TRABALHO TÉCNICO-CIENTÍFICO

DETECÇÃO INTELIGENTE DE PATOLOGIAS EM PAVIMENTOS RODOVIÁRIOS: APLICAÇÃO PRÁTICA COM REDES NEURAIS CONVOLUCIONAIS

Bruna Tamanini Barcelos[1]; Kaio Gefferson de Almeida Mesquita[1]; Frederico Rodrigues[1]; Ludmila Ferreira[1] & Aleandro Matteoni[1]

RESUMO

O sistema rodoviário brasileiro é o principal elo da infraestrutura de transporte nacional, responsável por mais de 60% do escoamento de cargas e 95% do transporte de passageiros. Contudo, grande parte dessa infraestrutura apresenta deficiências estruturais, como trincas, fissuras e buracos, que comprometem a segurança, o conforto e a durabilidade dos pavimentos. A identificação correta dessas patologias é essencial para o planejamento de intervenções de manutenção. No entanto, os métodos convencionais, baseados em inspeções visuais, são onerosos, demorados e sujeitos à subjetividade do avaliador. Neste contexto, este trabalho apresenta uma solução baseada em redes neurais convolucionais (CNN) para a detecção automatizada de defeitos em pavimentos rodoviários, a partir de imagens obtidas com câmeras embarcadas. O modelo foi desenvolvido com o algoritmo YOLOv7, utilizando arquitetura de múltiplos rótulos, permitindo a identificação simultânea de diferentes tipos de patologias. Foram treinados dois modelos distintos: um com classes agrupadas de defeitos e outro com classes desagrupadas. As classes foram definidas conforme terminologia da norma DNIT 005/2003-TER. A avaliação comparativa demonstrou que o modelo com classes agrupadas apresentou melhor desempenho, com aumento nos indicadores de acurácia, *recall* e média de precisão (mAP), superando o modelo desagrupado em até 91% no mAP. Isso evidencia que a redução da complexidade na terminologia dos defeitos melhora a capacidade preditiva da rede. Os resultados indicam que a ferramenta tem grande potencial de aplicação prática na gestão de pavimentos, proporcionando agilidade, padronização e redução de custos operacionais. A continuidade do trabalho prevê o uso de bancos de dados mais diversos, contemplando diferentes tipos de pavimento e regiões do país, para aprimorar a generalização do modelo em escala nacional.

PALAVRAS-CHAVE: Pavimentos; Detecção de Patologias; Inteligência Artificial; Visão Computacional.

ABSTRACT

The Brazilian road system is the main link in the national transportation infrastructure, responsible for more than 60% of cargo flow and 95% of passenger transport. However, a large part of this infrastructure has structural deficiencies, such as cracks, fissures and potholes, which compromise the safety, comfort and durability of the pavements. The correct identification of these pathologies is essential for planning maintenance interventions. However, conventional methods, based on visual inspections, are expensive, time-consuming and subject to the subjectivity of the evaluated person. In this context, this work presents a solution based on convolutional neural networks (CNN) for the automated detection of defects in road pavements, from images obtained with on-board cameras. The model was developed with the YOLOv7 algorithm, using a multi-label architecture, allowing the simultaneous identification of different types of pathologies. Two distinct models were trained: one with grouped classes of defects and another with ungrouped classes. The classes were defined according to the terminology of the DNIT 005/2003-TER standard. The comparative evaluation revealed that the model with grouped classes presented better performance, with an increase in the accuracy, recall and average precision (mAP) indicators, outperforming the ungrouped model by up to 91% in the mAP. This shows that the reduction in complexity in the defect terminology improved the predictive capacity of the network. The results indicate that the tool has great potential for practical application in pavement management, providing agility, standardization and reduction of operational costs. The continuation of the work foresees the use of more diverse databases, covering different types of pavement and regions of the country, to improve the generalization of the model on a national scale.

KEY WORDS: Pavements; Pathology Detection; Artificial Intelligence; Computer Vision.

[1] Imtraff Engenharia e Mobilidade, e-mail:bruna.barcelos@grupoimtraff.com.br; kaio.mesquita@grupoimtraff.com.br; frederico@grupoimtraff.com.br, ludmila.ferreira@grupoimtraff.com.br; aleandro.matteoni@grupoimtraff.com.br.

INTRODUÇÃO

As infraestruturas de transporte constituem um dos principais ativos estratégicos para o desenvolvimento econômico, social e logístico de um país. No Brasil, cuja matriz de transporte é fortemente dependente do modal rodoviário, responsável por 61% da movimentação de cargas e 95% do transporte de passageiros, a qualidade da malha viária é determinante para a competitividade nacional (ILOS, 2019). No entanto, apesar de o país contar com uma extensa rede rodoviária, apenas 12,4% encontra-se pavimentada. A Pesquisa CNT (2024) revela que mais da metade dos trechos pavimentados apresenta condições inadequadas de trafegabilidade, sendo 34,7% classificados como regulares, 16,3% como ruins e 5,9% como péssimos, o que evidencia a necessidade de soluções modernas e eficazes para inspeção e manutenção dos pavimentos.

Um dos principais desafios da infraestrutura rodoviária está na identificação precisa dos defeitos, fator que impacta diretamente a tomada de decisão quanto às ações de manutenção (DNIT, 2010). Predominantemente, as metodologias normatizadas no Brasil para avaliação de defeitos em pavimentos ainda se baseiam em inspeções presenciais, realizadas por técnicos especializados. Embora a norma mais recente permita o uso de equipamentos de escaneamento automatizado, seu alto custo limita a aplicação em larga escala.

Nesse contexto, a automação baseada em visão computacional surge como uma alternativa promissora para ampliar a cobertura e a precisão das avaliações (Carvalho, R. B. et al., 2022). Com os avanços tecnológicos, há uma tendência crescente de utilização de levantamentos automatizados com equipamentos que vão desde câmeras a sensores diversos, integrando técnicas de visão computacional (Massucatto, 2019).

Nos últimos oito anos, as redes neurais convolucionais (CNNs), uma categoria específica de redes neurais dentro do campo do aprendizado profundo (*Deep Learning* - DL), ganharam destaque, especialmente em aplicações de visão computacional. As CNNs são projetadas para classificar imagens, sendo capazes de analisar grandes volumes de dados e reconhecer padrões de forma contínua. Na engenharia de pavimentos, a aplicação dessas técnicas vem evoluindo com os avanços em *hardware* e *software* (Gopalakrishnan, 2018).

A partir de uma revisão da literatura sobre o uso de aprendizado profundo na gestão de pavimentos, observou-se que os principais focos têm sido a identificação de defeitos em rodovias pavimentadas. Os estudos concentram-se, principalmente, na classificação, detecção e segmentação de trincas e panelas (Gopalakrishnan et al., 2017; Araújo, 2023), com menor ênfase na detecção de outras patologias menos frequentes e na redução de erros de classificação. Verificou-se, assim, uma limitação metodológica nas tecnologias existentes: os modelos de identificação de defeitos, devido à grande quantidade de classes, tendem a perder precisão durante o treinamento e não generalizam bem para novos casos, resultando em um fenômeno conhecido como *overfitting* (Géron, 2019).

Diante desse cenário, o presente trabalho propõe um método de treinamento e aplicação de uma rede neural convolucional com múltiplas etiquetas, capaz de identificar patologias em pavimentos rodoviários em tempo real. O objetivo é reduzir os erros de classificação entre diferentes tipos de defeitos e automatizar o processo de diagnóstico. Foram comparados dois modelos: um com classes de defeitos agrupadas e outro com classes desagrupadas. Os modelos foram treinados com base na terminologia da norma DNIT 005/2003-TER, utilizando imagens reais de rodovias brasileiras. Os resultados foram comparados com classificações manuais, a fim de aferir os indicadores de acurácia, recall e média de precisão.

REFERENCIAL TEÓRICO

Avaliação Funcional de Pavimentos Rodoviários

Os pavimentos rodoviários desempenham um papel essencial na infraestrutura de qualquer sociedade, sendo responsáveis por garantir a mobilidade, a segurança e a integração econômica. No entanto, sua durabilidade pode ser comprometida por diversos fatores, como o tráfego intenso, o envelhecimento dos materiais e as variações climáticas. Esse cenário reforça a necessidade de estratégias eficazes de monitoração e gestão, cujo objetivo principal é assegurar a integridade estrutural, prolongar a vida útil do pavimento e preservar a segurança dos usuários (Cagle, 2003).

A detecção precoce de defeitos torna-se, portanto, uma etapa crítica. Trincas, buracos, desgastes e deformações estão entre as falhas mais recorrentes, cuja evolução pode comprometer significativamente a segurança viária e reduzir a longevidade do pavimento (Yang et al., 2020). O método mais comum para avaliação da superfície dos pavimentos ainda é a inspeção visual, normatizada e amplamente utilizada. Embora as normas técnicas ofereçam diretrizes consistentes para esse tipo de atividade, sua aplicação prática enfrenta diversos desafios: trata-se de um processo dispendioso, especialmente quando se consideram longas extensões de rodovias, além de expor os técnicos aos riscos inerentes ao ambiente rodoviário.

Adicionalmente, o julgamento humano introduz um componente subjetivo à avaliação, o que pode comprometer a padronização e a precisão dos resultados. Nesse contexto, e diante dos avanços recentes em tecnologia, observa-se um interesse crescente, tanto na indústria quanto na pesquisa acadêmica, pelo desenvolvimento de soluções automatizadas para a detecção de defeitos e patologias em pavimentos (Cha et al., 2017). Entre essas abordagens, aquelas baseadas em imagens destacam-se por sua acessibilidade, graças à ampla disponibilidade de câmeras e *smartphones* (Jahanshahi et al., 2009).

Avanços com Visão Computacional

Estudos de detecção de patologias em pavimentos com base em imagens são, em geral, classificados em abordagens manuais e abordagens automáticas baseadas na extração de características. Os primeiros trabalhos concentraram-se em métodos como detectores de bordas (Abdel-Qader et al., 2003), operações morfológicas (Cord e Chambon, 2011) e técnicas de limiarização (Oliveira e Correia, 2009). As características extraídas manualmente são então inseridas em algoritmos clássicos de aprendizado de máquina, como máquinas de vetores de suporte (SVM), *Random Forest* e redes neurais artificiais (NN).

Com o avanço da tecnologia, diversos estudos passaram a adotar técnicas baseadas em aprendizado profundo, os quais se dividem em classificação de imagens, reconhecimento de objetos e segmentação semântica (Hsien e Tsai, 2020). Os métodos baseados em aprendizado profundo focam na aprendizagem hierárquica de características, com níveis superiores formados por combinações de características de níveis inferiores. Essa estrutura permite a modelagem de funções complexas que relacionam diretamente os dados de entrada às saídas desejadas (Gopalakrishnan et al., 2017).

As CNNs, inspiradas no funcionamento do córtex visual humano, têm se destacado na tarefa de classificação de imagens (Massucatto, 2019). Essas redes aplicam filtros, aprendidos automaticamente, para identificar objetos nas imagens e distinguir entre diferentes classes (Academy, 2019).

Uma vantagem significativa das CNNs é a exigência reduzida de pré-processamento em comparação a outros algoritmos de classificação, o que torna o processamento computacional mais eficiente. A rede consegue reduzir as imagens sem perda significativa de informações essenciais. Entretanto, para que uma CNN atinja altos níveis de precisão nas previsões, geralmente são necessárias grandes quantidades de imagens etiquetadas, o que pode representar um desafio, especialmente pela complexidade e o tempo demandado na rotulagem manual.
Nesse contexto, a aprendizagem por transferência (*transfer learning*) surge como uma solução viável, permitindo o treinamento de modelos eficazes com menor volume de dados. Em vez de iniciar o processo do zero, essa abordagem utiliza modelos previamente treinados em grandes bases de dados, como o *ImageNet*, para resolver problemas semelhantes. Esses modelos empregam pesos pré-ajustados e uma taxa de aprendizagem adaptável ao longo do processo (Tajbakhsh, 2016).
A taxa de aprendizagem (*learning rate* – LR) é um parâmetro crucial nos algoritmos de otimização, pois determina o tamanho dos passos dados em direção ao mínimo de uma função de perda (Murphy, 2012). Smith (2017) propõe que, em vez de se adotar uma LR fixa ou com decaimento exponencial, essa taxa varie ciclicamente entre valores mínimo e máximo predefinidos. O autor sugere a realização de algumas épocas de treinamento iniciais para estimar esses limites, permitindo uma melhor adaptação do modelo durante o processo de otimização.

METODOLOGIA

Na metodologia deste trabalho foi proposto um desenvolvimento em 7 etapas, sendo elas: (i) Caracterização da rede neural; (ii) Separação do *dataset* de treinamento; (iii) Demarcação dos *labels* para treinamento; (iv) Treinamento do algoritmo; (v) Delimitação do controle manual; (vi) Aplicação do algoritmo; (vii) Extração dos indicadores e análise comparativa.
Todo o processo foi realizado em liguagem *python* 3.10 desenvolvido através do YOLO v7 (*You Only Look Once*), um *framework* implementado com aprendizado profundo de uma *darknet*. Baseado nesse algoritmo, foram realizadas implementações para otimização da velocidade e da detecção de várias patologias em uma mesma imagem, conforme será apresentado.

Rede Neural

A primeira etapa consiste na caracterização e descrição da rede neural convolucional. Posteriormente é necessário separar imagens amostrais para cada classe de patologia da rodovia em análise. Em seguida foi delimitado a demarcação dos *labels* para cada classe, de forma manual, depois de forma otimizada conforme o padrão de identificação. A ferramenta utilizada foi desenvolvida pelos autores com linguagem *python* e *framework pyqt5* para otimizar o processo.
O treinamento do algoritmo ocorreu em uma GPU de 16 núcleos e 32 GB de RAM, permitindo processamento paralelo e visualização dos indicadores enquanto ocorre o treinamento. O processo de treinamento neste caso se refere a calibração dos pesos de cada classe para cada função de ativação em cada camada da rede neural.
O processo é verificado através de uma função de perda (*Loss Function*), comparando a distribuição de probabilidade original com a distribuição treinada e diminuindo o erro a cada passagem (*epochs*) de imagens da amostra. Os pesos são atualizados com um processo de *backpropagation*, onde a derivada de cada função de ativação é calculada em relação a cada peso da função, de modo a tentar diminuir o erro na próxima passada, uma vez que a derivada tende a zero quando se encontra um ponto ótimo local ou global (Figura 2).

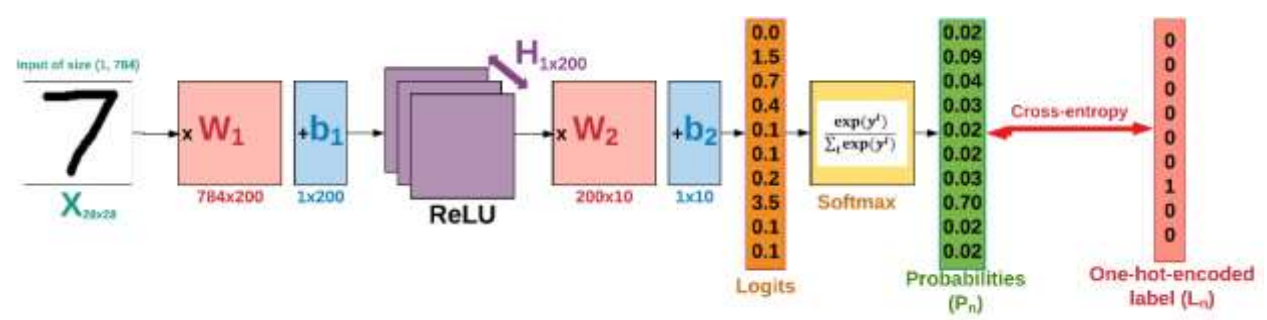

Figura 2. Estruturação do processo de identificação de objetos em uma imagem com CNN (Géron, 2019)

Para cada modelo aplicado foi selecionado um controle amostral de identificação manual, com a finalidade de comparar os resultados entre a percepção humana e da inteligência artificial, comparando a variação dos erros. Por fim, com o treinamento dos algoritmos foram extraídos os indicadores de *recall*, *acurracy* e média de precisão, e verificado a variação desses parâmetros.

O modelo de patologias agrupadas considerou 10 classes, já o modelo de patologias desagrupadas avaliou 25 classes. Essa variação se deve ao fato de que algumas classes apresentam variações, sendo muitas vezes difíceis de verificar até mesmo manualmente, portanto com a finalidade de verificar a hipótese de que classes agrupadas representam melhores resultados, agregou-se algumas classes e verificou-se os resultados de forma comparativa.

A escolha da estrutura em *darknet* se deu pelo fato do processamento de cada imagem chegar até 0,05 segundos e uma taxa de 30 *frames* por segundo. Uma das contribuições neste trabalho foi a utilização de pesos iniciais pré-treinados para identificação das patologias, antes mesmo do novo processo de treinamento. A inicialização não-aleatória dos pesos de cada camada, além de otimizar o processo de treinamento, pode garantir que o mesmo não estagne em um ótimo local, podendo alcançar o ponto ótimo global da função de ativação requisitada.

Estruturação do *Dataset* e Demarcação dos *Labels*

A base de dados de imagens é composta por registros fotográficos extraídos de vídeos captados em uma rodovia brasileira. Em conformidade com a Lei Geral de Proteção de Dados (LGPD), os detalhes específicos da rodovia não serão divulgados.

As gravações foram realizadas com uma câmera *GoPro Hero 10 Black*, posicionada frontalmente em um veículo em movimento. A resolução utilizada foi de 1920 × 1080 *pixels*, fixada em um veículo operando na velocidade da via 80km/h. Posteriormente, as imagens foram planificadas para a vertical de forma a possibilitar a visualização do pavimento, estando-a perpendicular ao mesmo (Figura 3).

Figura 3. *Dataset* e planificação das imagens (Autores, 2025)

Para a classe agrupada foram avaliadas 3886 imagens, para a classe desagrupada foram avaliadas 3910 imagens. A Figura 4 ilustra a demarcação dos *labels* no *dataset*. Os grupos demarcados para treinamento em cada modelo estão representados na Tabela 1, assim como o percentual de imagens por classe. A caracterização e siglas demarcadas nas patologias seguem a terminologia da norma DNIT 005/2003-TER, com exceção de TBE (trincas em bloco com erosão) devido ausência deste defeito na rodovia avaliada.

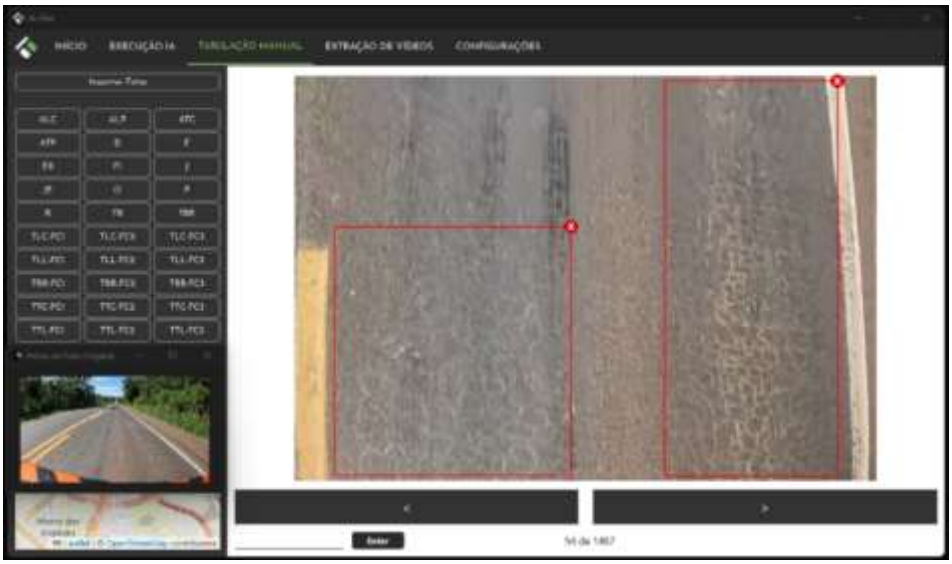

Figura 4. Exemplo de demarcação dos labels no dataset (Autores, 2025)

Tabela 1. Resumo dos dados de treinamento dos modelos agregado e desagregado (Autores, 2025)

Modelo Agrupado	Imagens	Proporção	Modelo Desagrupado	Imagens	Proporção
D	120	3,09%	D	120	3,07%
FI	341	8,78%	FI	341	8,72%
J	496	12,76%	J	496	12,69%
JE	61	1,57%	JE	61	1,56%
P	304	7,82%	P	304	7,77%
R	619	15,93%	R	619	15,83%
TTC	790	20,33%	TTC – FC1	416	10,64%
			TTC – FC2	348	8,90%
			TTC – FC3	26	0,66%
TTL	751	19,33%	TTL – FC1	399	10,20%
			TTL – FC2	312	7,98%
			TTL - FC3	40	1,02%
TLC	369	9,50%	TLC – FC1	120	3,07%
			TLC – FC2	239	6,11%
			TLC – FC3	10	0,26%
TLL	35	0,90%	TLL – FC1	4	0,10%
			TLL – FC2	24	0,61%
			TLL – FC3	7	0,18%
-			TB	2	0,05%
-			O	1	0,03%
-			E	11	0,28%
-			EX	5	0,13%
-			ALC	1	0,03%
-			ATC	4	0,10%

Treinamento da Rede Neural

O treinamento da rede neural convolucional envolve a aplicação de técnicas de otimização matemática por meio de uma função de custo (*loss function*), permitindo o ajuste automático dos vieses (*biases*) e dos pesos da rede. Nesse processo, o conjunto de dados de treinamento, com os rótulos esperados, é apresentado ao algoritmo de otimização com o objetivo de minimizar globalmente a função de perda.
O treinamento da CNN demandou elevado tempo computacional, o que tornou fundamental a utilização de pesos iniciais não aleatórios para acelerar o processo. Como alternativa às unidades de processamentos gráficos (GPUs), é possível empregar matrizes de portas programáveis em campo (FPGAs) ou circuitos integrados específicos de aplicação (ASICs), que oferecem soluções de *hardware* mais otimizadas para esse tipo de tarefa.
Em relação aos hiperparâmetros dos modelos, optou-se por mantê-los constantes em ambas as configurações, com o intuito de isolar e avaliar exclusivamente a influência dos dados de treinamento, sem interferência de fatores externos. Assim, foram definidos um *batch size* de 4 e um total de 350 épocas de treinamento.

Algoritmo em Detalhes e Indicadores

Neste estudo, as imagens foram processadas por meio da divisão em uma malha regular de 19 por 19 células. Cada célula é responsável por prever até cinco caixas delimitadoras (*bounding boxes*), permitindo a identificação de múltiplos objetos dentro de sua área. Para cada caixa, são atribuídas uma pontuação de confiança, que expressa a probabilidade da presença de um objeto, e valores de probabilidade para cada classe possível, funcionando também como classificador.
Ao todo, a malha gera 361 células, resultando em até 1805 caixas por imagem. No entanto, muitas dessas caixas apresentam baixa confiabilidade. Por essa razão, foi estabelecido um limite mínimo de 50% na pontuação de confiança para que as caixas fossem consideradas válidas, eliminando, assim, predições pouco confiáveis.
A escolha pelo algoritmo YOLO (*You Only Look Once*) deve-se à sua elevada eficiência: ele realiza a detecção de objetos em uma única etapa, sem a necessidade de fragmentar a imagem e aplicar classificadores repetidamente (Redmon, 2016). Essa abordagem torna o YOLO uma solução ágil e eficaz para tarefas de detecção em tempo real (Figura 5).

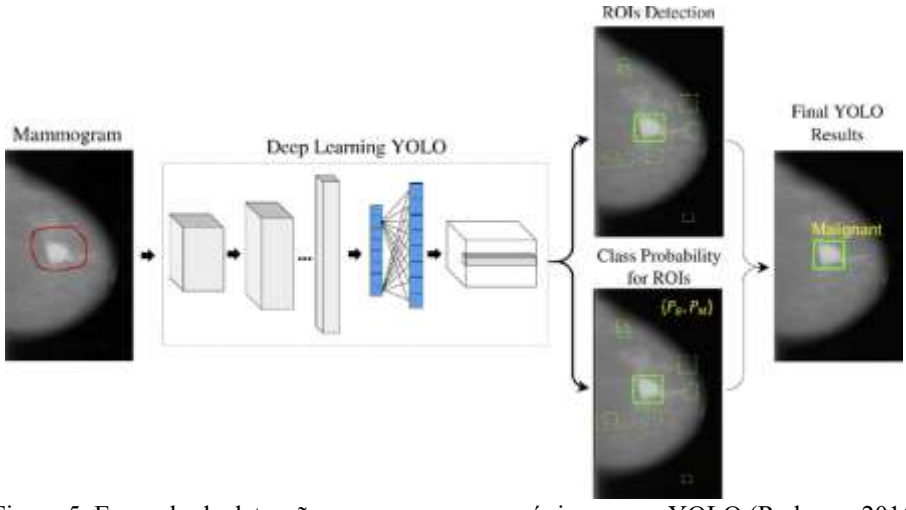

Figura 5. Exemplo de detecção em uma passagem única com o YOLO (Redmon, 2016)

Após a aplicação da rede neural, os modelos foram comparados, avaliando-se a capacidade preditiva dos modelos de classificação com os dados de teste, verificando se os modelos realmente conseguiram expandir aceitavelmente seus resultados de predição para conjuntos de dados desconhecidos. Os indicadores utilizados para avaliar o nível de previsão dos modelos de classificação foram o *precision, recall*, mAP e a acurácia.

O *recall* mede a capacidade do modelo de encontrar todos os exemplos positivos. É calculado como a proporção de previsões positivas corretas em relação ao número total de exemplos positivos no conjunto de dados. É a métrica a ser usada quando é importante minimizar falsos negativos. A precisão mede a capacidade do modelo de não rotular como positivo um exemplo negativo. É calculada como a proporção de previsões positivas corretas em relação ao número total de previsões positivas. Como diferentes classes de objetos podem ter diferentes níveis de dificuldade na detecção, o mAP calcula a média dos APs para todas as classes presentes no conjunto de dados. Isso dá uma visão geral do desempenho do modelo em todas as classes.

Além dos indicadores citados, também foi verificado a precisão dos modelos em relação a identificação manual para cada categoria (Jain, 1999; Sokolova, 2009; Powers, 2011).

RESULTADOS

As Figuras 6 e 7 foram geradas durante o processo de treinamento e expressam as métricas para avaliação da qualidade em identificar as patologias.

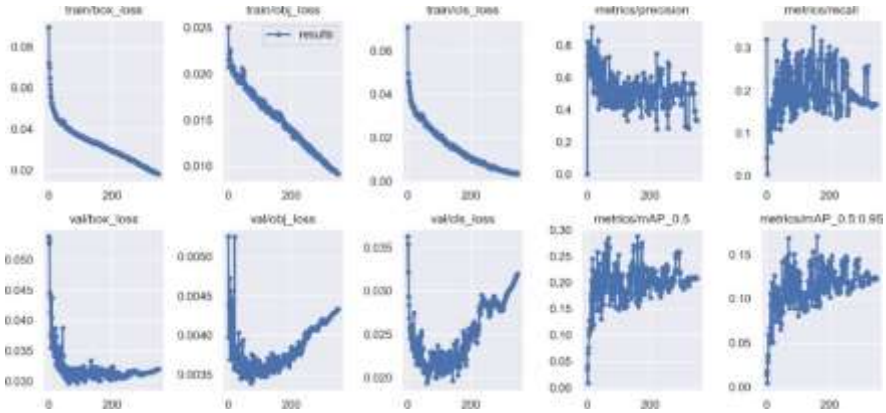

Figura 6. Resultado do treinamento no modelo desagrupado (Autores, 2025)

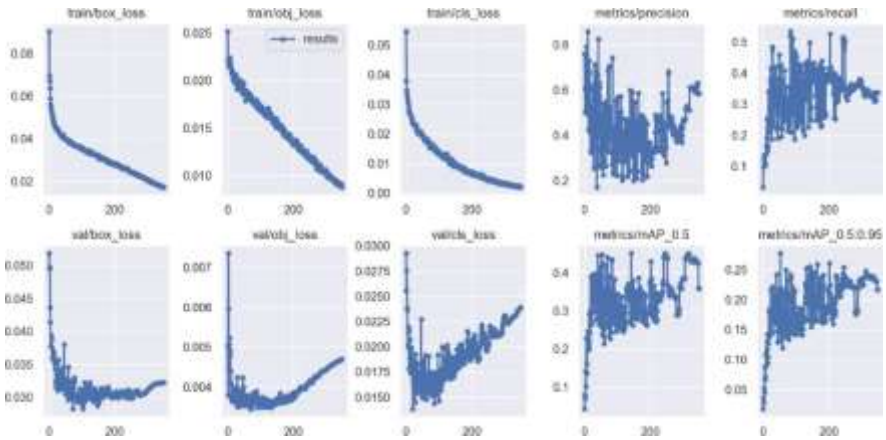

Figura 7. Resultado do treinamento no modelo agrupado (Autores, 2025)

Os resultados obtidos reforçam a hipótese de que o agrupamento de classes contribui significativamente para a melhoria do desempenho do modelo de detecção de patologias em pavimentos. De modo geral, as métricas avaliadas apresentaram valores superiores no modelo com classes agrupadas, em comparação àquele com classes desagrupadas.

Após 300 épocas de treinamento, observou-se estabilização do *recall* em torno de 0,18 para o modelo desagrupado e de 0,34 para o modelo agrupado, representando uma melhora relativa de aproximadamente 88%. Da mesma forma, a mAP apresentou incremento expressivo, com valores em torno de 0,12 para o modelo desagrupado e de 0,23 para o agrupado, um ganho de cerca de 91%. Esses resultados indicam que o agrupamento de subclasses semanticamente semelhantes, favorece o aumento da taxa de detecção correta de eventos positivos, ao minimizar ambiguidades entre categorias.

A aplicação prática dessa abordagem, no entanto, demanda a categorização prévia e manual dos tipos de defeitos, com foco direcionado por classe. Essa estratégia tende a reduzir a sobreposição de padrões visuais entre subclasses distintas e, consequentemente, a taxa de erros relacionados à similaridade entre manifestações patológicas.

Em relação à métrica de precisão, observou-se que o número de épocas utilizadas ainda não foi suficiente para sua estabilização em nenhum dos modelos. Apesar disso, os gráficos indicam uma tendência de crescimento na fase final do treinamento para o modelo com classes agrupadas, sugerindo potencial de melhoria com treinamento adicional.

A Figura 8 ilustra exemplos dos resultados obtidos na detecção por inteligência artificial realizada na prática, conforme a metodologia proposta. Verificou-se que a acurácia na identificação das categorias agrupadas foi superior em todas as classes avaliadas, exceto no caso da classe JE. Para essa classe específica, o modelo desagrupado apresentou desempenho superior, com uma redução de 38% na taxa de erro em relação ao controle.

Em particular, as classes Trincas Transversais e Trincas Longitudinais obtiveram taxas de acerto de 57% e 62%, respectivamente, evidenciando o potencial do modelo para a identificação eficaz dessas manifestações patológicas, mesmo em cenários de maior granularidade na classificação.

Figura 8. Aplicação do modelo (Autores, 2025)

A Tabela 2 apresenta os resultados comparativos entre os modelos de classificação por inteligência artificial, um utilizando categorias de patologias agrupadas e outro utilizando as categorias desagrupadas, confrontados com um grupo de controle obtido por contagem manual. Essa análise visa verificar os impactos práticos da estratégia de pré-processamento de classes sobre a performance dos modelos, especialmente quanto à acurácia e à viabilidade de aplicação em cenários reais.

Tabela 2. Resultados comparativos das predições (Autores, 2025)

Patologias	Modelo Agrupado	Manual	Patologias	Modelo Desagrupado	Manual
D	1	2	D	415	2
FI	2	2	FI	13	2
J	230	404	J	343	404
JE	58	10	JE	36	10
P	26	4	P	43	4
R	116	81	R	1447	81
TTC	9	3	TTC – FC1	4	1
			TTC – FC2	10	1
			TTC - FC3	1	1
TTL	3	4	TTL – FC1	13	2
			TTL – FC2	0	1
			TTL - FC3	4	1
TLC	66	105	TLC – FC1	218	94
			TLC – FC2	48	10
			TLC – FC3	1	1
TLL	125	67	TLL – FC1	100	55
			TLL – FC2	65	11
			TLL – FC3	1	1
-			TB	0	1
-			TBE	0	1
-			O	0	1
-			E	1	0
-			EX	6	1
-			ALC	0	0
-			ATC	0	0
-			ATC	0	0

Em resumo, identifica-se uma significativa melhora no modelo com classes agrupadas em relação ao modelo com classes desagrupadas. Utilizando o número total de patologias e excluindo aquelas que estão presentes apenas no modelo desagrupado para 682 contadas manualmente. Destacando o desvio entre a identificação do modelo e a contagem manual, soma-se 384 (agrupado) e 2213 (desagrupado).

Impacto do Agrupamento de Classes

A análise demonstra de forma contundente que o agrupamento das classes contribui significativamente para a melhoria da performance do modelo. O erro médio relativo entre os valores previstos e os valores manuais no modelo agrupado foi de aproximadamente 7%, enquanto no modelo desagrupado esse erro atingiu 407%, tornando-o completamente inadequado para aplicações práticas. Considerando o número total de ocorrências manualmente contadas nas classes comuns, excluindo as classes exclusivas do modelo desagrupado, obteve-se 682 ocorrências manuais.

O modelo agrupado identificou 384 ocorrências, resultando em um desvio absoluto de 298 ocorrências. Já o modelo desagrupado apresentou 2213 ocorrências, com um desvio de 1531 ocorrências em relação ao controle manual. Essa discrepância demonstra que o modelo desagrupado superestima significativamente a presença de patologias em quase todas as classes, o que pode gerar falsos positivos, alarmes indevidos e comprometer toda a cadeia de decisão técnica baseada na predição automatizada. Por exemplo, no caso da patologia D, o modelo agrupado previu apenas 1 ocorrência contra 2 reais, enquanto o modelo desagrupado apresentou 415 ocorrências sendo 413 classificadas como falso positivo.

Relação entre Volume de Dados e Desempenho

Em relação à hipótese de que o aumento na representatividade de certas classes no conjunto de dados melhoraria os resultados do modelo, não foi possível confirmar essa premissa nesta etapa. Apesar de algumas classes terem sido reforçadas com dados adicionais na amostra de treinamento, não houve melhoria perceptível nas métricas nem na acurácia das identificações. As taxas de erro permaneceram elevadas em diversas categorias, tanto para as com alta frequência quanto para as com baixa ocorrência, indicando que o simples aumento da amostragem não é suficiente para tornar o modelo mais preciso.

Análise específica por classes

A patologia J apresentou um comportamento discrepante, no modelo agrupado foram 230 ocorrências contra 404 manuais, enquanto no desagrupado foram 343. Apesar do melhor desempenho neste caso, trata-se de uma exceção, não uma regra.
Para as classes compostas, como TTC e TTL, o modelo desagrupado dividiu a predição em subclasses (FC1, FC2, FC3), o que, embora mais detalhado, aumentou a complexidade e a sobreposição de classificações, dificultando a correspondência direta com os dados de controle e contribuindo para o excesso de predições.
Observou-se também que o modelo desagrupado gerou predições em classes ausentes no conjunto de validação, como TB, TBE, O, E, EX, ALC, ATC, cujos valores esperados são nulos ou inexistentes. Esse comportamento reforça a hipótese de que o modelo sem agrupamento sofre de *overfitting* ou ruído excessivo, tornando-o instável em condições reais.

CONSIDERAÇÕES FINAIS

Os resultados confirmam que o agrupamento de classes melhora o desempenho do modelo de detecção de patologias em pavimentos. O modelo com classes agrupadas apresentou *recall* de 0,34 e 0,18 no desagrupado, e mAP de 0,23 no agrupado contra 0,12 no desagrupado, representando melhorias de 88% e 91%, respectivamente.
Essa abordagem reduz ambiguidades entre categorias semelhantes, elevando a taxa de acertos. Apesar de a precisão não ter estabilizado, observou-se tendência de crescimento no modelo agrupado. A aplicação prática exige categorização manual, mas favorece diagnósticos mais consistentes. As taxas de acerto para Trincas Transversais e Longitudinais foram de 57% e 62%, respectivamente. A única exceção foi a classe JE, na qual o modelo desagrupado teve desempenho superior.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDEL-QADER, I.; ABUDAYYEH, O.; KELLY, M.E. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. J. Comput. Civ. Eng. 2003, 17, 255–263.

ACADEMY, D. S. Deep learning book, 2019.

CAGLE, R.F. Infrastructure Asset Management: An Emerging Direction; AACE International Transactions: Morgantown, WV, USA, 2003; pp. PM21–PM26.

CARVALHO, R. B. et al. Detecção inteligente de falhas em pavimentações asfálticas com redes neurais convolucionais regionais. In: WBCI – Workshop Brasileiro de Cidades Inteligentes, 2022. DOI:10.5753/wbci.2022.222795.

CHA, Y.J.; CHOI, W.; SUH, G.; MAHMOUDKHANI, S.; BUYUKOZTUK, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Comput.-Aided Civ. Infrastruct. Eng. 2017, 33, 731–747.

CNT. Pesquisa CNT de Rodovias 2024. Brasília, 2024.

CORD, A.; CHAMBON, S. Automatic Road Defect Detection by Textural Pattern Recognition Based on AdaBoost. Comput.-Aided Civ. Infrastruct. Eng. 2011, 27, 244–259.

DNIT. NORMA DNIT 005/2003 - TER. Defeitos nos pavimentos flexíveis e semi-rígidos Terminologia. Rio de Janeiro, 2003.

DNIT. NORMA DNIT 154/2010 – ES. Pavimentação asfáltica – Recuperação de defeitos em pavimentos asfálticos. Rio de Janeiro, 2010.

GÉRON, A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media. v 1. 856 p.

GOPALAKRISHANAN, K.; KHAITAN, S. K.; CHOUDHARY, A. and AGRAWAL, A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. ConstrBuild Mater, vol. 157, pp. 322–330, 2017.

HSIEH, Y.A.; TSAI, Y.J. Machine Learning for Crack Detection: Review and Model Performance Comparison. J. Comput. Civ. Eng. 2020, 34, 04020038.

ILOS – Instituto de Logística e Supply Chain. Panorama da Logística Brasileira 2019. Rio de Janeiro: ILOS, 2019. Disponível em: https://www.ilos.com.br.

JAHANSHAHI, M.R.; KELLY, J.S.; MASRI, S.F.; SUKHATME, G.S. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. Struct. Infrastruct. Eng. 2009, 5, 455–486.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. (1999). Data clustering: a review. ACM Computing Surveys, 31(3), 264-323.

MASSUCATTO, J. D. P. Aplicação de conceitos de redes neurais convolucionais na classificação de imagens de folhas. B.S. thesis, Universidade Tecnologica Federal do Paraná, 2019.

MURPHY, K. P. Machine Learning: A Probabilistic Perspective. 2012.

OLIVEIRA, H.; CORREIA, P.L. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the 2009 17th European Signal Processing Conference, Glasgow, UK, 24– 28 August 2009; pp. 622–626.

REDMON, J. You Only Look Once Unified Real Time Object Detection. University of Washington, 2016.

SOKOLOVA, M.; LAPALME, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.

Convolutional Neural Networks for Medical Image Analysis : Full Training or Fine Tuning ?. vol. 35, no. 5, pp. 1299–1312, 2016.

YANG, F.; ZHANG, L.; YU, S.; PROKHOROV, D.; MEI, X.; LING, H. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. IEEE Trans. Intell. Transp. Syst. 2020, 21, 1525–1535.