalho (2009) define a inspeção visual como uma investigação *in loco*, a fim de, identificar as diversas manifestações patológicas. Dentre os métodos vigentes, a inspeção visual, destaca-se como um dos mais empregados. Segundo Mazer (2012), esse é um método não destrutivo, podendo ser executado em qualquer componente estrutural, além de possuir diversos outros benefícios, como sua simplicidade e baixo custo. Contudo, vale notar que esse método sempre requer a presença de um profissional qualificado no local.

Consequentemente, há a necessidade de explorar abordagens que permitam a realização desse processo por indivíduos não especializados ou que viabilizem análises à distância por parte de profissionais especializados na área. Com propósito de identificar como as manifestações patológicas se manifestam [Brito 2017].

Entretanto, o método tradicional esbarra em desafios ao inspecionar estruturas de maiores proporções, tornando locais elevados praticamente inacessíveis e impondo complexas demandas logísticas à inspeção predial. Como resposta a esses dilemas, emergiu a abordagem de inspeção assistida por Veículos Aéreos Não Tripulados (VANTs) [Melo & Costa 2015]. Tondelo (2019) argumenta que a utilização de VANTs acelera a avaliação das edificações, particularmente em zonas que antes não eram acessíveis aos profissionais.

Contudo, apesar desses avanços, persiste uma lacuna no processo, uma vez que a análise do relatório fotográfico gerado pelo VANT ainda requer intervenção profissional. Diante disso, o presente estudo busca automatizar a classificação de determinadas manifestações patológicas mediante a aplicação de conceitos atuais em visão computacional.

2.3. Abordagens baseadas em visão computacional

Enxergar, visualizar e distinguir desempenham papéis fundamentais para nós humanos [Marr, 2010]. Os olhos, atuam como dispositivos de captação, convertem a luz em sinais transmitidos ao cérebro através do nervo óptico, onde são assimilados para compreensão [Palmer 1999; Kandel et al., 2012]. Szeliski (2022) define a visão computacional como a tecnologia que utiliza sistemas computacionais para análise e interpretação visual de informações em tempo real, buscando a capacidade visual humana.

No âmbito da visão computacional, o maior desafio é descrever o mundo captado em imagens e reconstruir suas características, algo realizado cotidianamente por humanos e animais, todavia, para algoritmos de visão computacional, frequentemente encontramos erros nessas traduções [Boden 2006]. Para Szeliski (2022) a amplitude de aplicações da visão computacional é notável, abrangendo desde reconhecimento óptico de caracteres, inspeção de máquinas, setores de varejo e logística, imagens médicas, veículos autônomos, construção de modelos 3D, entre outras. No contexto deste estudo, que busca identificar manifestações patológicas em estruturas, a análise avançada de imagens por meio de algoritmos e técnicas computacionais emerge como uma ferramenta promissora para aprimorar as inspeções em estruturas.

A detecção e avaliação de defeitos em estruturas representa um campo de pesquisa ativo, com diversas abordagens disponíveis para resolver esse desafio complexo. Bai et al. (2022) introduziram uma rede neural profunda para detectar automaticamente danos estruturais e rachaduras em imagens. De forma similar, Chen et al. (2022) desenvolveram um método aprimorado baseado em CNNs para detectar e reconhecer algumas manifestações patológicas em estruturas. Cha et al. (2017) propuseram uma abordagem baseada em conceitos de visão computacional para detecção de fissuras em concreto, usando redes neurais convolucionais para melhorar a eficácia da detecção em condições do mundo real. Alipour e Harris (2020) contribuíram para detecção de rachaduras em diferentes materiais. Diniz et al. (2023) utilizaram redes neurais profundas para detectar diversas patologias em estruturas de concreto, incluindo rachaduras, fragmentação, eflorescências, manchas de corrosão e barras de aço expostas.

Em geral, os trabalhos existentes utilizam técnicas de *deep learning* e visão computacional. No entanto, ainda há espaço para avanços na diferenciação específica entre trincas, fissuras e rachaduras por análise de imagens.

2.4. Redes neurais convolucionais e suas aplicações

De acordo com Kim (2017), uma CNN possui múltiplas camadas ocultas, buscando imitar o processamento e o reconhecimento de imagens no cérebro. É assim que a CNN se diferencia em termos de conceito e funcionamento em relação às redes neurais prévias [Kim 2017].

De acordo com O'Shea (2015), em uma Rede Neural Artificial (RNA) os dados de entrada, frequentemente representados como um conjunto de números, são direcionados através de camadas ocultas, nestas camadas, a rede toma decisões com base nas informações da camada anterior e avalia como mudanças nos dados afetam o resultado final. Isso é identificado por O'Shea (2015) como o processo de aprendizado. Segundo O'Shea (2015), as CNNs compartilham semelhanças com as RNAs convencionais, uma vez que ambas são constituídas por neurônios que se otimizam por meio do aprendizado.

De acordo com Li et al. (2017), as CNNs desempenham um papel crucial no *deep learning*, pois têm a capacidade de aproveitar grandes volumes de dados para melhorar seus resultados. A principal diferença entre as CNNs e as RNAs convencionais é que as CNNs são principalmente aplicadas no reconhecimento de padrões em imagens [O'Shea 2015]. Desta forma, as CNNs ao simularem o complexo processo de processamento visual do cérebro humano, revelam-se uma ferramenta notável para a análise e interpretação de imagens. Por meio do seu uso, abrem-se portas para compreensão das manifestações patológicas em estruturas.

3. Metodologia

Nesta seção, descreve-se a metodologia utilizada para desenvolver e avaliar o modelo de classificação de patologias em estruturas usando redes neurais convolucionais.

3.1. Uso de redes neurais convolucionais para a detecção de padrões

Chauhan, R. et al. (2018) afirma que as redes neurais convolucionais são algoritmos de *deep learning* projetados para analisar imagens de entrada, aplicando operações de convolução com filtros ou núcleos para extrair características. As CNNs são amplamente utilizadas para a detecção de padrões, pois são capazes de aprender representações hierárquicas e invariantes dos dados, que facilitam a identificação de objetos, rostos, gestos e outros padrões complexos [Goodfellow et al., 2016]. De acordo com Chauhan, R. et al. (2018), as CNNs têm se mostrado eficazes para reconhecimento e detecção de imagens.

A arquitetura básica de uma CNN consiste em várias camadas, que podem ser divididas em três tipos principais: camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas [Nielsen, 2015].

As CNNs são treinadas por meio de algoritmos de aprendizado supervisionado, que ajustam os pesos dos filtros e das conexões da rede de acordo com um critério de otimização, que geralmente é baseado na minimização de uma função de custo (ou perda)

que mede o erro entre a saída da rede e a saída desejada (ou rótulo) para cada dado de entrada. Um dos algoritmos mais utilizados para o treinamento de CNNs é o método do gradiente descendente estocástico (SGD), que atualiza os pesos da rede em pequenos passos na direção oposta ao gradiente da função de custo [Ruder 2016].

As CNNs têm demonstrado um desempenho superior ao de outros métodos de aprendizado de máquina para a detecção de padrões em diversas áreas e aplicações, como visão computacional, processamento de linguagem natural, reconhecimento biométrico, análise médica, entre outras [Schmidhuber 2015]. E por isto elas foram escolhidas para realizar este trabalho.

3.2. Elaboração do *dataset*

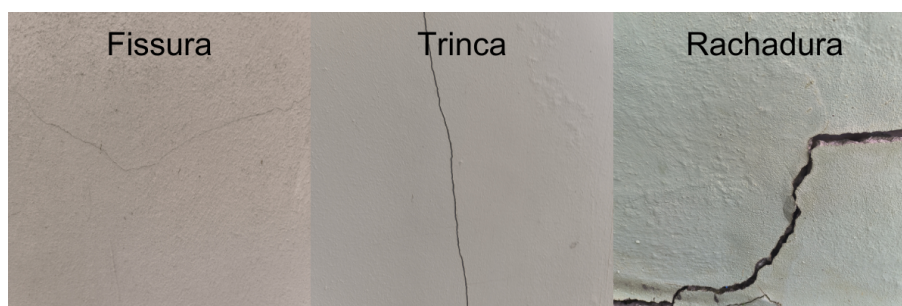
A elaboração do *dataset* compreendeu a preparação dos dados essenciais para o treinamento do modelo de CNN proposto. As imagens foram obtidas por meio de um relatório fotográfico conduzido em 70 edificações localizadas na Universidade Federal de Santa Maria. Posteriormente, as fotografias foram categorizadas manualmente mediante critérios técnicos conforme descrito no item 2.1, e validadas individualmente por um engenheiro civil especializado.

O *dataset* resultante é composto por um total de 2040 imagens, distribuídas entre 883 imagens de fissuras, 777 de trincas e 380 de rachaduras. Durante a captura das fotografias das patologias, foram considerados diferentes horários do dia, introduzindo variações de sombra e iluminação nas imagens. Além disso, o conjunto de dados inclui imagens capturadas sob diversos ângulos e distâncias das manifestações patológicas. Vale destacar que, em alguns casos, foram tiradas múltiplas fotos de uma mesma patologia para proporcionar variação de perspectiva, enriquecendo assim o treinamento do modelo e capacitando-o a classificar as patologias em uma ampla gama de cenários, levando em consideração variações de iluminação, distância e perspectiva.

3.3. Preparação dos dados

Primeiramente, foi importado o conjunto de dados formado por imagens pertencentes a diversas classes de patologias. As imagens coletadas foram organizadas em pastas correspondentes às suas respectivas classes. Cada pasta estava nomeada com o tipo de manifestação patológica nela presente. Esse procedimento facilitou o processo de rotulação das imagens, que consistiu em atribuir uma etiqueta a cada imagem conforme a sua classe. A figura 2 apresenta exemplo das três manifestações patológicas avaliadas.

Figura 2. Representação das imagens presentes no *dataset*

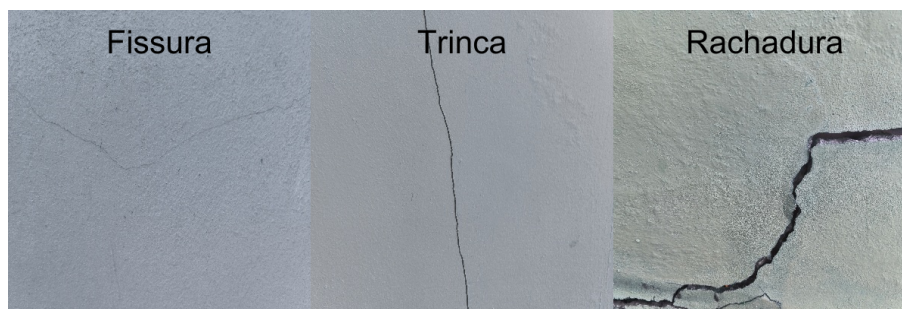


O *dataset* formado pelas imagens rotuladas foi dividido aleatoriamente pelo algoritmo em três subconjuntos: treinamento, validação e teste. O que justifica essa etapa é a possibilidade da avaliação do modelo em diferentes etapas do aprendizado. O subconjunto de treinamento foi utilizado para ajustar os parâmetros do modelo, o subconjunto de validação foi utilizado a fim de avaliar o resultado do modelo no decorrer das épocas e o subconjunto de teste foi utilizado para avaliar o desempenho do modelo final em dados não vistos anteriormente. A proporção entre os subconjuntos foi de 64%, 16% e 20%, respectivamente.

3.4. Pré-processamento das imagens

Primeiramente, todas as imagens foram ajustadas para um formato padronizado de 400x400 *pixels* com três canais de cor (RGB), para garantir a uniformidade das dimensões de entrada da CNN. Depois, a normalização da intensidade de *pixel* foi realizada, ajustando os valores dos *pixels* de todas as imagens para o intervalo [0, 1]. Essa etapa foi fundamental para garantir que os atributos das imagens estivessem na mesma faixa de valores, para possibilitar o processo de aprendizado da CNN. A figura 3 apresenta as imagens após o seu pré-processamento.

Figura 3. Imagens pré-processadas



Por fim, para lidar com o desbalanceamento de imagens por classe, visando evitar que o modelo fosse enviesado para as classes com mais exemplo, calculou-se a média de imagens por classe no conjunto de treinamento. Após verificar a necessidade do aumento de dados em alguma classe, selecionaram-se algumas imagens do conjunto já pré-processado de treinamento para aplicar algumas transformações, como espelhamento horizontal, rotação, e ajustes de brilho e variação de contraste. Essa técnica aumentou a diversidade das amostras nas classes sub-representadas, tornando o conjunto de treinamento mais rico.

3.5. Estratégias de aumento de dados para superar a escassez de imagens

A escassez de amostras em algumas classes do conjunto de dados foi um desafio para o treinamento da CNN. Para abordar esse problema, foram implementadas estratégias de aumento de dados para enriquecer o conjunto de treinamento. O aumento de dados consiste em aplicar transformações nas imagens originais, criando novas imagens com variações sutis, mas que mantêm as características relevantes para a classificação [Shorten and Khoshgoftaar 2019]. As estratégias de aumento de dados utilizadas neste trabalho foram:

- Rotação: Imagens foram aleatoriamente rotacionadas em diferentes ângulos para criar variações de perspectiva. A rotação foi aplicada usando uma função que gera um ângulo aleatório entre -15 e 15 graus e rotaciona a imagem de acordo com esse ângulo.
- Espelhamento horizontal: Imagens foram espelhadas horizontalmente. O espelhamento horizontal foi aplicado usando uma função que inverte a imagem ao longo do eixo horizontal.
- Ajustes de brilho e contraste: Foram aplicados ajustes aleatórios de brilho e contraste para criar variações nas intensidades de *pixel* das imagens. Os ajustes de brilho e contraste foram aplicados usando uma função que gera dois fatores aleatórios entre 0,8 e 1,2 e multiplica os valores dos *pixels* pelos fatores.
- Translação (deslocamento): Em alguns casos, as imagens foram deslocadas em pequenas quantidades verticalmente e horizontalmente. A translação foi aplicada usando uma função que gera dois valores aleatórios entre -10 e 10 *pixels* e desloca a imagem de acordo com esses valores.

Essas estratégias de aumento de dados foram aplicadas, melhorando a capacidade do modelo de generalização, tornando-o mais eficaz na classificação de estruturas em imagens.

3.6. Criação do modelo

Para a criação do modelo, adotou-se uma abordagem que envolveu a utilização da arquitetura ResNet50 como base. Essa escolha se fundamentou na eficácia da ResNet50, na extração de características complexas e hierárquicas de imagens. No entanto, para adaptação da arquitetura a essa tarefa específica, introduziram-se camadas personalizadas.

Primeiramente, acrescentou-se uma camada de *Global Average Pooling*, que desempenhou um papel importante na redução da dimensionalidade dos recursos extraídos pela ResNet50. Essa camada gerou um vetor unidimensional para cada imagem, facilitando a fase de classificação subsequente. Além disso, incluiu-se uma camada de *dropout*, que aplicou uma taxa de desligamento aleatório aos neurônios da camada anterior, a fim de combater o *overfitting*. O *dropout* é uma técnica de regularização que desativa aleatoriamente um percentual de unidades (neurônios) durante o treinamento. Impedindo que o modelo se torne excessivamente adaptado aos dados de treinamento e, em vez disso, o incentiva a aprender representações mais robustas. Essa medida teve o propósito de evitar um ajuste excessivo do modelo aos dados de treinamento, contribuindo para uma melhor generalização.

A camada de saída do modelo foi configurada de acordo com o número de classes na nossa tarefa de classificação, que envolveu cinco classes. A função de ativação *softmax* foi empregada para transformar os resultados em probabilidades associadas a cada classe.

Durante o processo de treinamento, algumas das camadas da ResNet50 permaneceram congeladas, mantendo seus pesos pré-treinados fixos. No entanto, permitimos que as camadas finais fossem treinadas especificamente, possibilitando a adaptação dos recursos aprendidos pela ResNet50 às características específicas das imagens.

Adicionalmente, configura-se um *callback* que monitorava a perda de validação durante o treinamento. Caso essa perda não demonstrasse melhorias significativas, o

callback ajustava dinamicamente a taxa de aprendizado, contribuindo para otimizar o processo de treinamento do modelo.

Por fim, o modelo foi configurado para aprendizado de transferência. Isso envolveu a liberação parcial das camadas da ResNet50 para poderem ser atualizadas durante o treinamento. As camadas personalizadas, por outro lado, foram treinadas com dados específicos da tarefa de classificação de patologias em estruturas. Essa estratégia aproveitou o conhecimento prévio da ResNet50 em tarefas de classificação de imagens, economizando tempo e recursos computacionais.

3.7. Arquitetura ResNet50: detalhes e características

Para a criação do modelo, é fundamental estabelecer uma arquitetura sólida e adaptá-la às demandas específicas da tarefa, nesse caso, a classificação de patologias em estruturas. Para essa finalidade, a arquitetura escolhida como base foi a ResNet50.

A ResNet50 é uma arquitetura de rede neural convolucional que foi desenvolvida pela Microsoft Research em 2015 [He et al. 2016]. É uma variante da popular arquitetura ResNet. O “50” no nome se refere ao número de camadas na rede, que tem 50 camadas de profundidade [Datagen, 2020]. Implicando na capacidade de aprender representações complexas dos dados, capturando características sutis e hierárquicas que são fundamentais para a identificação das fissuras, trincas e rachaduras.

4. Treinamento e avaliação

O treinamento do modelo seguiu a metodologia proposta na seção 3. O conjunto de dados, composto por imagens representativas de várias classes de patologias, foi importado e verificado quanto à qualidade e resolução. Em seguida, as imagens foram organizadas em pastas de acordo com suas respectivas classes e etiquetas foram atribuídas com base em critérios técnicos.

A compilação do modelo envolveu a definição de aspectos para o treinamento e a avaliação. Utilizou-se o otimizador Adam, reconhecido por ser um algoritmo eficiente e adaptativo para atualizar os pesos do modelo. A função de perda quantifica a discrepância entre as previsões geradas pelo modelo e os rótulos verdadeiros das imagens. Para avaliar o desempenho do modelo, recorreremos à acurácia e à precisão como métricas, fornecendo uma medida da qualidade das classificações realizadas. Também criamos uma matriz de confusão para examinar em detalhes as previsões do modelo em cada classe, identificando acertos e erros específicos.

5. Resultados e discussão

Após o treinamento da CNN, foi possível verificar as métricas de desempenho do modelo nos conjuntos de treinamento, validação e teste. Essas métricas fornecem uma visão geral do quão bem o modelo está executando sua tarefa de classificação. Os resultados estão apresentados na tabela 1.

A acurácia refere-se à medida geral de quão bem o modelo está realizando a tarefa de classificação. Ela é representada como a porcentagem de previsões corretas em relação ao total de previsões feitas pelo modelo. A acurácia de treinamento é uma métrica que quantifica a concordância entre as previsões do modelo e os dados de treinamento, geralmente resultando em um valor elevado, uma vez que o modelo é ajustado

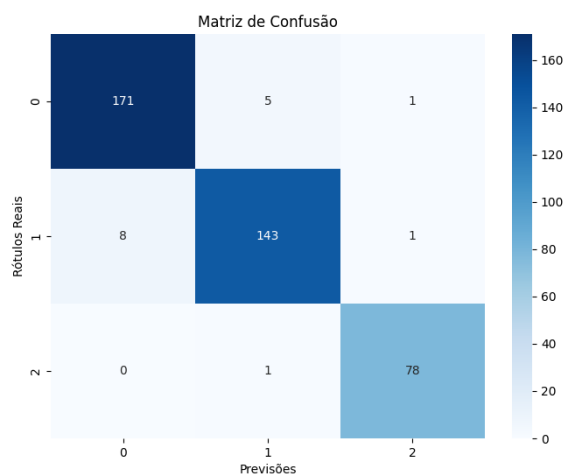
Tabela 1. Resultados obtidos após o treinamento

<i>Conjunto de Dados</i>	<i>Acurácia</i>
Treinamento	95,91%
Validação	90,51%

para minimizar o erro nesses dados específicos. Em contrapartida, a acurácia de validação avalia o desempenho do modelo em um conjunto de dados independente, não utilizado no processo de treinamento, desempenhando o papel de indicador de sua habilidade de generalização para novos dados. É comum que a acurácia de validação seja ligeiramente inferior, e sua importância reside na detecção do *overfitting*, uma condição na qual o modelo se adapta excessivamente aos dados de treinamento e não consegue generalizar eficazmente para dados desconhecidos. Ambas as métricas desempenham um papel crucial na avaliação abrangente do modelo, garantindo sua capacidade de realizar previsões precisas em instâncias inéditas. No contexto dessas métricas, é possível observar que no conjunto de treinamento a acurácia atingiu um valor de 95,91%, indicando que o modelo se ajustou bem aos dados de treinamento. Porém, ao avaliarmos a acurácia no conjunto de validação, essa métrica foi ligeiramente menor, registrando 90,51%.

A matriz de confusão detalha as previsões do modelo em relação às classes reais no conjunto de teste. Ela é organizada de tal forma que cada linha representa as instâncias da classe real e cada coluna representa as instâncias da classe prevista. O valor na célula na interseção da *i*-ésima linha e *j*-ésima coluna é o número de observações conhecidas por pertencerem à *i*-ésima classe, mas previstas pela *j*-ésima classe. Os elementos na diagonal principal da matriz de confusão correspondem às instâncias corretamente classificadas, enquanto os outros elementos correspondem às instâncias incorretamente classificadas. Portanto, um modelo perfeito teria todos os valores fora da diagonal principal como zero. Isso nos permite entender melhor como o modelo está se comportando em termos de falsos positivos, falsos negativos e acertos. A matriz de confusão pode ser visualizada na figura 4. O conjunto de testes da matriz de confusão foi separado antes da realização

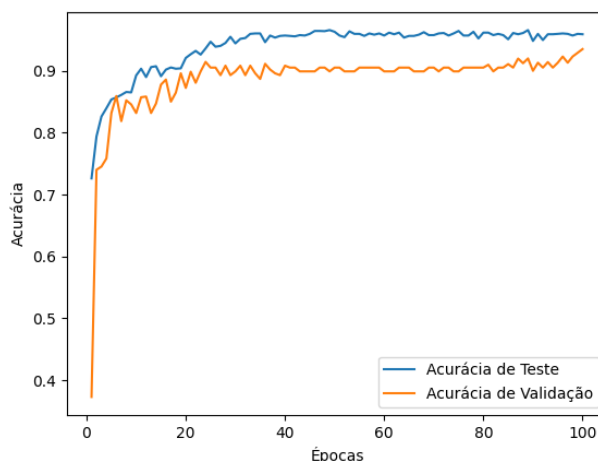
Figura 4. Matriz de confusão do modelo



do aumento de dados, portanto o teste não foi realizado com o mesmo quantitativo de cada imagem. Realizou-se o processo dessa forma, pois a utilização somente de imagens reais proporcionou o teste com imagens que realmente são capturadas e enviadas ao modelo, evitando um possível erro pelas manipulações realizadas nas imagens. Com base na matriz de confusão, observamos o número de imagens corretamente classificadas para cada classe. A precisão do conjunto de teste, com base na matriz de confusão, foi calculada em relação ao número de previsões corretas sobre o número de previsões possíveis, obtendo-se uma precisão de 96%.

Ao longo do treinamento do modelo, monitorou-se a evolução da precisão em cada época. A figura 5 ilustra essa evolução, mostrando como a precisão no conjunto de treinamento e no conjunto de validação se comportou ao longo das épocas. No acompa-

Figura 5. Gráfico de acurácia



nhamento do treinamento do modelo ao longo das épocas, nota-se uma certa tendência. O gráfico de evolução da acurácia revelou um aumento consistente dessa métrica com o passar das épocas. Isso indica que o modelo estava aprendendo eficazmente com os dados de treinamento, ajustando seus parâmetros para reduzir o erro. Embora a acurácia no conjunto de validação tenha permanecido abaixo da acurácia de treinamento, esse resultado estava dentro das expectativas, já que o conjunto de validação contém dados não vistos anteriormente pelo modelo. No entanto, a manutenção de uma acurácia de cerca de 93% no conjunto de validação demonstra que o modelo conseguiu generalizar bem para dados desconhecidos, afastando preocupações com *overfitting*.

Através dos resultados obtidos, evidenciados pelas métricas de desempenho, análise da matriz de confusão e pelo acompanhamento da evolução da acurácia ao longo do treinamento, pode-se avaliar a eficácia do modelo. Obteve-se uma acurácia de treinamento de 95,91% e uma acurácia de teste de 96%, demonstrando um ótimo desempenho. Além disso, a capacidade de generalização para o conjunto de validação, com uma acurácia de cerca de 93%, evidencia a ausência de *overfitting*. Esses resultados representam um passo importante no desenvolvimento de soluções eficazes para a classificação de patologias em estruturas, com implicações significativas para aplicações na engenharia civil.

6. Conclusão

Este estudo proporcionou uma compreensão sobre a aplicação das redes neurais convolucionais na inspeção estrutural, com foco na identificação de trincas, fissuras e rachaduras, representando uma abordagem promissora para aprimorar a identificação dessas falhas estruturais. Os resultados obtidos demonstram altos índices de precisão nas etapas de treinamento, teste e validação do modelo, enfatizando a capacidade desse método de diferenciar com precisão trincas, fissuras e rachaduras.

No entanto, enfatiza-se que este trabalho representa apenas o ponto de partida no desenvolvimento desse modelo. A expansão contínua do conjunto de dados, considera-se uma etapa essencial para aprimorar a qualidade do modelo. Além disso, a exploração de diferentes arquiteturas de CNNs sugere-se como um possível caminho para melhorar a precisão do modelo. A validação em cenários do mundo real é fundamental, permitindo avaliar o desempenho em condições práticas, aumentando, a confiabilidade dos resultados.

Assim, percebe-se um avanço na diferenciação de manifestações patológicas, com implicações práticas amplas, especialmente no que diz respeito à segurança e manutenção de estruturas. Estabelecendo uma base sólida para pesquisas futuras e iniciativas voltadas para o aprimoramento desta tecnologia. A abordagem proposta tem o potencial de mudar como abordamos a manutenção e a segurança das estruturas que nos cercam, por meio da combinação de visão computacional, aprendizado de máquina e inspeção estrutural. Resultando em reduções significativas nos riscos e custos associados às falhas estruturais, bem como na melhoria da eficiência dos processos de manutenção, permitindo intervenções mais oportunas e eficazes. Obtendo-se edificações mais seguras e confiáveis, beneficiando a sociedade como um todo.

Referências

- Alipour, M., & Harris, D. K. (2020). *Increasing the robustness of material-specific deep learning models for crack detection across different materials*. Engineering Structures, 206, 110157.
- Bai, Y., Zha, B., Sezen, H., & Yilmaz, A. (2022). *Engineering deep learning methods on automatic detection of damage in infrastructure due to extreme events*. Structural Health Monitoring, 22(1), 338-352.
- Boden, M. A. (2008). *Mind as machine: A history of cognitive science*. Oxford University Press.
- Brito, T. F. D. (2017). *Análise de manifestações patológicas na construção civil pelo método gut: estudo de caso em uma instituição pública de ensino superior*.
- Carvalho, N. F. De. (2009). *Verificação de patologias de elementos estruturais em concreto armado*. Revista Obras Civis, 1(1), 38-40.
- Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). *Deep learning-based crack damage detection using convolutional neural networks*. Computer-Aided Civil and Infrastructure Engineering, 32(5), 361-378.
- Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018). *Convolutional neural network (CNN) for image detection and recognition*. In 2018 first international conference on

- secure cyber computing and communication (ICSCCC) (pp. 278-282). IEEE.
- Chen, Z., Wang, C., Wu, J., Deng, C., & Wang, Y. (2022). *Deep convolutional transfer learning-based structural damage detection with domain adaptation*. *Applied intelligence*, 53(5), 5085-5099.
- Corsini, R. (2010). *Trinca ou fissura*. São Paulo: Técnica, 160.
- Datagen (2020) *Resnet-50: The Basics And A Quick Tutorial*. In: Datagen Blog. Disponível Em: <https://datagen.tech/guides/computer-vision/resnet-50/>. Acesso Em: 03 Set. 2023.
- Dias, A. P. L., do Amaral, I. A. R., & dos Santos Amarante, M. (2021). *Patologias das construções*. *Revista Pesquisa e Ação*, 7(1), 66-80.
- Diniz, J. de C. N., Paiva, A. C. de., Junior, G. B., Almeida, J. D. S. de., Silva, A. C., Cunha, A. M. T. da S., & Cunha, S. C. A. P. da S.. (2023). *A Method for Detecting Pathologies in Concrete Structures Using Deep Neural Networks*. 13(9), 16.
- Gomide, T. L. F., Neto, J. C. P. F., Gullo, M. A., & Della Flora, S. M. (2020). *Inspeção predial total*. Oficina de Textos.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Harris, S. Y. (2001). *Building pathology: deterioration, diagnostics, and intervention*. John Wiley & Sons.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., & Hudspeth, A. J. (2012). *Principles of Neural Science*. McGraw-Hill Education. New York.
- Kim, P. (2017). *Matlab deep learning. With machine learning, neural networks and artificial intelligence*. 130(21).
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). *A survey of convolutional neural networks: analysis, applications, and prospects*. *IEEE transactions on neural networks and learning systems*.
- Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- Mazer, W. (2012). *Inspeção e ensaios em estruturas de concreto*. Curitiba: UTFPR.
- Melo, R. R. S., & Costa, D. B. (2015). *Uso de veículo aéreo não tripulado (VANT) para inspeção de logística em canteiros de obra*. SIBRAGEC-ELAGEC, São Carlos: São Paulo (Brasil).
- Neumann, P. N., Cagol, A. C., Visoscki, P. C., & Edler, M. A. R. (2017). *Patologias nas edificações: uma nova concepção na construção civil*. *Revista Interdisciplinar de Ensino, Pesquisa e Extensão-RevInt*, 4(1).
- Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25, pp. 15-24). San Francisco, CA, USA: Determination press.

- O'Shea, K., & Nash, R. (2015). *An introduction to convolutional neural networks*. arXiv preprint arXiv:1511.08458.
- Oliveira, A. M. de. (2012). *Fissuras, trincas e rachaduras causadas por recalque de diferencial de fundações*.
- Palmer, Stephen E. (1999). *Vision Science: Photons To Phenomenology*. Mit Pres.
- Prabhu, S. R. (2023). *Introduction to Pathology*. In Textbook of General Pathology for Dental Students (pp. 1-4). Cham: Springer Nature Switzerland.
- Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. arXiv preprint arXiv:1609.04747.
- Schmidhuber, J. (2015). *Deep learning in neural networks: An overview*. Neural networks, 61, 85-117.
- Shorten, C., & Khoshgoftaar, T. M. (2019). *A survey on image data augmentation for deep learning*. Journal of big data, 6(1), 1-48.
- Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.
- Tondelo, P. G., & Barth, F. (2019). *Análise das manifestações patológicas em fachadas por meio de inspeção com VANT*. PARC Pesquisa em Arquitetura e Construção.