



UNIVERSIDADE FEDERAL DO PARÁ - UFPA  
INSTITUTO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Thabatta Moreira Alves de Araújo

**Detecção de Danos em Superfícies  
Geotécnicas com Redes Neurais  
Convolucionais de Baixa Complexidade**

Belém

2024

Thabatta Moreira Alves de Araújo

# **Detecção de Danos em Superfícies Geotécnicas com Redes Neurais Convolucionais de Baixa Complexidade**

Tese de Doutorado submetida à avaliação da Banca Examinadora aprovada pelo colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará para a obtenção do Grau de Doutora em Engenharia Elétrica, na área de Computação Aplicada.

Universidade Federal do Pará - UFPA

Orientador: Professor Dr. Carlos Renato Lisboa Francês

Belém

2024

Solicite sua ficha catalográfica em: <<http://bcficat.ufpa.br/>>

**“DETECÇÃO DE DANOS EM SUPERFÍCIES GEOTÉCNICAS COM REDES  
NEURAIIS CONVOLUCIONAIS DE BAIXA COMPLEXIDADE”**

AUTORA: THABATTA MOREIRA ALVES DE ARAÚJO

TESE DE DOUTORADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO  
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO  
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE DOUTORA EM ENGENHARIA  
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 29/05/2024

**BANCA EXAMINADORA:**



Documento assinado digitalmente

CARLOS RENATO LISBOA FRANCES

Data: 29/05/2024 19:13:04-0300

Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Carlos Renato Lisboa Francês**  
(Orientador – PPGEE/UFPA)

Documento assinado digitalmente



JASMINE PRISCYLA LEITE DE ARAUJO

Data: 29/05/2024 19:39:54-0300

Verifique em <https://validar.iti.gov.br>

---

**Prof.<sup>a</sup> Dr.<sup>a</sup> Jasmine Prsicyla Leite de Araújo**  
(Avaliadora Interna - PPGEE/UFPA)

Documento assinado digitalmente



JOAO CRISOSTOMO WEYL ALBUQUERQUE COS

Data: 29/05/2024 19:45:44-0300

Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. João Crisóstomo Weyl Albuquerque Costa**  
(Avaliador Interno - PPGEE/UFPA)

Documento assinado digitalmente



ANDRE CARLOS PONCE DE LEON FERREIRA DE C

Data: 03/06/2024 13:18:43-0300

Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. André Ponce de Leon Ferreira de Carvalho**  
(Avaliador Externo - ICMC/USP)

Documento assinado digitalmente



CLAUDIO ALEX JORGE DA ROCHA

Data: 30/05/2024 16:18:04-0300

Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Cláudio Alex Jorge da Rocha**  
(Avaliador Externo - IFPA)

**VISTO:**

---

**Prof. Dr. Diego Lisboa Cardoso**

(Coordenador do PPGEE/ITEC/UFPA)

*Dedico a todas as mães pesquisadoras este trabalho, na esperança de que ele sirva  
como um símbolo de reconhecimento da vossa (nossa) grandeza.*

# Agradecimentos

Primeiro e acima de tudo, meu agradecimento a Deus. A cada prece, a cada súplica, encontrei a força e a sabedoria para seguir em frente.

Ao Professor Carlos Renato Lisboa Francês: Mais do que um orientador, o senhor se tornou um guia inspirador e um exemplo de liderança. Agradeço por sua confiança, por ter respeitado minha área de atuação e autonomia, por sua paciência infinita, compreensão e justiça.

Ao IFPA e Professor Cláudio Alex da Rocha: Agradeço ao IFPA por me oportunizar desenvolver como profissional. Em especial na pessoa do Professor Cláudio Alex, por sua visão estratégica e compromisso com a interiorização da pós-graduação enquanto reitor. As parcerias proporcionadas como o Minter/Dinter entre UFPA/UNIFESSPA/IFPA são iniciativas que democratizam o acesso ao conhecimento e impulsionam o desenvolvimento de regiões distantes dos grandes centros, como sul do Pará, onde morei por tantos anos da minha vida.

Aos meus queridos filhos, Ana Luíza e André, e meu esposo Marcelo: Espero que um dia compreendam que as ausências também foram por vocês. Uma mãe e esposa, feliz e realizada, contribui para a felicidade da família. Agradeço por me motivarem a buscar sempre o meu melhor.

À UFPA, ao PPGEE e à Norte Energia S.A: Agradeço por proporcionarem um ambiente propício ao aprendizado e ao desenvolvimento profissional. A cada membro da equipe do projeto "Deep learning e machine learning para extração de padrões em estruturas críticas da UHE Belo Monte", agradeço pelo conhecimento compartilhado e pela colaboração. Em especial, ao meu amigo e colega, Carlos André pela sua amizade, parceria e apoio durante todos esses anos.

Ao CEFET-MG: por me acolher e me oportunizar mecanismos para continuar a minha pesquisa.

Em cada etapa desta jornada, encontrei apoio, incentivo e aprendizados. Agradeço a todos que, de alguma forma, contribuíram para a realização desta tese.

*'O que é Deus? É a inteligência suprema, a causa primária de todas as coisas.'*  
*(Allan Kardec, 1857)*

# Resumo

A maioria dos desastres naturais resulta de eventos geodinâmicos, como deslizamentos de terra e colapso de estruturas geotécnicas. Essas falhas causam catástrofes que impactam diretamente o meio ambiente e causam perdas financeiras e humanas. A inspeção visual é o principal método para detectar falhas em superfícies de estruturas geotécnicas. Todavia, as visitas no local podem ser arriscadas devido à possibilidade de solo instável. Além disso, o design do terreno e as condições de instalação hostis e remotas inviabilizam o acesso a essas estruturas. Quando uma avaliação rápida e segura é necessária, a análise por visão computacional torna-se uma alternativa. No entanto, estudos em técnicas de visão computacional ainda precisam ser explorados neste campo devido às particularidades da engenharia geotécnica, como dados públicos limitados, redundantes e escassos. Neste contexto, esta tese apresenta uma abordagem com Redes Neurais Convolucionais para a identificação de defeitos na superfície de estruturas geotécnicas com o objetivo de reduzir a dependência de inspeções no local conduzidas por humanos. Para tanto, foram coletadas imagens de taludes às margens de uma rodovia brasileira, com o auxílio de veículo aéreo não tripulado (VANT) e dispositivos móveis. Em seguida, foram exploradas arquiteturas de baixa complexidade para construir um classificador binário capaz de detectar em imagens falhas aparentes a olho nu humano. A arquitetura composta por 3 camadas convolucionais, cada uma com 32 filtros, seguidas por duas camadas densas de 128 neurônios cada, e saída com um neurônio apresentou acurácia de 94,26%. A avaliação de desempenho com o conjunto de teste obteve índice AUC de 0,99, matriz de confusão e precisão-revocação (AUPRC) que indicam desempenho robusto do classificador mesmo com desequilíbrio de classes, ao mesmo tempo que mantém uma baixa complexidade computacional, tornando-a adequada para aplicações práticas em campo. As contribuições da tese incluem a disponibilização de banco de imagens, a obtenção de um modelo de classificação adequado para dados escassos e desequilibrados e recursos computacionais limitados, e uma estratégia para automação da inspeção em estruturas geotécnicas.

**Palavras-chave:** Visão computacional. Redes Neurais Convolucionais. Arquitetura enxuta. Estruturas geotécnicas. Identificação de falhas. Inspeção remota.



# Abstract

Most natural disasters result from geodynamic events, such as landslides and collapse of geotechnical structures. These failures are catastrophic that directly impact the environment and cause financial and human losses. Visual inspection is the main method for detecting surface flaws in geotechnical structures. However, visits to the site can be risky due to the possibility of soil's instability. Furthermore, the terrain design, hostile environment and remote installation conditions make access to these structures impractical. When a quick and safe assessment is necessary, computer vision analysis becomes a potential alternative. However, studies on computer vision techniques still need to be explored in this field due to the particularities of geotechnical engineering, such as limited, redundant and scarce public data sets. In this context, this thesis presents a *redes neurais convolucionais*, do inglês *Convolutional Neural Network* (CNN) approach for identifying defects on the surface of geotechnical structures to reduce dependence on human-led on-site inspections. To this end, images of surface failure indicators were collected on slopes on the banks of a Brazilian highway, with the help of UAVs and mobile devices. Next, low-complexity CNN architectures were explored to build a binary classifier capable of detecting flaws apparent to the naked human eye in images. The architecture composed of three convolutional layers, each with 32 filters, followed by two fully connected layers, each composed of 128 neurons and output with one neuron, showed an accuracy of 94.26%. The performance evaluation of the model with the test set obtained AUC metrics of 0.99, confusion matrix, and a AUPRC curve that indicates robust performance of the classifier in detecting damage, while maintaining a low computational complexity, making it suitable for applications field practices. The contributions of the thesis include the provision of an image database, the obtaining of a classification model suitable for scarce data and limited computational resources, and the exploration of strategies for remote inspection and detection of signs of failure in geotechnical structures.

**Keywords:** Computer vision. Convolutional Neural Networks (CNNs). Suitable architecture. Geotechnical structures. Defect identification. Remote inspection.

# Lista de ilustrações

Figura 1.	Deslizamento às margens da rodovia 381, Brasil 2022. . . . .	28
Figura 2.	Erosão na superfície de taludes às margens da rodovia 381, Brasil 2022. . . . .	29
Figura 3.	Deslizamentos na superfície de taludes às margens da rodovia 381, Brasil 2022. . . . .	30
Figura 4.	Ação biológica de formigueiro em taludes às margens da rodovia 381, Brasil 2022. . . . .	32
Figura 5.	Vegetação típica em taludes às margens da rodovia 381, Brasil 2022. . . . .	33
Figura 6.	Descrição das tarefas de aprendizado de máquina por tipo e objetivo. . . . .	37
Figura 7.	Modelo de neurônio artificial - Perceptron. . . . .	38
Figura 8.	Estrutura típica da arquitetura de uma CNN. . . . .	41
Figura 9.	Representação de 32 filtros da camada de convolução, em escala de cinza, núcleo tamanho 3x3. . . . .	44
Figura 10.	Ação de 32 filtros na camada de convolução. . . . .	45
Figura 11.	Agrupamento máximo. . . . .	46
Figura 12.	Funções de ativação comuns em CNN. . . . .	47
Figura 13.	Efeitos da transformação geométrica em uma imagem. . . . .	58
Figura 14.	Estruturação de uma matriz de confusão. . . . .	64
Figura 15.	Curva <i>Receiver Operating Characteristic</i> (ROC). . . . .	65
Figura 16.	Curvas de precisão-revocação para diferentes limiares. . . . .	67
Figura 17.	Fluxograma para detecção de falhas de superfície em estruturas geotécnicas. . . . .	84
Figura 18.	Imagens de monitoramento de vulcões pelo GeoNet. . . . .	86
Figura 19.	Perigos geológicos avaliados pela NBMG. . . . .	87
Figura 20.	Efeito do clima nas imagens de satélite, São Sebastião- São Paulo, 2022. . . . .	89

Figura 21.	Imagens de satélite de grandes deslizamentos, São Sebastião-São Paulo, 2023. . . . .	89
Figura 22.	Exemplos de imagens anotadas como "com dano". . . . .	92
Figura 23.	Exemplos de imagens anotadas como "intacto". . . . .	92
Figura 24.	Número de parâmetros treináveis por arquitetura. . . . .	97
Figura 25.	Arquitetura proposta. . . . .	99
Figura 26.	Acurácia média de cinco execuções. . . . .	102
Figura 27.	Perda média de cinco execuções. . . . .	102
Figura 28.	Matriz de confusão. . . . .	104
Figura 29.	Curva ROC-AUC . . . . .	106
Figura 30.	Curva Precisão-Revocação. . . . .	107
Figura 31.	Imagem de entrada. . . . .	108
Figura 32.	Mapa de características na primeira camada de convolução. . . .	109
Figura 33.	Mapa de características na segunda camada de convolução. . . .	110
Figura 34.	Mapa de características na terceira camada de convolução. . . .	110
Figura 35.	Teste do modelo na detecção de danos visíveis em superfícies de taludes. . . . .	112

# Lista de tabelas

Tabela 1.	Compilado de trabalhos correlatos. . . . .	80
Tabela 2.	Divisão das amostras em treinamento, validação e teste. . . . .	93
Tabela 3.	Valores médios de acurácia e perda das arquiteturas executadas. . . . .	96
Tabela 4.	Informações das camadas do modelo CNN. . . . .	102

# Lista de siglas e abreviaturas

**Adam** *Adaptive Moment Estimation*

**ANN** redes neurais artificiais, do inglês *Artificial Neural Network*

**ANEEL** Agência Nacional de Energia Elétrica

**AUC** *Area Under the ROC Curve*

**AUPRC** *Area Under the Precision-Recall Curve*

**BVP** problemas de contorno, do inglês *Boundary Value Problem*

**CEFET-MG** Centro Federal de Educação Tecnológica de Minas Gerais

**CNN** redes neurais convolucionais, do inglês *Convolutional Neural Network*

**CILAMCE** *Ibero-Latin American Congress on Computational Methods in Engineering*

**CPU** unidade central de processamento, do inglês *Central Processing Unit*

**DA** aumento de dados, do inglês *Data Augmentation*

**DL** aprendizado profundo, do inglês *Deep Learning*

**DenseNet** *Densely Connected Convolutional Networks*

**FC** densa ou totalmente conectada, do inglês *Fully Connected*

**FEM** método dos elementos finitos, do inglês *Finite Element Method*

**FMEA** análise de modos de falha e seus efeitos, do inglês *Failure Mode and Effect Analysis*

**FN** Falso Negativo

**FP** Falso Positivo

**IA** Inteligência Artificial

**IFPA** Instituto Federal de Educação, Ciência e Tecnologia do Pará

**IoT** Internet das Coisas, do inglês *Internet of Things*

**GPU** unidade de processamento gráfico, do inglês *Graphics Processing Unit*

**MLP** redes neurais artificiais multicamadas, do inglês *multi-layer Perceptron*

**ML** aprendizado de máquina, do inglês *Machine Learning*

**NBMG** Departamento de Minas e Geologia de Nevada, do inglês *Nevada Bureau of Mines and Geology*

**PRC** *Precision-Recall Curve*

**RF** *Random Forest*

**ResNet** redes residuais, do inglês *Residual Networks*

**ReLU** *Rectified Linear Unit*

**ROC** *Receiver Operating Characteristic*

**RGB** vermelho, verde, e azul, do inglês *Red, Green, Blue*

**SVM** máquina de vetores de suporte, do inglês *Singular Vector Machine*

**TVP** Taxa de Verdadeiros Positivos

**TFP** Taxa de Falsos Positivos

**TC** Tomografia Computadorizada

**TL** transferência de aprendizado, do inglês

**UHE** Usina Hidrelétrica

**VANT** Veículo Aéreo Não Tripulado

**VC** Visão Computacional

**VGG** *Visual Geometry Group*

**VN** Verdadeiro Negativo

**VP** Verdadeiro Positivo

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Hipótese</b>	<b>20</b>
<b>1.2</b>	<b>Objetivos</b>	<b>21</b>
<b>1.3</b>	<b>Organização da Tese</b>	<b>21</b>
<b>2</b>	<b>DANOS EM SUPERFÍCIES GEOTÉCNICAS</b>	<b>23</b>
<b>2.1</b>	<b>Definição de Termos</b>	<b>23</b>
<b>2.2</b>	<b>Inspeção Visual</b>	<b>25</b>
<b>2.3</b>	<b>Danos Visíveis em Superfícies de Estruturas Geotécnicas</b>	<b>27</b>
2.3.1	Erosão	28
2.3.2	Deslizamento	30
2.3.3	Bioerosão e Bioturbação	31
2.3.4	Desvegetação	32
<b>3</b>	<b>REDES NEURAIS CONVOLUCIONAIS</b>	<b>35</b>
<b>3.1</b>	<b>Aprendizado de Máquina</b>	<b>35</b>
<b>3.2</b>	<b>Redes Neurais</b>	<b>37</b>
<b>3.3</b>	<b>Redes Neurais Convolucionais em Tarefas de Classificação de Imagens</b>	<b>39</b>
3.3.1	Convolução	41
3.3.2	Agrupamento	44
3.3.3	Função de Ativação	47
3.3.4	Linearização	49
3.3.5	Conectividade Total	50
<b>3.4</b>	<b>Desafios em Redes Neurais Convolucionais</b>	<b>51</b>
<b>3.5</b>	<b>Regularização</b>	<b>54</b>
3.5.1	Abandono	55
3.5.2	Aumento de Dados	56
3.5.3	Otimização Adaptativa	58



<b>3.6</b>	<b>Modelos de Referência em Redes Neurais Convolucionais . . .</b>	<b>60</b>
<b>3.7</b>	<b>Avaliação de Desempenho do Modelo . . . . .</b>	<b>62</b>
3.7.1	Acurácia e Perda . . . . .	62
3.7.2	Precisão, Revocação, Especificidade e Pontuação F1 . . . . .	63
3.7.3	Matriz de Confusão . . . . .	64
3.7.4	Curva ROC e índice AUC . . . . .	65
<b>4</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>68</b>
<b>4.1</b>	<b>Avaliação da Integridade Estrutural . . . . .</b>	<b>68</b>
<b>4.2</b>	<b>Visão Computacional . . . . .</b>	<b>72</b>
<b>4.3</b>	<b>Identificação de Lacunas . . . . .</b>	<b>78</b>
<b>5</b>	<b>CONSTRUÇÃO DE UM CLASSIFICADOR PARA A DETEC-</b>	
	<b>ÇÃO DE DANOS APARENTES EM TALUDES . . . . .</b>	<b>82</b>
<b>5.1</b>	<b>Visão Geral do Processo . . . . .</b>	<b>82</b>
<b>5.2</b>	<b>Ambiente Operacional . . . . .</b>	<b>84</b>
<b>5.3</b>	<b>Geração das Amostras . . . . .</b>	<b>85</b>
<b>5.4</b>	<b>Aumento de Dados . . . . .</b>	<b>91</b>
<b>5.5</b>	<b>Divisão das Amostras . . . . .</b>	<b>91</b>
<b>5.6</b>	<b>Arquitetura Proposta . . . . .</b>	<b>92</b>
<b>6</b>	<b>RESULTADOS E DISCUSSÃO . . . . .</b>	<b>100</b>
<b>6.1</b>	<b>Desempenho do Modelo . . . . .</b>	<b>101</b>
<b>6.2</b>	<b>Visualização das Ativações de Rede . . . . .</b>	<b>108</b>
<b>7</b>	<b>CONCLUSÃO . . . . .</b>	<b>115</b>
<b>7.1</b>	<b>Contribuições . . . . .</b>	<b>115</b>
<b>7.2</b>	<b>Limitações e Dificuldades . . . . .</b>	<b>118</b>
<b>7.3</b>	<b>Desdobramentos e Trabalhos Futuros . . . . .</b>	<b>120</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>122</b>

# 1 Introdução

De uma perspectiva geotécnica, a maioria das catástrofes naturais decorre de diversos eventos geodinâmicos, incluindo deslizamentos de terra, instabilidade de encostas e outros fenômenos. Esses desastres naturais podem ter impactos abrangentes, afetando o meio ambiente, causando perdas financeiras e humanas e impactos diretos em todo o ecossistema (SOLÓRZANO et al., 2022; LÖBMANN et al., 2020). Em ambientes onde a estabilidade estrutural é primordial, como é o caso das estruturas geotécnicas, é imperativo permanecer atento para quaisquer indicações de falha na superfície. Estes indicadores podem ser sinais de alerta precoce de um potencial compromisso estrutural que pode levar a uma falha catastrófica. A observação desses sinais de alerta possibilita a adoção de medidas proativas para evitar tais falhas. No entanto, prever falhas e prevenir desastres geotécnicos é uma tarefa desafiadora (HANDY, 2007) e multi abordagem.

A identificação dos danos na superfície da estrutura fornece informações essenciais sobre o comportamento do solo e das rochas de estruturas geotécnicas, tais como escavações, barragens, aterros, obras de contenção e formações geológicas naturais. Uma abordagem primária para identificar falhas é através de inspeção visual durante visitas ao local. As vistorias em campo desempenham um papel crucial na engenharia geotécnica e geologia de engenharia, pois fornecem informações essenciais sobre o desempenho da estrutura, (HANDY, 2007; LIM et al., 2021; HAN et al., 2022). Sinais aparentes de dano como erosão, deslizamento, ausência de cobertura do solo, infiltração, surgência, recalques podem ser facilmente identificados pelo olho humano.

Embora eficiente, a inspeção visual é um procedimento fortemente dependente do trabalho de especialistas, e sujeito a subjetividades (PAN; CHEN, 2015). Por isso, com a finalidade de registrar e levantar informações relevantes para análise sobre a segurança das estruturas dados digitais são produzidos, na forma de imagens, áudio e vídeo. No rol de tecnologias para a produção desses dados, são empregados satélites e câmeras, e mais recentemente, os Veículo Aéreo Não

Tripulado (VANT) e robôs (GREENWOOD et al., 2019; KANELLAKIS; NIKOLAKOPOULOS, 2017; JANG; KIM; AN, 2019; LEI et al., 2020; TIAN et al., 2023; CHA; CHOI; BüyüKöZTürk, 2017; DI et al., 2014). A utilização dessas tecnologias aliada ao avanço da computação tem despertado o interesse na automatização da inspeção, às luzes da visão computacional.

Nesse sentido, o estado da arte apresenta soluções promissoras para a vistoria remota e extração de padrões em imagens digitais (SPENCER; HOSKERE; NARAZAKI, 2019). Dentre os modelos de Visão Computacional (VC), as arquiteturas de redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN), têm se destacado com sucesso na identificação de falhas em superfícies de estruturas de concreto, como destacado nos estudos de (CHA; CHOI; BüyüKöZTürk, 2017; LEI et al., 2020; LI et al., 2019b; SPENCER; HOSKERE; NARAZAKI, 2019). No campo da visão computacional as CNNs são reconhecidas pela robustez em tarefas de processamento de imagens devido a arquitetura multicamadas capazes de extrair automaticamente recursos de imagens durante a fase de treinamento (LECUN; BENGIO; HINTON, 2015; LI et al., 2022). O que as coloca em vantagem em relação a técnicas tradicionais de processamento de imagens, as máquina de vetores de suporte, do inglês *Singular Vector Machine* (SVM), e *Random Forest* (RF), que embora robustas, exigem extração manual das características (MYAGILA; KILAVO, 2022; SUDHA et al., 2019).

Embora pesquisas recentes indiquem o potencial de tecnologias como visão computacional, aprendizado de máquina, aprendizado profundo e *big data* na engenharia, particularmente no contexto de estruturas geotécnicas de solo e rocha, sua aplicação ainda é relativamente incipiente (ZHANG; LI; LI, 2021; PHOON; ZHANG, 2023). Na prática, observa-se uma proporção menor de estudos neste campo que empregam essas tecnologias em comparação com outras áreas (ZHANG; LI; LI, 2021). Isso abre um leque promissor de oportunidades para pesquisa e aplicação de modelos, desde os clássicos até os mais avançados, no âmbito das estruturas de solo e rocha.

A desproporção observada nos estudos sobre identificação de danos em superfícies de estruturas geotécnicas pode ser atribuída às particularidades inerentes às imagens nesse contexto. Considerando que as imagens representam a principal

fonte de dados para modelos de CNNs em tarefas de visão computacional, torna-se elementar levar em conta as características intrínsecas desses dados. Pois a qualidade dos dados de treinamento impacta significativamente o desempenho dos modelos, exigindo uma análise criteriosa para garantir a efetividade da solução proposta.

A aquisição de imagens com danos visíveis em estruturas geotécnicas, por si, é uma limitação intrínseca aos modelos de visão computacional, pois eventos como falhas e catástrofes são raros, indesejáveis e imprevisíveis por natureza. Essa eventual escassez de dados leva a um desequilíbrio significativo entre imagens com e sem indicativos de danos em conjuntos de dados de taludes, o que gera o problema de desequilíbrio de classes. Essa disparidade conduz a modelos de classificação que tendem a favorecer a classe majoritária (BARELLA; GARCIA; CARVALHO, 2019), representada pelas imagens sem danos, e que negligenciam a identificação precisa das classes minoritárias, que correspondem aos danos visíveis nos taludes.

Além disso, há particularidades visuais dos taludes. Os taludes, componentes essenciais das estruturas geotécnicas, moldam a paisagem e fornecem informações relevantes sobre a estabilidade do solo (TERZAGHI; PECK; MESRI, 1996). A cobertura vegetal uniforme em condições saudáveis, embora esteticamente agradável, representa um desafio para modelos de visão computacional. Nesse sentido, apesar do grande volume de imagens coletadas por VANT e dispositivos móveis nas inspeções rotineiras, além de desequilíbrio de classe tem-se a redundância nas imagens dos taludes saudáveis, que é considerável. A similaridade das imagens pode levar ao sobreajuste, limitando a capacidade dos modelos de identificar e classificar com precisão as nuances das características geotécnicas.

Sendo assim, a abordagem adequada para classificação de imagens é um fator determinante para enfrentar as escassez e desequilíbrio de dados e obter o sucesso na tarefa de visão computacional nesse campo da engenharia. Principalmente em cenários com restrições extremas, como locais hostis e remotos, que exigem a detecção de falhas em tempo real para prevenção de desastres causados pelo colapso de infraestrutura crítica. Nesses ambientes, as severas limitações de *hardware* e energia impõem desafios adicionais à seleção de modelos adequados.

Para superar os desafios computacionais em ambientes geotécnicos, diversas

estratégias de otimização e transferência de aprendizado podem ser exploradas. A adaptação de redes pré-treinadas, a aplicação de técnicas de compressão de imagem, a investigação do uso de modelos neurais leves já existentes, e a construção de arquiteturas personalizadas são exemplos de estratégias. No contexto da arquitetura de rede, a literatura elenca CNNs de referência como *Visual Geometry Group* (VGG), Inception e redes residuais, do inglês *Residual Networks* (ResNet) (YADAV; JADHAV, 2019; ABEDALLA et al., 2021), como robustas e complexas para diversas aplicações.

A implementação dessas estratégias pode ser desafiadora em alguns contextos, pois exige mais recursos computacionais à medida que a rede se aprofunda (BRIGATO; IOCCHI, 2021; DARAPANENI; KRISHNAMURTHY; PADURI, 2020). Modelos complexos, que demandam alto poder de processamento para treinamento e execução, podem ser inviáveis, pois muitas estruturas geotécnicas estão localizadas em áreas com restrições de recursos. Para lidar com conjuntos de dados escassos e recursos limitados em modelos CNN, o estado da arte propõe arquiteturas mais enxutas e de menor complexidade, demonstrando sua eficácia na promoção da eficiência computacional sem comprometer a precisão (BRIGATO; IOCCHI, 2021; FOROUGH; CHEN; WANG, 2021; LIMÃO; ARAÚJO; FRANCES, 2023).

## 1.1 Hipótese

A hipótese central desta pesquisa busca responder à seguinte pergunta "em cenários desafiadores de escassez de dados, desequilíbrio de classes e limitação de recursos computacionais para experimentação e avaliação dos modelos, é possível desenvolver uma arquitetura de CNN robusta para automatizar a inspeção, com auxílio de VANT, capaz de detectar danos em superfícies de estruturas geotécnicas com alta precisão?".

Para testar a viabilidade da hipótese principal, esta tese se debruça sobre os seguintes desdobramentos:

- I. As bases de dados públicas de imagens de taludes são insuficientes ou pouco representativas para análise de dados em superfícies de estruturas geotécnicas.

- II. É possível representar um fenômeno multiclasse por meio de uma classificação binária em engenharia geotécnica.
- III. Arquiteturas enxutas e personalizadas de CNN apresentam alto desempenho diante da escassez de amostras e restrição recursos computacionais.
- IV. A inspeção visual pode ser automatizada pela integração de tecnologias, como VANT e dispositivos geradores de imagem, a técnicas de extração de padrões de imagem.

## 1.2 Objetivos

A presente tese tem como principal objetivo investigar a aplicação da integração VANT a arquiteturas de Redes Neurais Convolucionais (CNN) na identificação de danos em superfícies de estruturas geotécnicas, no contexto da visão computacional e considerando um cenário de escassez de amostras e desequilíbrio de classes. Para tal, foram coletadas imagens de um problema do mundo real, com auxílio de VANT e dispositivos móveis. A partir dessas imagens, foi construído um classificador binário e comparado o desempenho de diferentes arquiteturas na tarefa de identificar danos visíveis a olho nu em imagens de taludes.

## 1.3 Organização da Tese

Considerando a natureza multidisciplinar desta tese, que abrange temas de diversas áreas da engenharia, matemática, computação, geologia e geotecnia, os capítulos foram estruturados de forma a fornecer os principais conceitos necessários para a compreensão dos objetivos propostos e dos resultados alcançados.

A obra é composta por sete capítulos, organizados da seguinte maneira:

- Capítulo 2 - Danos em Superfícies Geotécnicas

Este capítulo introduz termos técnicos e a definição de dano em estruturas geotécnicas, focando em anomalias visíveis. A compreensão desses conceitos é essencial para entender o modelo de visão computacional proposto.

- Capítulo 3 - Redes Neurais Convolucionais

O capítulo fundamenta os modelos de visão computacional utilizados na tese, com foco nas Redes Neurais Convolucionais. Aborda os princípios do aprendizado de máquina, a concepção de redes neurais, a arquitetura típica de CNN e sua aplicação na classificação de imagens.

- Capítulo 4 - Trabalhos Relacionados

Este capítulo apresenta uma revisão sistemática da literatura para identificar pesquisas existentes sobre aprendizado de máquina, aprendizado profundo e redes convolucionais. A revisão visa estabelecer uma base sólida para a metodologia proposta nesta tese, que se utiliza de análise de imagens para a inspeção automatizada dessas estruturas.

- Capítulo 5 - Construção de um Classificador para a Detecção de Danos Aparentes em Taludes

Este capítulo detalha a metodologia da pesquisa, desde a coleta e pré-processamento dos dados até o treinamento e validação do modelo de Redes Neurais Convolucionais. A seleção de hiperparâmetros e a busca por uma arquitetura otimizada visam alcançar os objetivos propostos.

- Capítulo 6 - Resultados e Discussão Este capítulo demonstra os resultados da pesquisa sobre a detecção de falhas em estruturas geotécnicas utilizando Redes Neurais Convolucionais. A análise detalhada dos resultados, incluindo medidas de desempenho e comparação com métodos tradicionais, demonstra a efetividade e confiabilidade do modelo proposto.

- Capítulo 7 - Conclusão Este capítulo apresenta as contribuições da pesquisa para o campo de estudo, os desafios encontrados durante a pesquisa, e os desdobramentos e as perspectivas para trabalhos futuros.

## 2 Danos em Superfícies Geotécnicas

### Considerações Iniciais

Este capítulo apresenta uma breve descrição de alguns termos técnicos relevantes para a análise de falhas em superfícies de estruturas geotécnicas. A definição de dano, nesta seção, é abordada de forma restrita, focando nos aspectos relacionados às anomalias em superfícies identificáveis a olho nu. No entanto, é importante reconhecer que essa definição não captura a totalidade das falhas que podem comprometer a segurança e a confiabilidade das obras. A compreensão dos termos e da definição de dano restrita, em conjunto com o conhecimento de técnicas de visão computacional, será fundamental para a apreensão do modelo proposto nos capítulos subsequentes, que visa automatizar a identificação de falhas em superfícies de estruturas geotécnicas.

### 2.1 Definição de Termos

A mecânica dos solos é uma área da engenharia que se ocupa a investigar o comportamento e a resistência dos solos e rochas. O termo solos por si é um termo abrangente, do qual se enquadram materiais com diferentes comportamentos e propriedades, (TERZAGHI; PECK; MESRI, 1996; HANDY, 2007). É pertinente a conceituação de alguns termos, que ao longo desta tese, são abordados.

Considerando as definições apresentadas por (VARNES, 1978):

- a) Rocha: Massa dura ou firme de material inorgânico.
- b) Solo: Agregado de partículas sólidas, tipicamente minerais e rochas, que foram transportados ou formados pelo intemperismo das rochas. Os gases ou líquidos que preenchem os poros do solo tornam-se parte do solo.
- c) Terra: Material em que 80% ou mais das partículas são menores que



2mm, o limite superior de partículas do tamanho de areia.

- d) Lama: descreve material heterogêneo, fluído, em que 80% ou mais das partículas são menores que 0,06 mm.

As estruturas geotécnicas podem ser entendidas como construções, ou elementos de construções, cuja concepção e dimensionamento são essencialmente controlados pelo comportamento mecânico e hidráulico dos maciços terrosos ou rochosos onde estão implantadas, (TERZAGHI; PECK; MESRI, 1996). São exemplos deste tipo de estrutura as fundações, as barragens, as estruturas de contenção de terras, os taludes naturais ou artificiais, os aterros e as obras subterrâneas, entre outras. E podem ser definidos como:

- a) Fundação: são estruturas normalmente enterradas, usualmente em concreto ou metal, com a finalidade de prover sustentação às construções. Podem ser rasas ou profundas, conforme a carga transmitida ao terreno, o que delimita também o aspecto construtivo.
- b) Barragens: as barragens são estruturas geotécnicas construídas com as finalidades diversas, tais como o armazenamento de água, geração de energia elétrica, entre outros, (ICOLD, 2023). Podem ser construídas utilizando-se solo compactado, concreto armado, enrocamento ou a parte granular do próprio material a ser contido, em diferentes métodos construtivos e diferentes configurações.
- c) Contenções: estruturas com a finalidade de prover estabilidade contra a ruptura de maciços de terra ou rocha. Como por exemplo os muros de arrimo ou muros de gravidade, constituídos de concreto ciclópico, concreto armado, cortina de tirantes, gabiões, solo cimento ensacado, muros em forma de cortina com perfis metálicos com painéis pré-moldados, estacas pranchas, entre outros.
- d) Taludes: Os taludes ou encostas naturais são superfícies inclinadas de maciços terrosos, rochosos ou mistos (solo e rocha). Podem ser naturais ou artificiais (construídos pela ação do homem)

Destaca-se que a maioria dessas estruturas complexas são divididas em elementos que trabalham em conjunto para suportar cargas e garantir a estabilidade.

Em alguns casos, o talude em si pode ser considerado uma estrutura geotécnica completa, especialmente quando projetado e construído com técnicas específicas para garantir sua estabilidade. Isso ocorre em situações de taludes criados em áreas montanhosas para construção de estradas, ferrovias ou de aterros compactados, taludes artificiais construídos com materiais compactados para elevar o nível do terreno ou criar diques (FERNANDES, 2018).

## 2.2 Inspeção Visual

As inspeções estruturais constituem procedimentos de fundamental importância para a avaliação do estado atual das infraestruturas. A identificação visual de falhas se configura como um componente vital e primário na análise do solo, possibilitando a descrição metódica e detalhada da fotografia da estrutura, conforme destacado por (DAS, 2011). Essa relevância se justifica pelo objetivo primordial das inspeções, que reside na garantia da segurança e da facilidade de manutenção das infraestruturas, mediante a detecção de danos visíveis a olho nu humano. Nesse sentido, agências governamentais investem consideráveis recursos financeiros na realização de inspeções periódicas em infraestruturas críticas, como pontes e barragens, conforme atesta (SPENCER; HOSKERE; NARAZAKI, 2019).

A detecção de danos em infraestruturas, como taludes e encostas, configura-se como um processo essencial para a avaliação do estado de saúde dessas estruturas e a tomada de medidas preventivas contra colapsos e desastres. Essa atividade envolve a identificação e o diagnóstico de patologias, considerando três aspectos fundamentais (FARRAR; LIEVEN, 2007):

- a) Tipo de Dano: A natureza da patologia deve ser claramente definida, permitindo a classificação em diferentes categorias, como erosão, rachaduras, deslizamentos, entre outras.
- b) Localização do Dano: A identificação precisa da localização do dano é essencial para a implementação de medidas corretivas direcionadas e eficazes. Essa localização pode ser expressa em coordenadas geográficas, referências topográficas ou outros sistemas de posicionamento adequados.

- c) Taxa de Crescimento do Dano: A avaliação da taxa de crescimento do dano permite a estimativa da evolução da patologia ao longo do tempo, possibilitando a tomada de decisões preventivas com base em cenários probabilísticos.

A detecção de danos em estruturas geotécnicas por meio da inspeção visual geralmente requer uma rigorosa avaliação técnica assistida por procedimentos e ferramentas especializadas (DAS, 2017; VARNES, 1978). A identificação de uma única causa exata de falha é muitas vezes inatingível. Pois uma combinação de elementos, geralmente contribui para o desencadeamento das falhas (VOLKWEIN et al., 2011; KHAN et al., 2021). São elementos a saber:

- a) Parâmetros Morfológicos: Geometria (forma e das dimensões da infraestrutura, incluindo inclinações, relevos e outros aspectos topográficos)
- b) Parâmetros Geológicos: Tipo de solo e rochas (Materiais geológicos que compõem a infraestrutura), intempéries (ação de fatores climáticos, como chuva, vento e temperatura, etc.), variabilidade da estrutura geológica (heterogeneidade das formações geológicas subjacentes à infraestrutura),
- c) Parâmetros Hidrogeológicos: Nível de água subterrânea (nível freático), fluxo de água superficial,
- d) Fraturamento.
- e) Propriedades Mecânicas do Solo e das Rochas: Coesão (capacidade do solo de se manter unido, resistindo ao cisalhamento), Ângulo de atrito interno (ângulo de atrito entre as partículas do solo),
- f) Movimento de Massa: O deslocamento de grandes volumes de solo ou rochas, como deslizamentos e fluxos de detritos.

A análise conjunta desses parâmetros, com base em técnicas de monitoramento adequadas, permite a detecção precisa e abrangente de danos em infraestruturas, possibilitando a implementação de medidas preventivas eficazes e a garantia da segurança e da confiabilidade dessas estruturas. Para esse monitoramento a engenharia apresenta estratégias multi abordagem baseadas em: inspeções no local (PAN; CHEN, 2015; MASCAREÑAS et al., 2021), instrumentação de campo

(JUNG; BERGES; GARRETT, 2014; SALAJEGHEH; MEYMAND; KERMANI, 2018) e ensaios de laboratório (FIGUEIREDO et al., 2011). Apoiadas por modelos numéricos ou analíticos, como elementos finitos, elementos de contorno, equação de diferenças, e as teorias da mecânica dos solos e rochas (TERZAGHI; PECK; MESRI, 1996; DAS, 2011; SHI et al., 2022).

No âmbito das estruturas de solo e rocha, a inspeção visual se configura como um elemento fundamental no processo de avaliação de sua condição e estabilidade. Essa prática, consolidada na engenharia de estruturas, conforme atestado (HAN et al., 2022; SPENCER; HOSKERE; NARAZAKI, 2019), assume ainda mais relevância em estruturas de solo e rocha. Nestas, observa-se um quantitativo significativo de indicadores de falha que não podem ser diagnosticados diretamente por instrumentos de campo, ou por ensaios em laboratório.

Embora a inspeção visual seja uma ferramenta valiosa, é importante reconhecer que essa metodologia apresenta diversas fontes de variabilidade que podem levar a interpretações ambíguas (MASCAREÑAS et al., 2021; PAN; CHEN, 2015), principalmente devido à dependência do grau de experiência do profissional que a realiza (SPENCER; HOSKERE; NARAZAKI, 2019; MASCAREÑAS et al., 2021).

### 2.3 Danos Visíveis em Superfícies de Estruturas Geotécnicas

É válido ressaltar que existem diferenças consideráveis na identificação de falhas em estruturas. O material (concreto, metálicas, terra e rocha), o ambiente de instalação (urbano, rural, residencial, entre outros) e as relações constitutivas (relação de tensão e deformação) variam substancialmente conforme a categoria da estrutura. O que implica em abordagens específicas para a análise de falhas conforme a categoria em que a estrutura se enquadra. Por exemplo, as teorias sobre fraturas em estruturas de concreto são distintas das teorias aplicadas a fraturas em rochas e solos.

A análise da superfície do terreno configura-se como um método eficaz para a detecção de falhas geológicas e geotécnicas. A ausência de cobertura vegetal, em particular, se destaca como um indicador recorrente e representativo de irregularidades no solo, como evidenciado em cenários de erosão e deslizamentos, comumente

Figura 1. Deslizamento às margens da rodovia 381, Brasil 2022.



Fonte: Autora.

encontrados em áreas suscetíveis a tais fenômenos (BENZER, 2010; NOGUEIRA et al., 2020), como evidencia a Figura 21.

De forma que, o mapeamento de solo sem cobertura do solo, deslizamentos, e erosão, contribui para a solução de problemas de solos (TERZAGHI; PECK; MESRI, 1996). As próximas seções abordam algumas das falhas visíveis mais comuns em superfícies de estruturas geotécnicas, facilmente identificáveis a olho nu.

### 2.3.1 Erosão

A erosão é o desgaste superficial das partículas do solo ou aprisionamento por água da chuva, vento, mudanças de temperatura ou outros agentes geológicos, incluindo processos como o arrasto gravitacional (BORRELLI et al., 2021), como demonstra a Figura 2.

Os processos erosivos de taludes e encostas são controlados por fatores naturais e antrópicos. Dos quais pode-se elencar a erosividade da chuva, a erodibilidade (vulnerabilidade) dos solos, a natureza da cobertura vegetal, as características do

Figura 2. Erosão na superfície de taludes às margens da rodovia 381, Brasil 2022.



Fonte: Autora.

talude e os tipos de uso e ocupação do solo (HUDSON, 1961).

Ressalta-se ainda que o terreno irregular é mais vulnerável à erosão da água. Os respingos, o escoamento superficial e o transporte têm efeitos mais pronunciados em encostas íngremes. Além disso, as características e propriedades do solo influenciam a vulnerabilidade do solo, pois impactam a velocidade de infiltração, a permeabilidade e a capacidade de absorção de água e indicam a dispersão, abrasão e chuva e forças de transporte de inundação (NGUYEN; INDRARATNA, 2020; COOKE; DOORNKAMP et al., 1990).

A erosão do solo, um problema ambiental crescente, se configura como um indicador preciso da instabilidade do solo, acarretando na redução de sua resistência (NOGUEIRA et al., 2020). Diversos fatores contribuem para esse processo, como a diminuição do estresse de sucção da matriz, a presença de discontinuidades, a modificação da estrutura em solos suscetíveis, a liquefação de areia saturada e a perda de coesão (COOKE; DOORNKAMP et al., 1990).

A segurança estrutural está diretamente ligada à estabilidade de taludes, que, por sua vez, é determinada pelo grau de dano sofrido por eles devido à erosão visível (NOGUEIRA et al., 2020). Os taludes como elementos que possuem inclinação são áreas fortemente propensas à erosão em relação à terrenos não inclinados (BENZER, 2010). O que ressalta a vulnerabilidade desse elemento estrutural geotécnico e a sua importância na análise de falhas.

### 2.3.2 Deslizamento

O termo deslizamento é usado para se referir a um modo de falha caracterizado por movimentos de massa, onde há uma zona distinta de fraqueza que separa o material de deslizamento do material subjacente mais estável (VARNES, 1978; VOLKWEIN et al., 2011), como ilustra a Figura 21.

Figura 3. Deslizamentos na superfície de taludes às margens da rodovia 381, Brasil 2022.



Fonte: Autora.

É um fenômeno complexo, que é causado por vários fatores, incluindo geológicos, topográficos, climáticos, humanos e outros, que envolvem a geometria,

qualidade e umidade do solo, precipitação, aspecto, índice de vegetação, construção naquela região, carga na área superficial e distância das estradas e rios, (KHAN et al., 2021; VOLKWEIN et al., 2011; AZIZ et al., 2021).

Os movimentos de massa possuem como principal característica o aumento nas tensões de cisalhamento e ou uma diminuição na resistência ao cisalhamento do material de inclinação. Podem ser classificados por tipo de material e tipo de movimento. Geralmente, o material é rocha ou solo. Os dois tipos principais de movimento são movimento rotatório e translacional. Ou a combinação de dois ou mais tipos de movimento, (VARNES, 1978).

### 2.3.3 Bioerosão e Bioturbação

A biologia desempenha um papel vital na estabilização de estruturas de solo (OADES, 1993). Plantas e animais exploram a superfície terrestre para alimento e abrigo e, no processo, afetam o solo e o seu entorno de muitas maneiras diferentes (GABET; REICHMAN; SEABLOOM, 2003). A ação de uma variedade de elementos da fauna, como formigas, capivaras, minhocas e cupins, como ilustra a Figura 4, de forma continuada e em maiores proporções pode resultar em uma quebra lenta da estabilidade estrutural.

A perda dessa estabilidade estrutural decorre de processos geomórficos de superfície e da alteração das propriedades fundamentais do terreno causadas por esses agentes biológicos, como a distribuição granulométrica, a porosidade, o teor de carbono e outros nutrientes e a taxa de fluxo de fluência (movimento do solo) (WILKINSON; RICHARDS; HUMPHREYS, 2009; OADES, 1993). Essas transformações, de forma individual ou combinada, podem alcançar patamares críticos e induzir a eventual instabilidade e levar a falha da estrutura (WILKINSON; RICHARDS; HUMPHREYS, 2009; OADES, 1993).

Nesse sentido, destaca-se que esse processo biológico pode impactar em uma variedade de escalas espaciais e temporais da estrutura, que se estendem além do solo para a evolução da paisagem. Desse modo, a identificação da ação de agentes biológicos, por meio de técnicas visuais, é capaz, inicialmente, de demonstrar a extensão dos biovazios nos solos (RICHARDS, 2009). De forma que, a presença



Figura 4. Ação biológica de formigueiro em taludes às margens da rodovia 381, Brasil 2022.



Fonte: Autora.

de ninhos, tocas de animais, é cenário sugestivo de processos de biorerosão ou bioturbação do solo.

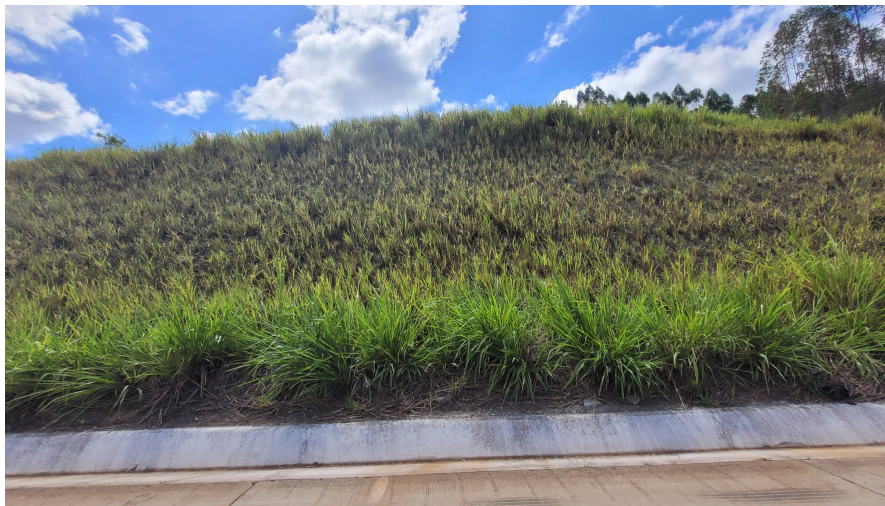
#### 2.3.4 Desvegetação

O desgaste do solo causado pela ação direta da água, vento ou outras forças naturais pode ser amenizado pela proteção vegetal da superfície. Ou seja, a vegetação atua de forma mecânica na proteção do solo contra os processos erosivos, formando uma espécie de cobertura que reduz o impacto direto da chuva e do vento sobre o terreno (LÖBMANN et al., 2020).

Além disso, as raízes das plantas ajudam a aumentar a coesão do solo,

tornando-o mais resistente à erosão e à movimentação. Conseqüentemente, o movimento das massas é restringido, evitando deslizamentos e deslocamentos indesejados (CAZZUFFI; CRIPPA, 2005). Nesse sentido, a vegetação natural ou plantada, como mostrado na Figura 5, é frequentemente considerada em análises de estabilização de estruturas geotécnicas (GREENWOOD; NORRIS; WINT, 2004).

Figura 5. Vegetação típica em taludes às margens da rodovia 381, Brasil 2022.



Fonte: Autora.

No contexto da identificação da erosão, a vegetação desempenha um papel significativo para a ocorrência e indicação das falhas (LÖBMANN et al., 2020). A cobertura vegetal, ao atuar como um elemento protetor do solo, pode revelar áreas suscetíveis à erosão, especialmente em solos com baixa coesão ou em áreas com declive acentuado. A ausência ou o declínio da vegetação em tais locais pode ser interpretado como um sinal de alerta para a possível presença de erosão subjacente.

Além de servir como um indicador de falha, a proteção vegetal também se configura como uma estratégia eficaz na prevenção da degradação do solo, demonstrando alto potencial para evitar o carreamento do solo sob a ação das intempéries (BENZER, 2010).

## Considerações Finais

Os conceitos básicos da mecânica dos solos apresentados neste capítulo são essenciais para a compreensão dos mecanismos de falha em superfícies geotécnicas. A delimitação do escopo de dano, por sua vez, serviu como base para o desenvolvimento do modelo de visão computacional proposto nos capítulos seguintes, que se concentra na detecção automática de anomalias visíveis em imagens digitais de tais superfícies.

## 3 Redes Neurais Convolucionais

### Considerações Iniciais

Neste capítulo, apresentam-se os conceitos fundamentais que embasam os modelos de visão computacional utilizados no desenvolvimento desta tese. Considerando que as principais abordagens de processamento de imagens se apoiam em técnicas de aprendizado de máquina, apresentam-se, primeiramente, os princípios fundamentais nos quais esses algoritmos se baseiam. Aborda-se a concepção das redes neurais, seguida pela introdução do modelo de redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN) e sua caracterização como uma técnica de aprendizado de máquina, do inglês *Machine Learning* (ML) e aprendizado profundo, do inglês *Deep Learning* (DL). Por fim, descrevem-se os principais elementos que compõem a arquitetura de uma CNN na execução de tarefas de classificação de imagens.

### 3.1 Aprendizado de Máquina

Uma gama de soluções que envolvem tarefas como análise de redes sociais, extração de conhecimento de dados biológicos, detecção de fraudes, diagnóstico médico, biometria, sistemas de recomendação, fundamentadas em AM têm atendido satisfatoriamente à que foram propostas. O aprendizado de máquina, do inglês *Machine Learning* (ML), pode ser entendido como o conjunto de algoritmos, de origem interdisciplinar, capazes induzir modelos de forma automatizada, subsidiados por métodos matemáticos, com a finalidade de organizar conjuntos de dados. Os algoritmos de ML podem se enquadrar em categorias (CARVALHO, 2016), conforme o método, em:

- Baseados em distâncias: consiste na proximidade dos dados do conjunto

- Probabilísticos: baseiam-se na relação entre as classes e na probabilidade de ocorrência de determinado evento.
- Baseados em procura: consiste em procurar no espaço de soluções, com auxílio de uma função de avaliação, de hipóteses de representação de um problema.
- Baseados em otimização: formulação baseada na otimização de uma função objetivo (maximização ou minimização).

Cada um desses métodos possui parâmetros próprios e configuráveis, que são inerentes a cada um dos paradigmas e impactam na estrutura dos algoritmos. O processo de ajuste destes parâmetros nos algoritmos de ML é denominado aprendizagem, e representa a memória do fenômeno modelado, adquirida a partir da apresentação de conjuntos de dados, (ARAUJO, 2015). Este ajuste dependerá, fundamentalmente, do tipo de tarefa a ser realizada. Um esquema proposto por (CARVALHO, 2016) subdivide e especifica as características dos tipos de tarefas executáveis por ML, de acordo com a finalidade de aplicação, em preditivas e descritivas. Em tarefas tipo preditivas os algoritmos de ML induzem modelos capazes de prever o valor de uma variável de saída a partir dos valores de variáveis de entrada. Enquanto as tarefas tipo descritivas algoritmos de AM descrevem as principais características de um conjunto de dados (CARVALHO, 2016), como mostra a Figura 6.

Sendo assim, determinar a natureza da tarefa (tipo e objetivo) é fundamental para indicar a abordagem em ML, uma vez que a aprendizagem, ou seja o ajuste de parâmetros do modelo, pode assumir formas distintas conforme o tipo de tarefa. Essa aprendizagem pode ocorrer na forma supervisionada ou não supervisionada. O aprendizado supervisionado ocorre pela minimização do erro entre o valor final apresentado pela rede e o valor esperado, fornecido pelos rótulos das amostras de treinamento. Enquanto que, o aprendizado não supervisionado caracteriza-se pelo mapeamento bem definido entre classes e atributos. Comumente são aplicados a tarefas descritivas e fornecem, para cada entrada, a saída desejada, (CARVALHO, 2016). Nos algoritmos não supervisionados não há classes definidas. As variáveis estruturam-se em nuvens de pontos e uma dada condição de entrada é avaliada como

Figura 6. Descrição das tarefas de aprendizado de máquina por tipo e objetivo.



Fonte: Adaptado de (CARVALHO, 2016).

pertencente, ou não, ao domínio, por meio de métricas em relação à delimitação do contorno da solução.

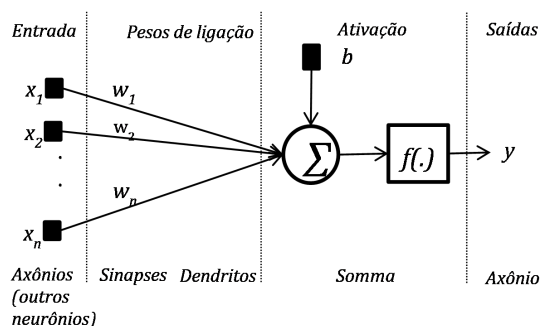
## 3.2 Redes Neurais

As redes neurais artificiais, do inglês *Artificial Neural Network* (ANN), são uma classe de algoritmos de aprendizado de máquina inspirados na estrutura e funcionamento do cérebro humano. O primeiro modelo matemático baseado na estrutura e funcionamento do neurônio humano foi proposto por (MCCULLOCH; PITTS, 1943). O modelo de neurônio artificial era capaz de representar funções booleanas, similar a portas lógicas. No entanto, o modelo apresentou limitações no ajuste de parâmetros (pesos). O que comprometia substancialmente o desempenho dos resultados esperados como saída em problemas de natureza não linear.

Nesse contexto, surgiu o Perceptron, uma evolução proposta por (ROSENBLATT, 1958), que transformou o neurônio artificial em um classificador binário capaz de aprender por meio do ajuste de seus pesos a partir de exemplos de amostras

de treinamento. Logo, uma saída  $y$  pôde ser representada pela soma ponderada das entradas  $x_i$  por pesos  $w_i$ , como argumento de uma função de ativação  $f(\cdot)$ , como mostra a Figura 7. Mesmo diante dos avanços, o Perceptron apresentou-se inapto ao classificar padrões não linearmente separáveis. O que levou ao desinteresse em relação às redes neurais durante a década de 1970.

Figura 7. Modelo de neurônio artificial - Perceptron.



Fonte: (ARAUJO, 2015).

Estudos envolvendo arranjos de neurônios em várias camadas somente foram impulsionados por (RUMELHART; HINTON; WILLIAMS, 1986), após a apresentação do algoritmo de treinamento denominado de retropropagação, conhecido também como *backpropagation*. No cerne do algoritmo reside a utilização de um método supervisionado para ajustar os parâmetros de uma rede neural artificial. Esse ajuste é realizado por meio da minimização de erros entre pares de entrada e saída. Para cada exemplo de treinamento apresentado à entrada, a rede busca aproximar sua saída da saída desejada, também conhecida como etiqueta. Adotando esse método de treinamento, redes neurais artificiais ANN compostas por diversas camadas de neurônios, chamadas de redes multicamadas perceptron redes neurais artificiais multicamadas, do inglês *multi-layer Perceptron* (MLP), puderam realizar a classificação de conjuntos não linearmente separáveis.

Embora efetivas, o desempenho das MLP para execução de tarefas com dados complexos e não estruturados não é trivial. Nesse contexto, emergiram estudos sobre redes neurais em diversas configurações, incluindo os modelos de aprendizado profundo cuja arquitetura fundamenta-se na ampliação no número de camadas entre entrada e saída da rede. Na seção seguinte, discorre-se sobre um dos

modelos mais populares de aprendizado profundo, as redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN).

### 3.3 Redes Neurais Convolucionais em Tarefas de Classificação de Imagens

O aprendizado profundo, do inglês *Deep Learning* (DL), são modelos computacionais que tiveram sua evolução a partir das redes neurais artificiais. No rol de modelos de aprendizado profundo, as redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN), uma categoria de redes neurais artificiais, tornaram-se populares pela sua robustez diante de uma gama de problemas. As CNNs diferenciam-se das ANN clássicas não somente pela estrutura em múltiplas camadas, mas também pela capacidade de convolução das classes pelas camadas intermediárias, (GARBIN; ZHU; MARQUES, 2020).

As redes neurais convolucionais surgiram inspiradas no funcionamento das células nervosas do córtex visual animal, (HUBEL; WIESEL, 1962). Nesse tipo de rede, inserem-se camadas de convolução constituídas de filtros, capazes de extrair padrões a partir dos dados apresentados na entrada, similar ao modo como ocorre o processamento visual nos mamíferos. Assim, a CNN é estruturada em um arranjo sequencial de camadas combinadas que resultam como saída mapas de atributos que representam características básicas da entrada, como orientação e bordas (BASHA et al., 2020). Dessa forma, o aprendizado ocorre de forma hierárquica, à medida que são extraídas características cada vez mais complexas e abstratas em relação à camada anterior. O que torna modelos de Redes Neurais Convolucionais com grande potencial de generalização, além de classificá-los como aprendizado de máquina e aprendizado profundo (SANTOS et al., 2020; MHASKAR; LIAO; POGGIO, 2016; LECUN; BENGIO; HINTON, 2015).

A dinâmica de ajuste automático dos parâmetros dos filtros pela CNN tem se mostrado atrativa em atividades que envolvem a interpretação e o processamento de imagens por meio de algoritmos, como ocorre no campo da Visão Computacional (VC). Diante de uma imagem, torna-se possível determinar a probabilidade de



representação de um objeto de interesse, o que viabiliza a classificação desses objetos em uma ou mais classes. A classificação de imagens é considerada um problema clássico na área da visão computacional. Nesse tipo de tarefa, a rede convolucional fornece como saída as coordenadas espaciais que definem a sub-região e um vetor com as probabilidades de classificação para cada objeto (ZHAO et al., 2019).

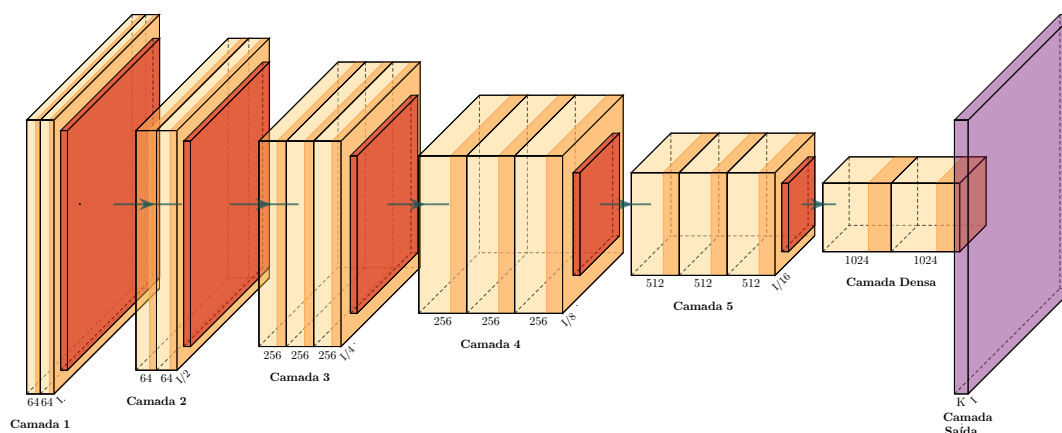
Por essa razão, em alguns casos as CNNs são consideradas mais vantajosas em comparação com outros métodos de aprendizado de máquina, uma vez que projetar manualmente filtros eficazes para extrair características relevantes dos dados é uma tarefa complexa (AFFONSO et al., 2017; MYAGILA; KILAVO, 2022; CHAUDHARI, 2018). No contexto do processamento de imagens, o uso de múltiplas camadas convolucionais contribui para evitar a degradação do desempenho em tarefas de detecção (AKHTAR; RAGAVENDRAN, 2020). Além disso, a operação de convolução em CNNs reduz a sensibilidade do modelo a distorções e transformações de imagem, o que contribui para o bom desempenho desse tipo de rede em tarefas de reconhecimento de imagem.

Ao contrário das redes neurais rasas, que possuem apenas uma ou duas camadas intermediárias, as redes neurais profundas apresentam diversas camadas empilhadas umas sobre as outras (MHASKAR; LIAO; POGGIO, 2016) entre a camada de entrada e a camada de saída. Essa profundidade arquitetônica permite que a rede aprenda com níveis crescentes de abstração, capaz de extrair informações cada vez mais complexas dos dados (LECUN; BENGIO; HINTON, 2015). E apresentando eficácia na solução de diversos problemas, com destaque para as tarefas de classificação de imagens.

A estrutura típica de uma rede convolucional é composta por camadas de convolução, camadas de agrupamento (*pooling*) e camadas densas ou densa ou totalmente conectada, do inglês *Fully Connected* (FC), como ilustra a Figura 8, (ABEDALLA et al., 2021).

Nas próximas seções, são abordados os conceitos e a teoria que fundamenta a constituição das camadas presentes em uma arquitetura CNN, com foco no caso particular dos elementos típicos de classificadores de imagens. Além disso, apresenta-se as funções de ativação conhecidas, como *Rectified Linear Unit* (ReLU),

Figura 8. Estrutura típica da arquitetura de uma CNN.



Fonte: Adaptado de (IQBAL, 2018).

sigmoide e Softmax, que são aplicadas às camadas para possibilitar a aprendizagem de representações mais complexas dos dados. Também se aborda técnicas para regularização, como abandono, que é comumente utilizada para prevenir o sobreajuste e melhorar a generalização do modelo. Por fim, conclui-se com a descrição do otimizador *Adaptive Moment Estimation* (Adam), frequentemente empregado no ajuste de pesos da rede durante o treinamento.

### 3.3.1 Convolução

A camada de convolução é responsável por convolver regiões (*patches*) da imagem com objetivo de extrair características universais e cada vez mais complexas. Para isso, é realizada uma operação de convolução entre filtros inicializados aleatoriamente e os dados apresentados como entrada da camada. Essa operação é realizada em toda a imagem deslizando os filtros e encontrando o produto ponto a ponto elemento a elemento entre o filtro e as partes da imagem de entrada.

Considerando uma imagem de entrada  $I$  com dimensões de altura  $H$ , largura  $W$  e  $C$  canais de cor. Cada *pixel* da imagem é representado por uma matriz onde cada dimensão corresponde a um canal de cor que varia conforme o modelo de cor

da imagem. Um dos modelos de cor mais comum em imagens digitais é o RGB, vermelho, verde, e azul, do inglês *Red, Green, Blue* (RGB). No modelo de cor RGB há três canais ( $C = 3$ ) e cada canal armazena um valor de intensidade. A combinação do valor presente em cada canal representa a cor do *pixel*. Ou seja, os canais de cor referem-se à quantidade de dados separados usados para representar a cor em cada *pixel*. No processamento de imagem também são utilizados diferentes modelos de cor, como os monocromáticos em escala de cinza  $C = 1$ , modelos multiespectrais que possuem múltiplos canais além do espectro visível como o infravermelho, ultravioleta, e outros espectros.

Juntamente com essa entrada  $I$ , há um conjunto de  $K$  filtros de convolução, cada um com altura e largura ( $h \times w$ ) (também conhecido como tamanho do filtro ou núcleo (do inglês *kernel*). O filtro  $K$  desliza pela imagem de entrada  $I$ , visitando diferentes regiões da imagem. Para cada posição do filtro, os valores dos *pixels* da imagem na região de sobreposição são alinhados com os pesos correspondentes do filtro na mesma posição. Cada valor de pixel da imagem é multiplicado pelo peso do filtro correspondente. A operação de convolução em uma única camada da CNN pode ser definida na equação 3.1.

$$Z(x, y) = \sum_{c=1}^C \sum_{i=-(Fw-1)}^{Fw-1} \sum_{j=-(Fh-1)}^{Fh-1} I(x+i, y+j, c) \cdot K(i, j, c) \quad (3.1)$$

Onde  $Z(x, y)$  é a saída da operação de convolução em uma localização específica  $(x, y)$  no mapa de características de saída;  $c$  é o índice de canal, representando um dos  $C$  canais de cor na imagem de entrada e o canal correspondente no filtro;  $i$  é o deslocamento horizontal do filtro em relação à localização atual  $(x, y)$  na imagem de entrada, e varia de  $-\left(\frac{Fw-1}{2}\right)$  a  $\left(\frac{Fw-1}{2}\right)$ , onde  $Fw$  é a largura do filtro;  $j$  é o deslocamento vertical do filtro em relação à localização atual  $(x, y)$  na imagem de entrada, e varia de  $-\left(\frac{Fh-1}{2}\right)$  a  $\left(\frac{Fh-1}{2}\right)$ , onde  $Fh$  é a altura do filtro.  $I(x+i, y+j, c)$  é o valor do pixel de entrada na posição  $(x+i, y+j)$  no canal  $c$ ;  $K(i, j, c)$  é o valor do filtro na posição  $(i, j)$  no canal  $c$ ; e  $b_i$  é o termo de viés associado ao filtro  $c$ .

O somatório  $Z(x, y)$  pondera a importância de cada pixel para a ativação do filtro naquele canal específico, levando em consideração a posição relativa do pixel

e do peso. Os resultados das multiplicações entre os *pixels* da imagem e os pesos do filtro na mesma posição são somados. O processo de multiplicação, ponderação e soma é repetido para todas as posições do filtro na imagem e para todos os canais do filtro.

A quantidade e o tamanho dos filtros são hiperparâmetros que podem ser configurados durante o treinamento da CNN para otimizar o desempenho na tarefa específica. A Figura 9 ilustra 32 filtros de tamanho  $3 \times 3$  ( $h = 3 \times w = 3$ ) em escala de cinza  $C = 1$ . Ao iniciar a rede, os valores presentes em cada um dos canais de cor são inicializados aleatoriamente e tem os seus valores ajustados à medida que a rede é exposta à exemplos.

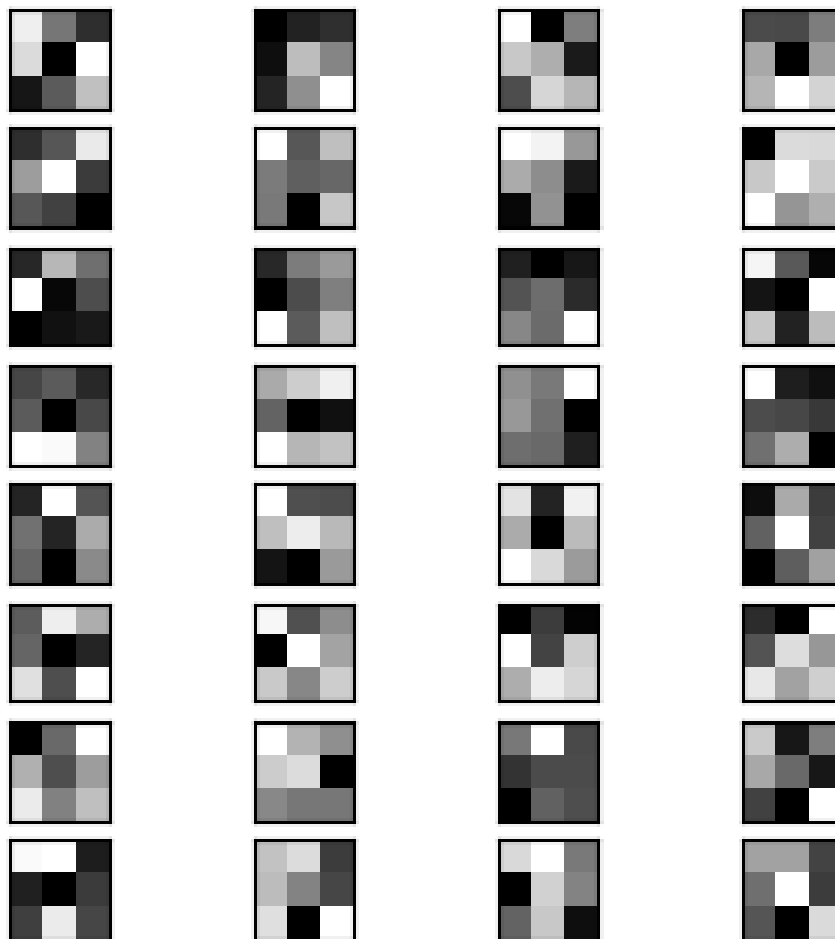
Quando  $Z(x, y)$  é utilizada como argumento na função  $f$ , produz-se o mapa de características  $A(x, y)$ , que representa a saída da camada convolucional no canal. Essencialmente, a aplicação da função de ativação  $f$  sobre a ativação do filtro  $Z(x, y)$  resulta no mapa de características conforme a equação 3.2

$$A(x, y) = f(Z(x, y)) \quad (3.2)$$

O mapa de características, ilustrado pela Figura 10, exemplifica como o resultado da ação dos filtros na camada de convolução. Esses mapas sintetizam características como cantos, bordas ou características da imagem e é lido por outras camadas para que possam aprender as características restantes da imagem (SZEGEDY et al., 2015).

A convolução aplicada aos *pixels* da imagem permite automaticamente a identificação e análise de informações visuais complexas. Quando aplicada em pelo menos uma camada da rede possibilita a extração automática de características locais nos dados (AFFONSO et al., 2017). Diferentemente, de SVM ou RF, entre outras técnicas clássicas de processamento de imagens, que geralmente exigem métodos manuais ou semi automatizados para extrair características como implementação de filtros de borda, filtros de textura ou histograma (MYAGILA; KILAVO, 2022; CHAUDHARI, 2018).

Figura 9. Representação de 32 filtros da camada de convolução, em escala de cinza, núcleo tamanho 3x3.



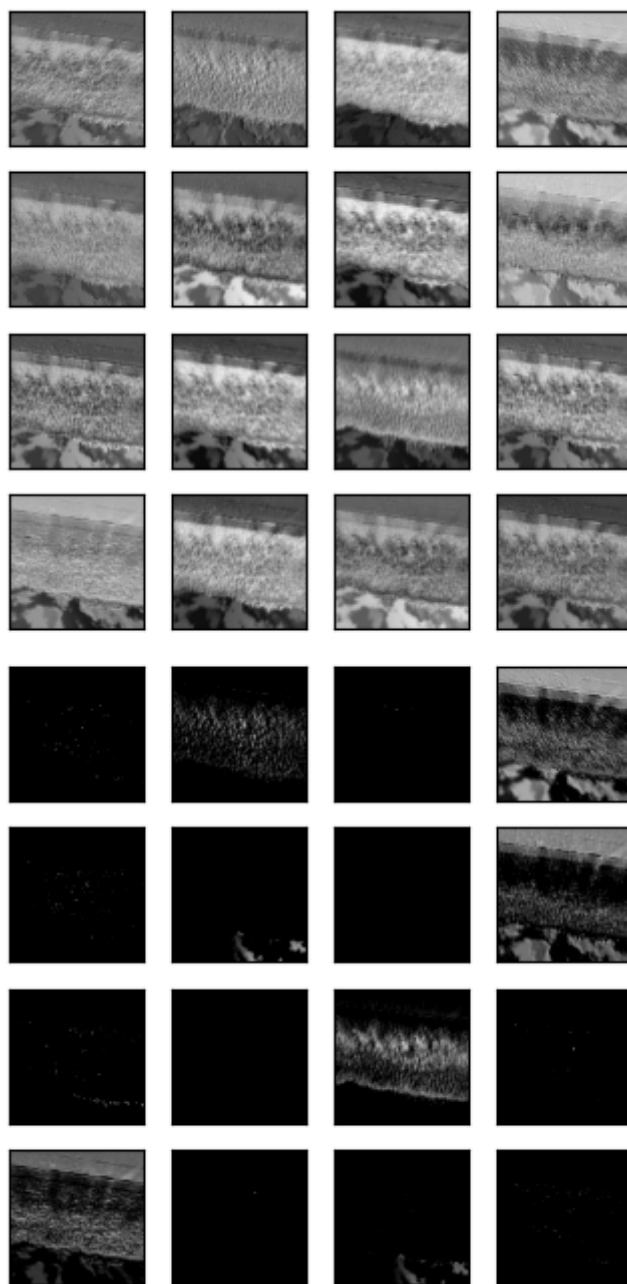
Fonte: Autora.

### 3.3.2 Agrupamento

O agrupamento é outro componente crucial de uma arquitetura CNN. A operação de agrupamento (do inglês *pooling*) é responsável por redimensionar os mapas de características resultantes da convolução na camada anterior. O processo de agrupamento elementar envolve a amostragem do mapa de características perfurando em sub-regiões, (AKHTAR; RAGAVENDRAN, 2020).

Existem diferentes critérios para o processo de redução da dimensionalidade.

Figura 10. Ação de 32 filtros na camada de convolução.



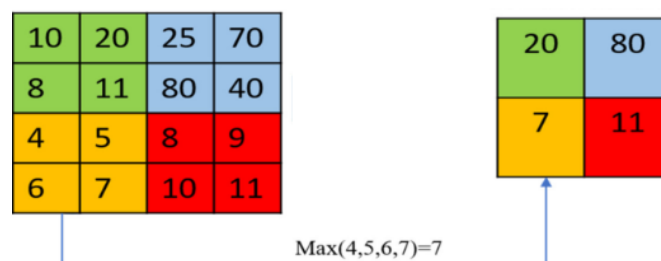
Fonte: Autora.

Uma desses critérios, frequentemente utilizado para processamento de representa-

ções dependentes de estatísticas de contagem, é agrupamento pelo valor máximo do *pixel*, ou '*max pooling*' (LECUN; BENGIO; HINTON, 2015; ZAFAR et al., 2022; MURRAY; PERRONNIN, 2014; SCHERER; MÜLLER; BEHNKE, 2010).

Este tipo de operação envolve deslizar uma janela, de tamanho  $k \times k$ , a determinado passo (do inglês, *stride*) ao longo do mapa de representações. A redução da dimensionalidade por meio do agrupamento máximo consiste em reconstruir uma nova matriz a partir do valor máximo obtido em cada janela. Esse processo é ilustrado pela Figura 11.

Figura 11. Agrupamento máximo.



Fonte: Adaptado de (ZAFAR et al., 2022).

Algebricamente, a técnica pode ser descrita por criar invariância de posição sobre regiões locais maiores (ZAFAR et al., 2022) e reduzir a dimensão do mapa de características  $A$  por um fator de  $P$  ao longo de cada eixo, como descrito pela equação (3.3), gerando um novo mapa.

$$\text{Max\_Pooling}(A, x, y) = \max_{0 \leq p, q < P} A[x \cdot P + p, y \cdot P + q] \quad (3.3)$$

Onde  $A$  representa o mapa de características de entrada,  $x$  e  $y$  indicam as coordenadas espaciais dentro do mapa de características. A janela de agrupamento, com um tamanho de  $P \times P$ , é movida ao longo do mapa de características. Para cada posição  $(x, y)$ , a operação de agrupamento máximo seleciona o valor máximo dentro da janela do mapa de características  $A$ . Este processo reduz as dimensões do mapa de características de entrada por um fator de  $P$  em cada eixo, resultando no mapa de agrupamento de saída. Neste contexto, o passo  $S$  (*stride*) é definido igual ao tamanho da janela de agrupamento  $P$  para garantir que não haja sobreposição entre

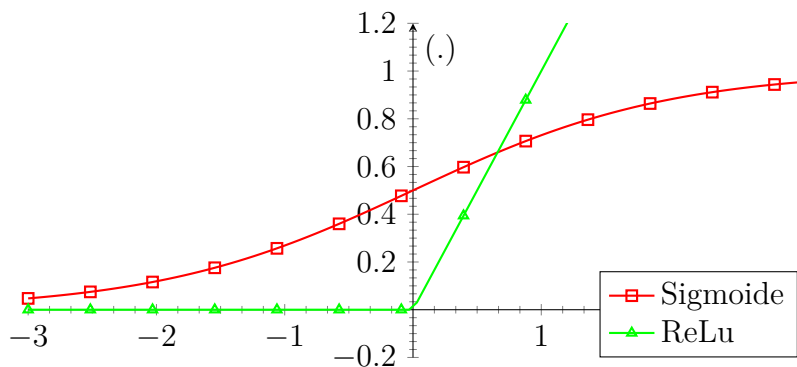
as janelas adjacentes. Conseqüentemente, a cada passo, a janela de agrupamento se move  $P$  unidades horizontalmente e  $P$  unidades verticalmente.

O agrupamento pode conduzir a uma taxa de convergência mais rápida ao reduzir os mapas de características. A seleção de características invariantes superiores também que melhora o desempenho de generalização (NAGI et al., 2011; MURRAY; PERRONNIN, 2014). Nesse contexto, o agrupamento desempenha um papel vital nas tarefas de reconhecimento visual, detecção e segmentação para superar as preocupações como sobreajuste, tempo de computação e precisão de reconhecimento (AKHTAR; RAGAVENDRAN, 2020; ZAFAR et al., 2022).

### 3.3.3 Função de Ativação

Outro hiperparâmetro que impacta no desempenho da rede é a função de ativação. As funções de ativação  $f(\cdot)$ , como ilustradas na Figure 12 são comumente usadas após cada camada de convolução para inserir um grau de não linearidade na saída do neurônio (LECUN; BENGIO; HINTON, 2015), uma vez que os dados da imagem não são linearmente separáveis (WANG et al., 2020).

Figura 12. Funções de ativação comuns em CNN.



Fonte: Autora.

Para classificar os dados de entrada, é necessário ter uma camada para prever classes, que geralmente está localizada na última camada da arquitetura da CNN (CHA; CHOI; BüyüKöZTürk, 2017; TAN; GUO; POH, 2021). Uma das funções de ativação que se destaca nesse tipo de tarefa é a Softmax. Trata-se



de uma função de normalização utilizada para converter um vetor de números reais em um vetor de probabilidades que somam 1. Isto é, a função garante que a soma das probabilidades de todas as classes seja igual a 1, o que é essencial para interpretar a saída como uma distribuição de probabilidade sobre todas as classes possíveis. É frequentemente usada em problemas de classificação multiclasse, como reconhecimento de padrões em imagens (CHA; CHOI; BüyüKöZTürk, 2017; FOROUGHI; CHEN; WANG, 2021; SPENCER; HOSKERE; NARAZAKI, 2019).

Para calcular a saída da função Softmax para o  $i$ -ésimo elemento de um vetor  $z$  com  $n$  elementos, tem-se em 3.4

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (3.4)$$

Uma particularização da Softmax frequentemente empregada em modelos de classificação é a função Sigmoid. A função sigmoide é uma função mapeia os valores de entrada para um intervalo entre 0 e 1 e comumente usada na camada de saída de modelos de Redes Neurais Convolucionais para tarefas de classificação binária (LI et al., 2022). A função sigmoide é definida como 3.5,

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (3.5)$$

No entanto, é importante observar que a função Softmax e Sigmoid podem apresentar o problema de saturação dos gradientes (WANG et al., 2020), especialmente em regiões onde os valores de entrada estão longe de zero. Na retropropagação a derivada pode se tornar muito pequena para determinados valores de entrada e conduzir o treinamento ao desaparecimento do gradiente (LECUN; BENGIO; HINTON, 2015; THEODORIDIS, 2020). Em consequência, as atualizações nos pesos e bias nas camadas anteriores da rede se tornam muito pequenas ou até mesmo zero. Isso pode conduzir o treinamento a problemas de convergência ou estagnação do aprendizado (CHA; CHOI; BüyüKöZTürk, 2017).

Em muitos casos, a função de ativação Sigmoid e a Softmax são substituídas pela função *Rectified Linear Unit* (ReLU) na maioria das camadas da CNN. Na função ReLU o valor de entrada para a função for positivo, retorna o mesmo valor de

entrada, caso contrário, retorna zero. Logo, a derivada de ReLu é 1 (para entradas positivas) ou 0 (para entradas negativas). Matematicamente, a ReLu pode ser expressa como,

$$\text{ReLu}(z) = \begin{cases} 0, & \text{se } z < 0 \\ z, & \text{se } z \geq 0 \end{cases} \quad (3.6)$$

Apesar da seção linear na parte inicial da função, como mostra a Figura 12, a ReLu é considerada uma função não linear. O que possibilita que Redes Neurais Convolucionais aprendam relações complexas nos dados que não podem ser capturadas por modelos lineares simples. Essa característica, aliada à sua derivada suave, evita o problema do desaparecimento do gradiente durante a retropropagação (THEODORIDIS, 2020; CHA; CHOI; BüyüKöZTürk, 2017) pois é insaturada (WANG et al., 2020).

As especificidades da ReLu a tornam uma função extremamente popular para uso nas camadas intermediárias de modelos CNN (LECUN; BENGIO; HINTON, 2015; ZAFAR et al., 2022; ADNAN et al., 2022; MHASKAR; LIAO; POGGIO, 2016) inclusive em aplicações na análise das falhas em estruturas (COFRE-MARTEL et al., 2019; NOGUEIRA et al., 2020; CHA; CHOI; BüyüKöZTürk, 2017). Além disso, a ReLu normalmente aprende muito mais rápido em redes com muitas camadas, permitindo o treinamento de uma rede supervisionada profunda sem pré-treinamento (LECUN; BENGIO; HINTON, 2015).

Após a ativação ao final de cada bloco de convolução, finalmente os mapas de características são transformados em vetores para servirem de entrada para a camada densa (LI et al., 2018).

### 3.3.4 Linearização

A saída das camadas de convolução e agrupamento produzem vetores multidimensionais que requerem um processo de transformação para alimentar as camadas totalmente conectadas. Esse processo de linearização, do inglês *flatten*, converte os mapas de características multidimensionais das camadas de convolução

anteriores  $M$  em um vetor unidimensional  $I$ . A operação de linearização pode ser formalizada pela equação 3.7.

$$\text{Flatten}(\mathbf{h}_{\text{drop}}) = \mathbf{X}_{\text{flat}}, \quad (3.7)$$

Onde  $\mathbf{h}_{\text{drop}}$  são os mapas de características de entrada provenientes das camadas convolucionais da imagem de entrada  $I$ , e  $\mathbf{X}_{\text{flat}}$  é o vetor de saída linearizado.

A camada de linearização restaura os mapas de características de entrada provenientes das camadas de convolução de uma imagem em um vetor linear, facilitando assim o processamento adicional nas camadas totalmente conectadas.

### 3.3.5 Conectividade Total

Após a linearização, o vetor unidimensional  $X$  resultante é dado como entrada para a primeira camada totalmente conectada, ou densa, FC. Na arquitetura de Redes Neurais Convolucionais, as camadas densa ou totalmente conectada, do inglês *Fully Connected* (FC), também conhecidas como camada densa, assumem um papel fundamental no processo de aprendizado e classificação. Essa camada opera por meio da multiplicação de matrizes da sua entrada com uma matriz de peso, seguida da aplicação de uma função de ativação, como descrito por (BASHA et al., 2020).

O vetor  $X$  é dado como entrada para a primeira camada densa,  $FC_1$ . Após a primeira camada densa, os neurônios nas camadas subsequentes ( $FC_i$ , onde  $i > 1$ ) recebem como entrada os valores gerados pela camada de neurônios anterior (BASHA et al., 2020). Após a primeira camada densa, os neurônios nas camadas subsequentes ( $FC_i$ , onde  $i > 1$ ) recebem como entrada os valores gerados pela camada de neurônios anterior (BASHA et al., 2020). Em outras palavras, os neurônios nas camadas  $FC_i$  subsequentes utilizam como entrada as saídas dos neurônios da camada  $FC_{i-1}$ . Cada neurônio na camada  $FC_i$  recebe como entrada um conjunto de valores provenientes das saídas dos neurônios da camada anterior. Esses valores são então processados pelos neurônios na camada  $FC_i$  para produzir suas próprias saídas. Desta forma, a conectividade total é mantida nas camadas

$F$ , e as conexões têm pesos  $a$  (parâmetros)  $\in \mathcal{R}^d$ , o que significa que os pesos pertencem ao espaço de números reais  $d$ -dimensional e há um total de  $d$  pesos para cada conexão.

Um único neurônio de uma camada totalmente conectada pode ser expresso pela equação 3.8,

$$Y_i = f \left( \sum_{j=1}^N a_{ij} \cdot X_j + b_i \right), \quad (3.8)$$

Onde  $Y_i$  representa a saída do neurônio  $i$  na camada totalmente conectada.  $f$  é a função de ativação.  $a_{ij}$  é o peso que liga o neurônio  $i$  à entrada  $j$ .  $X_j$  representa as características que, se a camada for a primeira camada densa, são as entradas do modelo; caso contrário,  $X_j$  é a saída da camada FC anterior.  $N$  é o número total de características processadas, e  $b_i$  é o viés do neurônio  $i$ .

A última camada densa é também a saída da rede. Na camada de saída  $i$  representa o número de classes do modelo. Isto é, à quantidade de variáveis de saída do problema. Logo, em um classificador multiclasse, cada neurônio  $i$  representa uma classe. Para os classificadores binários há somente um ou dois neurônios na saída, uma vez que somente duas classes possíveis de saída, uma classe positiva e uma classe negativa.

### 3.4 Desafios em Redes Neurais Convolucionais

Embora as Redes Neurais Convolucionais ofereçam inúmeras vantagens nas tarefas de visão computacional, elas também apresentam desafios significativos. Um desses desafios proeminentes é o fenômeno do sobreajuste, frequentemente observado em arquiteturas CNN (GARBIN; ZHU; MARQUES, 2020; SHORTEN; KHOSHGOFTAAR, 2019; CHA; CHOI; BÜYÜKÖZTÜRK, 2017). O sobreajuste ocorre quando o modelo se ajusta excessivamente aos dados de treinamento, resultando em dificuldades para generalizar para novos dados. Isto é, a rede categoriza com êxito um conjunto de dados de treinamento, mas não consegue generalizar, levando a baixa validação e desempenho de dados de teste.

Dentre as causas do sobreajuste, destaca-se a complexidade dos modelos. CNNs são arquiteturas de dimensão escalável em termos de camadas, canais e resolução (ABEDALLA et al., 2021). O grande número de parâmetros, à medida que a rede se escala, permite que a rede aprenda características complexas. No entanto, em alguns casos, pode haver dificuldades na generalização, e conseqüente sobreajuste do modelo, em decorrência do excesso de parâmetros de treinamento (XU et al., 2019; GARBIN; ZHU; MARQUES, 2020). Devido a esse excesso, a propensão ao sobreajuste agrava-se diante de conjuntos de dados menores (FOROUGHI; CHEN; WANG, 2021; BRIGATO; IOCCHI, 2021).

Em relação ao tamanho de dados, outra problemática relevante no contexto dos modelos CNN é o desequilíbrio classes em tarefas de classificação (BARELLA; GARCIA; CARVALHO, 2019). Esse desequilíbrio ocorre quando a quantidade de exemplos disponíveis para cada classe varia consideravelmente, geralmente com uma classe majoritária e diversas classes minoritárias. Nesse cenário, as CNNs tendem a favorecer a classe majoritária durante o treinamento, ignorando características importantes das classes minoritárias. Isso resulta em um modelo com baixo desempenho na classificação das classes menos representadas, mesmo que apresente um bom desempenho geral.

Em relação ao número de parâmetros e a complexidade do modelo, arquiteturas muito profundas podem tornar a CNN impraticável, especialmente em cenários de recursos computacionais limitados (BRIGATO; IOCCHI, 2021; DARAPANENI; KRISHNAMURTHY; PADURI, 2020). Em teoria, aumentar o número de camadas (camadas mais profundas), ou aumentar os canais e a resolução da imagem, deve melhorar o desempenho da rede. Por outro lado, em algumas configurações, o estado da arte pode ser melhorado usando modelos de baixa complexidade, com menos camadas, canais ou resolução (BRIGATO; IOCCHI, 2021; DARAPANENI; KRISHNAMURTHY; PADURI, 2020).

Outro ponto importante é sobre o ajuste adequado dos hiperparâmetros (DARAPANENI; KRISHNAMURTHY; PADURI, 2020; OTTONI; AMORIM; NOVO, 2022), que pode ser uma tarefa complexa e exigir experiência ou métodos de otimização automatizados. Apesar do sucesso dos modelos de aprendizagem profunda, a compreensão teórica sobre Redes Neurais Convolucionais permanece limitada

(BASHA et al., 2020). O que dificulta a configuração e seleção de valores adequados dos hiperparâmetros. Nesse contexto, a definição das arquiteturas CNN ainda é um processo empírico e não existem regras precisas para a otimização de hiperparâmetros, embora haja arquiteturas de referência recomendadas para determinadas aplicações.

Além disso, o desaparecimento do gradiente emerge como um desafio no treinamento de arquiteturas CNN. Esse fenômeno se caracteriza pela diminuição gradual ou pelo aumento descontrolado dos valores dos gradientes durante o treinamento (ADNAN et al., 2022; ABEDALLA et al., 2021; ZHANG, 2019). O problema está relacionado à propagação do gradiente através da rede durante o processo de retropropagação. Esses gradientes podem diminuir significativamente à medida que são propagados para trás através das várias camadas da rede. Isso ocorre porque a multiplicação repetida de gradientes menores que 1 pode fazer com que eles se tornem muito pequenos, resultando em uma contribuição insignificante para a atualização dos pesos das camadas mais antigas da rede. Como resultado, as camadas iniciais da rede podem não receber gradientes significativos o suficiente para aprender efetivamente os padrões nos dados de entrada.

Diversos fatores podem contribuir para o desaparecimento do gradiente em CNNs, como o uso de funções de ativação saturáveis, a presença de camadas profundas e a ausência de técnicas de normalização adequadas. As consequências do desaparecimento do gradiente podem incluir a estagnação do treinamento, a instabilidade do modelo e a obtenção de um desempenho insatisfatório. O que pode ser percebido pelo treinamento mais lento ou mesmo a uma estagnação do processo de aprendizado (THEODORIDIS, 2020). Na estagnação do aprendizado as camadas iniciais da rede não conseguem aprender padrões úteis nos dados devido à falta de informações relevantes transmitidas pelos gradientes.

Para lidar com o desaparecimento do gradiente em CNNs para classificação de imagens, várias técnicas podem ser empregadas, como o uso de inicialização de pesos adequada, arquiteturas consideradas mais eficientes para conter o desaparecimento do gradiente, como as *Residual Networks* (ResNet), e o uso de funções de ativação recomendadas para determinadas tarefas. Essas técnicas, denominadas técnicas de regularização, são empregadas com sucesso para mitigar os obstáculos

enfrentados pelos modelos CNN.

### 3.5 Regularização

Na Seção 3.4, foram apresentados diversos desafios comumente encontrados em arquiteturas de Redes Neurais Convolucionais. Entre eles, destacam-se a generalização, o sobreajuste, o decaimento do gradiente, a estagnação do aprendizado e as limitações computacionais. As causas desses desafios são diversas e podem estar relacionadas às características do conjunto de dados utilizado, à complexidade da arquitetura da rede ou à necessidade de técnicas de regularização adequadas.

No âmbito do aprendizado de máquina e do aprendizado profundo, a regularização desponta como uma estratégia para combater os fatores que comprometem o desempenho dos modelos. Essa estratégia se caracteriza por técnicas suplementares que visam aprimorar a generalização dos modelos, capacitando-os a apresentar um bom desempenho em dados não vistos durante o treinamento (KUKAČKA; GOLKOV; CREMERS, 2017; NAGPAL et al., 2015).

No contexto específico das CNNs, diversas estratégias de regularização se concentram na arquitetura do modelo em si. O que originou arquiteturas de referência progressivamente mais complexas, como AlexNet, VGG, ResNet, Inception e DenseNet (SHORTEN; KHOSHGOFTAAR, 2019).

No entanto, outras soluções são consideradas funcionais para melhorar a capacidade de generalização dos modelos CNN. Dentre essas soluções para a regularização, o estado da arte destaca o abandono (GARBIN; ZHU; MARQUES, 2020; THEODORIDIS, 2020; ADNAN et al., 2022), o aumento de dados (ABEDALLA et al., 2021; SHORTEN; KHOSHGOFTAAR, 2019; OTTONI; AMORIM; NOVO, 2022; MIKOŁAJCZYK; GROCHOWSKI, 2018), e a otimização adaptativa (ZHANG, 2019; GARBIN; ZHU; MARQUES, 2020; ZHANG; LI; LI, 2021), como promissoras para mitigar o sobreajuste e a estagnação do aprendizado. Nas sessões seguintes, discorre-se sobre a teoria que abarca essas soluções.

### 3.5.1 Abandono

Em Redes Neurais Convolucionais típicas, a extração de recursos globais e ruidosos é um processo inevitável, mas nem todos esses recursos se revelam úteis (XU et al., 2019). Para amenizar essa extração indesejada, uma estratégia eficaz consiste na inserção de uma camada de abandono, também conhecida como *dropout*, após a etapa de agrupamento.

As camadas de abandono são implementadas após o processo de convolução, com o objetivo principal de superar o problema do sobreajuste em modelos de aprendizado profundo (CANO et al., 2021; GARBIN; ZHU; MARQUES, 2020; KUKAČKA; GOLKOV; CREMERS, 2017). O princípio fundamental do abandono reside na interrupção aleatória de conexões entre neurônios em camadas interconectadas, como detalhado em (GARBIN; ZHU; MARQUES, 2020; XU et al., 2019).

O abandono atua de forma estratégica, quebrando aleatoriamente as conexões entre neurônios em camadas interconectadas com uma taxa predeterminada. Essa ruptura aleatória das conexões visa minimizar as adaptações entre os neurônios, permitindo que a rede generalize amostras de treinamento de forma mais eficaz, conforme evidenciado em (ZAFAR et al., 2022).

Durante o treinamento em uma camada de CNN, a saída após a aplicação do abandono ( $h_{\text{drop}}$ ) é obtida pela multiplicação elemento a elemento com uma máscara  $r(x, y)$ . Esta técnica aplica a máscara de abandono ao resultado da operação de agrupamento, efetivamente desativando um subconjunto aleatório de elementos e reduzindo suas ativações, conforme representado pela equação (3.9).

$$h_{\text{drop}}(x, y) = r(x, y) \cdot (\text{Max\_Pooling}(A, x, y)) \quad (3.9)$$

A máscara de abandono é uma matriz do mesmo tamanho que o mapa de características de saída, onde cada elemento  $r(x, y)$  representa um valor amostrado aleatoriamente (0 ou 1) retirado de uma distribuição de Bernoulli. A probabilidade de cada elemento ser 1 é indicada por  $d$  (taxa de abandono), que controla a proporção geral de neurônios a serem desativados durante o treinamento.



$$h_{\text{drop}}(x, y) = \begin{cases} r(x, y) \cdot (\text{Max\_Pooling}(A, x, y)) & \text{com probabilidade } 1 - d \\ 0 & \text{com probabilidade } d \end{cases} \quad (3.10)$$

### 3.5.2 Aumento de Dados

No âmbito do aprendizado de máquina, o pré-processamento de dados se configura como uma etapa fundamental para o sucesso do processo de aprendizado. Para realizar o pré-processamento de dados de forma eficaz e mitigar problemas relacionados à qualidade e ao volume dos dados de treinamento, o aumento de dados surge como uma ferramenta para promover o desempenho dos modelos, conforme evidenciado em (OTTONI; AMORIM; NOVO, 2022; EKUNDAYO; VIRIRI, 2021; CANO et al., 2021; BUSHARA; KUMAR, 2022).

O aumento de dados, do inglês *Data Augmentation* (DA) atua na raiz do problema do sobreajuste, o conjunto de dados de treinamento, abordando-o de forma direta (SHORTEN; KHOSHGOFTAAR, 2019). Como um pré-processamento, pode ser uma estratégia exitosa para promover desempenho dos modelos e impulsionar os resultados precisos, por meio do conjunto de dados (OTTONI; AMORIM; NOVO, 2022; BUSHARA; KUMAR, 2022).

O DA consiste em realizar transformações randomizadas nas imagens, seja com o intuito de aumentar o número de amostras (ABEDALLA et al., 2021; FOROUGH; CHEN; WANG, 2021; BUSHARA; KUMAR, 2022) ou melhorar a qualidade e robustez dessas amostras (SHORTEN; KHOSHGOFTAAR, 2019; MIKOŁAJCZYK; GROCHOWSKI, 2018; ZAFAR et al., 2022). As modificações são rápidas e fáceis de implementar e provaram ser um bom método para aumentar o conjunto de dados de treinamento (MIKOŁAJCZYK; GROCHOWSKI, 2018).

Uma das principais estratégias para implementar o DA em modelos de aprendizado de máquina reside nas transformações geométricas. Essas transformações consistem em modificações na imagem original, realizadas por meio de diversas técnicas, que consistem no movimento em relação ao eixo da imagem, mudanças na forma dos elementos e alterações no tamanho e posição dentro da imagem

(OTTONI; AMORIM; NOVO, 2022; CANO et al., 2021; MIKOŁAJCZYK; GROCHOWSKI, 2018; BUSHARA; KUMAR, 2022). São transformações geométricas frequentemente empregadas no aumento de dados:

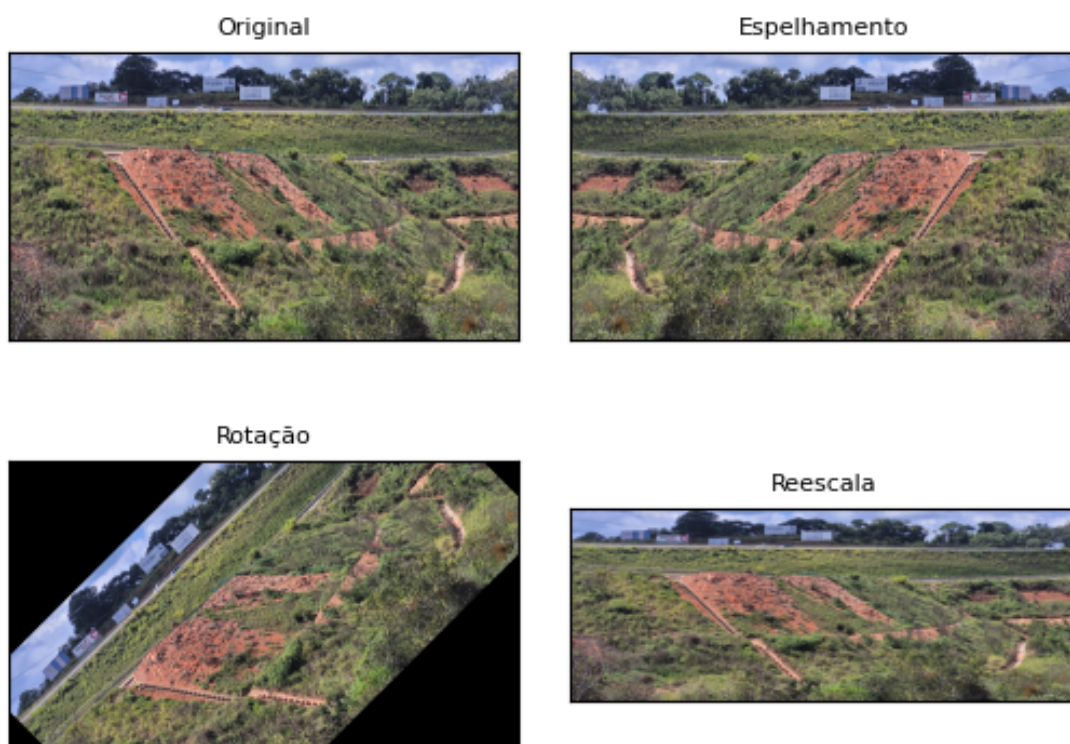
- a) Rotação : é um movimento circular da imagem em torno de um ponto fixo.
- b) Espelhamento: As imagens são espelhadas aleatoriamente na direção horizontal (esquerda-direita) ou na direção vertical (para cima).
- c) Cisalhamento: é a distorção da imagem ao longo de um eixo para criar ou retificar os ângulos de percepção.
- d) Ampliação : é realizado um aumento aleatório da imagem adicionando novos valores de pixels.
- e) Escala: é o ajuste das dimensões ou escala dos dados para que se adequem a uma determinada faixa ou formato.

A Figura 13 mostra um exemplo prático do uso de transformações geométricas no aumento de dados de imagens.

Já o aumento de dados por transformações fotométricas age sobre a composição pictórica da imagem. Os valores de pixel de uma imagem em cada canal de cor são alterados por filtros que modificam as características do espaço de cor de uma imagem (SHORTEN; KHOSHGOFTAAR, 2019). Dentre os métodos difundidos para o aumento de dados por transformações pictóricas estão (MIKOŁAJCZYK; GROCHOWSKI, 2018): histograma, equalização, aumento do contraste ou brilho, balanceamento de branco, nitidez e desfocagem. As modificações pictóricas são especialmente úteis na solução de desafios de iluminação, frequentes para problemas de reconhecimento de imagem (SHORTEN; KHOSHGOFTAAR, 2019).

O aumento de dados por meio de transformações geométricas e fotométricas se destaca como uma técnica eficaz para combater o sobreajuste em modelos de aprendizado de máquina. No entanto, a implementação dessa técnica exige cautela e atenção a diversos fatores. É crucial garantir que as transformações aplicadas à imagem preservem seu rótulo, especialmente em tarefas complexas como o reconhecimento de dígitos (SHORTEN; KHOSHGOFTAAR, 2019). Além

Figura 13. Efeitos da transformação geométrica em uma imagem.



Fonte: Autora.

disso, o impacto em tarefas específicas, como a análise de sentimento em imagens, deve ser considerado, pois alterações na cor podem afetar a detecção de emoções (SHORTEN; KHOSHGOFTAAR, 2019).

Além disso, o aumento de dados pode aumentar a quantidade de dados a serem processados, impactando na memória, nos custos de transformação e no tempo de treinamento. A seleção adequada das transformações, guiada pelas características do problema e do conjunto de dados, é fundamental para otimizar o desempenho do modelo.

### 3.5.3 Otimização Adaptativa

A otimização adaptativa é uma estratégia de regularização comumente utilizada com a finalidade de ajustar dinamicamente a taxa de aprendizado durante

o processo de treinamento de Redes de Aprendizado Profundo (KUKAČKA; GOLKOV; CREMERS, 2017), adaptando-se às características específicas do conjunto de dados e da tarefa em questão. Trata-se de algoritmos de aprendizado de máquina baseados em otimização, que quando aplicados à Redes Neurais Artificiais, procuram os pesos ideais que minimizam a função de erro, com vistas à redução do erro global de rede. Ao atualizar os vetores de peso, leva a um desempenho de generalização significativamente melhorado, uma vez que permite um controle mais preciso da direção no processo de aprendizado (ZHANG, 2019).

Dentre os otimizadores conhecidos, o Adam, do inglês *Adaptive Moment Estimation*, é frequentemente empregado como otimizador no ajuste dos pesos durante o processo de treinamento de modelos de Redes Neurais Convolucionais (ADNAN et al., 2022; BUSHARA; KUMAR, 2022; KUKAČKA; GOLKOV; CREMERS, 2017). É recomendado em modelos CNN por resolver problemas com gradientes heterogêneos de forma eficaz e estável diante de ruídos (ZHANG, 2019).

O Adam adapta a taxa de aprendizado de forma individual para cada parâmetro do modelo, calculando estimativas dos momentos de primeira e segunda ordem dos gradientes das variáveis do modelo. Ou seja, para alcançar a direção e a magnitude do gradiente, a técnica usa médias móveis exponenciais dos gradientes (primeiro momento) e médias móveis exponenciais dos quadrados dos gradientes (segundo momento).

Considerando os hiperparâmetros  $\alpha$  como a taxa de aprendizado,  $\beta_1$  como o fator de decaimento exponencial para a primeira estimativa de momento,  $\beta_2$  como o fator de decaimento exponencial para a segunda estimativa de momento, e  $\epsilon$  uma constante pequena para evitar divisão por zero. O otimizador Adam para uma Rede Neural Convolucional (CNN) pode ser descrito em quatro etapas:

a) Etapa 1: Inicialização de parâmetros

Inicialmente, a primeira estimativa de momento ( $m_t$ ), a segunda estimativa de momento ( $v_t$ ) e a iteração ( $t$ ) são definidas como zero.

b) Etapa 2: Atualização do primeiro momento 3.11 e segundo momento 3.12

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad \text{e} \quad (3.11)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (g_t^2) \quad \text{onde } g_t \text{ é o gradiente atual.} \quad (3.12)$$

- c) Etapa 3: Correção de bias do primeiro momento 3.13 e segundo momento 3.14

$$m_t = \frac{m_t}{1 - \beta_1^t} \quad \text{e} \quad (3.13)$$

$$v_t = \frac{v_t}{1 - \beta_2^t} \quad (3.14)$$

- d) Etapa 4: Atualização de pesos do primeiro momento 3.15 e segundo momento 3.16

$$W_{t+1} = W_t - \alpha \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \quad \text{e} \quad (3.15)$$

$$b_{t+1} = b_t - \alpha \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \quad (3.16)$$

Onde  $\alpha$  é a taxa de aprendizado,  $W$  é o peso,  $b$  é as bias, e  $\epsilon$  é uma constante pequena para evitar divisão por zero.

## 3.6 Modelos de Referência em Redes Neurais Convolucionais

Esta seção apresenta o resumo de alguns modelos populares de CNN utilizados em tarefas de visão computacional. Esses modelos incluem AlexNet, VGG, Inception V2, Resnet e DenseNet (DARAPANENI; KRISHNAMURTHY; PADURI, 2020; ABEDALLA et al., 2021; EKUNDAYO; VIRIRI, 2021; YADAV; JADHAV, 2019; JANG; KIM; AN, 2019).

O AlexNet é uma arquitetura de aprendizagem profunda que compreende 5 camadas convolucionais e 3 camadas totalmente conectadas (ABEDALLA et al., 2021). Ele emprega a função de ativação ReLu e outras técnicas, como abandono

e aumento de dados, para mitigar o sobreajuste. Além disso, aborda a questão do desaparecimento do gradiente, que é um problema comum na aprendizagem profunda.

O *Visual Geometry Group* (VGG) é outra arquitetura CNN popular. O arranjo consiste em 16 ou 19 camadas convolucionais usando filtros 3x3 (EKUNDAYO; VIRIRI, 2021; YADAV; JADHAV, 2019). O modelo segue uma abordagem uniforme na arquitetura, que pode exigir poder de processamento significativo da unidade de processamento gráfico, do inglês *Graphics Processing Unit* (GPU) ou unidade central de processamento, do inglês *Central Processing Unit* (CPU) devido a um grande número de camadas e parâmetros.

O Inception V2, ou GoogLeNet, é um modelo de aprendizagem profunda distinguido pela arquitetura "Inception"(JANG; KIM; AN, 2019). Essa arquitetura emprega filtros de diferentes tamanhos, camadas de agrupamento e convoluções 1x1 antes de convoluções 3x3 e 5x5 maiores para reduzir a dimensionalidade (YADAV; JADHAV, 2019). Essa abordagem alivia a necessidade de escolher o tamanho do filtro manualmente.

Uma ResNet, ou Redes Residuais, é um modelo de aprendizagem profunda que aborda a questão do desaparecimento de gradiente usando blocos residuais para treinar redes profundas (ABEDALLA et al., 2021). Este modelo permite o treinamento de redes com centenas de camadas, que anteriormente era uma tarefa desafiadora (DARAPANENI; KRISHNAMURTHY; PADURI, 2020).

Por fim, o *Densely Connected Convolutional Networks* (DenseNet) é um modelo de CNN caracterizado por conectividade entre camadas convolucionais. Cada camada recebe ativação de todas as camadas anteriores, permitindo que cada camada acesse diretamente os recursos aprendidos em todas as camadas anteriores (DARAPANENI; KRISHNAMURTHY; PADURI, 2020). Esta abordagem difere dos métodos tradicionais e tem mostrado resultados promissores em várias tarefas de visão computacional.

## 3.7 Avaliação de Desempenho do Modelo

Esta seção apresenta as métricas de desempenho mais comuns para avaliar o desempenho de redes neurais convolucionais em tarefas de classificação de imagens, detecção de objetos, segmentação semântica e outras aplicações. A avaliação do desempenho de CNNs é fundamental para determinar a eficácia e a capacidade de generalização desses modelos nas tarefas propostas.

### 3.7.1 Acurácia e Perda

Ao avaliar modelos de classificação, a acurácia, do inglês *accuracy*, é uma métrica para expressar o quão bem um modelo apresenta na previsão das etiquetas de classe para um determinado conjunto de dados (EKUNDAYO; VIRIRI, 2021; ADNAN et al., 2022). A acurácia pode ser apresentada como uma pontuação percentual entre 0 e 100%, sendo que 100% indica que todas as previsões estão corretas e 0 significa que nenhuma das previsões está correta, como mostrado na equação (3.17).

$$\text{Acurácia (\%)} = \frac{\text{Número de Previsões Corretas}}{\text{Total de Previsões}} \times 100 \quad (3.17)$$

A perda, do inglês *loss*, é uma métrica que avalia a disparidade entre as previsões do modelo e as anotações reais dos dados. O propósito principal da função de perda é medir a precisão das previsões do modelo (EKUNDAYO; VIRIRI, 2021; ADNAN et al., 2022; FOROUGH; CHEN; WANG, 2021). Minimizar a perda ao longo do processo de treinamento é imperativo, pois reflete o quanto as previsões do modelo se aproximam das etiquetas reais.

Em tarefas de classificação o cálculo de perda pode ser realizado por meio da função entropia cruzada binária, descrita pela equação (3.18).

$$\text{Perda} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))) \quad (3.18)$$

Na equação 3.18,  $N$  é o número total de exemplos no conjunto de dados,  $y_i$  é a classe verdadeira, por exemplo,  $i$ , e  $p(y_i)$  é a probabilidade prevista pelo modelo

de que o exemplo  $i$  pertence à classe positiva. A função de entropia cruzada é uma função interessante para a avaliação da perda no treinamento de classificadores (TAN; GUO; POH, 2021) com bases de dados pequenas (FOROUGH; CHEN; WANG, 2021).

### 3.7.2 Precisão, Revocação, Especificidade e Pontuação F1

As medidas obtidas pela precisão, revocação ou sensibilidade, especificidade e pontuação F1 ( *F1-Score*) auxiliam na elaboração de elementos gráficos para a avaliação do desempenho do modelo com base nos resultados das previsões. Os resultados de detecção podem ser categorizados como Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN) para detecção correta das etiquetas, Falso Positivo (FP) para detecção incorreta de objetos inexistentes, e Falso Negativo (FN) para objetos não detectados (HAN et al., 2022; ADNAN et al., 2022; CANO et al., 2021). Essas categorias auxiliam na elaboração da matriz de confusão e na obtenção dos índices *Area Under the ROC Curve* (AUC) e *Area Under the Precision-Recall Curve* (AUPRC). Essas métricas são aceitas como métodos padrão para avaliar o desempenho de modelos preditivos das tarefas de classificação de imagens (HAN et al., 2022; CANO et al., 2021), incluindo problemas com dados desequilibrados (EPPENHOF; PEÑA-CASTILLO, 2019; BOWD et al., 2022).

A precisão, do inglês *precision*, descrita na equação (3.19), é a razão de verdadeiros positivos VP para todos os positivos previstos feitos por um modelo.

$$\text{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}} = \frac{\text{VP}}{\text{n}^\circ \text{ de previsões positivas}} \quad (3.19)$$

Na equação (3.20) a métrica revocação, do inglês *recall*, também conhecida como sensibilidade, mede a proporção de previsões verdadeiras positivas VP entre todas as instâncias positivas reais e indica a Taxa de Verdadeiros Positivos (TVP).

$$\text{Revocação} = \frac{\text{VP}}{\text{VP} + \text{FN}} = \frac{\text{VP}}{\text{n}^\circ \text{ de exemplos positivos reais}} \quad (3.20)$$

As situações em que exemplos negativos são classificados incorretamente como positivos FP também são utilizados na avaliação dos modelos. A Taxa de



Falsos Positivos (TFP) ou Especificidade, indica a tendência de um modelo em classificar os exemplos negativos como positivos.

$$\text{Especificidade} = \frac{\text{FP}}{\text{FP} + \text{VN}} = \frac{\text{FP}}{\text{n}^\circ \text{ de exemplos positivos reais}} \quad (3.21)$$

A pontuação F1, dada pela equação (3.22), é uma métrica que equilibra precisão e a revocação e é usada para avaliar ambas simultaneamente.

$$\text{Pontuação F1} = \frac{2 \cdot \text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (3.22)$$

### 3.7.3 Matriz de Confusão

A matriz de confusão sintetiza as informações sobre a funcionalidade do modelo, particularmente em relação a falsos positivos, falsos negativos e verdadeiros positivos (LU; TAN; JIANG, 2021). A matriz de confusão é uma ferramenta capaz de representar visualmente o desempenho do modelo em operação, uma vez que sumariza amostras em diferentes categorias (AZIMI; PEKCAN, 2020).

A estrutura da matriz de confusão é composta por linhas e colunas que representam as classes reais e previstas, respectivamente. Cada célula da matriz representa o número de amostras que foram classificadas corretamente, isto é, quando a classe real e a classe prevista coincidem, e classificadas incorretamente, quando a classe real e a classe prevista não coincidem. A Figura 14 ilustra como é estruturada a matriz de confusão.

Figura 14. Estruturação de uma matriz de confusão.

	Predito Positivo	Predito Negativo
Classe Positivo	VP	FN
Classe Negativo	FP	VN

Fonte: Autora.

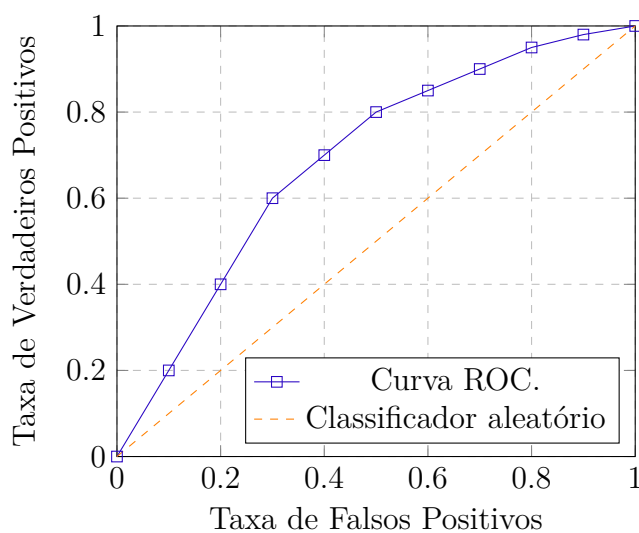
Na diagonal da matriz, localizam-se as previsões corretas do modelo: Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN). As demais posições da matriz

indicam incongruências entre a predição do modelo e a classe real da imagem, representadas pelos Falso Positivo (FP) e Falso Negativo (FN).

### 3.7.4 Curva ROC e índice AUC

Traçar a curva *Receiver Operating Characteristic* (ROC) também é uma estratégia de avaliação do modelo. Trata-se de uma representação visual de como o modelo prevê etiquetas verdadeiras em um classificador. A curva ROC é criada traçando a taxa de verdadeiros positivos no eixo  $y$  (vertical) TVP e a taxa de falsos positivos no eixo  $x$  TFP, conforme demonstrado na Figura 15.

Figura 15. Curva ROC.



Fonte: Autora.

A curva em laranja representa o desempenho de um classificador aleatório. Quanto mais próxima a curva ROC estiver do canto superior esquerdo do gráfico, maior será a área sob a curva *Area Under the ROC Curve* (AUC), indicando um melhor desempenho do modelo. Essa região representa o equilíbrio entre a alta taxa de verdadeiros positivos e a baixa taxa de falsos positivos para uma ampla gama de limiares.

A métrica AUC fornece índices que mensuram a qualidade geral da curva ROC e resumem brevemente a capacidade do modelo de discriminar entre classes

em diferentes limiares de classificação (AZIMI; PEKCAN, 2020). Quanto maior a área sob a curva, melhor o desempenho do modelo. Uma pontuação AUC de 1 significa que o modelo é capaz de separar perfeitamente as classes em todos os limiares de classificação, enquanto uma pontuação menor que 0,5 indica que o modelo tem um desempenho equivalente ao de um classificador aleatório (ADNAN et al., 2022; CANO et al., 2021; AZIMI; PEKCAN, 2020).

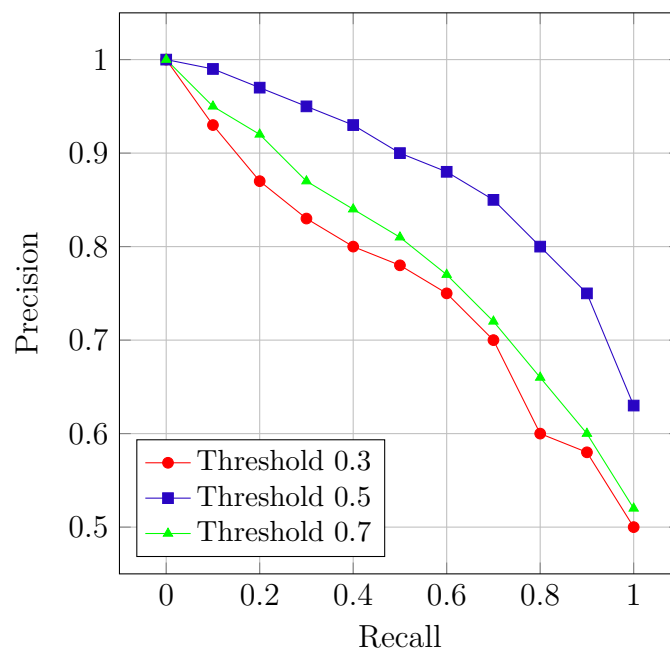
## Curva Precisão-Revocação e índice AUPRC

Um dos grandes desafios dos modelos de CNN é obter boa generalização quando uma classe apresenta-se mais frequente em detrimento a outra no conjunto de amostras de treinamento e validação/teste. Em tarefa de classificação com modelos de ML é frequente a utilização de valores limiares (*thresholds*), também chamado de ponto de corte, para delimitar a fronteira entre os valores de probabilidade que correspondem a determinada classe. Ao variar esse limiar, a proporção de classificações positivas e negativos que o modelo classifica é impactada e pode induzir a métricas de desempenho enganosas.

Desse modo, as métricas comumente empregadas para avaliar esses modelos, como acurácia por exemplo, podem não detectar o privilégio entre as classes e indicar resultados enganosos de desempenho. No contexto do desequilíbrio de classes, a relação entre a precisão e a revocação torna-se interessante pois é capaz de expressar o quanto o modelo é sensível à variação nos limiares, diferentemente de ROC-AUC.

Por meio da relação entre a precisão e a revocação da qual é possível obter um valor denominado AUPRC, baseado na área sob a curva *Precision-Recall Curve* (PRC), que mensura o desempenho dos modelos de classificação diante de classes desbalanceadas em dados (BOWD et al., 2022; EPPENHOF; PEÑACASTILLO, 2019). O valor de AUPRC indica a capacidade do modelo de identificar corretamente as amostras verdadeiras positivas, o que é elementar em tarefas onde os exemplos de determinada classe são raros ou custosos de obter.

Figura 16. Curvas de precisão-revocação para diferentes limiares.



Fonte: Autora.

## Considerações Finais

Neste capítulo, explorou-se o estado da arte sobre Redes Neurais Convolucionais, como uma técnica de aprendizado de máquina. Ao longo do capítulo, foram apresentados os princípios fundamentais dos modelos CNN, sua arquitetura e funcionamento, desafios e estratégias de aprimoramento dos modelos. Também se abordou brevemente sobre arquiteturas de referência, com a finalidade de promover uma visão ampla das técnicas disponíveis na literatura para modelagem e solução de problemas complexos. E por fim, apresentou-se as métricas comumente empregadas para avaliar o desempenho dos modelos de aprendizado de máquina, incluindo a métrica AUPRC recomendada na avaliação de modelos com desequilíbrio de classes.

## 4 Trabalhos Relacionados

### Considerações Iniciais

A avaliação da integridade estrutural de infraestruturas geotécnicas, como barragens, taludes e fundações, é fundamental para garantir a segurança pública e o desempenho adequado dessas estruturas. Nesse contexto, técnicas de ML, aprendizado profundo, do inglês *Deep Learning* (DL) e visão computacional VC vêm se mostrando ferramentas promissoras para a identificação de padrões em imagens e dados que indicam a presença de falhas e anomalias. Este capítulo apresenta uma revisão sistemática da literatura sobre o estado da arte na aplicação de aprendizado de máquina, aprendizado profundo e visão computacional para a identificação de falhas em estruturas geotécnicas. A revisão abrange alguns estudos que abordam a identificação de padrões em diferentes tipos de estruturas, incluindo solos, rochas e concreto, e que contribuíram para o desenvolvimento da pesquisa.

### 4.1 Avaliação da Integridade Estrutural

O monitoramento da integridade estrutural assume um papel relevante na avaliação e acompanhamento das propriedades e comportamentos de estruturas que podem comprometer sua segurança, confiabilidade ou desempenho (SPENCER; HOSKERE; NARAZAKI, 2019). Estudos que integram o monitoramento de estruturas com técnicas de aprendizado de máquina para detectar condições anormais e danos são cada vez mais comuns. De modo que, uma ampla gama de estruturas, incluindo pontes, barragens, ferrovias, edifícios, túneis, fundações e estruturas metálicas, entre outras, têm sido contempladas por abordagens integradas de monitoramento. Além do aprendizado de máquina, aprendizado profundo e visão computacional, também têm sido extensivamente empregadas e obtido êxito no monitoramento e na avaliação da integridade estrutural.

Nesse contexto, nos parágrafos seguintes são elencados estudos da literatura

que demonstram aplicações práticas de monitoramento de estruturas e análise de falhas com auxílio de técnicas de aprendizado de máquina. A revisão inicia com as abordagens clássicas de aprendizado de máquina que utilizam dados, principalmente, das leituras dos instrumentos de campo em estruturas de concreto. Em seguida, apresenta-se a aplicação de modelos de aprendizado profundo, como um subconjunto de técnicas mais complexas de aprendizado de máquina, e que possuem potencial de investigação no campo da análise da integridade das estruturas.

Tradicionalmente, a avaliação da integridade das estruturas se baseia na identificação de falhas em conjuntos de dados provenientes de medições realizadas por instrumentos de campo nas por vistorias *in locu* por profissional da área. Um exemplo de identificação de falhas provenientes medições é o estudo de (JUNG; BERGES; GARRETT, 2014). No estudo, o método de regressão foi empregado com a finalidade de detectar erosão interna em barragens a partir da análise das relações entre leituras piezométricas e níveis do reservatório. O modelo numérico desenvolvido por (JUNG; BERGES; GARRETT, 2014) alcançou uma precisão de detecção de 98,5%, e demonstrou efetividade na identificação de erosão interna em barragens. A abordagem possibilitou aos autores entenderem como os processos físicos refletem nos níveis piezométricos em reservatórios, e o quão é difícil para os engenheiros reconhecerem os problemas que geralmente ocorrem dentro de barragens e sem sinais visuais.

A modelagem de níveis piezométricos em barragens é usual no monitoramento da segurança e da estabilidade das estruturas. Nesse contexto, técnicas de aprendizado de máquina se apresentam como ferramentas promissoras para a construção de modelos precisos e eficazes. Em (SALAJEGHEH; MEYMAND; KERMANI, 2018) foi realizado um estudo comparativo para avaliar a efetividade de diferentes métodos de aprendizado de máquina na modelagem de níveis piezométricos na Barragem de Jiroft, localizada na província de Kerman, Irã. O estudo analisou os resultados de cinco métodos distintos: Regressão Linear Múltipla (Um método tradicional de modelagem estatística), Árvore de Decisão de Regressão (Um método de aprendizado de máquina que utiliza árvores de decisão para prever valores contínuos), Árvore de Decisão de Classificação (Um método de aprendizado de máquina que utiliza árvores de decisão para classificar dados em categorias),

Redes Neurais Artificiais (Um tipo de rede neural artificial que utiliza o algoritmo Levenberg-Marquardt para o treinamento), Redes Neurais Artificiais com Otimização de Enxame de Partículas (Um tipo de rede neural artificial que utiliza o algoritmo de otimização de enxame de partículas para o treinamento), e por fim, Redes Neurais Artificiais com Algoritmos de Aprendizado de Pesquisa de Harmonia (Um tipo de rede neural artificial que utiliza o algoritmo de aprendizado de pesquisa de harmonia para o treinamento). O estudo de (SALAJEGHEH; MEYMAND; KERMANI, 2018) não identificou um método superior absoluto para a modelagem de níveis piezométricos na Barragem de Jiroft. Todos os métodos apresentaram resultados comparáveis em termos de precisão e confiabilidade.

Um método para projetar um sistema de detecção e monitoramento de irregularidades em pontes rodoviárias é proposto por (DANG et al., 2022). O sistema combina tecnologias de Internet das Coisas, do inglês *Internet of Things* (IoT) e Inteligência Artificial (IA) para realizar um monitoramento remoto. Arquitetura do Sistema é composta por uma camada de aquisição de dados, camada de processamento de dados e uma camada de armazenamento. Na camada de aquisição, um Raspberry Pi 4 atua como *gateway* IoT, coletando dados de deslocamentos e deformações da ponte por meio de sensores. Na camada de processamento, uma CNN embarcada no Raspberry Pi 4 extrai padrões dos dados brutos de medições em tempo real. E na última camada, os padrões extraídos são armazenados na nuvem do Google Firebase para análise e visualização posteriores. O sistema permitiu o monitoramento simultâneo e estável de pontes remotas, possibilitando a detecção precoce de irregularidades. E rede convolucional embarcada extraiu padrões complexos dos dados de medições com alta precisão. O armazenamento dos dados na nuvem do Google Firebase facilitou o acesso e a análise dos resultados por diferentes usuários e especialistas.

Uma abordagem para o monitoramento da integridade estrutural, utilizando Redes Neurais Convolucionais e técnicas de transferência de aprendizado, do inglês (TL) foi proposta por (AZIMI; PEKCAN, 2020). A metodologia proposta explora uma forma de dados de resposta compactados obtidos por meio de técnicas de TL, possibilitando a identificação e localização de danos em sistemas estruturais realistas de larga escala. Modelos CNN foram treinados e avaliados utilizando

históricos de resposta de aceleração e dados de resposta compactados em termos de histogramas discretos. A utilização de dados de resposta compactados reduz significativamente o custo computacional do processo de monitoramento. Para mensurar o desempenho da CNN, (AZIMI; PEKCAN, 2020) utilizou a acurácia e a matriz de confusão como métrica. A metodologia proposta demonstra alta capacidade de identificar e localizar danos em sistemas estruturais de larga escala. e boa capacidade de generalização para diferentes tipos de sensores, abrindo caminho para aplicações em diversos cenários.

Dentre as práticas de engenharia para o monitoramento, a inspeção visual se destaca como um procedimento primário na avaliação da integridade de estruturas, especialmente em situações onde a instrumentação não está disponível (MASCAREÑAS et al., 2021). Essa prática garante a segurança e confiabilidade de diversas estruturas, desde pontes e barragens até edifícios e túneis. Apesar de sua importância, a inspeção visual apresenta desafios inerentes que precisam ser considerados. Um dos desafios está relacionado a relatividade e relevância dos parâmetros de inspeção. Os parâmetros utilizados por diferentes profissionais durante a inspeção podem apresentar discrepâncias significativas, influenciando diretamente os resultados da avaliação. Outro desafio é a dependência da experiência do profissional. A experiência individual do inspetor impacta consideravelmente na qualidade da avaliação, tornando-a suscetível à subjetividade. Para minimizar os desafios da inspeção visual, (PAN; CHEN, 2015) propõe uma abordagem baseada em aprendizado de máquina, em que utiliza da teoria de conjuntos *fuzzy* para a modelagem de eventuais incertezas e ambiguidades presentes na avaliação da integridade estrutural. O modelo define funções de associação da faixa semântica, que classificam as condições estruturais em diferentes categorias. Dentre as contribuições do estudo estão a redução da subjetividade, a confiabilidade e a padronização da inspeção de estruturas.

Para além de métodos tradicionais como instrumentação e inspeção visual, a combinação de métodos numéricos clássicos e aprendizado de máquina têm sido empregados para a avaliação da integridade estrutural. Na pesquisa desenvolvida por (GUAN et al., 2023) se propõe o treinamento de uma CNN com dados tridimensionais de tensão-deformação do solo obtidos através do método dos elementos



finitos, do inglês *Finite Element Method* (FEM). Essa CNN foi aplicada com sucesso na resolução de problemas de contorno, do inglês *Boundary Value Problem* (BVP) em análises de meios infinitos, como solos. A pesquisa demonstrou a efetividade da metodologia em três a BVPs com diferentes geometrias, malhas e condições de contorno, obtendo resultados satisfatórios sem a necessidade de teorias constitutivas complexas e com maior eficiência computacional em relação aos métodos tradicionais. Similarmente, (SHI et al., 2022) empregam CNN para prever o comportamento de rochas sob diferentes condições de carga. Essa pesquisa abre caminho para o desenvolvimento de ferramentas inovadoras para o monitoramento e a análise de estabilidade de taludes e outras estruturas geotécnicas. Outro trabalho nessa área é o de (ARAUJO, 2015), que propõe uma integração entre FEM e redes neurais clássicas. Modelos analíticos e numéricos, como o FEM, são utilizados para gerar amostras de deslocamentos e deformações ao longo da geometria das estruturas, permitindo a obtenção do comportamento para qualquer ponto da estrutura.

## 4.2 Visão Computacional

Um dos potenciais do aprendizado de máquina, relacionados à inteligência artificial, é de fornecer modelos baseados em imagens capazes de caracterizar e identificar a variabilidade dos objetos em imagens como padrões bidimensionais (VEDALDI et al., 2014). Nesse contexto, este trabalho inicia a revisão com aplicações de modelos clássicos de ML para tarefas de visão computacional. Tais modelos já são empregados para avaliar padrões em imagens de superfícies, possibilitam a identificação de falhas em diversas categorias de estruturas. Em seguida, evidencia-se o potencial das CNNs no processamento de imagens, por meio de estudos que demonstram a viabilidade de CNN na resolução de problemas em diversos setores, como indústria, saúde e agricultura. Por fim, são apresentados estudos que empregam modelos CNN para detectar possíveis danos em estruturas por meio de imagens de superfícies, bem como estudos que demonstram soluções baseadas na integração dos modelos de redes convolucionais a dispositivos modernos de aquisição de imagens, como VANTs.

Em (DI et al., 2014) é apresentada uma estratégia clássica de processamento

de imagem para a detecção automática de crateras em Marte utilizando dados de imagem e topográficos. A metodologia consiste em duas etapas principais. Na primeira etapa, um algoritmo de impulsionamento é empregado para identificar regiões quadradas com potencial presença de crateras. Essa etapa filtra as áreas irrelevantes e concentra o processamento em regiões com maior probabilidade de conter crateras. Na segunda etapa, uma nova variante de características Haar-like é proposta para aprimorar a classificação de crateras. A nova variante combina características Haar-like tradicionais com características de padrão binário local para obter uma representação mais completa das crateras. Essa combinação de descritores visuais melhora a capacidade do algoritmo de discriminar entre crateras e outras características da superfície marciana. Experimentos com dados topográficos de Marte demonstraram que a abordagem proposta apresenta um desempenho superior em comparação a métodos tradicionais de detecção de crateras. A nova variante de características Haar-like, em conjunto com a combinação de descritores visuais, contribuiu para a detecção precisa e eficiente de crateras, mesmo em locais com terreno desafiador.

Em (RAMANDI et al., 2022) uma análise propõe um algoritmo para identificação automática de fraturas em rochas baseado em imagens de tomografia computadorizada 3D em escala de cinza a partir da aplicação de filtragem Gaussiana. O experimento foi realizado em duas etapas. A primeira etapa do algoritmo realiza uma visão inicial de baixo nível das imagens. Essa etapa utiliza técnicas clássicas de processamento de imagem, com a extração manual de característica por filtros, para identificar possíveis áreas de fraturas na rocha. Na segunda etapa, um processo interpretativo de alto nível é aplicado para extrair eficientemente as fraturas planares das imagens digitais. Essa etapa utiliza técnicas de aprendizado de máquina para classificar as áreas de fraturas identificadas na etapa anterior e determinar suas características geométricas, como orientação, tamanho e forma. A identificação precisa das fraturas em rochas é fundamental para entender sua estrutura geológica e comportamento mecânico e auxiliar na avaliação da estabilidade do terreno e na prevenção de deslizamentos.

No âmbito do aprendizado de máquina e da visão computacional, as Redes Neurais Convolucionais se consolidaram como a vanguarda tecnológica, por

superarem consistentemente, em alguns casos, os métodos clássicos baseados em descritores de recursos manuais. Essa supremacia se deve a diversos fatores intrínsecos à arquitetura das CNNs, que as tornam ferramentas atrativas para a extração e o aprendizado de características relevantes em imagens e dados visuais. Nesse sentido, os modelos CNN têm apresentado desempenho notável em muitas tarefas de visão computacional (SHORTEN; KHOSHGOFTAAR, 2019), uma vez que são capazes de interpretar com sucesso o mundo ao seu redor por meio de tarefas de detecção, classificação, segmentação, e reconhecimento de padrões em um conteúdo visual (LINDSAY, 2021). O que atribui um caráter versátil às aplicações da visão computacional em áreas diversas como saúde, indústria, sensoriamento, entre outras.

Na agricultura, as CNNs demonstram grande potencial para a identificação precisa e precoce de doenças em plantas, permitindo aos agricultores tomar medidas de controle mais eficazes e reduzir perdas na produção. Em (LU; TAN; JIANG, 2021), por exemplo, uma solução baseada em CNNs foi desenvolvida para classificar doenças de plantas com alta precisão, superando métodos tradicionais baseados em imagens manuais. Estudos como o de (KATTENBORN et al., 2021) também demonstram o potencial das CNNs para a análise de imagens de satélite e outras fontes de dados para identificar doenças em grandes áreas agrícolas.

Na indústria, um estudo desenvolvido por (AFFONSO et al., 2017) investigou a viabilidade do aprendizado profundo para a classificação da qualidade de placas de madeira a partir de imagens. Para tal, os autores compararam o desempenho do aprendizado profundo com técnicas tradicionais de extração de padrões, como as máquina de vetores de suporte, do inglês *Singular Vector Machine* (SVM). Os autores utilizaram um conjunto de dados de imagens de placas de madeira com diferentes níveis de qualidade. As imagens foram pré processadas e então utilizadas para treinar modelos de aprendizado profundo e SVM. O desempenho dos modelos foi avaliado através de métricas como precisão, sensibilidade e pontuação F1. Os resultados do estudo demonstraram que o aprendizado profundo superou significativamente as técnicas tradicionais de classificação na tarefa de avaliar a qualidade das placas de madeira. Essa superioridade foi particularmente evidente em cenários de alta complexidade, onde as características das imagens eram mais

sutis e difíceis de serem identificadas por métodos tradicionais.

Na saúde, o potencial da visão computacional com modelos CNN tem sido empregado na identificação e prevenção de doenças, que utilizam exames de imagem como principal base de treinamento. Uma aplicação de CNN para a detecção automática de lesões de tuberculose em imagens de tomografia computadorizada pulmonar 3D é demonstrada por (KALINOVSKY; LIAUCHUK; TARASAU, 2017). O diagnóstico precoce da tuberculose é importante para o controle da doença, e a Tomografia Computadorizada (TC) pulmonar é um método diagnóstico importante, embora a análise manual das imagens possa ser demorada e suscetível a erros. Utilizando um conjunto de dados de 338 tomografias computadorizadas 3D de pacientes com tuberculose, com lesões marcadas manualmente, os algoritmos baseados em CNNs foram treinados e testados. Os resultados demonstram que as CNNs superam algoritmos baseados em métodos convencionais de análise de imagens, com a melhor performance obtida por um algoritmo que utiliza uma arquitetura CNN profunda com 16 camadas convolucionais. De acordo com os resultados apresentados, as CNNs podem ser uma ferramenta eficaz para auxiliar os radiologistas na análise de imagens de TC, reduzindo o tempo de diagnóstico e aumentando a precisão da detecção de lesões.

Solução similar é apresentada por (JIANG et al., 2018). Um modelo de Rede Neural Convolucional foi proposto para a segmentação semântica de imagens médicas. A CNN foi treinada em um conjunto de dados de imagens médicas, aprendendo a identificar e segmentar diferentes estruturas anatômicas. Um dos principais desafios na segmentação de imagens médicas é a disponibilidade limitada de dados. Hospitais e organizações de pesquisa geralmente não possuem grandes conjuntos de dados, dificultando o treinamento de modelos de aprendizado profundo robustos. Para superar esse desafio, (JIANG et al., 2018) exploraram técnicas para treinar redes neurais em conjuntos de dados pequenos. Os experimentos qualitativos e quantitativos do estudo demonstraram que a CNN proposta foi capaz de lidar com sucesso com o problema de segmentação semântica de imagens médicas, mesmo com conjuntos de dados limitados. As conclusões do estudo indicam que as CNNs podem ser uma ferramenta promissora para a segmentação semântica de imagens médicas, mesmo em cenários com escassez de dados.

No campo da avaliação da integridade estrutural, abordagens baseadas em modelos CNN têm sido exploradas para automatizar a identificação de padrões em imagens. O modelo YOLO, do inglês *You Only Live Once*, um tipo de arquitetura de rede convolucional, foi considerado uma ferramenta promissora para a detecção de trincas em superfícies de concreto, especialmente em barragens. Um estudo de (LI et al., 2019a) propõe a utilização do YOLO para automatizar a detecção de trincas, buscando superar as limitações dos métodos tradicionais, que geralmente são ineficientes, dispendiosos e dependentes da experiência dos profissionais. Os resultados apresentados demonstram a viabilidade do YOLO para a detecção rápida e precisa de trincas em imagens de concreto. O modelo foi capaz de identificar a localização das trincas com alta precisão, tanto no conjunto de dados original quanto em um conjunto de dados expandido por meio de técnicas de aumento de dados.

Em (CHA; CHOI; BÜYÜKÖZTÜRK, 2017) uma arquitetura CNN personalizada para tarefa de visão computacional foi proposta com a finalidade de detecção de rachaduras em imagens de concreto. As imagens para treinamento, validação e teste foram capturadas com uma câmera de mão, totalizando 332 imagens. Essas imagens foram posteriormente divididas em 277 para treinamento e validação, e 55 para teste. Para aumentar a quantidade de imagens disponíveis as imagens foram segmentadas em imagens menores, cada uma representando uma área específica. Assim, o conjunto de dados de treinamento e validação ampliou para 40.000 imagens. A rede CNN treinada com 32.000 imagens obteve acurácia de 98,22%, enquanto a rede validada com 8.000 imagens alcançou 97,95%. O estudo também gerou mapas de rachaduras, evidenciando as áreas afetadas nas imagens. Os resultados do teste demonstraram um desempenho consistente e superior aos métodos tradicionais de detecção de bordas, como Canny e Sobel, especialmente na detecção de rachaduras finas em condições de iluminação desafiadoras. O método proposto por (CHA; CHOI; BÜYÜKÖZTÜRK, 2017) apresentou diversos benefícios em comparação aos métodos tradicionais, dentre eles a extração automática de características.

Devido às particularidades dos fenômenos de falhas em infraestruturas, tópicos relacionados à base de dados tem sido preocupação na aplicação área da visão computacional por modelos CNN uma vez que, em alguns contextos, esses

dados são escassos.

Nesse contexto, (LIM et al., 2021) exploraram o aumento de dados em imagens híbridas para detecção de corrosão superficial em pontes de aço usando o Faster R-CNN. O Faster R-CNN é combina arquiteturas VGG e métodos clássicos para estimar a probabilidade de uma região conter um objeto e para ajustar as caixas delimitadoras das regiões propostas. Partindo de um conjunto inicial de apenas 72 amostras, o estudo demonstrou o impacto positivo do aumento de dados através da combinação de imagens no espectro visível e imagens termográficas, elevando para 202 o total de amostras. Essa estratégia resultou em uma identificação mais precisa de áreas corroídas, mesmo em condições de iluminação desafiadoras.

Além da diversificação do tipo de imagens digitais, estratégias envolvendo dispositivos para aquisição de imagens remotas por VANT também tem sido explorada. Em (JANG; KIM; AN, 2019) para a detecção de trincas em estruturas de concreto utilizou um conjunto de imagens híbridas, compostas por imagens do espectro visível e imagens termográficas, coletadas por VANT, para treinar um modelo de CNN. As 200 imagens vetoriais foram convertidas e tiveram suas propriedades modificadas pela técnica de aumento de dados para gerar 20.000 amostras. Os resultados apontam que o modelo foi capaz de identificar com maior precisão as características das trincas, superando os resultados obtidos com o uso de apenas um tipo de imagem.

Uma solução similar foi desenvolvida para viabilizar a automação de inspeções visuais, permitindo a identificação precisa e eficiente de falhas em estruturas. A avaliação de desempenho da classificação *pixel-wise* baseada em modelos de referência VGG e ResNet, para a detecção de danos em superfícies de infraestruturas, é demonstrada por (SPENCER; HOSKERE; NARAZAKI, 2019). O algoritmo foi treinado em um conjunto de dados, que foi aumentado artificialmente de 339 para 1.695 imagens para expandir a diversidade das imagens e aprimorar a generalização do modelo. Os resultados do estudo demonstraram a capacidade do modelo de reconhecer com precisão seis tipos diferentes de danos em superfícies de concreto, incluindo diferentes tipos de fraturas e corrosão. A saída do algoritmo consiste em uma imagem segmentada, onde a área que representa o dano é destacada e classificada como uma das categorias de dano treinadas.

O estado da arte demonstrou que modelos CNN podem identificar rachaduras em concreto com acurácia superior a 95%, superando métodos tradicionais de análise de imagens. Nesta perspectiva, os modelos de Redes Neurais Convolucionais estão emergindo como uma abordagem promissora na engenharia geotécnica para a identificação e quantificação de danos em estruturas, elementos de estrutura e solos a partir de imagens. Sua capacidade de analisar dados com alta precisão e eficiência os eficazes e com desempenho superior, em alguns casos, em relação aos métodos tradicionais de análise de imagens. A implementação de modelos CNN na engenharia geotécnica tem o potencial de aprimorar a segurança pública, prevenir colapsos e otimizar a gestão de infraestruturas geotécnicas.

Um método proposto detecção de fissuras em solos argilosos utilizando redes neurais convolucionais em imagens digitais é avaliado por (HAN et al., 2022) . O método, baseado na arquitetura Mask R-CNN, supera significativamente os métodos tradicionais de detecção de fissuras, como o algoritmo U-Net, em termos de precisão e eficiência. O Mask R-CNN atingiu uma precisão média de localização de 64,14% e uma precisão média de segmentação de 47,59% na detecção de fissuras em solos argilosos. Esses resultados demonstram a superioridade do Mask R-CNN sobre o U-Net, que obteve precisões de 52,36% e 38,92%, respectivamente.

Em (LIMÃO; ARAÚJO; FRANCES, 2023) uma arquitetura de Rede Neural Convolucional foi projetada especificamente para a classificação de imagens de erosão em taludes. A CNN proposta atingiu uma acurácia de 96,67% na classificação de imagens de erosão em taludes. A alta precisão da CNN proposta permite a identificação precisa e aprimorada de sinais de instabilidade em taludes, facilitando a detecção precoce de riscos de colapso e permitindo a implementação de ações de remediação imediatas ou medidas preventivas eficazes, contribuindo para a segurança de estruturas geotécnicas e a mitigação de desastres.

### 4.3 Identificação de Lacunas

A análise da literatura demonstra a existência de diversas pesquisas relacionadas aos temas desta tese. Para identificar as lacunas de conhecimento que este estudo busca preencher, foi realizado um mapeamento dos temas abordados em

trabalhos anteriores, como apresentado na Tabela 4.3. A análise da literatura e a identificação das lacunas de conhecimento demonstram a relevância e o potencial desta tese.

A intersecção entre os temas expressos na Tabela 4.3 indica que a aplicação de visão computacional e CNN no monitoramento da integridade estrutural já é uma realidade, com destaque para sua utilização em estruturas de concreto, demonstrando o potencial dessa tecnologia para a área. A revisão também revelou pesquisas que analisam a cobertura do solo, entretanto com foco em aplicações agrícolas com a utilização de imagens de satélite, o que evidencia a maturidade dessa tecnologia em outras áreas.

Diante do exposto, identifica-se uma lacuna de pesquisa no potencial de integração entre a aquisição de imagens de alta resolução por meio de VANTs e modelos de visão computacional baseados em CNNs para a identificação de falhas em imagens de superfícies de solo e rocha. Essa integração pode ser aplicada ao monitoramento de diversas infraestruturas críticas, como diques, rodovias e aterros, onde a detecção precoce de danos nas imagens pode auxiliar na prevenção de acidentes e na otimização de recursos de manutenção. O que abre uma oportunidade para pesquisas no campo prático nesta área. Uma vez que essa área apresenta grande potencial de aplicação, considerando a importância dessas estruturas para diversos setores, como construção civil, mineração e agricultura.

Adicionalmente, observa-se outra lacuna na literatura sobre a disponibilidade de dados para o treinamento de modelos CNNs. Essa lacuna é evidenciada pela frequência com que a técnica de aumento de dados é empregada em pesquisas com CNNs. O uso da técnica sugere a escassez de dados nesta área da engenharia. Essa lacuna destaca a necessidade de pesquisas que explorem técnicas de coleta, pré-processamento e aumento de dados específicos para essa área de aplicação, a fim de aprimorar o desempenho dos modelos CNNs e a confiabilidade dos resultados, especialmente, em estruturas geotécnicas de solo e rocha.



Tabela 1. Compilado de trabalhos correlatos.

Referências	Tópicos de Interesse								
	CNN	ML	VC	Instrumentação	Inspecção	VANT	Solo	Concreto	DA
(SPENCER; HOSKERE; NARAZAKI, 2019);									
(LIM et al., 2021);									
(JANG; KIM; AN, 2019)	x		x		x		x	x	x
(CHA; CHOI; BüyüKÖZTürk, 2017)	x		x		x				x
(AZIMI; PEKCAN, 2020);									
(DANG et al., 2022)	x			x					x
(WANG et al., 2023)	x			x			x		
(OTTONI; AMORIM; NOVO, 2022)	x		x		x			x	x
(KATTENBORN et al., 2021)	x		x				x		x
(ZHANG; TONG, 2023)	x		x		x		x		x
(MOUNTRAKIS; HEYDARI, 2023)	x		x		x		x		
(HAN et al., 2022)	x		x		x				x
(LIMÃO; ARAÚJO; FRANCES, 2023)	x		x				x		
(KHAN et al., 2021)		x		x			x		
(PAN; CHEN, 2015)		x			x		x		
(LEI et al., 2020)		x	x		x		x		x
(XU et al., 2019);									
(FIGUEIREDO et al., 2011);									
(SALAJEGHEH; MEYMAND; KERMANI, 2018)		x		x				x	
(SHI et al., 2016)		x	x						x
(NOGUEIRA et al., 2020)		x	x				x		
(JUNG; BERGES; GARRETT, 2014)		x		x			x		
(JR.; ADELI, 2018)		x	x		x				x

Fonte: Autora.

## Considerações Finais

A revisão identificou um número crescente de estudos que aplicam aprendizado de máquina, aprendizado profundo e visão computacional com potencial de aplicação na identificação de falhas em estruturas geotécnicas. Os resultados demonstram que essas técnicas podem ser eficazes na detecção de diversos tipos de falhas, como rachaduras, erosões, deslizamentos e instabilidades. No entanto, a revisão também constatou que a aplicação dessas técnicas em engenharia geotécnica ainda é relativamente recente e em desenvolvimento. Há necessidade de mais pesquisas para aprimorar a precisão e robustez dos métodos, além de ampliar sua aplicação para diferentes tipos de estruturas e condições de solo e rocha. O que abre um campo promissor para futuras investigações.

Cabe ressaltar que as técnicas clássicas de processamento de imagem, como segmentação de imagens, filtragem e extração de características, ainda são ferramentas eficazes na visão computacional e em alguns casos mais eficientes por serem consideradas interpretáveis, requerem menos recursos computacionais e conjuntos de dados menores.

No que tange as abordagens usando CNN demonstram a aplicação de várias camadas convolucionais para processar efetivamente o conjunto de dados disponível e extrair informações relevantes dos padrões de imagem (LINDSAY, 2021; TAN; GUO; POH, 2021; BARI et al., 2021). Exemplos de aplicações de visão computacional na medicina (YADAV; JADHAV, 2019; CANO et al., 2021; LIAUCHUK et al., 2017; ABEDALLA et al., 2021), na agricultura (KATTENBORN et al., 2021; LU; TAN; JIANG, 2021), sensoriamento (LI et al., 2018; MOUNTRAKIS; HEYDARI, 2023), fabricação (AFFONSO et al., 2017) e reconhecimento facial (EKUNDAYO; VIRIRI, 2021; ZHAO et al., 2019; WANG et al., 2020). Além disso, destaca-se as análises de desempenho dos modelos diante de cenários de dados escassos (FOROUGH; CHEN; WANG, 2021; BRIGATO; IOCCHI, 2021), restrição de recursos computacionais (FOROUGH; CHEN; WANG, 2021) e arquiteturas enxutas (DARAPANENI; KRISHNAMURTHY; PADURI, 2020).

# 5 Construção de um Classificador para a Detecção de Danos Aparentes em Taludes

## Considerações Iniciais

Neste capítulo, apresenta-se a metodologia geral da pesquisa, o que abrange as etapas de aquisição e pré-processamento dos dados, treinamento, teste e validação. Discutem-se os fatores considerados na definição dos hiperparâmetros para o treinamento do modelo de Redes Neurais Convolucionais na busca por uma arquitetura que atenda aos objetivos propostos. Ressalta-se que, neste estudo, o termo "CNN de baixa complexidade" refere-se a arquiteturas projetadas com o objetivo de minimizar o número de parâmetros sem comprometer significativamente a acurácia do modelo.

Na presente tese, define-se dano os sinais de solo exposto perceptíveis em imagens a olho nu, sem o auxílio de ampliação ou ferramentas adicionais. Considerando as características das estruturas geotécnicas, optou-se pela abordagem de taludes, elementos estruturais extremamente importantes na composição da paisagem da estrutura e, portanto, capazes de representar e fornecer informações relevantes sobre a fotografia da estrutura.

### 5.1 Visão Geral do Processo

A segurança e confiabilidade de estruturas geotécnicas, como barragens, taludes e fundações, são essenciais para a prevenção de catástrofes e desastres. A inspeção visual constitui o método tradicional para detectar falhas em tais estruturas. No entanto, a realização de visitas presenciais pode ser arriscada devido à instabilidade do solo, expondo os profissionais a perigos e limitando a frequência das inspeções.

Visando superar os desafios da inspeção visual tradicional, propôs-se a automação do processo por meio da utilização de Veículo Aéreo Não Tripulado (VANT) e dispositivos móveis para a coleta de dados de imagens das superfícies das estruturas geotécnicas. Esses dados foram utilizados para alimentar um modelo de classificação binária, capaz de identificar a deterioração da superfície da estrutura com base no reconhecimento de áreas com solo exposto. Destaca-se que as imagens foram coletadas de taludes, em virtude da relevância desse elemento na área da geotecnia. Os taludes possuem grande potencial de representar a fotografia do terreno, uma vez que constitui-se como um elemento da estrutura ou como a própria estrutura, em grande parte das infraestruturas.

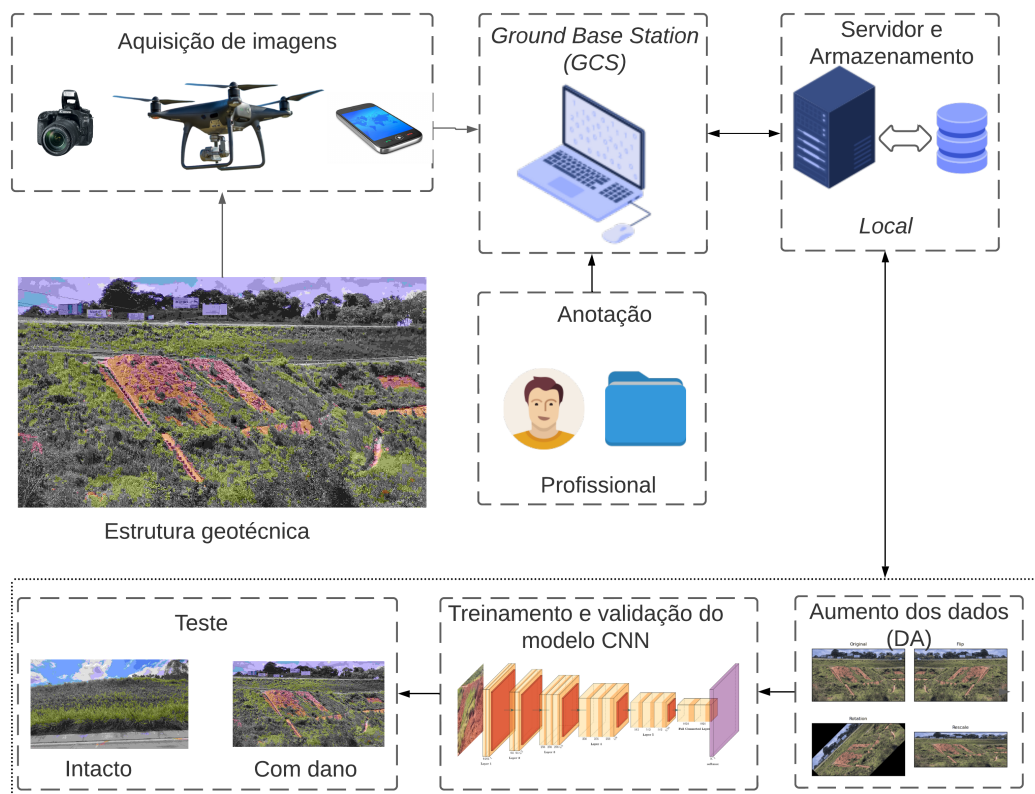
Para garantir a robustez e a generalização do modelo de classificação binária, é fundamental que ele seja treinado com um conjunto de dados diversificado e abrangente. Nesse sentido, aplicou-se técnicas de aumento de dados, que consistiram na geração artificial de novas imagens a partir das imagens originais coletadas. Essa técnica permitiu ampliar o volume e a qualidade das amostras de treinamento, com a finalidade de expor o modelo a diferentes condições de iluminação, ângulos de captura e características da superfície da estrutura, uma vez que foi identificada redundância nos dados adquiridos em inspeção.

Para a detecção de danos em imagens de estruturas geotécnicas, propôs-se um modelo personalizado baseado em redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN). Visando otimizar o desempenho computacional e a eficiência do modelo para implementação em dispositivos embarcados e *smartphones*, optou-se por uma arquitetura compacta em relação aos modelos de referência. Essa escolha se justifica pelo uso frequente de dispositivos embarcados com conexão remota no monitoramento de estruturas geotécnicas, especialmente em áreas remotas e hostis com severas limitações de infraestrutura, como descrito em (KHAN et al., 2021).

A construção de uma visão geral abrangente do processo de detecção de danos em superfícies de taludes por meio de imagens coletadas remotamente envolveu diversas etapas elementares. Essas etapas, interligadas e interdependentes, foram fundamentais para a definição da metodologia completa, como se discorre nas sessões subsequentes. A Figura 17 apresenta um resumo do processo proposto para

a inspeção de estruturas geotécnicas com a integração entre VANTs, dispositivos móveis e uma CNN.

Figura 17. Fluxograma para detecção de falhas de superfície em estruturas geotécnicas.



Fonte: Autora.

## 5.2 Ambiente Operacional

Para a configuração do ambiente do trabalho, foram utilizadas as seguintes tecnologias:

- Hardware: Windows 11 Home x64, 16GB de RAM e Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30GHz.
- Linguagem de programação: Python 3.9.16.

- Plataforma de computação interativa baseada na Web: Project Jupyter.
- Principais Bibliotecas Python: tensorflow 2.8.0, Keras 2.12.0, Pandas 1.4.1, NumPy 1.22.3, Matplotlib 3.5.1 e Scikit-learn 1.0.2.

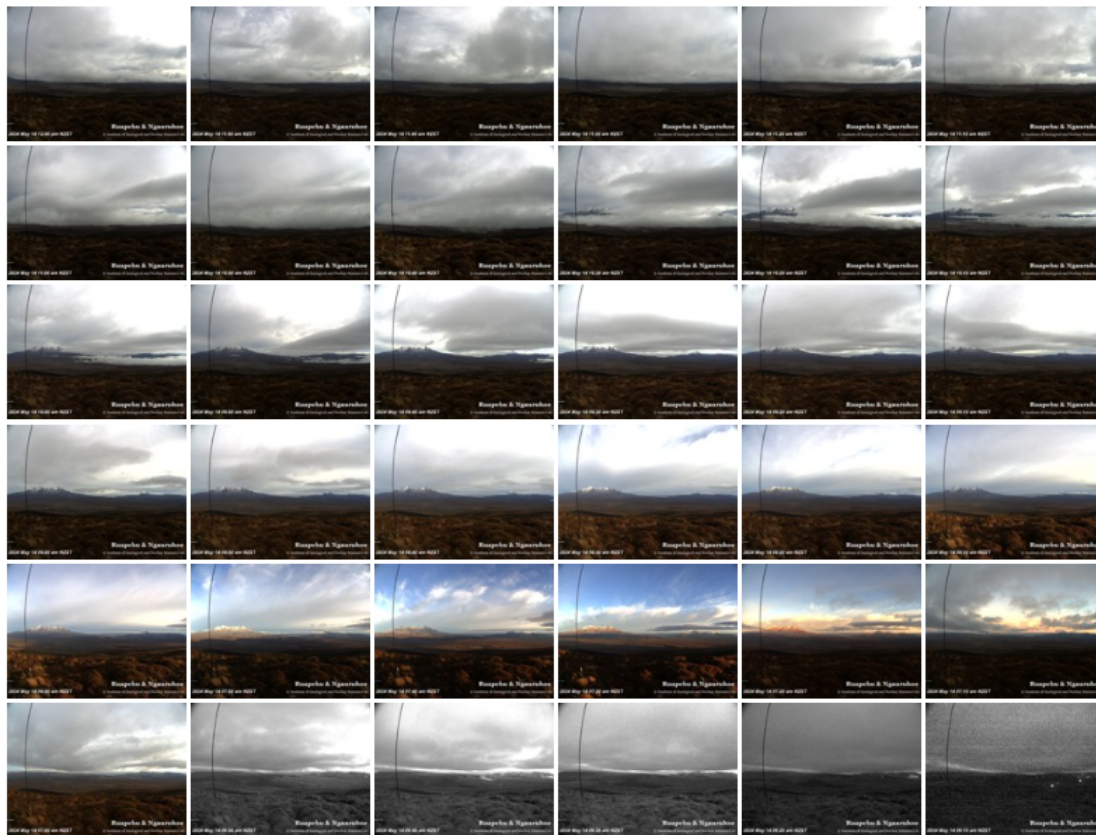
### 5.3 Geração das Amostras

Na literatura sobre visão computacional, o emprego de bases de dados públicas para experimentação e validação de modelos é uma prática recorrente. Nesse contexto, realizou-se uma pesquisa extensiva sobre fontes de domínio público de imagens de danos em estruturas geotécnicas. Na pesquisa, foram identificadas três bases: o Instituto Americano de Geologia (AMERICAN... , 2024), o Geonet (GEONET, 2022) e Departamento de Minas e Geologia de Nevada (NBMG, 2022b). Durante a análise dessas bases de dados, foram observadas algumas limitações em relação às imagens disponibilizadas.

O *Earth Science World Image Bank* é um serviço comunitário prestado pelo Instituto Americano de Geociências para fornecer imagens publicamente para educadores, pesquisadores em geociências e o público em geral. A página indica que oferece dados de vários países, em todos os continentes. Porém, a pesquisa na base os dados não é intuitiva e há dificuldades de acessar às imagens.

Já o Geonet é um serviço de monitoramento de perigos geológicos na Nova Zelândia. A infraestrutura de monitoramento compreende uma rede de instrumentos geofísicos (sensores e câmeras), aplicações de software automatizadas e profissionais para detectar, analisar e atuar frente a terremotos, atividade vulcânica, grandes deslizamentos de terra e tsunamis (GEONET, 2022). Os dados não estão em sua totalidade disponíveis na página do Geonet, mas podem ser solicitadas para fins de pesquisa. As imagens disponibilizadas são de câmeras fixas distribuídas em regiões consideradas estratégicas para o monitoramento, com destaque ao monitoramento de atividade vulcânica. As limitações observadas nas imagens para a aplicação neste estudo incluem a baixa resolução devido a distância das câmeras em relação à superfície e a forte interferência de condições climáticas. Todas essas limitações inviabilizam o uso dessas imagens para representar a detecção de danos aparentes em superfícies de taludes, como pode ser constatado pelas imagens na Figura 18.

Figura 18. Imagens de monitoramento de vulcões pelo GeoNet.



Fonte: (GEONET, 2022).

O Departamento de Minas e Geologia de Nevada, do inglês *Nevada Bureau of Mines and Geology* (NBMG) é uma unidade de pesquisa e serviço público estadual da Universidade de Nevada. Os cientistas da NBMG realizam pesquisas e publicam relatórios sobre recursos minerais, geologia de engenharia, geologia ambiental, hidrogeologia e mapeamento geológico. Os projetos de geologia de engenharia, geologia ambiental e hidrogeologia incluem investigações de riscos de terremotos e aspectos relacionados da neotectônica, exame de questões geológicas envolvidas no descarte de resíduos nucleares e perigosos, mapeamento de características geomórficas, e estudos de risco (NBMG, 2022b). As imagens disponibilizadas publicamente pouco contribuem para representar danos em estruturas geotécnicas, como mostra a Figura 19. A pasta intitulada "Perigos Geológicos", por exemplo, inclui apenas 13 imagens (NBMG, 2022a).

Figura 19. Perigos geológicos avaliados pela NBMG.



Fonte: (NBMG, 2022a).

A análise inicial das imagens pesquisadas nas bases mencionadas revelou que estas eram pouco representativas dos danos superficiais em taludes, tanto em termos de quantidade quanto de qualidade. Os deslizamentos de terra e erosões presentes nas imagens abrangiam diferentes categorias de estruturas geotécnicas, como vulcões e água subterrânea, que não se encontram no escopo desta tese. Em outras palavras, as imagens pouco ou nada contribuíam para a descrição do problema de danos aparentes em taludes. A construção de uma base de dados representativa é fundamental para que o modelo possa representar adequadamente o fenômeno em questão (KATTENBORN et al., 2021).

Ademais, a maioria das imagens presentes nessas bases são oriundas de satélites, o que gera diversos obstáculos à sua utilização. Um desses obstáculos está relacionado ao eventual custo para o acesso às imagens. Usuários precisam se inscrever ou comprar o acesso (NBMG, 2022b), o que pode inviabilizar economicamente a utilização desse tipo de fonte de dados.

Além disso, devido à falta de controle do usuário sobre a aquisição das imagens pelos satélites, as condições climáticas, como nuvens, neblina e precipitação, podem afetar significativamente a imagem e mascarar falhas, como pode ser observado na Figura 20, extraída da base de imagens disponibilizada pela (NASA,



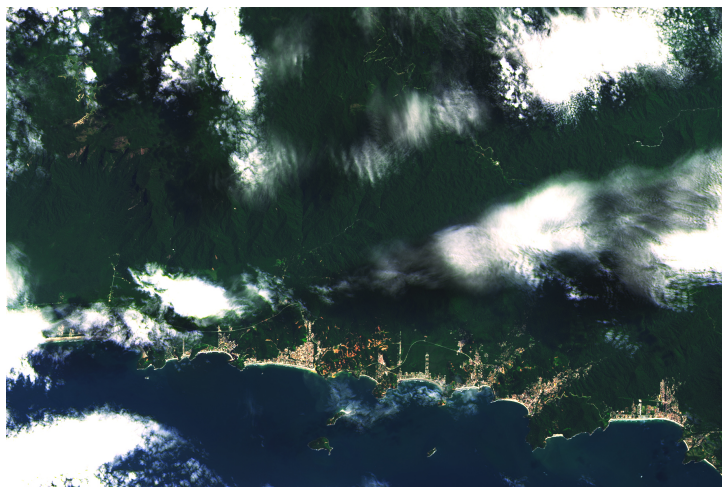
2022). Essa situação é agravada pela resolução limitada das imagens, que nem sempre são capazes de capturar as especificidades do solo devido a fatores como a distância da órbita terrestre (ZHONG et al., 2020; SOLÓRZANO et al., 2022). O que pode ser visto na Figura 21, que revela timidamente nas imagens o deslizamento generalizado que ocorreu em São Sebastião, estado de São Paulo, Brasil, no ano de 2023.

A pesquisa por imagens em plataformas de busca como o Google por exemplo, também foi explorada na tentativa de se obter um conjunto de dados para a pesquisa. No entanto, essa abordagem demandou um esforço manual considerável para encontrar imagens condizentes com a problemática de engenharia em questão. Adicionalmente, questões relacionadas a direitos autorais de imagem inviabilizaram a utilização de diversas imagens obtidas por esse método.

Apesar do crescimento dos dados de monitoramento geotécnico ao longo do tempo, sua quantidade ainda é considerada insuficiente em comparação com outros campos, conforme evidenciado por (PHOON; ZHANG, 2023; ZHANG; LI; LI, 2021). Diante das limitações das bases públicas de dados, que resultaram em escassez de amostras adequadas para a construção do modelo, optou-se pela criação de um conjunto de dados autêntico, derivado de uma aplicação do mundo real. As bases públicas pesquisadas apresentaram baixa relevância para o presente estudo, o que justifica a escolha por essa abordagem.

Para a construção do banco de imagens foram empregados um VANT modelo *Dji Phantom 3 Professional* e um celular *smarthphone* modelo *Samsung S20FE*. A altura de vôo do VANT em relação à superfície foi de aproximadamente 10 metros, controlada manualmente, e câmera com inclinação de 45°. As fotografias adquiridas pelo *smarthphone* foram tiradas em diversos ângulos e distâncias. A estratégia de aquisição das imagens proporcionou diversos benefícios, dos quais se destaca o controle sobre a captura das imagens ( dia, hora, local, altura em relação à superfície) e a obtenção de imagens de alta precisão. Vantagens também relatadas nos estudos de (SPENCER; HOSKERE; NARAZAKI, 2019; GREENWOOD et al., 2019; KANELLAKIS; NIKOLAKOPOULOS, 2017; LEI et al., 2020; LIM et al., 2021).

Figura 20. Efeito do clima nas imagens de satélite, São Sebastião- São Paulo, 2022.



Fonte: (NUSSBAUM, 2022).

Figura 21. Imagens de satélite de grandes deslizamentos, São Sebastião- São Paulo, 2023.



Fonte: (NUSSBAUM, 2023).

As imagens foram obtidas em inspeção em taludes localizados às margens da rodovia BR 381, no trecho entre Belo Horizonte e João Monlevade, durante o

período de janeiro a maio de 2022. Vale salientar que este período coincidiu com o registro do maior número de deslizamentos de terra nos últimos 23 anos em Minas Gerais, Brasil, conforme dados de (MACEDO; SANDRE, 2022).

Foram capturados 8 gigabytes de dados em imagens (vídeos e fotos). Dentre esses dados, foram selecionadas manualmente, com o auxílio de um especialista, 337 imagens digitais tipo RGB consideradas representativas para a modelagem de danos aparentes em taludes. Essas imagens foram disponibilizadas publicamente na plataforma de compartilhamento Figshare e podem ser acessadas em (ARAUJO, 2023).

O conjunto de dados utilizado neste estudo, composto por 337 imagens, pode ser considerado limitado em comparação com bases de dados massivas utilizadas em modelos CNN de última geração, que frequentemente apresentam milhares ou até milhões de imagens (ABEDALLA et al., 2021; AFFONSO et al., 2017; MOUNTRAKIS; HEYDARI, 2023; WANG et al., 2020; CHA; CHOI; BüyüKöZTürk, 2017; HAN et al., 2022; COFRE-MARTEL et al., 2019). No entanto, é importante ressaltar que o tamanho ideal da base de dados depende de diversos fatores, como a complexidade do problema, a natureza dos dados, os recursos disponíveis e os objetivos da aplicação.

Para além da limitação no volume de imagens, constatou-se um desequilíbrio no número de exemplos por classe. O fenômeno de danos visuais em taludes apresenta natureza multiclasse, que podem ser categorizadas em diversas classes. Para este estudo foram consideradas como danos aparentes as falhas cujo solo exposto é comum, como nas categorias de: deslizamentos, erosão, desvegetação e a presença de tocas de animais, como formigueiros e cupinzeiros. Se o modelo aprender apenas a partir de um número limitado de exemplos de uma determinada classe, há menor probabilidade de que ele realize previsões corretas em dados não vistos (SHORTEN; KHOSHGOFTAAR, 2019; LU; TAN; JIANG, 2021). Visando simplificar a arquitetura da rede, optou-se por uma estratégia de classificação binária para mitigar problemas relacionados ao desbalanceamento de classes. Essa decisão se justifica pelo número significativamente maior de imagens com superfície aparentemente saudável em comparação com as categorias de danos.

Além da simplificação da estrutura do modelo, adotou-se estratégias de regularização para evitar problemas associados a pequenos volumes de conjuntos de treinamento em Redes Neurais Convolucionais, como baixa precisão e o sobreajuste.

Deste modo, as imagens anotadas como danificadas apresentam como fator comum a ausência de proteção do solo. Diante disso, as imagens foram manualmente anotadas em duas classes: "damage", com danos visíveis, e "intact" com superfície aparentemente saudável. Exemplos dessas categorias incluem imagens com deslizamentos de terra, desvegetação e erosão (dados negativos), enquanto as imagens sem esses danos configuram os dados positivos. As Figuras 22 e 23 ilustram essa definição.

## 5.4 Aumento de Dados

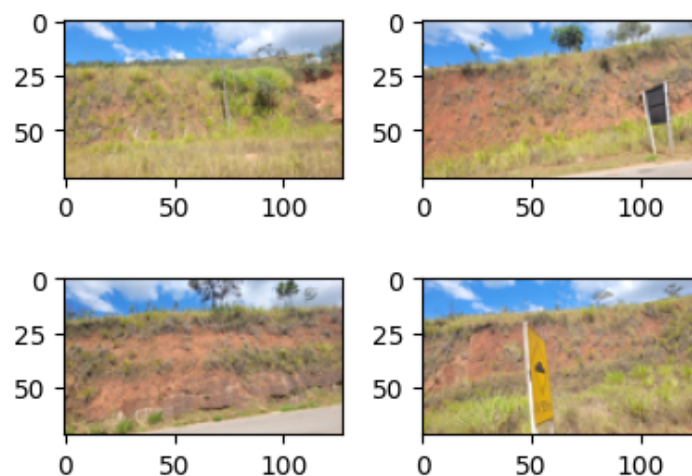
Para o treinamento de um classificador CNN, imagens brutas de superfícies de taludes danificadas foram coletadas por meio de câmeras e dispositivos diversos, incluindo celular e um VANT. As imagens abrangeram exemplos com ampla gama de variações, como iluminação e sombras.

No entanto, em virtude das limitações inerentes ao problema de engenharia geotécnica, o conjunto de dados de treinamento foi submetido a técnicas de aumento de dados. O objetivo do aumento de dados foi expandir e diversificar as imagens no conjunto de treinamento, visando aprimorar o desempenho do modelo. Para tal, o número de imagens foi duplicado, de 337 para 674 imagens. As propriedades geométricas de inversão, rotação e redimensionamento foram aplicadas às imagens de treinamento. Além disso, o filtro de suavização Gaussiano foi empregado como estratégia de composição pictórica.

## 5.5 Divisão das Amostras

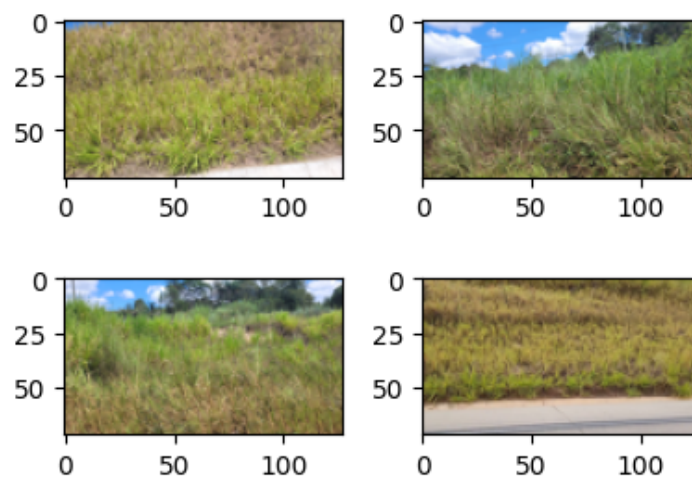
Após a duplicação das imagens, as amostras foram divididas aleatoriamente em conjuntos de treinamento, validação e teste, conforme apresentado na Tabela 2.

Figura 22. Exemplos de imagens anotadas como "com dano".



Fonte: Autora.

Figura 23. Exemplos de imagens anotadas como "intacto".



Fonte: Autora.

## 5.6 Arquitetura Proposta

A definição de modelos de Redes Neurais Convolucionais ainda é um processo empírico. Em decorrência disso, não existem regras precisas para a otimização de hiperparâmetros, o que dificulta a configuração e a seleção de valores adequados, como taxas de aprendizado e parâmetros de regularização. Nesse contexto, a seleção

Tabela 2. Divisão das amostras em treinamento, validação e teste.

Classe	Treinamento		Validação		Teste	
	Imagens	(%)	Imagens	(%)	Imagens	(%)
Dano	312	46,29	53	7,86	53	7,86
Intacto	192	28,49	32	4,75	32	4,75
Total	504	74,78	85	12,61	85	12,61

Fonte: Autora.

cuidadosa da largura da rede, número de filtros nas camadas de convolução e profundidade da rede desempenha um papel vital no projeto de CNN para obter um melhor desempenho (BASHA et al., 2020). No que se refere à profundidade do modelo CNN, à medida que a complexidade e o tamanho das redes aumentam (profundidade, largura e resolução), recursos computacionais consideráveis são necessários para treinamento e validação, pois o número de parâmetros também aumenta consideravelmente (DARAPANENI; KRISHNAMURTHY; PADURI, 2020).

No estado da arte, as arquiteturas mais profundas tem sido preferidas em relação às arquiteturas rasas para treinar o modelo CNN usando conjuntos de dados mais profundos (BASHA et al., 2020). Isto é, quando os conjuntos de dados são mais profundos, possivelmente mais complexos ou detalhados, as arquiteturas de rede mais profundas tendem a produzir resultados melhores (LECUN; BENGIO; HINTON, 2015). No entanto, há uma inversão dessa tendência quando se trata de conjuntos de dados mais amplos, possivelmente com menos complexidade ou detalhamento. Nestes casos, as arquiteturas de redes mais rasas, ou menos complexa em termos de quantidade de camadas de convolução, neurônios ou resolução, mostraram um desempenho superior em comparação com os modelos CNN mais profundos (BASHA et al., 2020).

Nesta perspectiva, os experimentos foram realizados com o objetivo de desenvolver uma arquitetura CNN personalizada e de baixa complexidade, buscando o equilíbrio entre um número mínimo de camadas e um desempenho satisfatório. Estudos demonstram que modelos CNN de baixa complexidade podem apresentar desempenho comparável ou superior a arquiteturas de última geração em cenários com conjuntos de dados de treinamento limitados (BRIGATO; IOCCHI, 2021;

FOROUGH; CHEN; WANG, 2021; LIMÃO; ARAÚJO; FRANCES, 2023), como é o caso do problema da engenharia geotécnica abordado nesta tese.

Os parâmetros de simulação das arquiteturas CNN foram divididos em três grupos principais, conforme sugerido por (MOUNTRAKIS; HEYDARI, 2023). O primeiro grupo considera as características de entrada, seguido pela estrutura da rede (número de camadas, neurônios por camada, etc.) e, por fim, pelas configurações de otimização da rede (tamanho do lote de treinamento, tipo de otimizador, taxa de aprendizado, etc.).

Na camada de entrada de uma rede, a preparação da imagem para o processamento foi realizada. A camada de entrada foi configurada com uma resolução de  $128 \times 128 \times 3$  *pixels*. Cada dimensão representa, respectivamente, a altura, a largura e o número de canais de cor da imagem. Os canais de cor contêm as intensidades de cada cor primária (vermelho, verde e azul) em uma escala de 0 a 255, onde 0 indica a ausência total da cor e 255 a intensidade máxima.

No que concerne à estrutura da rede, foi adotada uma padronização das camadas, composta por: camada de entrada, convolução, ativação, agrupamento, abandono, linearização e saída. São três blocos de convolução que compreendem uma sequência de camadas de convolução, ativação, agrupamento e abandono.

Foram realizadas execuções com arquiteturas compostas por 3 blocos de convolução. Modelos de convolução com dois blocos de convolução apresentaram resultados insatisfatórios em termos de precisão e perdas em cenários similares ao abordado nesta tese, conforme demonstrado por (LIMÃO; ARAÚJO; FRANCES, 2023).

O número de filtros foi mantido fixo em 32 ou 64, com tamanho de núcleo  $3 \times 3$ , para todas as camadas de convolução. No estado da arte é comum encontrar núcleos de tamanho  $2 \times 2$ ,  $3 \times 3$ ,  $5 \times 5$  (FOROUGH; CHEN; WANG, 2021; LU; TAN; JIANG, 2021; AKHTAR; RAGAVENDRAN, 2020; LIU et al., 2020), bem como 32 e 64 filtros nas camadas de convolução (FOROUGH; CHEN; WANG, 2021; BUSHARA; KUMAR, 2022; BASHA et al., 2020). Na camada totalmente conectada, foram testadas configurações com uma e 2 camadas intermediárias, contendo 64 ou 128 neurônios, respectivamente, e uma camada de saída com um neurônio.

Na camada de agrupamento, a técnica de agrupamento máximo foi empregada por apresentar desempenho satisfatório para classificação de imagens, quando comparada a outras técnicas, como o agrupamento médio (SCHERER; MÜLLER; BEHNKE, 2010; LU; TAN; JIANG, 2021). Após as camadas de convolução e agrupamento, foi aplicada uma camada de abandono com taxa fixa de 0,25 em todas os blocos de convolução. Valores entre 0,2 e 0,4 para a taxa de abandono foram considerados satisfatórios em aplicações de sensoriamento remoto (MOUNTRAKIS; HEYDARI, 2023).

A função de ativação ReLu foi empregada após cada bloco de convolução, por ser avaliada eficiente em modelos CNN (LECUN; BENGIO; HINTON, 2015), conforme descrito no Capítulo 3. Para a camada de saída do modelo empregou-se a função de ativação Sigmóide, uma vez que esse tipo de função é capaz de mapear números reais para valores entre 0 e 1, tornando-a útil em tarefas de classificação binária (WANG et al., 2020). Com essa função de ativação na camada de saída do modelo tornou-se possível prever se cada dado de entrada é uma superfície danificada ou intacta, após os processos de convolução. Depois de passar pela ativação Sigmóide na última camada, a CNN gera um número entre 0 (0%) e 1 (100%), que indica a probabilidade de pertencimento da imagem a uma das classes.

Para a otimização do aprendizado, recorreu-se ao otimizador Adam. Este otimizador é amplamente explorado em cenários de aprendizagem profunda, como é o caso dos modelos CNN. A capacidade do otimizador de Adam para adaptar suas taxas de aprendizagem aborda o desafio do desaparecimento de gradiente (LI et al., 2022), que é frequentemente observado em arquiteturas CNN de baixa complexidade.

O principal critério para a escolha da arquitetura foram os valores médios de acurácia e perda, de cinco execuções de cada arranjo, como realizado nos experimentos por (MHASKAR; LIAO; POGGIO, 2016). Esse procedimento objetivou assegurar uma comparação de desempenho justa entre as arquiteturas e minimizar o impacto de fatores aleatórios, como inicialização de peso e variabilidade de treinamento, típico em modelos de Redes Neurais.

Portanto, fixado o número de camadas convolucionais em 3, os experimentos



foram realizados variando-se:

- Número de filtros em cada camada de convolução: 32 ou 64 filtros.
- Número de camadas intermediárias densa ou totalmente conectada, do inglês *Fully Connected* (FC) : 1 ou 2 camadas.
- Número de neurônios em cada camada totalmente conectada FC: 64 ou 128.

Os resultados apresentados para as arquiteturas CNN experimentadas são demonstrados na Tabela 3.

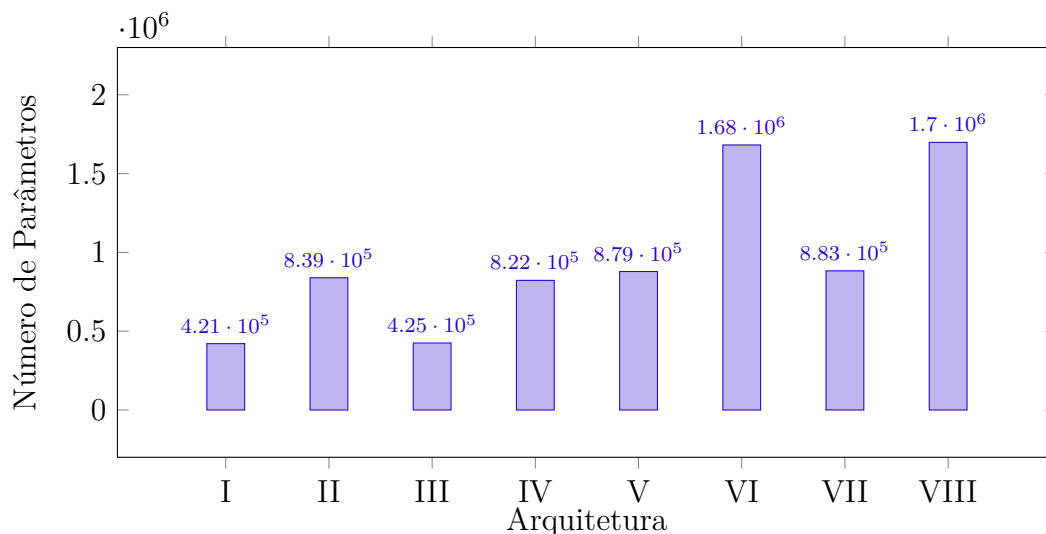
Tabela 3. Valores médios de acurácia e perda das arquiteturas executadas.

Arquitetura	Treinamento		Validação		Teste	
	Acurácia (%)	Perda ( $\times 10^{-2}$ )	Acurácia (%)	Perda ( $\times 10^{-2}$ )	Acurácia (%)	Perda ( $\times 10^{-2}$ )
I: 32 filtros, 1 FC 64 neurônios	<b>94,65</b>	15,87	90,00	33,65	84,71	44,56
II: 32 filtros, 1 FC 128 neurônios	94,26	<b>15,17</b>	90,00	22,88	87,06	53,69
III: 32 filtros, 2 FC 64 neurônios	88,71	26,96	90,00	26,41	80,00	58,95
IV: 32 filtros, 2 FC 128 neurônios	94,26	15,62	<b>94,00</b>	24,72	<b>92,94</b>	<b>26,38</b>
V: 64 filtros, 1 FC 64 neurônios	61,98	66,47	66,47	66,31	62,35	66,31
VI: 64 filtros, 1 FC 128 neurônios	94,06	15,57	<b>94,00</b>	20,36	<b>90,59</b>	<b>25,86</b>
VII: 64 filtros, 2 FC 64 neurônios	<b>95,45</b>	<b>11,66</b>	90,00	25,11	89,41	52,80
VIII: 64 filtros, 2 FC 128 neurônios	86,53	33,56	84,00	40,77	76,47	47,99

Fonte: Autora.

Como esperado, a taxa de sucesso da CNN em termos de acurácia no treinamento, validação e teste foi semelhante na maioria das combinações. Em negrito, destaca-se os dois melhores valores de cada métrica ( maior acurácia, menor perda) para cada conjunto de amostras ( treinamento, validação e testes), afim de auxiliar na identificação do arranjo de camadas com o melhor desempenho médio. Em termos de acurácia e perda, a combinação de camadas intermediárias (IV) composta de 32 filtros, 2 camadas densas intermediárias de 128 neurônios, e a combinação (VI) de 64 filtros e 1 camada densa intermediária de 128 neurônios alcançaram desempenho similar.

Figura 24. Número de parâmetros treináveis por arquitetura.



Fonte: Autora.

Considerando desempenho satisfatório em termos de profundidade e largura de rede na maioria das combinações, o critério para escolha da rede baseou-se no critério de uma arquitetura mais leve, uma vez que busca-se viabilizar o processamento por ambientes com limitação de recursos computacionais. Portanto, levantou-se o número de parâmetros treináveis de cada combinação. O número de parâmetros é capaz de refletir a profundidade de uma rede e está relacionado à demanda de recurso computacional para processamento. Arquiteturas muito profundas, possuem mais parâmetros e obviamente requerem mais capacidade de processamento. Conforme demonstrado na Figura 24 a arquitetura (IV) apresenta menos parâmetros em relação à (VI).

Os resultados expressos pela Figura 24 demonstram que a largura da rede, representada pela quantidade de neurônios na camada densa FC, aumenta significativamente o número de parâmetros treináveis. O número de parâmetros aproximadamente duplica da arquitetura (II) em relação a (I), (IV) em relação a (III), (VI) em relação à (V) e (VIII) em relação à (VII). Em relação à profundidade, o número de camadas densas pouco impactou em arquiteturas com a mesma quantidade de neurônios. Como pode ser constatado pelas arquiteturas (I) e (III),

(V) e (VII), de 1 camada densa. E nas arquiteturas (II) e (IV), e (VI) e (VIII), com 2 camadas densas.

Diante da necessidade de otimizar o uso dos recursos computacionais disponíveis e minimizar o risco de sobreajuste do modelo, optou-se pela arquitetura (IV) em detrimento da arquitetura (VI), que apresentava um número maior de parâmetros treináveis. Essa escolha é coerente com as recomendações de (BRIGATO; IOCCHI, 2021; FOROUGH; CHEN; WANG, 2021), que alertam para os perigos do sobreajuste em modelos treinados com escassez de amostras.

Portanto, a arquitetura CNN detecção de danos geotécnicos em imagens pode ser resumida como:

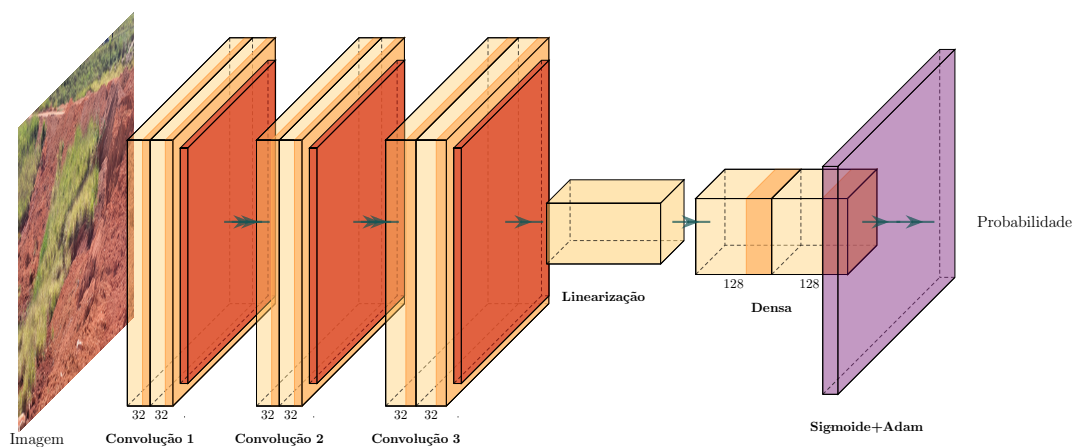
- Camadas convolucionais: 3 camadas convolucionais; 32 filtros; agrupamento máximo; abandono de 0,25; Função de ativação ReLu;
- Camadas totalmente conectadas: 2 camadas totalmente conectadas de 128 neurônios; Função de ativação ReLu ; 1 camada totalmente conectada de 1 neurônio; Função de ativação sigmoide; Algoritmo de otimização Adam.

A Figura 25 apresenta um resumo visual da estrutura da arquitetura proposta, detalhando a disposição e as características de cada camada.

## Considerações Finais

O presente capítulo abordou a tarefa de detecção de danos em imagens de taludes, desde a coleta e pré-processamento dos dados até a escolha da arquitetura de rede neural convolucional mais adequada para o problema. Na etapa de coleta de dados, foi utilizado drone (VANT) e *smartphones* para capturar imagens de diferentes taludes em diversas condições. As imagens coletadas foram pré-processadas por meio de aumento de dados e normalização dos dados, com a finalidade de garantir a qualidade e a consistência do conjunto de dados. Em seguida, foram exploradas diversas arquiteturas de CNNs para a tarefa de classificação de imagens com danos, com ênfase em modelos enxutos em termos de quantidade de camadas. As redes foram treinadas e avaliadas utilizando métricas como acurácia e perda.

Figura 25. Arquitetura proposta.



Fonte: Adaptado de (IQBAL, 2018).

Após uma análise criteriosa dos resultados, a arquitetura CNN escolhida foi a 3 camadas convolucionais (convolução, agrupamento, ativação e abandono), que apresentou o melhor desempenho geral na tarefa de classificação de acordo com as métricas de acurácia e perda.

## 6 Resultados e Discussão

### Considerações Iniciais

A inspeção remota e a detecção de danos em estruturas geotécnicas a partir de falhas visíveis em imagens de superfícies é um desafio para a segurança e o gerenciamento das infraestruturas. Este capítulo apresenta uma solução para este problema utilizando aquisição de dados por tecnologias emergentes, como VANT e celulares, e modelos de Redes Neurais Convolucionais de baixa complexidade.

É importante destacar que a complexidade de um modelo de CNN é um conceito amplo que engloba diversos fatores, como o tamanho da entrada, a utilização de modelos pré-treinados e técnicas de transferência de aprendizagem. Neste trabalho, a complexidade é abordada como a arquitetura de rede que busca um equilíbrio entre um número mínimo de camadas (menor profundidade) e a capacidade de superar os desafios típicos da análise de falhas em imagens de superfícies geotécnicas. Neste contexto, a severa restrição de recursos computacionais do ambiente operacional dos experimentos da tese também é considerada.

- a) Arquitetura personalizada: Propõe-se uma arquitetura CNN enxuta, projetada para lidar com um problema de classificação de imagem em um cenário de dados escassos sem comprometer o desempenho do modelo.
- b) Combate ao desbalanceamento de classes: Para lidar com o desbalanceamento de classes, modela-se o problema multiclasse como um classificador binário. Assim transforma-se o problema em uma tarefa mais simples, permitindo uma arquitetura mais enxuta.
- c) Estratégias para mitigar o sobreajuste: Empregam-se diversas técnicas para combater o sobreajuste, frequente em modelos treinados com escassez de amostras, incluindo a redução de camadas de convolução e a utilização de técnicas de regularização e otimização.

- d) Análise de resultados e métricas de desempenho: Apresenta-se uma análise dos resultados obtidos, com foco nas métricas de desempenho que avaliam a eficácia do modelo na identificação de falhas.

## 6.1 Desempenho do Modelo

Para avaliar o desempenho do modelo proposto na detecção de danos em imagens de estruturas geotécnicas, foi utilizada uma abordagem de validação que dividiu o conjunto de dados de treinamento em três subconjuntos: conjunto de treinamento (para treinar o modelo) e conjunto de validação (para avaliar o desempenho do modelo em dados não vistos durante o treinamento) e conjunto de testes (para visualizar a saída do modelo).

A arquitetura específica da Rede Neural Convolutacional selecionada para testes do modelo foi composta pelas camadas de entrada, três camadas de convolucionais, duas camadas intermediárias FC e uma camada de saída. A Tabela 4 detalha individualmente a configuração empregada em cada uma das camadas:

Reconhecendo a importância da robustez dos resultados, foi adotado um processo que incluiu um mínimo de cinco execuções para cada arquitetura de rede e seus respectivos hiperparâmetros. Essa abordagem multirreplicada objetivou garantir que o desempenho observado não seja um mero acaso, mas sim um reflexo consistente da capacidade da arquitetura proposta. A combinação selecionada apresentou o melhor desempenho médio em termos de acurácia e perda durante o treinamento e validação.

A média de múltiplas execuções da arquitetura encontrada para acurácia (*accuracy*) alcançou patamares de 94,26% para treinamento e 92% para validação, respectivamente, como demonstra a Figura 26. Esse resultado foi obtido em um tempo relativamente curto, com menos de 50 épocas (*epochs*).

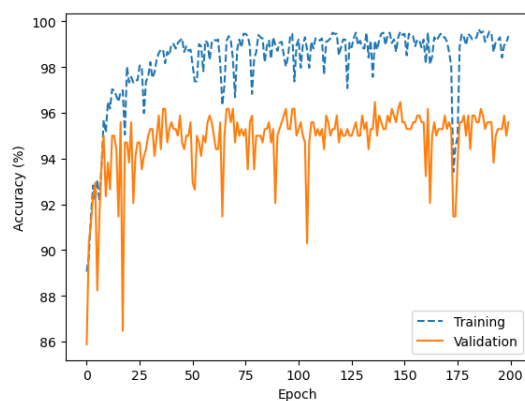
O treinamento com lotes menores em um conjunto de dados pequeno contribui para a generalização do modelo, apesar de aumentar sua sensibilidade, evidenciada pelas oscilações nas curvas de aprendizagem. Lotes menores expõem a rede a menos exemplos por atualização, aumentando a frequência de atualização

Tabela 4. Informações das camadas do modelo CNN.

Camadas	Hiperparâmetros	
Entrada	Tamanho da imagem	128x128
	Numero de canais	3
Convolutacional 1	Número de filtros	32
	Tamanho do filtro	3x3
	Função de ativação	ReLu
	Agrupamento	Máximo 2x2
Convolutacional 2	Número de filtros	32
	Tamanho do filtro	3x3
	Função de ativação	ReLu
	Agrupamento	Máximo 2x2
Convolutacional 3	Número de filtros	32
	Tamanho do filtro	3x3
	Função de ativação	ReLu
	Agrupamento	Máximo 2x2
Densa 1	Número de neurônios	128
	Função de ativação	ReLu
Densa 2	Número de neurônios	128
	Função de ativação	ReLu
Densa Saída	Número de neurônios	1
	Função de ativação	Sigmóide

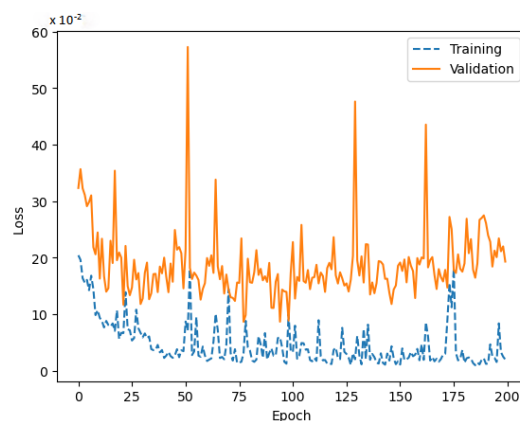
Fonte: Autora.

Figura 26. Acurácia média de cinco execuções.



Fonte: Autora.

Figura 27. Perda média de cinco execuções.



Fonte: Autora.

com novos exemplos, o que promove a aprendizagem de padrões mais robustos.

As oscilações nas curvas de aprendizagem são um indicador natural dessa maior sensibilidade.

A convergência das curvas de acurácia e perda de treinamento e validação demonstra que o modelo não está memorizando os dados de treinamento em excesso e generaliza bem para novos dados. No sobreajuste, as curvas continuariam subindo indefinidamente, de forma distorcida, ao longo do treinamento. A ausência de distorção entre as curvas, sugere a efetividade das técnicas de regularização na prevenção do sobreajuste e na promoção da generalização.

A acurácia é uma métrica que mensura o desempenho da rede baseada em exemplos classificados corretamente, independente da classe. Em um cenário de escassez de amostras e desequilíbrio de classes, valores de acurácia elevados podem desconsiderar o privilégio que as classes majoritárias podem exercer durante o treinamento. Assim, nem sempre uma alta acurácia representa a capacidade do modelo de generalizar os resultados para todas as classes. No caso do problema em questão, a acurácia é útil para avaliar a possibilidade de sobreajuste, evidenciada pelas tendências das curvas, e o tempo médio gasto na convergência durante o treinamento. No entanto, para assegurar a robustez da solução, foram empregadas outras formas de avaliação de desempenho.

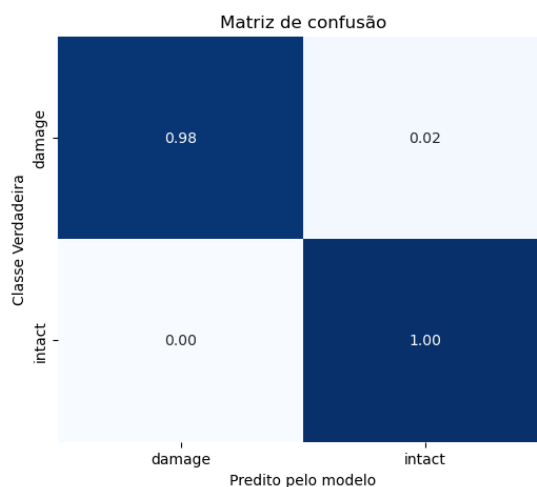
Portanto, a avaliação por meio de outras métricas, como índices AUC e AUPRC em um conjunto de teste independente foi realizada para avaliar a influencia do desequilíbrio de classes no desempenho do modelo. Essa abordagem é fundamental para verificar se o modelo generaliza bem para dados não vistos durante o treinamento e para tanto considera as relações estabelecidas entre as predições e as etiquetas reais das imagens.

Para sintetizar os resultados de predição do modelo com dados do conjunto de testes foi construída a matriz de confusão. A matriz relaciona a previsão (predito pelo modelo) pelo modelo e que representa a probabilidade saída da rede, com as classes verdadeiras da imagem anotada. O quantitativo de instâncias verdadeiras e falsas positivas e negativas obtido é demonstrado pela Figura 28.

Os resultados expressos na matriz indicam que todas as imagens pertencentes à classe "intacto", foram precisamente classificadas com a classe correta. Isso



Figura 28. Matriz de confusão.



Fonte: Autora.

demonstra a capacidade do modelo em distinguir com clareza imagens que não apresentam danos. Além disso, os resultados apresentados pela matriz também sugerem uma confusão do modelo ao identificar corretamente as imagens com algum dano aparente. Embora baixa, essa probabilidade de confusão indica que o modelo pode classificar erroneamente uma região com dano aparente como saudável, negligenciando possíveis falhas.

No entanto, não se pode atribuir esses resultados da matriz ao desequilíbrio de classes, pois a classe majoritária no modelo é a negativa ("com dano"). Em problemas desbalanceados, os modelos inadequados tendem a priorizar a classe majoritária. Portanto, em um modelo inadequado, a probabilidade de classificar uma superfície intacta como danificada deveria ser maior do que a de classificar uma superfície danificada como intacta.

Uma hipótese examinada nesta tese é a capacidade do modelo de apresentar bom desempenho em cenários com escassez de amostras e desequilíbrio de classes. Embora a acurácia e a matriz de confusão indiquem resultados satisfatórios, é importante ressaltar que o problema de detecção foi modelado por um classificador binário. Nesse modelo, o número de amostras de imagens com superfícies aparentemente danificadas ( classe majoritária) se sobrepõe ao número de imagens saudáveis

( classe minoritária). Ao reduzir um fenômeno multiclasse a um modelo binário, agrupando as imagens com dano visível a olho nu, obteve-se um quantitativo maior de imagens na classe negativa ("com dano") em relação à classe positiva ("saudável"). Nesses casos de desequilíbrio de classes, as métricas tradicionais podem ser enganosas, pois o modelo pode estar favorecendo a classe com "com dano".

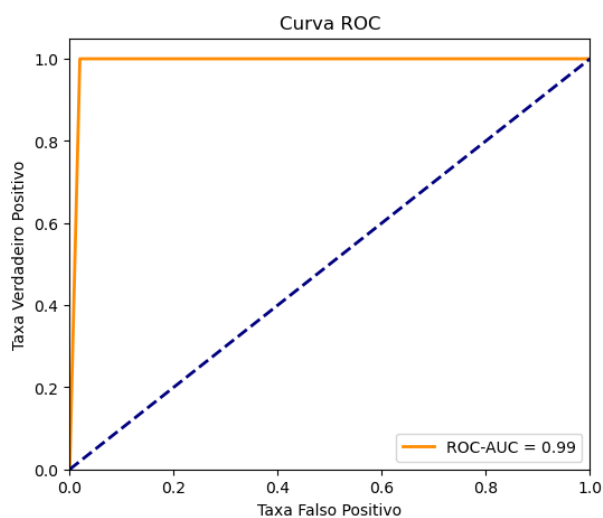
É válido destacar que o valor de limiar (*threshold*), também chamado de ponto de corte, define a fronteira entre as classes. Valores próximos de 0 indicam maior probabilidade de pertencimento à classe negativa ("com dano"), enquanto valores próximos de 1 indicam maior probabilidade de pertencimento à classe positiva ("intacto"). As previsões da rede neural, que variam entre 0 e 1, precisam ser categorizadas como uma das classes do modelo. Na classificação binária, a saída é um número entre 0 e 1 que representa a probabilidade de pertencimento a uma das duas classes, "com dano" ou "intacto". Ao se variar esse limiar, são produzidos diferentes valores de VP, VN, FP e FN, que podem não ser percebidos em sua totalidade pela acurácia.

Por outro lado, quando se relaciona a Taxa de Verdadeiros Positivos (TVP) e Taxa de Falsos Positivos (TFP) obtém-se uma curva, denominada ROC, que possibilita visualizar a qualidade do modelo por meio da relação entre positivos reais e falsos. Essa relação é descrita pela curva ROC-AUC na Figura 29.

A curva ROC-AUC demonstra que, quando todos os exemplos pertencem à classe positiva (classe minoritária, "intacto"), a TVP tende a 1, enquanto a TFP tende a 0. Conseqüentemente, obtém-se um AUC de 0,99, o que indica a alta capacidade do modelo de distinguir entre imagens de superfície saudável e com danos aparentes. Um valor de AUC próximo a 1 evidencia a efetividade do modelo em diferenciar as classes de forma satisfatória. Isso significa que o modelo é capaz de distinguir com alta precisão os exemplos positivos (classe "intacto") dos exemplos negativos (danos causados por falta de cobertura vegetal, deslizamentos de terra e erosão).

Entretanto, AUC não é uma métrica não é capaz de captar a interferência do desbalanço de classes no modelo. Para avaliar o impacto do desbalanço no desempenho, foi traçada a curva PRC. O índice AUPRC obtido a partir dessa

Figura 29. Curva ROC-AUC



Fonte: Autora.

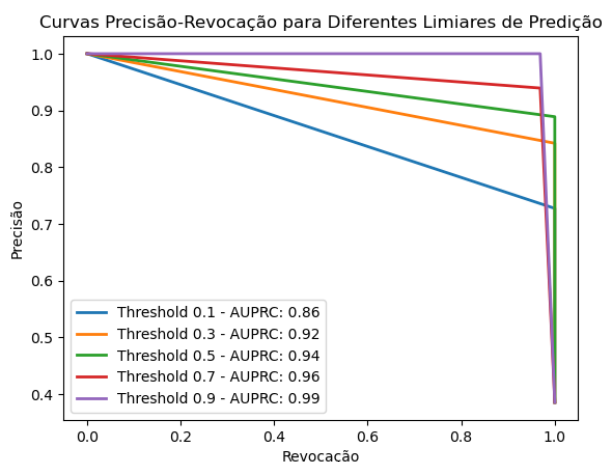
curva auxilia na interpretação do desempenho do modelo.

A métrica *Area Under the Precision-Recall Curve* (AUPRC) é interessante pois mede o desempenho do modelo na recuperação de amostras relevantes, com ênfase na classe minoritária. Esse desempenho pode ser verificado com a alteração dos limiares de classificação, uma vez que os resultados de precisão e revocação avaliam principalmente a ação da classe positiva (minoritária, "intacto"). A curva precisão-revocação, em diferentes limiares de classificação, é descrita pela Figura 30.

Para os valores de limiar, é importante considerar que:

- Limiar em 0.1: quando a saída da rede fornecer um valor menor ou igual a 0.1, a etiqueta predita é "com dano", e valores maiores que 0.1, "intacto".
- Limiar em 0.3: quando a saída da rede fornecer um valor menor ou igual a 0.3, a etiqueta predita é "com dano", e valores maiores que 0.3, "intacto".
- Limiar em 0.5: quando a saída da rede fornecer um valor menor ou igual a 0.5, a etiqueta predita é "com dano", e valores maiores que 0.5, "intacto".

Figura 30. Curva Precisão-Revocação.



Fonte: Autora.

- Limiar em 0.7: quando a saída da rede fornecer um valor menor ou igual a 0.7, a etiqueta predita é "com dano", e valores maiores que 0.7, "intacto".
- Limiar em 0.9: quando a saída da rede fornecer um valor menor ou igual a 0.9, a etiqueta predita é "com dano", e valores maiores que 0.9, "intacto".

Os resultados da AUPRC indicam um desempenho muito bom na classificação de instâncias como positivas ("intacto"), que são a classe minoritária do modelo. O menor valor de AUPRC encontrado foi de 0,86 para um limiar que define como "intacto" todas as probabilidades acima de 0,1. Apesar de esse limiar aumentar a probabilidade de falsos positivos (superfícies com danos aparentes classificadas como saudáveis), o desempenho do modelo ainda é considerado bom. E sinaliza a robustez da arquitetura uma vez que valores mais elevados de AUPRC revelam a capacidade do modelo de identificar a classe minoritária, "intacto".

No que se refere à análise de falhas, este limiar pode ser considerado perigoso, pois pode levar à classificação de instâncias da classe "com dano" como saudáveis, negligenciando situações de risco. Ao aumentar este limiar, além de melhorar o desempenho do índice AUPRC, também se aumenta a rigidez na identificação de falhas, pois o modelo só classifica como "intacto" instâncias com alta probabilidade de serem realmente "intacto". Por outro lado, imagens de entrada com etiqueta

"intacto" podem ser classificadas como "com dano" se a saída do modelo for menor ou igual a 0,9.

## 6.2 Visualização das Ativações de Rede

As CNNs processam imagens através de etapas de filtragem, ativação e redimensionamento. Essas etapas permitem que a CNN extraia características cada vez mais relevantes da imagem à medida que a rede se aprofunda. Analisando a ativação máxima em cada camada da rede, podemos identificar quais regiões da imagem contribuem mais para a classificação de falha. Para ilustrar o processo, utilizou-se uma imagem do conjunto de dados de teste (Figura 31), anotada como "com dano" por apresentar deslizamento aparente como falha na superfície.

Figura 31. Imagem de entrada.

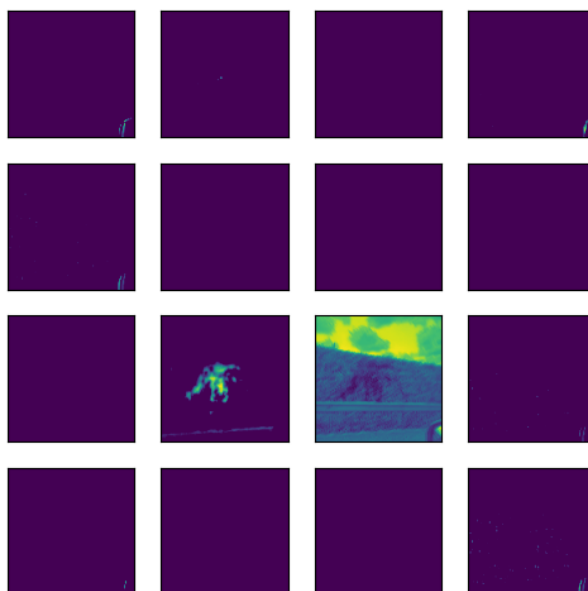


Fonte: Autora.

Mapas de características extraídas pelas camadas convolucionais são apresentados nas Figuras 32, 33 e 34. Cada mapa exibe 16 elementos do mapa de características de resultantes da convolução, que utiliza 32 filtros, ativação e redimensionamento da imagem de referência. Esses mapas representam atributos

extraídos da imagem de referência pelos três blocos convolucionais, cada um composto por operações de convolução, agrupamento e ativação não lineares.

Figura 32. Mapa de características na primeira camada de convolução.

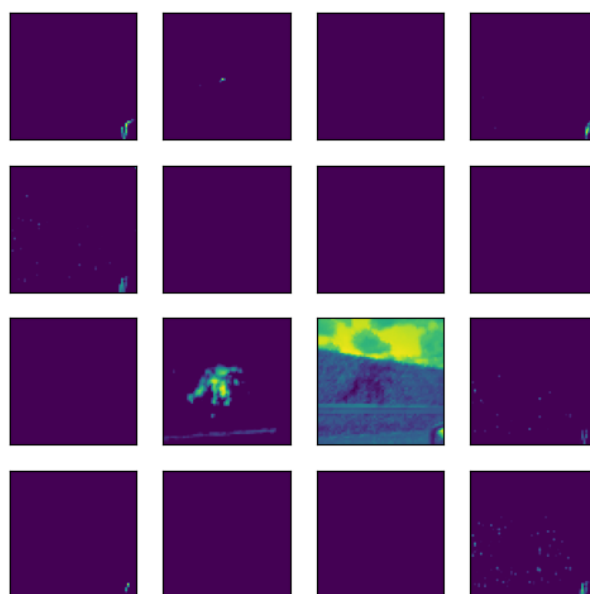


Fonte: Autora

As visualizações dos mapas de ativação revelam que os indicadores de dano se destacam nitidamente nas primeiras camadas da rede neural, diferenciando-se do restante da superfície da imagem. Essa análise dos mapas de ativação é crucial para entender como a rede neural aprende a distinguir áreas danificadas nas imagens e, conseqüentemente, como classifica uma imagem como "intacta" ou "com dano". Embora a interpretação detalhada dos mapas de características dos atributos aprendidos seja complexa, compreender como a rede distingue áreas danificadas nas imagens ao longo do aprofundamento da arquitetura nos permite entender melhor o processo de classificação da rede.

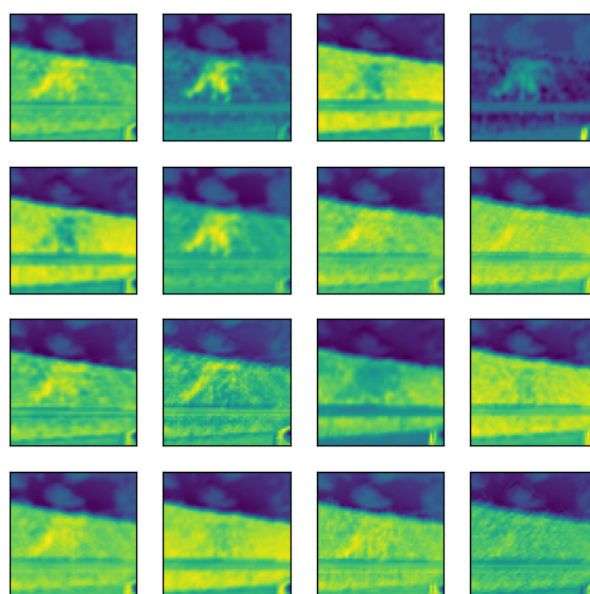
A última densa ou totalmente conectada, do inglês *Fully Connected* (FC) utiliza a função de ativação Sigmoide, que gera valores de saída entre 0 e 1. O único neurônio na camada de saída gera um valor de probabilidade, indicando a chance de a imagem de entrada pertencer a uma determinada classe. Neste classificador binário, a classe "intact" representa imagens com superfícies saudáveis

Figura 33. Mapa de características na segunda camada de convolução.



Fonte: Autora.

Figura 34. Mapa de características na terceira camada de convolução.



Fonte: Autora.

e é codificada como 1, enquanto a classe "damage", que representa imagens com danos, é codificada como 0.

Para visualização da classificação das imagens, um limiar de 0,5 é aplicado à probabilidade de saída. Se a probabilidade for maior ou igual a 0,5, a imagem é classificada como "intact". Caso contrário, com probabilidade inferior a 0,5, a imagem é classificada como "damage".

Em síntese, as predições do modelo consistem em três elementos principais:

- Predição: A predição da classe da imagem, que pode ser "damage" (indicando a presença de erosões, deslizamentos ou desvegetação) ou "intact" (indicando a superfície do solo aparentemente saudável).
- Probabilidade: Um valor numérico que representa a confiança do modelo na sua predição. Esse valor varia entre 0 e 1, onde 0 indica que o modelo está confiante de que a imagem apresenta danos e 1 indica que o modelo está confiante de que a imagem não apresenta danos.
- Classe: A classe real da imagem, utilizada para avaliar o desempenho do modelo. Essa classe foi definida durante a anotação e preparação dos dados e representa a classe verdadeira da imagem, independentemente da predição do modelo.

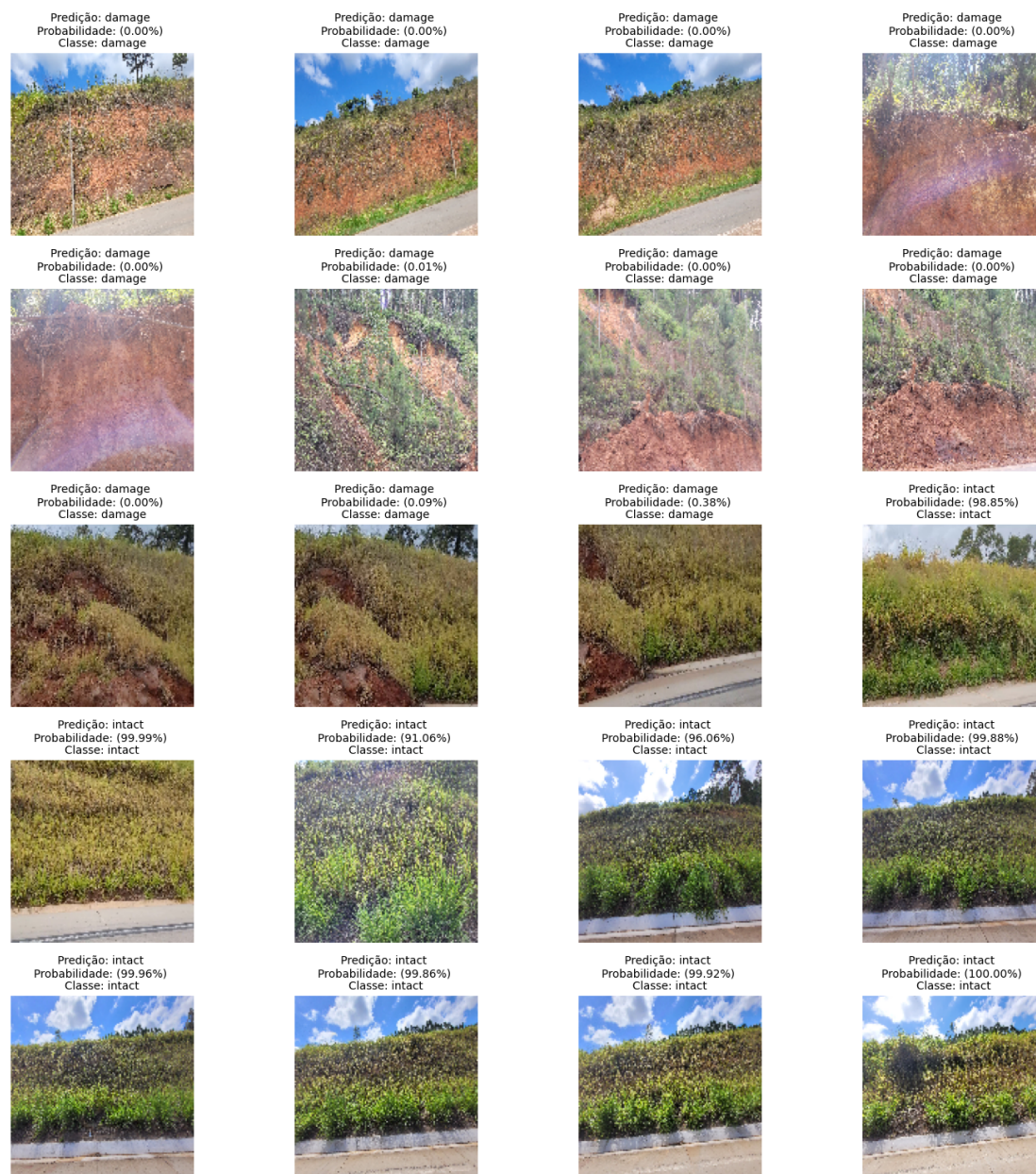
Sendo assim, de acordo com o modelo proposto, ao receber uma imagem, a resposta é a probabilidade de a imagem ser classificada na classe "damage". A Figura 35 demonstra a predição do modelo para algumas imagens do conjunto de teste.

O modelo apresentou-se assertivo nas 20 imagens testadas, confirmando os resultados positivos da avaliação de desempenho da solução proposta. Isso representa um índice de um resultado notável que corrobora os resultados promissores da avaliação de desempenho e reforça o potencial da solução proposta.

Embora a CNN personalizada tenha apresentado bom desempenho na detecção de danos em superfícies geotécnicas, conforme demonstrado em (CHA; CHOI; BÜYÜKÖZTÜRK, 2017; LIMÃO; ARAÚJO; FRANCES, 2023), ainda se



Figura 35. Teste do modelo na detecção de danos visíveis em superfícies de taludes.



Fonte: Autora.

faz necessária a comparação de seu desempenho com modelos populares de CNN, como VGG, Inception e ResNet. Estudos recentes sugerem que arquiteturas CNN personalizadas mais simples podem fornecer resultados comparáveis a esses modelos

populares, como em (BRIGATO; IOCCHI, 2021; FOROUGH; CHEN; WANG, 2021). No entanto, o problema específico desta tese exige investigações e análises mais aprofundadas, pois há desafios singulares no campo de engenharia geotécnica (PHOON; ZHANG, 2023).

Além dos desafios relacionados à construção do conjunto de dados, a experimentação da arquitetura também enfrentou limitações. Um dos principais desafios foi o tempo computacional significativo exigido pelo treinamento dos modelos CNN. Teoricamente, é verdade que quanto menor o número de amostras, mais rápida será a taxa de convergência ao usar uma CPU de núcleo único (LI et al., 2022). Entretanto, o tempo de treinamento estimado em um processador CPU foi de aproximadamente 14 a 20 horas por conjunto de cinco experimentos. Esse tempo computacional elevado para uma CPU única limitou a possibilidade de experimentação de arquiteturas avançadas, que exigem recursos computacionais robustos. Uma das maneiras de reduzir o tempo de treinamento é através do processamento paralelo baseado em GPU (ALZUBAIDI et al., 2021; GARBIN; ZHU; MARQUES, 2020; CANO et al., 2021), o que pode ser uma implementação futura.

Apesar do custo de implementação do modelo CNN ser relativamente alto, seu desempenho geral pode ser considerado superior ao de outras técnicas, conforme demonstrado em (GIRSHICK, 2015). Nas CNNs, as técnicas de extração de características são desnecessárias, pois o modelo aprende automaticamente as características a partir dos dados de entrada. Essa vantagem pode economizar significativamente o esforço na implementação da detecção de padrões em imagens, o que também se aplica ao presente estudo. Ademais, essa característica configura-se como uma vantagem em relação a outras técnicas de processamento de imagens, como o SVM.

## Considerações Finais

A análise dos resultados revelou diversas vantagens e limitações do modelo proposto. Em destaque, sua capacidade de alcançar resultados satisfatórios com uma abordagem de baixa complexidade. Essa característica o torna uma ferramenta viável para cenários onde a simplicidade e a eficiência computacional são elemen-

tares. Para avaliar o desempenho do modelo proposto, foram utilizadas diversas métricas de desempenho, acurácia, perda, matriz de confusão e curva ROC-AUC. Os resultados obtidos demonstraram um desempenho satisfatório do modelo, com valores de acurácia acima de 90% em diversos cenários. Além disso, mesmo com desequilíbrio de classes, os resultados das curvas PRC em diferentes limiares de classificação indicaram índices AUPRC satisfatórios, mesmo em um cenário cujo limiar 0,1 favorece a falsos positivos.

As características do modelo proposto, como sua simplicidade o tornam uma ferramenta viável para diversas aplicações práticas em cenários de inspeção de infraestruturas. Além das aplicações práticas, o estudo também apresenta implicações relevantes para a área da engenharia geotécnica. O desenvolvimento e a validação do modelo proposto abrem caminho para novas pesquisas na área, com potencial para aprimorar os métodos de inspeção, monitoramento e manutenção de infraestruturas.

## 7 Conclusão

Nesta tese, propôs-se uma abordagem baseada em visão computacional para detectar automaticamente danos visíveis em imagens superfícies de estruturas geotécnicas. A integração do sistema VANT-CNN mostrou-se viável como estratégia para inspeção remota, e inclui benefícios como a minimização da necessidade de inspeções físicas pelos profissionais em locais remotos e hostis, e consequente garantia da segurança desses profissionais responsáveis pelo monitoramento de tais locais, a baixo custo computacional.

Ademais, a detecção precoce de falhas na superfície de taludes é fundamental para a segurança das infraestruturas adjacentes. A abordagem proposta viabiliza a aplicação prática para a identificação precoce de indicadores de falha, possibilitando a tomada de medidas preventivas e evitando catástrofes. E que pode ser aplicada além da de rodovias, mas também em estruturas como barragens, diques e aterros.

Nas seções subsequentes, serão abordados os desafios e as oportunidades encontrados durante a realização do trabalho de tese. Primeiramente, serão apresentadas as contribuições para o campo da solução proposta e os resultados obtidos. Em seguida, serão discutidas as limitações e as dificuldades enfrentadas durante o desenvolvimento do trabalho. Por fim, serão apresentados desdobramentos e perspectivas para trabalhos futuros.

### 7.1 Contribuições

Esta tese avaliou com sucesso a hipótese central desta pesquisa "Em cenários desafiadores de escassez de dados, desequilíbrio de classes e limitação de recursos computacionais para experimentação e avaliação dos modelos, é possível desenvolver uma arquitetura de CNN robusta para automatizar a inspeção, com auxílio de VANTs, com o objetivo de detectar danos em superfícies de estruturas geotécnicas com alta precisão".

Em resumo, as principais contribuições desta tese são:

- a) A participação no projeto "Deep learning e machine learning para extração de padrões em estruturas críticas da Usina Hidrelétrica (UHE) Belo Monte" apoiado pela Norte Energia S.A no âmbito do Projeto de Pesquisa e Desenvolvimento (P&D) da Agência Nacional de Energia Elétrica (ANEEL) código PD-07427-0321/2021.
- b) Construção de base de imagens representativas para danos visíveis em taludes: Imagens representativas de danos visíveis na superfície de taludes, caracterizadas pelo solo exposto, e superfícies saudáveis foram coletadas, anotadas, e disponibilizadas na plataforma de compartilhamento Figshare (ARAUJO, 2023).
- c) Proposta de uma abordagem para a solução de problemas multiclasse desequilibrados em imagens de danos através da classificação binária: um modelo para classificação binária de imagens foi implementado e avaliado, apresentando desempenho satisfatório. O código está disponível na plataforma de compartilhamento Figshare (ARAUJO, 2023).
- d) Desenvolvimento de um modelo de baixa complexidade para resolução de problemas com dados escassos: Foi demonstrado que CNNs personalizadas, com menos camadas convolucionais, combinadas com técnicas de regularização, otimização da aprendizagem e aumento de dados, alcançam desempenho satisfatório em conjuntos de dados redundantes e menores.
- e) Demonstração do desempenho de CNNs de baixa complexidade: Experimentos realizados com severas restrições de recursos computacionais para processamento (CPU única) sinalizam uma arquitetura enxuta atrativa em cenários de limitação computacional e dados de treinamento escassos.
- f) Exploração de uma estratégia para inspeção remota e detecção de danos: Utiliza aquisição de dados por VANT e dispositivos móveis para inspeção remota e detecção de danos visíveis a olho nu na superfície de estruturas geotécnicas por meio de CNN. Essa integração de tecnologias pode reduzir a necessidade de inspeções manuais e otimizar o processo de

monitoramento de infraestrutura. De modo a contribuir para a otimização da manutenção da integridade das estruturas e prevenção de desastres.

g) Publicização de resultados em periódicos e eventos, nacionais e internacionais.

- Aceite do artigo intitulado "*Enhancing geotechnical damage detection with deep learning: a Convolutional Neural Network approach*", pelo *Journal Peerj Computer Science*, Qualis A1, Fator de Impacto 3,8, e *citescore* 4,2.
- Publicação do artigo Limão, C. H. E., Araújo, T. M. A. de, & Frances, C. R. L. (2023). *Deep learning-based slope erosion detection. IAES International Journal of Artificial Intelligence (IJ-AI)*, 12(3), 1428. <https://doi.org/10.11591/ijai.v12.i3.pp1428-1438>

h) Formação de recursos humanos relativos às áreas de abrangência desta tese:

- Coordenação e orientação do projeto de iniciação científica júnior, no Instituto Federal de Educação, Ciência e Tecnologia do Pará (IFPA), intitulado "Estudo de redes de aprendizado profundo *Deep Learning* aplicadas à diagnósticos de danos estruturais por imagem", do estudante João Pedro Lima Brandão, 2018-2019.
- Coordenação e orientação do projeto de iniciação científica júnior, no IFPA, intitulado "Estudo comparativo de algoritmos de *machine learning* aplicados à classificação de danos em estruturas habitacionais", do estudante Eric Victor da Silva Alves, 2019-2020.
- Coordenação e orientação do projeto de iniciação científica júnior, no IFPA, intitulado "Estudo de redes de aprendizado profundo *Deep Learning* aplicadas à diagnósticos de danos estruturais por imagem", do estudante João Pedro Lima Brandão, 2018-2019.
- Membro da banca examinadora da dissertação de mestrado intitulada "Clusterização de padrões espaço-temporais de precipitação

na Amazônia via *Deep Convolutional Autoencoder*", de Vander Augusto Oliveira da Silva, julho de 2023.

- Orientação do Projeto Interdisciplinar no curso técnico em Informática para internet do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), intitulado "Aplicação *web* de análise de modo de falhas em barragens de terra baseado no método análise de modos de falha e seus efeitos, do inglês *Failure Mode and Effect Analysis* (FMEA)", dos estudantes Arthur Augusto e Gabriel Silva Galvão, concluído em 2º/2023.
- Orientação (em andamento) do Trabalho de Conclusão de Curso no curso de engenharia de computação do CEFET-MG, intitulado "Proposta de uma Arquitetura de Monitoramento Baseado em IoT para Barragens de Terra e Enrocamento", do estudante Diego Henrique Almeida Trindade, com previsão de conclusão em 1º/2024.
- Aceite do trabalho intitulado "*Moisture Detector in Concrete using Convolutional Neural Networks*" CEFET-MG, submetido ao *Ibero-Latin American Congress on Computational Methods in Engineering* (CILAMCE) 2024.

## 7.2 Limitações e Dificuldades

Ao longo da jornada de pesquisa que culminou nesta tese, diversos desafios se apresentaram, desde a concepção e proposição da solução até sua implementação e experimentação. Estes obstáculos, porém, serviram como oportunidades e moldaram o trabalho, contribuindo para o seu aprimoramento. Como se discorre nos parágrafos subsequentes.

As imagens são as principais fontes de dados para o desenvolvimento de soluções baseadas em visão computacional. Uma pesquisa abrangente foi realizada para identificar possíveis bases públicas de dados. Cada base foi cuidadosamente analisada para selecionar amostras que serviriam como base para o modelo. Essa etapa consumiu um tempo significativo, pois foi necessário avaliar a relevância e a qualidade das contribuições de cada base para a pesquisa. No entanto, a

análise concluiu que, na abordagem proposta, essas bases apresentam limitações que as tornam inviáveis para o problema de detecção de danos em superfícies de estruturas geotécnicas. Essas limitações incluem volume insuficiente de imagens, diretos autorais sobre as imagens, falta de diversidade e baixa representatividade de estruturas que se enquadram ao escopo desta tese.

A construção de um banco de dados robusto para um problema do mundo real apresentou desafios consideráveis. A escassez de dados públicos relevantes motivou a realização de inspeções *in loco* nas estruturas para coletar imagens que representassem cenários com superfícies saudáveis e com falhas aparentes. No entanto, a natureza excepcional das falhas nas estruturas exigiu paciência e persistência na busca por oportunidades de registrar imagens que caracterizassem adequadamente o problema de danos. Além disso, o planejamento meticuloso das inspeções foi crucial para garantir a segurança da equipe durante a coleta de dados, mesmo com o auxílio de VANTs. Em alguns casos, a verificação presencial das falhas foi imprescindível para complementar as informações obtidas remotamente.

A concepção da arquitetura do modelo de aprendizado de máquina também apresentou desafios. Apesar do volume significativo de imagens coletadas durante as inspeções, uma análise aprofundada revelou redundância considerável, especialmente em relação a taludes. Além disso, ao analisar as imagens para anotação de classes, observou-se um desequilíbrio significativo entre imagens de superfícies saudáveis e aquelas com danos visíveis. Visando reduzir a complexidade computacional do modelo, foi realizada uma investigação meticulosa sobre como construí-lo. A ideia de classificação binária surgiu após um levantamento detalhado de fatores comuns à maioria dos danos visíveis e a seleção das imagens que verdadeiramente são capazes de representar o problema. Essa abordagem permitiu que diversas classes fossem agrupadas em uma única classe, baseada em um fator comum: a presença de solo exposto. Tal decisão exigiu um esforço considerável e conhecimento técnico especializado para avaliar sua viabilidade do ponto de vista da engenharia geotécnica.

Somam-se às dificuldades já mencionadas os desafios impostos pelas severas restrições computacionais do ambiente operacional utilizado para a avaliação das arquiteturas, conforme detalhado no Capítulo 5. Apesar de sua reconhecida eficácia,



os modelos CNN são notoriamente exigentes em termos de recursos computacionais durante o treinamento. Nesse contexto, a proposta de uma arquitetura enxuta, com um número reduzido de camadas, visou não apenas atender às limitações do ambiente operacional, mas também preencher uma lacuna observada na literatura, que carece de abordagens CNN menos complexas e mais adequadas a cenários com restrições computacionais como o desta tese.

### 7.3 Desdobramentos e Trabalhos Futuros

O estudo apresentado abre caminho para diversas oportunidades de pesquisa futura que visam aprimorar a detecção de danos em superfícies geotécnicas utilizando modelos de aprendizado de máquina, incluindo arquiteturas CNN, baseados em visão computacional. Nesta seção, destaca-se algumas das principais áreas de investigação que merecem atenção:

- Coletar e incorporar um conjunto de dados expandido com milhares de imagens, incluindo imagens artificiais e reais, que representem diferentes tipos de falhas em superfícies geotécnicas. Isso pode incluir falhas em solos, rochas e estruturas geotécnicas em geral. Nesse sentido, é importante assegurar a inclusão de imagens coletadas em diferentes condições climáticas, como chuva, neve, neblina e iluminação precária, e condições climáticas extremas, como desastres naturais e tragédias ambientais, para testar a capacidade do modelo de detectar danos em situações críticas. Cabe também investigar técnicas de geração de dados artificiais, como GANs (*Generative Adversarial Networks*), para aumentar a diversidade e o volume de dados de treinamento, especialmente para cenários com baixa representatividade no banco de dados original.
- Diante da expansão do conjunto de dados, implementar o treinamento e a inferência do modelo em GPUs para aproveitar seu poder de processamento paralelo e acelerar o tempo de execução. Existem *frameworks* de CNN para distribuir o treinamento do modelo em múltiplos computadores ou GPUs, o

que possibilitará o treinamento de modelos ainda mais complexos e conjuntos de dados ainda maiores.

- Com a disponibilidade de recursos computacionais avançados para processamento, explorar a utilização de modelos pré-treinados e de transferência de aprendizado em tarefas de visão computacional. Realizar uma análise comparativa do desempenho do modelo proposto de arquiteturas CNN populares e bem estabelecidas, como VGG e ResNet, e também técnicas clássicas de processamento de imagens, como SVM, RF e outras, para identificar suas vantagens e desvantagens na detecção de danos em superfícies geotécnicas.
- E por fim, para uma abordagem completa de identificação de falhas em estruturas geotécnicas, investigar a viabilidade de construir um classificador multiclasse para identificar diferentes tipos de danos em superfícies de solos, como solo exposto, surgência e vegetação anormal, além dos danos já abordados na tese. Explorar a combinação de imagens com dados de sensores para detecção de falhas. A integração de dados de sensores como acelerômetros, inclinômetros e piezômetros pode fornecer informações complementares e aprimorar a precisão da detecção, especialmente em condições onde a visibilidade é limitada.

## Referências

ABEDALLA, A.; ABDULLAH, M.; AL-AYYOUB, M.; BENKHELIFA, E. Chest x-ray pneumothorax segmentation using u-net with efficientnet and resnet architectures. *PeerJ Computer Science*, PeerJ Inc., v. 7, p. 1–36, 2021. ISSN 23765992. Citado 10 vezes nas páginas 20, 40, 52, 53, 54, 56, 60, 61, 81 e 90.

ADNAN, M.; ALAROOD, A. A. S.; UDDIN, M. I.; REHMAN, I. ur. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Computer Science*, v. 8, p. e803, fev. 2022. ISSN 2376-5992. Disponível em: <<https://doi.org/10.7717/peerj-cs.803>>. Citado 7 vezes nas páginas 49, 53, 54, 59, 62, 63 e 66.

AFFONSO, C.; ROSSI, A. L. D.; VIEIRA, F. H. A.; CARVALHO, A. C. P. de Leon Ferreira de. Deep learning for biological image classification. *Expert Systems with Applications*, Elsevier Ltd, v. 85, p. 114–122, 11 2017. ISSN 09574174. Citado 5 vezes nas páginas 40, 43, 74, 81 e 90.

AKHTAR, N.; RAGAVENDRAN, U. Interpretation of intelligence in cnn-pooling processes: a methodological survey. *Neural Computing and Applications*, v. 32, p. 879–898, 2 2020. ISSN 0941-0643. Citado 4 vezes nas páginas 40, 44, 47 e 94.

ALZUBAIDI, L.; ZHANG, J.; HUMAIDI, A. J.; AL-DUJAILI, A.; DUAN, Y.; AL-SHAMMA, O.; SANTAMARÍA, J.; FADHEL, M. A.; AL-AMIDIE, M.; FARHAN, L. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data 2021 8:1*, SpringerOpen, v. 8, p. 1–74, 3 2021. ISSN 2196-1115. Citado na página 113.

AMERICAN Geosciences Institute. 2024. <<http://www.earthscienceworld.org/images/search/browserresults.html>>. Citado na página 85.

ARAUJO, T. Software, *Code\_damage\_detection\_CNN*. 2023. Figshare. Disponível em: <<https://doi.org/10.6084/m9.figshare.24247171>>. Citado na página 116.

ARAUJO, T. M. A. *Redes neurais incorporadas a métodos numéricos para solução de problemas de engenharia estrutural*. Dissertação (Dissertação de mestrado) — Universidade Federal de Alagoas, Maceió, AL, Brasil, 2015. Citado 3 vezes nas páginas 36, 38 e 72.

ARAUJO, T. M. A. de. *Superficial\_gotecnical\_damage.2023.Disponívelem* : <>. Citado 2 vezes nas páginas 90 e 116.

AZIMI, M.; PEKCAN, G. Structural health monitoring using extremely compressed data through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, Blackwell Publishing Inc., v. 35, p. 597–614, 6 2020. ISSN 14678667. Citado 5 vezes nas páginas 64, 66, 70, 71 e 80.

AZIZ, F.; KHAN, R.; YOUSAF, S.; HASEEB, A.; UDDIN, M. I. Exploring a design of landslide monitoring system. *Complexity*, v. 2021, p. 5552417, 2021. Disponível em: <<https://doi.org/10.1155/2021/5552417>>. Citado na página 31.

BARELLA, V.; GARCIA, L.; CARVALHO, A. d. The influence of sampling on imbalanced data classification. In: IEEE. *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.], 2019. p. 210–215. Citado 2 vezes nas páginas 19 e 52.

BARI, B. S.; ISLAM, M. N.; RASHID, M.; HASAN, M. J.; RAZMAN, M. A. M.; MUSA, R. M.; NASIR, A. F. A.; MAJEED, A. P. A. A real-time approach of diagnosing rice leaf disease using deep learning-based faster r-cnn framework. *PeerJ Computer Science*, v. 7, p. e432, apr 2021. ISSN 2376–5992. Disponível em: <<https://doi.org/10.7717/peerj-cs.432>>. Citado na página 81.

BASHA, S. S.; DUBEY, S. R.; PULABAIGARI, V.; MUKHERJEE, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, v. 378, p. 112–119, 2 2020. ISSN 09252312. Citado 5 vezes nas páginas 39, 50, 53, 93 e 94.

BENZER, N. Using the geographical information system and remote sensing techniques for soil erosion assessment. *Polish Journal of Environmental Studies*, v. 19, n. 5, p. 881–886, 2010. ISSN 1230-1485. Disponível em: <<https://www.pjoes.com/Using-the-Geographical-Information-System-r-nand-Remote-Sensing-Techniques-for-Soil,88460,0,2.html>>. Citado 3 vezes nas páginas 28, 30 e 33.

BORRELLI, P.; ALEWELL, C.; ALVAREZ, P.; ANACHE, J. A. A.; BAARTMAN, J.; BALLABIO, C.; BEZAK, N.; BIDDOCU, M.; CERDà, A.; CHALISE, D.; CHEN, S.; CHEN, W.; GIROLAMO, A. M. D.; GESSESSE, G. D.; DEUMLICH, D.; DIODATO, N.; EFTHIMIOU, N.; ERPUL, G.; FIENER, P.; FREPPAZ, M.; GENTILE, F.; GERICKE, A.; HAREGEWEYN, N.; HU, B.; JEANNEAU, A.; KAFFAS,

K.; KIANI-HARCHEGANI, M.; VILLUENDAS, I. L.; LI, C.; LOMBARDO, L.; LÓPEZ-VICENTE, M.; LUCAS-BORJA, M. E.; MÄRKER, M.; MATTHEWS, F.; MIAO, C.; MIKOŠ, M.; MODUGNO, S.; MÖLLER, M.; NAIPAL, V.; NEARING, M.; OWUSU, S.; PANDAY, D.; PATAULT, E.; PATRICHE, C. V.; POGGIO, L.; PORTES, R.; QUIJANO, L.; RAHDARI, M. R.; RENIMA, M.; RICCI, G. F.; RODRIGO-COMINO, J.; SAIA, S.; SAMANI, A. N.; SCHILLACI, C.; SYRRIS, V.; KIM, H. S.; SPINOLA, D. N.; OLIVEIRA, P. T.; TENG, H.; THAPA, R.; VANTAS, K.; VIEIRA, D.; YANG, J. E.; YIN, S.; ZEMA, D. A.; ZHAO, G.; PANAGOS, P. Soil erosion modelling: A global review and statistical analysis. *Science of The Total Environment*, v. 780, p. 146494, 8 2021. ISSN 00489697. Citado na página 28.

BOWD, C.; BELGHITH, A.; ZANGWILL, L. M.; CHRISTOPHER, M.; GOLDBAUM, M. H.; FAN, R.; REZAPOUR, J.; MOGHIMI, S.; KAMALIPOUR, A.; HOU, H. et al. Deep learning image analysis of optical coherence tomography angiography measured vessel density improves classification of healthy and glaucoma eyes. *American journal of ophthalmology*, Elsevier, v. 236, p. 298–308, 2022. Citado 2 vezes nas páginas 63 e 66.

BRIGATO, L.; IOCCHI, L. A close look at deep learning with small data. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2021. p. 2490–2497. Citado 7 vezes nas páginas 20, 52, 81, 93, 94, 98 e 113.

BUSHARA, A. R.; KUMAR, R. S. Deep learning-based lung cancer classification of ct images using augmented convolutional neural networks. *Electronic Letters on Computer Vision and Image Analysis*, Universitat Autònoma de Barcelona, v. 21, p. 130–142, 2022. ISSN 15775097. Citado 4 vezes nas páginas 56, 57, 59 e 94.

CANO, E.; MENDOZA-AVILÉS, J.; AREIZA, M.; GUERRA, N.; MENDOZA-VALDÉS, J. L.; ROVETTO, C. A. Multi skin lesions classification using fine-tuning and data-augmentation applying nasnet. *PeerJ Computer Science*, v. 7, p. e371, jun 2021. ISSN 2376-5992. Disponível em: <<https://doi.org/10.7717/peerj-cs.371>>. Citado 7 vezes nas páginas 55, 56, 57, 63, 66, 81 e 113.

CARVALHO, A. C. P. de Leon Ferreira de. *Computação e interdisciplinaridade*.

2016. Disponível em: <www.sbc.org.br>. Citado 3 vezes nas páginas 35, 36 e 37.

CAZZUFFI, D.; CRIPPA, E. Contribution of Vegetation to Slope Stability: An Overview of Experimental Studies Carried Out on Different Types of Plants. In: *Erosion of Soils and Scour of Foundations*. Austin, Texas, United States: American Society of Civil Engineers, 2005. p. 1–12. ISBN 978-0-7844-0781-3. Disponível em: <<http://ascelibrary.org/doi/abs/10.1061/40781%28160%299>>. Citado na página 33.

CHA, Y.-J.; CHOI, W.; BüyüKöZTürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, v. 32, p. 361–378, 5 2017. ISSN 10939687. Citado 9 vezes nas páginas 18, 47, 48, 49, 51, 76, 80, 90 e 111.

CHAUDHARI, K. G. Comparative analysis of cnn models to diagnose breast cancer. *International Journal of Innovative Research in Science, Engineering and Technology*, v. 7, n. 10, p. 8180–8187, 2018. Citado 2 vezes nas páginas 40 e 43.

COFRE-MARTEL, S.; KOBRICH, P.; DROGUETT, E. L.; MERUANE, V. Deep convolutional neural network-based structural damage localization and quantification using transmissibility data. *Shock and Vibration*, Hindawi Limited, v. 2019, 2019. ISSN 10709622. Citado 2 vezes nas páginas 49 e 90.

COOKE, R.; DOORNKAMP, J. C. et al. *Geomorphology in environmental management: a new introduction*. [S.l.]: Oxford University Press (OUP), 1990. Citado na página 29.

DANG, X.-K.; HO, L. A.-H.; NGUYEN, X.-P.; MAI, B.-L. Applying artificial intelligence for the application of bridges deterioration detection system. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, v. 20, p. 149, 2 2022. ISSN 2302-9293. Citado 2 vezes nas páginas 70 e 80.

DARAPANENI, N.; KRISHNAMURTHY, B.; PADURI, A. R. Convolution neural networks: A comparative study for image classification. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2020. p. 327–332. ISBN 9781728185248. Citado 6 vezes nas páginas 20, 52, 60, 61, 81 e 93.

DAS, B. M. *Geotechnical engineering handbook*. [S.l.]: J. Ross Pub, 2011. ISBN 9781932159837. Citado 2 vezes nas páginas 25 e 27.

DAS, B. M. *Principles of Geotechnical Engineering*. [S.l.]: Cengage Learning, 2017. Citado na página 26.

DI, K.; LI, W.; YUE, Z.; SUN, Y.; LIU, Y. A machine learning approach to crater detection from topographic data. *Advances in Space Research*, Elsevier Ltd, v. 54, p. 2419–2429, 12 2014. ISSN 18791948. Citado 2 vezes nas páginas 18 e 72.

EKUNDAYO, O.; VIRIRI, S. Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation. *PeerJ Computer Science*, v. 7, p. e736, nov 2021. ISSN 2376–5992. Citado 5 vezes nas páginas 56, 60, 61, 62 e 81.

EPPENHOF, E. J.; PEÑA-CASTILLO, L. Prioritizing bona fide bacterial small rnas with machine learning classifiers. *PeerJ*, v. 7, p. e6304, jan. 2019. ISSN 2167-8359. Disponível em: <<https://doi.org/10.7717/peerj.6304>>. Citado 2 vezes nas páginas 63 e 66.

FARRAR, C. R.; LIEVEN, N. A. Damage prognosis: The future of structural health monitoring. *Philos. Trans. Royal Soc. London, Ser. A*, v. 365, n. 1851, p. 623–632, 2007. Citado na página 25.

FERNANDES, M. d. M. *Mecânica dos Solos: Volume 2 - Introdução à Engenharia Geotécnica*. Lisboa, Portugal: Editora Lidel, 2018. ISBN 978-989-555-824-2. Citado na página 25.

FIGUEIREDO, E.; PARK, G.; FARRAR, C. R.; WORDEN, K.; FIGUEIRAS, J. Machine learning algorithms for damage detection under operational and environmental variability. *Structural Health Monitoring*, v. 10, n. 6, p. 559–572, 2011. Citado 2 vezes nas páginas 27 e 80.

FOROUGHI, F.; CHEN, Z.; WANG, J. A cnn-based system for mobile robot navigation in indoor environments via visual localization with a small dataset.

*World Electric Vehicle Journal*, MDPI, v. 12, 2021. ISSN 20326653. Citado 11 vezes nas páginas 20, 48, 52, 56, 62, 63, 81, 93, 94, 98 e 113.

GABET, E. J.; REICHMAN, O.; SEABLOOM, E. W. The effects of bioturbation on soil processes and sediment transport. *Annual Review of Earth and Planetary Sciences*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 31, n. 1, p. 249–273, 2003. Citado na página 31.

GARBIN, C.; ZHU, X.; MARQUES, O. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, Springer, v. 79, p. 12777–12815, 5 2020. ISSN 15737721. Citado 6 vezes nas páginas 39, 51, 52, 54, 55 e 113.

GEONET. 2022. <<https://www.geonet.org.nz/>>. Acesso em 27 de janeiro de 2022. Citado 2 vezes nas páginas 85 e 86.

GIRSHICK, R. Fast r-cnn. In: . [S.l.]: IEEE, 2015. p. 1440–1448. ISBN 978-1-4673-8391-2. Citado na página 113.

GREENWOOD, J. R.; NORRIS, J. E.; WINT, J. Assessing the contribution of vegetation to slope stability. *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, v. 157, n. 4, p. 199–207, out. 2004. ISSN 1353-2618, 1751-8563. Disponível em: <<https://www.icevirtuallibrary.com/doi/10.1680/geng.2004.157.4.199>>. Citado na página 33.

GREENWOOD, W. W.; ASCE, A. M.; LYNCH, J. P.; ASCE, M.; ZEKKOS, D. Applications of uavs in civil infrastructure. *Journal of Infrastructure Systems*, American Society of Civil Engineers, v. 25, p. 04019002, 1 2019. ISSN 1076-0342. Citado 2 vezes nas páginas 18 e 88.

GUAN, Q. Z.; YANG, Z. X.; GUO, N.; HU, Z. Finite element geotechnical analysis incorporating deep learning-based soil model. *Computers and Geotechnics*, Elsevier Ltd, v. 154, 2 2023. ISSN 18737633. Citado na página 71.

HAN, X.-L.; JIANG, N.-J.; YANG, Y.-F.; CHOI, J.; SINGH, D. N.; BETA, P.; DU, Y.-J.; WANG, Y.-J. Deep learning based approach for the instance segmentation of



clayey soil desiccation cracks. *Computers and Geotechnics*, v. 146, p. 104733, 2022. ISSN 0266-352X. Citado 6 vezes nas páginas 17, 27, 63, 78, 80 e 90.

HANDY, R. L. *Geotechnical engineering: soil and foundation principles and practice*. [S.l.]: McGraw-Hill Education, 2007. Citado 2 vezes nas páginas 17 e 23.

HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, v. 160, p. 106–154, 1962. ISSN 1469-7793. Citado na página 39.

HUDSON, N. An introduction to the mechanics of soil erosion under conditions of subtropical rainfall. *Rhodesia Science Association Proceedings*, v. 49, p. 14–25, 1961. Citado na página 29.

ICOLD. *International Commission on Large Dams*. 2023. Disponível em: <<https://www.icold-cigb.org/GB/dams/dams.asp>>. Citado na página 24.

IQBAL, H. *HarisIqbal88/PlotNeuralNet v1.0.0*. Zenodo, 2018. Disponível em: <<https://doi.org/10.5281/zenodo.2526396>>. Citado 2 vezes nas páginas 41 e 99.

JANG, K.; KIM, N.; AN, Y.-K. Deep learning-based autonomous concrete crack evaluation through hybrid image scanning. *Structural Health Monitoring*, v. 18, p. 1722–1737, 11 2019. ISSN 1475-9217. Citado 5 vezes nas páginas 18, 60, 61, 77 e 80.

JIANG, F.; GRIGOREV, A.; RHO, S.; TIAN, Z.; FU, Y.; JIFARA, W.; KHAN, A.; LIU, S. Medical image semantic segmentation based on deep learning. *Neural Computing and Applications*, v. 29, 03 2018. Citado na página 75.

JR., G. F. S.; ADELI, H. Infrared thermography for detecting defects in concrete structures. *Journal of civil engineering and management*, v. 24, p. 508–515, 11 2018. ISSN 1392-3730. Citado na página 80.

JUNG, I.-S.; BERGES, M.; GARRETT, J. H. *Anomaly Detection on Piezometer Data Collected from Embankment Dams Using Physical Model-Based Simulation*. [S.l.]: International Conference on Computing in Civil and Building Engineering, 2014. Citado 3 vezes nas páginas 27, 69 e 80.

KALINOVSKY, A.; LIAUCHUK, V.; TARASAU, A. Lesion detection in ct images using deep learning semantic segmentation technique. In: . [S.l.: s.n.], 2017. Citado na página 75.

KANELLAKIS, C.; NIKOLAKOPOULOS, G. Survey on computer vision for uavs: Current developments and trends. *Journal of Intelligent and Robotic Systems: Theory and Applications*, Springer Netherlands, v. 87, p. 141–168, 7 2017. ISSN 15730409. Citado 2 vezes nas páginas 18 e 88.

KATTENBORN, T.; LEITLOFF, J.; SCHIEFER, F.; HINZ, S. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, Elsevier B.V., v. 173, p. 24, 3 2021. ISSN 0924-2716. Citado 4 vezes nas páginas 74, 80, 81 e 87.

KHAN, R.; YOUSAF, S.; HASEEB, A.; UDDIN, M. I. Exploring a design of landslide monitoring system. *Complexity*, Hindawi Limited, v. 2021, 2021. ISSN 10990526. Citado 4 vezes nas páginas 26, 31, 80 e 83.

KUKAČKA, J.; GOLKOV, V.; CREMERS, D. *Regularization for Deep Learning: A Taxonomy*. 2017. Citado 3 vezes nas páginas 54, 55 e 59.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, p. 436–444, 5 2015. ISSN 14764687. Citado 9 vezes nas páginas 18, 39, 40, 46, 47, 48, 49, 93 e 95.

LEI, B.; REN, Y.; WANG, N.; HUO, L.; SONG, G. Design of a new low-cost unmanned aerial vehicle and vision-based concrete crack inspection method. *Structural Health Monitoring*, v. 19, p. 1871–1883, 11 2020. ISSN 1475-9217. Citado 3 vezes nas páginas 18, 80 e 88.

LI, L.; ZHANG, H.; PANG, J.; HUANG, J. Dam surface crack detection based on deep learning. In: *Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence*. New York, NY, USA: Association for Computing Machinery, 2019. (RICAI '19), p. 738–743. ISBN 9781450372985. Disponível em: <<https://doi.org/10.1145/3366194.3366327>>. Citado na página 76.

LI, S.; SONG, W.; FANG, L.; CHEN, Y.; GHAMISI, P.; BENEDIKTSSON, J. A. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, v. 57, p. 6690–6709, 9 2019. ISSN 0196-2892. Citado na página 18.

LI, Y.; ZHANG, H.; XUE, X.; JIANG, Y.; SHEN, Q. Deep learning for remote sensing image classification: A survey. *WIREs Data Mining and Knowledge Discovery*, v. 8, 11 2018. ISSN 1942-4787. Citado 2 vezes nas páginas 49 e 81.

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, Institute of Electrical and Electronics Engineers Inc., v. 33, p. 6999–7019, 12 2022. ISSN 21622388. Citado 4 vezes nas páginas 18, 48, 95 e 113.

LIAUCHUK, V.; KOVALEV, V.; KALINOVSKY, A.; TARASAU, A. Examining the ability of convolutional neural networks to detect lesions in lung ct images (deep learning) the tb portals: An open-access, web-based platform for global drug-resistant tuberculosis data sharing and analysis view project uav: back to base problem view project. *International Congress on Computer Assisted Radiology and Surgery - CARS-2017*, 2017. Disponível em: <<https://www.researchgate.net/publication/313900859>>. Citado na página 81.

LIM, H. J.; HWANG, S.; KIM, H.; SOHN, H. Steel bridge corrosion inspection with combined vision and thermographic images. *Structural Health Monitoring*, v. 20, p. 3424–3435, 11 2021. ISSN 1475-9217. Citado 4 vezes nas páginas 17, 77, 80 e 88.

LIMÃO, C. H. E.; ARAÚJO, T. M. A. de; FRANCES, C. R. L. Deep learning based slope erosion detection. *IAES International Journal of Artificial Intelligence (IJ-AI)*, v. 12, p. 1428, 9 2023. ISSN 2252-8938. Citado 6 vezes nas páginas 20, 78, 80, 93, 94 e 111.

LINDSAY, G. W. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience*, v. 33, p. 2017–2031, 9 2021. ISSN 0898-929X. Citado 2 vezes nas páginas 74 e 81.

LIU, S.; SU, Z.; LI, M.; SHAO, L. Slope stability analysis using elastic finite element stress fields. *Engineering Geology*, v. 273, p. 105673, 8 2020. ISSN 00137952. Citado na página 94.

LÖBMANN, M. T.; GEITNER, C.; WELLSTEIN, C.; ZERBE, S. The influence of herbaceous vegetation on slope stability – A review. *Earth-Science Reviews*, v. 209, p. 103328, out. 2020. ISSN 00128252. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0012825220303743>>. Citado 3 vezes nas páginas 17, 32 e 33.

LU, J.; TAN, L.; JIANG, H. Review on convolutional neural network (cnn) applied to plant leaf disease classification. *Agriculture*, v. 11, p. 707, 7 2021. ISSN 2077-0472. Citado 6 vezes nas páginas 64, 74, 81, 90, 94 e 95.

MACEDO, E. S. d.; SANDRE, L. H. Deaths from landslides in brazil: 1988 to 2022. *Instituto de Pesquisas Tecnológicas do estado de São Paulo – IPT*, São Paulo, Brazil, 2022. Citado na página 90.

MASCAREÑAS, D. D.; BALLOR, J. P.; MCCLAIN, O. L.; MELLOR, M. A.; SHEN, C.-Y.; BLECK, B.; MORALES, J.; YEONG, L.-M. R.; NARUSHOF, B.; SHELTON, P.; MARTINEZ, E.; YANG, Y.; CATTANEO, A.; HARDEN, T. A.; MOREU, F. Augmented reality for next generation infrastructure inspections. *Structural Health Monitoring*, v. 20, p. 1957–1979, 7 2021. ISSN 1475-9217. Citado 3 vezes nas páginas 26, 27 e 71.

MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115–133, 1943. Citado na página 37.

MHASKAR, H.; LIAO, Q.; POGGIO, T. Learning functions: when is deep better than shallow. *arXiv preprint arXiv:1603.00988*, 2016. Citado 4 vezes nas páginas 39, 40, 49 e 95.

MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2018. p. 117–122. ISBN 9781538661437. Citado 3 vezes nas páginas 54, 56 e 57.

MOUNTRAKIS, G.; HEYDARI, S. S. Harvesting the landsat archive for land cover land use classification using deep neural networks: Comparison with traditional classifiers and multi-sensor benefits. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier B.V., v. 200, p. 106–119, 6 2023. ISSN 09242716. Citado 5 vezes nas páginas 80, 81, 90, 94 e 95.

MURRAY, N.; PERRONNIN, F. Generalized max pooling. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, p. 2473–2480, 9 2014. ISSN 10636919. Citado 2 vezes nas páginas 46 e 47.

MYAGILA, K.; KILAVO, H. A comparative study on performance of svm and cnn in tanzania sign language translation using image recognition. *Applied Artificial Intelligence*, Taylor & Francis, v. 36, n. 1, p. 2005297, 2022. Citado 3 vezes nas páginas 18, 40 e 43.

NAGI, J.; DUCATELLE, F.; CARO, G. A. D.; CIRESAN, D.; MEIER, U.; GIUSTI, A.; NAGI, F.; SCHMIDHUBER, J.; GAMBARDELLA, L. M. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: . [S.l.]: IEEE, 2011. p. 342–347. ISBN 978-1-4577-0242-6. Citado na página 47.

NAGPAL, S.; SINGH, M.; VATSA, M.; SINGH, R. Regularizing deep learning architecture for face recognition with weight variations. In: *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE Press, 2015. p. 1–6. Disponível em: <<https://doi.org/10.1109/BTAS.2015.7358791>>. Citado na página 54.

NASA. *National Aeronautics and Space Administration (NASA) Visible Earth*. 2022. Acesso em 20 de janeiro de 2022. Disponível em: <<https://visibleearth.nasa.gov/>>. Citado na página 88.

NBMG. *Geologic Hazards*. 2022. Acesso em 20 de janeiro de 2022. Disponível em: <<https://goo.gl/photos/PA4HpPGzRfpoqAUg8>>. Citado 2 vezes nas páginas 86 e 87.

NBMG. *Nevada Bureau of Mines and Geology - Photos*. 2022. <<https://nbgm.unr.edu/Photos.html>>. Acesso em 29 de Janeiro 2022. Citado 3 vezes nas páginas 85, 86 e 87.

NGUYEN, T. T.; INDRARATNA, B. The energy transformation of internal erosion based on fluid-particle coupling. *Computers and Geotechnics*, Elsevier Ltd, v. 121, 5 2020. ISSN 18737633. Citado na página 29.

NOGUEIRA, K.; MACHADO, G. L. S.; GAMA, P. H. T.; SILVA, C. C. V. da; BALANIUK, R.; SANTOS, J. A. dos. Facing erosion identification in railway lines using pixel-wise deep-based approaches. *Remote Sensing*, v. 12, p. 739, 2 2020. ISSN 2072-4292. Citado 5 vezes nas páginas 28, 29, 30, 49 e 80.

NUSSBAUM, A. *NASA Earth Observatory images, using Landsat data from the U.S. Geological Survey. Story by Kathryn Hansen*. 2022. Data acquired October 29, 2022. Sources: Landsat 8 > OLI, Landsat 9 > OLI-2, Collection. Citado na página 89.

NUSSBAUM, A. *NASA Earth Observatory images, using Landsat data from the U.S. Geological Survey. Story by Kathryn Hansen*. 2023. Data acquired February 26, 2023. Sources: Landsat 8 > OLI, Landsat 9 > OLI-2, Collection. Citado na página 89.

OADES, J. The role of biology in the formation, stabilization and degradation of soil structure. *Geoderma*, v. 56, n. 1, p. 377–400, 1993. ISSN 0016-7061. International Workshop on Methods of Research on Soil Structure/Soil Biota Interrelationships. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0016706193901233>>. Citado na página 31.

OTTONI, A.; AMORIM, R. de; NOVO, M. e. a. Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. *International Journal of Machine Learning and Cybernetics*, v. 14, p. 171–186, 2022. Citado 5 vezes nas páginas 52, 54, 56, 57 e 80.

PAN, N. H.; CHEN, K. Y. Applying fuzzy theory to develop a model for inspecting and assessing soil and water conservation facilities. *Journal of Civil Structural*

*Health Monitoring*, Springer Verlag, v. 5, p. 441–455, 9 2015. ISSN 21905479. Citado 5 vezes nas páginas 17, 26, 27, 71 e 80.

PHOON, K.-K.; ZHANG, W. Future of machine learning in geotechnics. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, Taylor Francis, v. 17, n. 1, p. 7–22, 2023. Disponível em: <<https://doi.org/10.1080/17499518.2022.2087884>>. Citado 3 vezes nas páginas 18, 88 e 113.

RAMANDI, H. L.; IRTZA, S.; SIROJAN, T.; NAMAN, A.; MATHEW, R.; SETHU, V.; ROSHAN, H. Fracdetect: A novel algorithm for 3d fracture detection in digital fractured rocks. *Journal of Hydrology*, Elsevier, v. 607, p. 127482, 4 2022. ISSN 0022-1694. Citado na página 73.

RICHARDS, P. J. Aphaenogaster ants as bioturbators: Impacts on soil and slope processes. *Earth-Science Reviews*, v. 96, n. 1, p. 92–106, 2009. ISSN 0012-8252. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0012825209000877>>. Citado na página 31.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 6, p. 386–408, 1958. Citado na página 37.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986. Citado na página 38.

SALAJEGHEH, R.; MEYMAND, A. M.; KERMANI, M. Z. Evaluating performance of meta-heuristic algorithms and decision tree models in simulating water level variations of dams' piezometers. *Journal of Hydraulic Structures*, v. 4, p. 60–80, 2018. Citado 4 vezes nas páginas 27, 69, 70 e 80.

SANTOS, F. P. D.; ZOR, C.; KITTLER, J.; PONTI, M. A. Learning image features with fewer labels using a semi-supervised deep convolutional network. *Neural Networks*, Elsevier, v. 132, p. 131–143, 2020. Citado na página 39.

SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. *Lecture Notes in Computer*

*Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, Berlin, Heidelberg, v. 6354 LNCS, p. 92–101, 2010. ISSN 03029743. Citado 2 vezes nas páginas 46 e 95.

SHI, L. L.; ZHANG, J.; ZHU, Q. Z.; SUN, H. H. Prediction of mechanical behavior of rocks with strong strain-softening effects by a deep-learning approach. *Computers and Geotechnics*, Elsevier Ltd, v. 152, 12 2022. ISSN 18737633. Citado 2 vezes nas páginas 27 e 72.

SHI, P.; FAN, X.; NI, J.; WANG, G. A detection and classification approach for underwater dam cracks. *Structural Health Monitoring*, SAGE Publications Ltd, v. 15, p. 541–554, 9 2016. ISSN 17413168. Citado na página 80.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, SpringerOpen, v. 6, 12 2019. ISSN 21961115. Citado 7 vezes nas páginas 51, 54, 56, 57, 58, 74 e 90.

SOLÓRZANO, J.; MORANTE-CARBALLO, F.; MONTALVÁN-BURBANO, N.; BRIONES-BITAR, J.; CARRIÓN-MERO, P. A systematic review of the relationship between geotechnics and disasters. *Sustainability*, Multidisciplinary Digital Publishing Institute, v. 14, p. 12835, 10 2022. ISSN 2071-1050. Citado 2 vezes nas páginas 17 e 88.

SPENCER, B. F.; HOSKERE, V.; NARAZAKI, Y. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, v. 5, n. 2, p. 199–222, 2019. ISSN 2095-8099. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2095809918308130>>. Citado 8 vezes nas páginas 18, 25, 27, 48, 68, 77, 80 e 88.

SUDHA, S.; JAYANTHI, K.; RAJASEKARAN, C.; SUNDER, T. Segmentation of roi in medical images using cnn-a comparative study. In: IEEE. *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. [S.l.], 2019. p. 767–771. Citado na página 18.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convoluti-



ons. In: . [S.l.]: IEEE, 2015. p. 1–9. ISBN 978-1-4673-6964-0. Citado na página 43.

TAN, Y.-F.; GUO, X.; POH, S.-C. Time series activity classification using gated recurrent units. *International Journal of Electrical and Computer Engineering (IJECE)*, v. 11, p. 3551, 8 2021. ISSN 2722-2578. Citado 3 vezes nas páginas 47, 63 e 81.

TERZAGHI, K.; PECK, R. B.; MESRI, G. *Soil Mechanics in Engineering Practice*. [S.l.]: John Wiley & Sons, 1996. Citado 5 vezes nas páginas 19, 23, 24, 27 e 28.

THEODORIDIS, S. Chapter 18 - neural networks and deep learning. In: THEODORIDIS, S. (Ed.). *Machine Learning (Second Edition)*. Second edition. [S.l.]: Academic Press, 2020. p. 901–1038. ISBN 978-0-12-818803-3. Citado 4 vezes nas páginas 48, 49, 53 e 54.

TIAN, Y.; LUO, L.; YU, Z.; XU, H.; NI, F. Noncontact vision-based impact force reconstruction and spatial-temporal deflection tracking of a flexible barrier system under rockfall impact. *Computers and Geotechnics*, Elsevier, v. 153, p. 105070, 1 2023. ISSN 0266-352X. Citado na página 18.

VARNES, D. J. Slope movement types and processes. In: SCHUSTER, R. L.; KRIZEK, R. J. (Ed.). *Landslides, Analysis and Control*. [S.l.], 1978. (Special Report, 176), p. 11–33. Citado 4 vezes nas páginas 23, 26, 30 e 31.

VEDALDI, A.; MAHENDRAN, S.; TSOGLAS, S.; MAJI, S.; GIRSHICK, R.; KANNALA, J.; RAHTU, E.; KOKKINOS, I.; BLASCHKO, M. B.; WEISS, D.; TASKAR, B.; SIMONYAN, K.; SAPHRA, N.; MOHAMED, S. Understanding objects in detail with fine-grained attributes. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 3622–3629. Citado na página 72.

VOLKWEIN, A.; SCHELLENBERG, K.; LABIOUSE, V.; AGLIARDI, F.; BERGER, F.; BOURRIER, F.; DORREN, L. K.; GERBER, W.; JABOYEDOFF, M. *Rockfall characterisation and structural protection - A review*. 2011. 2617-2651 p. Citado 3 vezes nas páginas 26, 30 e 31.

WANG, L.; WU, C.; YANG, Z.; WANG, L. Deep learning methods for time-dependent reliability analysis of reservoir slopes in spatially variable soils. *Computers and Geotechnics*, v. 159, p. 105413, 2023. ISSN 0266-352X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0266352X23001702>>. Citado na página 80.

WANG, Y.; LI, Y.; SONG, Y.; RONG, X. The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, v. 10, n. 5, 2020. ISSN 2076-3417. Citado 6 vezes nas páginas 47, 48, 49, 81, 90 e 95.

WILKINSON, M. T.; RICHARDS, P. J.; HUMPHREYS, G. S. Breaking ground: Pedological, geological, and ecological implications of soil bioturbation. *Earth-Science Reviews*, v. 97, n. 1, p. 257–272, 2009. ISSN 0012-8252. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0012825209001470>>. Citado na página 31.

XU, Q.; ZHANG, M.; GU, Z.; PAN, G. Overfitting remedy by sparsifying regularization on fully-connected layers of cnns. *Neurocomputing*, v. 328, p. 69–74, 2019. ISSN 0925-2312. Chinese Conference on Computer Vision 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231218309524>>. Citado 3 vezes nas páginas 52, 55 e 80.

YADAV, S. S.; JADHAV, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, v. 6, p. 113, 12 2019. ISSN 2196-1115. Citado 4 vezes nas páginas 20, 60, 61 e 81.

ZAFAR, A.; AAMIR, M.; NAWI, N. M.; ARSHAD, A.; RIAZ, S.; ALRUBAN, A.; DUTTA, A. K.; ALMOTAIRI, S. A comparison of pooling methods for convolutional neural networks. *Applied Sciences (Switzerland)*, MDPI, v. 12, 9 2022. ISSN 20763417. Citado 5 vezes nas páginas 46, 47, 49, 55 e 56.

ZHANG, D.; TONG, J. Robust water level measurement method based on computer vision. *Journal of Hydrology*, v. 620, p. 129456, 5 2023. ISSN 00221694. Citado na página 80.

ZHANG, W.; LI, H.; LI, Y. e. a. Application of deep learning algorithms in geotechnical engineering: a short critical review. *Artificial Intelligence Review*, v. 54, p. 5633–5673, 2021. Citado 3 vezes nas páginas 18, 54 e 88.

ZHANG, Z. Improved adam optimizer for deep neural networks. *2018 IEEE/ACM 26th International Symposium on Quality of Service, IWQoS 2018*, Institute of Electrical and Electronics Engineers Inc., 1 2019. Citado 3 vezes nas páginas 53, 54 e 59.

ZHAO, Z.-Q. Q.; ZHENG, P.; XU, S.-T. T.; WU, X. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, IEEE, v. 30, n. 11, p. 3212–3232, 2019. ISSN 2162-2388. Citado 2 vezes nas páginas 40 e 81.

ZHONG, C.; LIU, Y.; GAO, P.; CHEN, W.; LI, H.; HOU, Y.; NUREMANGULI, T.; MA, H. Landslide mapping with remote sensing: challenges and opportunities. *International Journal of Remote Sensing*, Taylor & Francis, v. 41, n. 4, p. 1555–1581, 2020. Citado na página 88.