



UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”

---

FCT - Faculdade de Ciências e Tecnologia  
DMC - Departamento de Matemática e Computação  
Bacharelado em Ciência da Computação

Trabalho de Conclusão de Curso  
(Modalidade Trabalho Acadêmico)

Revisão bibliográfica

**Aprendizado profundo na detecção de anomalias em dados ferroviários**

Orientador: Prof. Dr. Cassio Machiaveli Oishi

Autor: Miguel de Campos Rodrigues Moret

Presidente Prudente, 20 de novembro de 2025

## **Aprendizado profundo na detecção de anomalias em dados ferroviários**

Revisão bibliográfica apresentado ao curso de Ciência da Computação do Departamento de Matemática e Computação da Universidade Estadual Júlio de Mesquita Filho - Faculdade de Ciências e Tecnologia como requisito para a aprovação na disciplina de Projeto Científico I.

Presidente Prudente – São Paulo  
2025

# **Aprendizado profundo na detecção de anomalias em dados ferroviários**

Trabalho aprovado, Presidente Prudente, 2025:

---

**Prof. Dr. Cassio Machiaveli Oishi**

Orientador

---

**Prof. Dr. Nome Avaliador**

Convidado 1

Presidente Prudente – São Paulo

2025

## **RESUMO**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla tempor velit enim, vitae imperdiet ligula finibus eget. Vestibulum tristique aliquet lectus sit amet fermentum. Praesent sapien eros, porttitor vitae dictum eget, posuere sed nunc. Vestibulum nec purus augue. Aliquam erat volutpat. Aenean imperdiet sapien scelerisque, sagittis mi id, cursus nisl. Integer vel arcu vitae nibh porta ultricies. Sed sem nunc, ornare lobortis maximus ornare, convallis vitae lectus. Cras vitae nunc dictum, auctor urna non, posuere felis. Donec vulputate enim magna, vitae euismod turpis pellentesque ac.

## **PALAVRAS-CHAVES**

Palavra, Palavra, Palavra, Palavra, Palavra, Palavra, Palavra.

## **ABSTRACT**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla tempor velit enim, vitae imperdiet ligula finibus eget. Vestibulum tristique aliquet lectus sit amet fermentum. Praesent sapien eros, porttitor vitae dictum eget, posuere sed nunc. Vestibulum nec purus augue. Aliquam erat volutpat. Aenean imperdiet sapien scelerisque, sagittis mi id, cursus nisl. Integer vel arcu vitae nibh porta ultricies. Sed sem nunc, ornare lobortis maximus ornare, convallis vitae lectus. Cras vitae nunc dictum, auctor urna non, posuere felis. Donec vulputate enim magna, vitae euismod turpis pellentesque ac.

## **KEYWORDS**

Word, Word, Word, Word, Word, Word, Word, Word.

## **LISTA DE FIGURAS**

1	Imagen mostrando as múltiplas camadas de um modelo de aprendizado profundo	9
2	Exemplificação de transferencia de aprendizado . . . . .	10
3	Exemplificação de uma rede neural convolucional . . . . .	11
4	Exemplificação do funcionamento de um transformador visual . . . . .	12
5	Conjunto de imagens do segundo dataset . . . . .	13
6	Conjunto de imagens do dataset em grade 2x2 . . . . .	14
7	Imagen mostrando o exemplo de uma matriz de confusão . . . . .	16

## SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>8</b>
1.1	Justificativa . . . . .	8
1.2	Objetivos . . . . .	8
<b>2</b>	<b>Fundamentação teórica</b>	<b>8</b>
2.1	Aprendizado profundo . . . . .	8
2.2	Transferencia de aprendizado . . . . .	9
2.3	Rede neural convolucional . . . . .	10
2.4	Transformadores visuais . . . . .	11
<b>3</b>	<b>Metodologia</b>	<b>12</b>
3.1	Datasets . . . . .	12
3.1.1	informações sobre os datasets . . . . .	13
3.2	Modelos . . . . .	14
3.2.1	CNN pré-treinada . . . . .	14
3.2.2	SVM . . . . .	15
3.2.3	ViT e DeiT . . . . .	15
3.3	Resultados Esperados . . . . .	15
3.4	Avaliação dos Resultados . . . . .	15
<b>4</b>	<b>Referências</b>	<b>17</b>

## 1 Introdução

### 1.1 Justificativa

Diante desses avanços, este trabalho concentra-se na investigação de métodos de aprendizado profundo aplicados à detecção automática de anomalias em trilhos ferroviários a partir de imagens. O objetivo é avaliar diferentes modelos, analisar seu desempenho e discutir suas potencialidades como ferramentas de apoio à manutenção preventiva da infraestrutura ferroviária. Ao reunir e analisar a literatura existente, busca-se destacar tendências, limitações e oportunidades de pesquisa, contribuindo para o desenvolvimento de soluções mais eficientes e acessíveis para o setor.

### 1.2 Objetivos

Sendo assim, este trabalho tem como objetivo investigar o uso de técnicas de aprendizado profundo para a detecção de anomalias em dados ferroviários, buscando desenvolver e avaliar modelos capazes de identificar automaticamente falhas estruturais em trilhos por meio de imagens. Para isso, serão conduzidas etapas de pré-processamento e análise exploratória dos dados, implementação de diferentes arquiteturas de aprendizado profundo, avaliação de desempenho por meio de métricas apropriadas e comparação dos resultados com estudos relacionados da literatura.

## 2 Fundamentação teórica

### 2.1 Aprendizado profundo

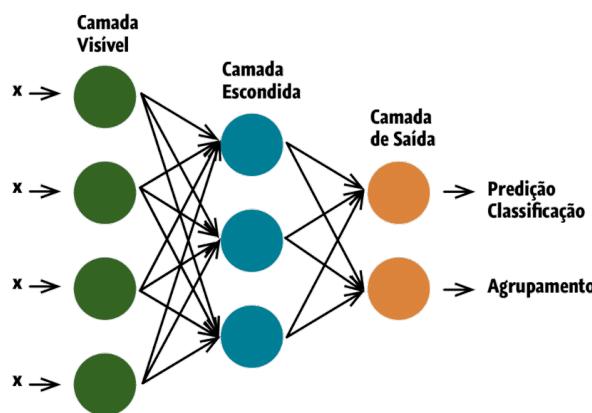
O aprendizado profundo é um subcampo do aprendizado de máquina que se baseia no uso de redes neurais artificiais com múltiplas camadas, capazes de modelar relações complexas em dados. Ao contrário de algoritmos tradicionais de aprendizado de máquina, que frequentemente dependem de extração manual de características, o aprendizado profundo consegue aprender automaticamente representações hierárquicas dos dados.

No contexto de imagens, por exemplo, redes profundas podem aprender a reconhecer padrões visuais em diferentes níveis de abstração: camadas iniciais detectam bordas e contornos, camadas intermediárias capturam formas e texturas, e camadas mais profundas identificam objetos completos ou padrões de defeitos. Fazendo com que o modelo identifique até mesmo

defeitos sutis nas imagens.

As redes profundas são compostas por várias unidades chamadas neurônios artificiais, organizadas em camadas. Cada neurônio realiza operações matemáticas sobre os dados de entrada, aplicando pesos e funções de ativação não lineares, o que possibilita à rede modelar relações complexas entre variáveis. O treinamento de uma rede profunda envolve a otimização desses pesos, geralmente por meio do algoritmo de retropropagação, que ajusta os parâmetros para minimizar o erro entre a saída prevista e a saída real.

Figura 1: Imagem mostrando as múltiplas camadas de um modelo de aprendizado profundo



Fonte: Scientific Figure on ResearchGate (s.d.)

## 2.2 Transferencia de aprendizado

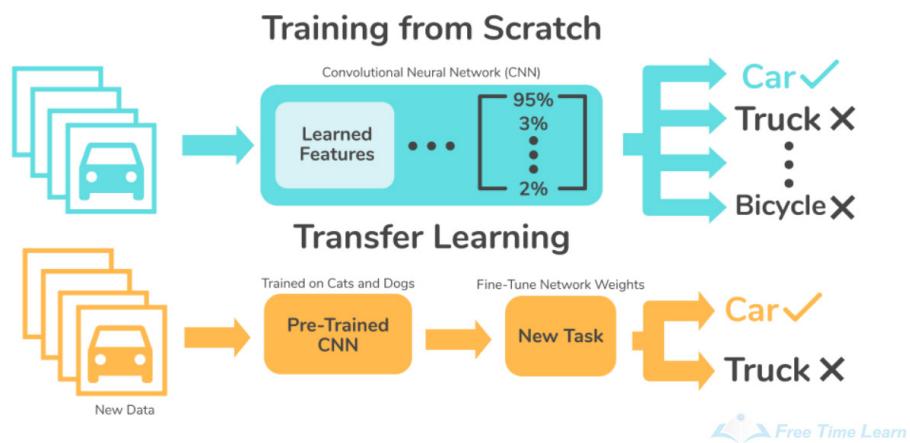
Um dos desafios em aprendizado profundo é a necessidade de grandes quantidades de dados rotulados para treinar redes neurais com bom desempenho. Em tarefas específicas, como a detecção de defeitos em trilhos ferroviários, os datasets podem ser pequenos, tornando difícil o treinamento de redes do zero. Nesse contexto a transferência de aprendizado acaba se tornando uma solução eficiente.

A transferência de aprendizado consiste em aproveitar o conhecimento adquirido por um modelo previamente treinado em um grande dataset genérico, como ImageNet ou COCO, e aplicá-lo a uma nova tarefa com dados limitados. O modelo pré-treinado já aprendeu a identificar padrões visuais fundamentais, como bordas, formas e texturas, que podem ser úteis em diversos domínios visuais.

Na prática, o funcionamento da transferência de aprendizado pode ser dividido em um processo em duas etapas:

1. **Reaproveitamento do conhecimento existente:** o modelo pré-treinado mantém sua capacidade de reconhecer padrões aprendidos anteriormente, funcionando como uma base de conhecimento para outros modelos.
2. **Adaptação à nova tarefa:** o modelo é ajustado para identificar padrões específicos da nova aplicação. Permitindo que o modelo aprenda o que é relevante para a nova tarefa sem precisar começar do zero, fazendo com que ele aprenda mais, mesmo possuindo poucos dados.

Figura 2: Exemplificação de transferencia de aprendizado



Fonte: FreeTime Learning (s.d.)

### 2.3 Rede neural convolucional

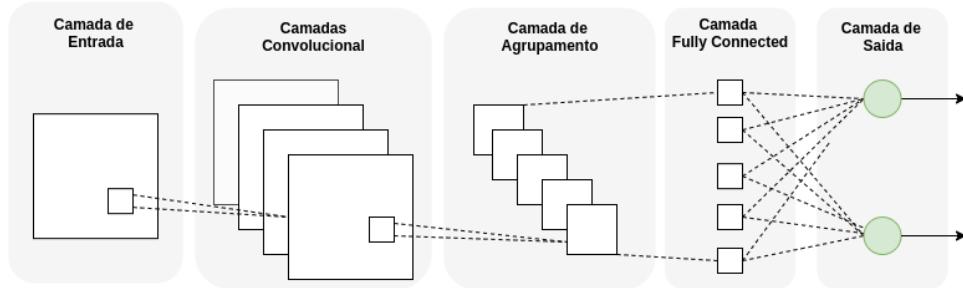
As Redes Neurais Convolucionais (CNNs) são uma arquitetura específica de aprendizado profundo projetada para lidar com dados estruturados espacialmente, como imagens. Diferentemente de redes totalmente conectadas, onde cada neurônio se conecta a todos os neurônios da camada anterior, as CNNs utilizam camadas convolucionais que aplicam filtros (kernels) a regiões locais da imagem, extraíndo características relevantes com compartilhamento de pesos.

Uma CNN típica é composta por três tipos de camadas principais:

- **Camadas Convolucionais:** Aplicam filtros que detectam padrões locais, como bordas, curvas e texturas. Cada filtro responde a um tipo específico de característica, produzindo mapas de ativação que destacam essas informações.

- **Camadas de Agrupamento:** Reduzem a dimensionalidade dos mapas de ativação, preservando informações importantes e tornando o modelo menos sensível a pequenas variações na posição dos objetos.
- **Camadas Totalmente Conectadas:** Recebem as características extraídas e produzem a saída final.

Figura 3: Exemplificação de uma rede neural convolucional



Fonte: Barbosa *et al.* (2021)

No caso da detecção de anomalias em trilhos ferroviários, as CNNs permitem que o modelo aprenda automaticamente padrões de falhas, como deformações na alma do trilho, fissuras ou ausência de dormentes, sem a necessidade de extração manual de características. A capacidade de capturar hierarquias de features locais torna as CNNs particularmente eficientes para imagens de alta resolução, mesmo em datasets de tamanho moderado.

## 2.4 Transformadores visuais

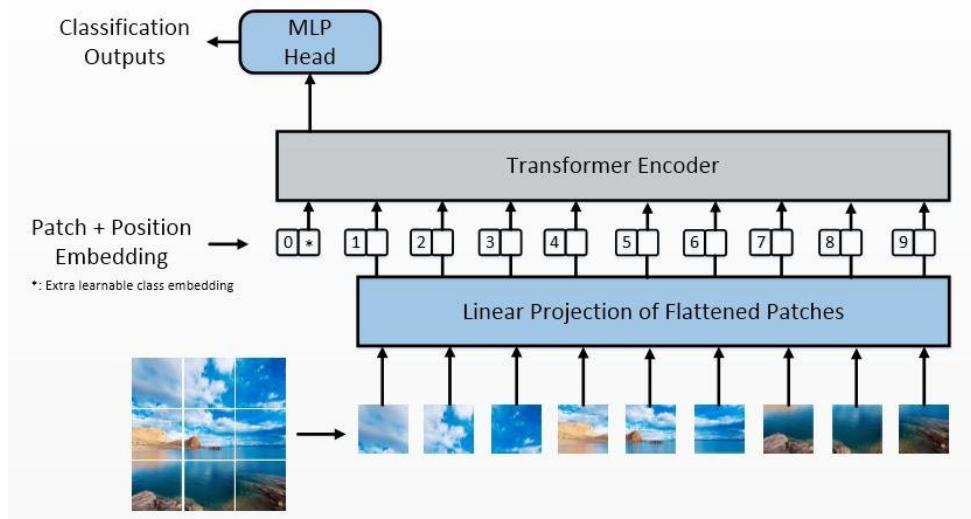
Os Transformadores Visuais (ViTs) representam uma abordagem mais recente no processamento de imagens, inspirada nos Transformers utilizados originalmente em tarefas de linguagem natural. Enquanto as CNNs focam em padrões locais, os transformadores capturam relações globais na imagem por meio de mecanismos de atenção, permitindo que o modelo considere o contexto completo ao tomar decisões.

O funcionamento básico de um Vision Transformer envolve:

- Divisão da imagem em patches menores.
- Transformação de cada patch em vetores de embedding, que codificam informações visuais.

- Passagem dos embeddings por camadas de autoatenção, que ponderam a importância relativa de cada patch em relação aos outros, integrando contexto global.
- Classificação ou detecção baseada nos embeddings processados, permitindo identificar padrões complexos e sutis, como pequenas fissuras ou deformações.

Figura 4: Exemplificação do funcionamento de um transformador visual



Fonte: Liang, Wang e Ling (2021)

Transformadores visuais têm se mostrado altamente eficazes em tarefas onde a variabilidade de cenários é grande ou os datasets são limitados, pois conseguem aproveitar melhor a informação contextual e evitar confusões causadas por objetos próximos ou fundos complexos.

Como observado no trabalho de Shahin et al, modelos híbridos, combinando CNNs e ViTs, também têm sido utilizados para aproveitar a extração local de CNNs e a consciência global dos transformadores, alcançando desempenho superior em detecção de defeitos ferroviários. Isso é feito, treinando CNNs em grandes datasets de imagens e utilizando-as para treinar os ViTs.

### 3 Metodologia

#### 3.1 Datasets

Um dos datasets utilizados neste trabalho será *Railway Track fault Detection Resized* (224 X 224), disponibilizado na plataforma Kaggle por Gerry *et al.* (2022). Trata-se de uma versão re-

dimensionada do conjunto original *Railway Track Fault Detection* disponibilizado por Hossain *et al.* (2021), também na plataforma Kaggle.

As imagens originais possuem alta resolução, o que torna o pré-processamento custoso, por esse motivo o dataset foi recriado com todas as imagens reduzidas para 244x244x3, facilitando seu uso, também foi incluído um arquivo *rails.csv*, que simplifica o carregamento e a organização das amostras (Gerry *et al.*, 2022).

Também será utilizado o dataset *Railway Track Fault Detection | Dataset2(Fastener)*, disponibilizado na plataforma Kaggle por Adnan (2021). Porém, ao contrário do conjunto anterior, esse conjunto contém apenas imagens de pinos de fixação e não foi redimensionado.

### 3.1.1 informações sobre os datasets

Ambos os datasets são compostos pelo mesmo número de imagens com e sem anomalias.

O conjunto de Gerry *et al.* (2022) contém 384 imagens de trilhos divididas entre 300 imagens de treinamento, 22 imagens de teste e 62 imagens de validação. Esse conjunto contém três tipos de anomalias: ausência de pinos de fixação, deformações e rachaduras nos trilho e danos na fundação do trilho (parte de madeira transversal ao trilho).

Enquanto que o conjunto de Adnan (2021) contém 1399 imagens de pinos de fixação divididas entre 980 imagens de treinamento, 140 imagens de teste 280 imagens de validação

Figura 5: Conjunto de imagens do segundo dataset

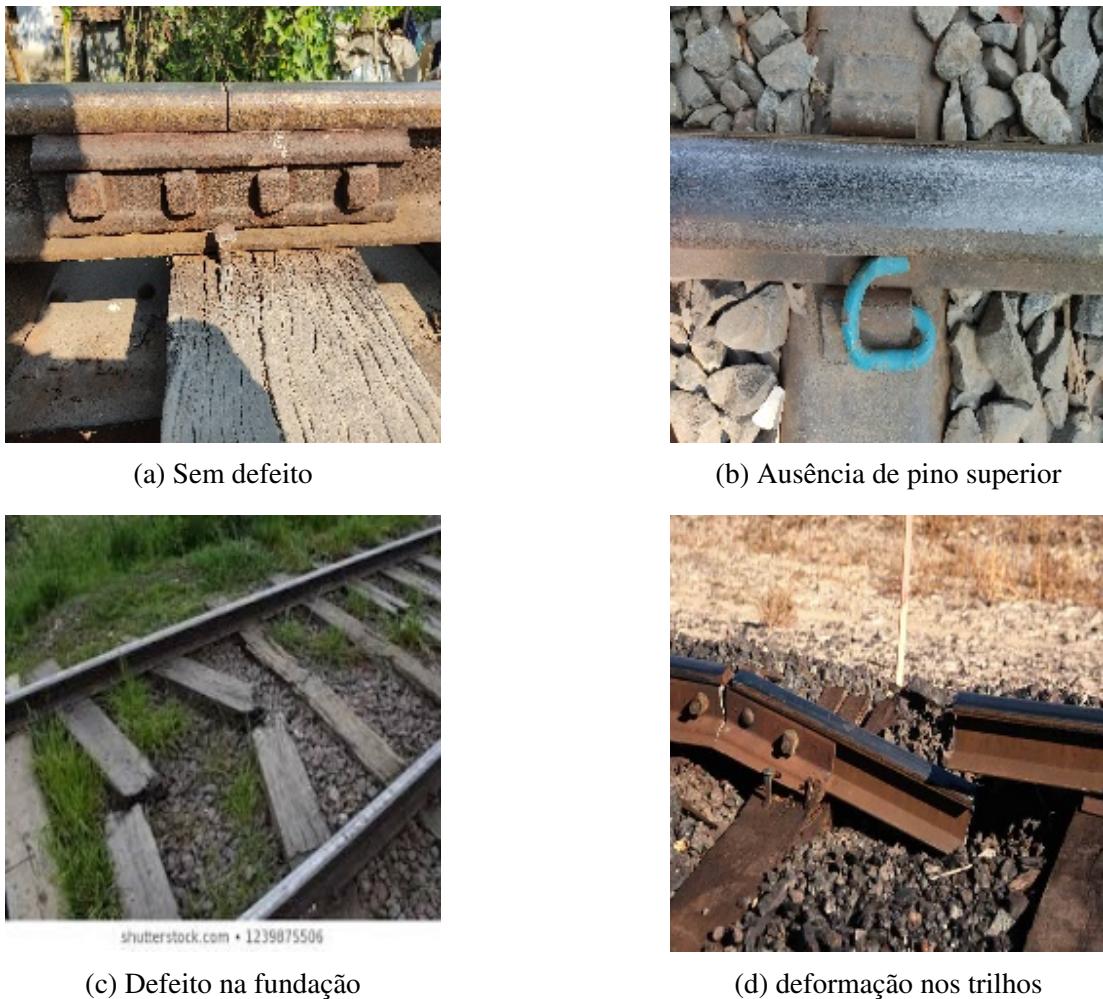


(a) Sem defeito



(b) Ausência de pino superior

Figura 6: Conjunto de imagens do dataset em grade 2x2



Fonte: Gerry *et al.* (2022)

### 3.2 Modelos

#### 3.2.1 CNN pré-treinada

Será utilizada uma rede neural convolucionais pré-treinada para aproveitar o conhecimento adquirido por grandes modelos treinados em bases extensas como ImageNet e minimizar os impactos causados pelo uso de um dataset reduzido.

A arquitetura escolhida foi a MobileNetV3, sendo esta uma melhoria do mobileNetV2 utilizada por Zheng *et al.* (2021) em seu trabalho de identificação de defeitos em pinos de fixação e amplamente recomendada para tratar de datasets pequenos, como é o caso.

### 3.2.2 SVM

Neste modelo híbrido, uma CNN pré-treinada será utilizada como extratora de características. As últimas camadas convolucionais produzem vetores representativos do conteúdo visual da imagem, que são utilizados como entrada para a SVM.

### 3.2.3 ViT e DeiT

## 3.3 Resultados Esperados

Espera-se que os modelos desenvolvidos sejam capazes de identificar com elevada precisão os diferentes tipos de defeitos presentes nas imagens do conjunto de dados. Em particular, os resultados esperados incluem:

- Os modelos baseados em aprendizado profundo obtenham desempenho superior aos métodos clássicos.
- As arquiteturas que utilizam transferência de aprendizado apresentem resultados mais estáveis e eficientes mesmo com um conjunto de dados limitado.
- Os modelos sejam capazes de generalizar para imagens de contextos distintos, mantendo desempenho consistente mesmo diante de variações de iluminação, ângulo ou ruído.
- A análise comparativa permita identificar quais modelos são mais adequados para detecção automática de defeitos ferroviários e justificar suas vantagens técnicas.

## 3.4 Avaliação dos Resultados

Os resultados dos testes serão obtidos através de uma matriz de confusão, que organiza as predições do modelo em quatro categorias fundamentais: verdadeiros positivos (TP) falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). Essa estrutura permite avaliar não apenas o desempenho geral do classificador, mas também entender em quais situações ele tende a errar.

Figura 7: Imagem mostrando o exemplo de uma matriz de confusão

		Valores Reais	
		Positivo (1)	Negativo (0)
Valores Previstos	Positivo (1)	$TP$ (Verdadeiro Positivo)	$FP$ (Falso Positivo)
	Negativo (0)	$FN$ (Falso Negativo)	$TN$ (Verdadeiro Negativo)

Fonte: Lopes (2023)

A partir da matriz de confusão, quatro métricas amplamente utilizadas na avaliação de modelos de classificação serão calculadas:

- **Acurácia** (Accuracy)

Mede a proporção de classificações corretas em relação ao total de previsões realizadas, sendo utilizada para analisar o desempenho dos modelos de forma geral.

$$A = \frac{TP + TN}{TP + FP + FN + TN}$$

- **Precisão** (Precision)

Mede a proporção de positivos verdadeiros entre todas as previsões positivas, sendo utilizada para avaliar o grau de confiabilidade das detecções positivas.

$$P = \frac{TP}{TP + FP}$$

- **Revocação** (Recall)

Mede a capacidade do modelo de recuperar os positivos reais, avaliando o quanto o modelo deixa de detectar casos positivos.

$$R = \frac{TP}{TP + FN}$$

- **F1-score**

Média harmônica entre precisão e revocação. Resume ambas as métricas em um único

valor equilibrado.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

Após a obtenção dos resultados, será realizada uma comparação com os estudos presentes na literatura, analisando como os diferentes modelos se comportaram no contexto deste trabalho em relação ao desempenho observado em suas aplicações originais. Essa análise permitirá identificar em que medida fatores como domínio domínio dos dados, condições de captura das imagens e complexidade das anomalias influenciam o desempenho das arquiteturas, além de evidenciar quais modelos apresentam maior capacidade de generalização para o cenário de detecção de falhas em trilhos ferroviários.

#### 4 Referências

ADNAN, Ashik. **Railway Track Fault Detection Dataset (Fastener Version)**. [S. l.: s. n.], 2021. Kaggle. Disponível em: <https://www.kaggle.com/datasets/ashikadnan/railway-track-fault-detection-dataset2fastener>. Acesso em: 20 nov. 2025.

BARBOSA, Guilherme *et al.* Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina. In: [s. l.: s. n.], out. 2021. p. 145–189. ISBN 9786587003658. DOI: 10.5753/sbc.7165.8.4.

FREETIME LEARNING. **Explain Transfer Learning in the Context of Deep Learning**. [S. l.: s. n.]. <https://www.freetimelearning.com/software-interview-questions-and-answers.php?Explain-transfer-learning-in-the-context-of-deep-learning.&id=4184>. Accessed: 2025-11-20.

GERRY *et al.* **Railway Track Fault Detection (Resized 224x224)**. [S. l.: s. n.], 2022. Kaggle Dataset. Accessed: 20 Nov. 2025. Disponível em: <https://www.kaggle.com/datasets/gpiosenka/railway-track-fault-detection-resized-224-x-224>.

HOSSAIN, Shahriar *et al.* **Railway Track Fault Detection**. [S. l.]: Kaggle, 2021. Kaggle Dataset. Accessed: 20 Nov. 2025. DOI: 10.34740/KAGGLE/DSV/1884733. Disponível em: <https://www.kaggle.com/dsv/1884733>.

LIANG, Jingxin; WANG, Dong; LING, Xufeng. Image Classification for Soybean and Weeds Based on ViT. **Journal of Physics: Conference Series**, v. 2002, p. 012068, ago. 2021. DOI: 10.1088/1742-6596/2002/1/012068.

LOPES, André. **Medidas de performance de modelos de classificação**. [S. l.: s. n.], 2023. <https://brains.dev/2023/medidas-de-performance-modelos-de-classificacao/>. Acessado em: 20 Nov 2025.

SCIENTIFIC FIGURE ON RESEARCHGATE. **Inteligência Artificial: Uma era de abundância ou o fim da espécie humana? – Visão geral das redes neurais no aprendizado profundo.** [S. l.: s. n.]. [https://www.researchgate.net/figure/Figura-4-Visao-geral-das-redes-neurais-no-aprendizado-profundo\\_fig4\\_319212925](https://www.researchgate.net/figure/Figura-4-Visao-geral-das-redes-neurais-no-aprendizado-profundo_fig4_319212925). Acessado em: 20 Nov 2025.

ZHENG, Danyang *et al.* A Defect Detection Method for Rail Surface and Fasteners Based on Deep Convolutional Neural Network. **Computational Intelligence and Neuroscience**, v. 2021, n. 1, p. 2565500, 2021. DOI: <https://doi.org/10.1155/2021/2565500>. eprint: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2021/2565500>.