



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
COLEGIADO DE ENGENHARIA MECÂNICA

Lorenzo Marques Trevizani

**Detecção e localização de dano estrutural
baseado na resposta à vibração: uma abordagem
supervisionada**

Vitória, ES

2023

Lorenzo Marques Trevizani

Deteccção e localização de dano estrutural baseado na resposta à vibração: uma abordagem supervisionada

Trabalho de Conclusão de Curso apresentado ao Departamento de Engenharia Mecânica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Mecânico.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Departamento de Engenharia Mecânica

Orientador: Prof. Dr. Luis Gustavo Giacon Villani

Coorientador: Eng. Maria Clara Assunção Viana

Vitória, ES

2023

Lorenzo Marques Trevizani

Detecção e localização de dano estrutural baseado na resposta à vibração: uma abordagem supervisionada/ Lorenzo Marques Trevizani. – Vitória, ES, 2023-
71 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Luis Gustavo Giacon Villani

Trabalho de Conclusão de Curso – Universidade Federal do Espírito Santo – UFES
Centro Tecnológico
Departamento de Engenharia Mecânica, 2023.

1. Dano estrutural. 2. Apredizado de máquina. 3. Vibração. I. Villani, Luis Gustavo. II. Universidade Federal do Espírito Santo. IV. Detecção e localização de dano estrutural baseado na resposta à vibração: uma abordagem supervisionada

CDU 02:141:005.7

Lorenzo Marques Trevizani

Detecção e localização de dano estrutural baseado na resposta à vibração: uma abordagem supervisionada

Trabalho de Conclusão de Curso apresentado ao Departamento de Engenharia Mecânica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Mecânico.

Trabalho aprovado. Vitória, ES, 25 de setembro de 2018:

Prof. Dr. Luis Gustavo Giacon Villani
Orientador

**Prof. Dr. Guilherme Fabiano
Mendonça dos Santos**
Convidado 1

Prof. Dr. Wilian Hiroshi Hisatugu
Convidado 2

Vitória, ES
2023

Agradecimentos

Conforme chegamos ao desfecho deste ciclo, é oportuno dedicar um momento de reflexão à jornada que me trouxe até este ponto. Ao longo dessa trajetória, deparei-me com desafios, celebrei conquistas e adquiri preciosos aprendizados, os quais compartilhei com aqueles que estiveram ao meu lado. Gostaria, sinceramente, de expressar minha profunda gratidão a cada uma dessas pessoas que, de maneiras singulares, deixaram uma marca significativa nesta caminhada. Seu apoio e presença fizeram toda a diferença, e sou imensamente grato por compartilhar essa jornada com vocês.

Agradeço primeiramente à Deus por Sua infinita bondade e graça em minha vida, por todas as bênçãos e oportunidades que Ele tem proporcionado a cada dia. Seu amor incondicional e orientação constante são um farol que ilumina meu caminho nos momentos de alegria e desafio. Se cheguei onde estou foi Ele que me trouxe, com Ele e por Ele eu sigo. Obrigado, Deus, por ser meu refúgio e minha fortaleza, e por tudo o que és em minha vida.

A minha família, vocês têm sido a base sólida que me sustenta em todos os momentos da minha vida. Agradeço por estarem sempre presentes, oferecendo seu amor incondicional, apoio inabalável e compreensão compassiva. Vocês são meu porto seguro, meu refúgio nos momentos difíceis sendo minha fonte de alegria. Agradeço por cada gesto de carinho, cada palavra de encorajamento e cada sacrifício feito em prol do meu bem-estar e sucesso. Sua confiança em mim e seu apoio incansável são meus verdadeiros tesouros que carrego no peito. Obrigado por serem meu maior tesouro.

Agradeço imensamente à minha parceira de vida, melhor amiga e namorada, Angélica. Por sua presença constante, amor incondicional e apoio inabalável. Sua dedicação, compreensão e inspiração provocam o melhor em mim. Sou grato por cada momento compartilhado, cada risada verdadeira, cada lágrima de choro e cada palavra de encorajamento. Você é minha fonte de conforto, minha confidente e minha melhor amiga. Eu te amo!

Aos meus amigos, em especial, que viveram essa jornada comigo: Ana Sofia, Arthur, Gabriela, Claudio, João Victor, Maria Clara e Wendell. O desfecho desse ciclo só é possível graças a vocês, tenho muito orgulho de ser vosso amigo. Aos meus amigos de longa data: Alexandre, Arthur, Bruno, Danielly, Gabriella, Marcos Yulo, Ranna, Thiago, Vitor e William. Sua amizade sincera e leal é um tesouro precioso que valorizo imensamente. Sou grato por cada um de vocês mas também a todos as outras amizades que cultivei ao longo desses anos. Obrigado por estarem ao meu lado.

Ao PET Engenharia Mecânica pela jornada profissional. Durante o tempo como petiano ativo, fui abençoado com oportunidades de crescimento, aprendizado e desenvolvi-

mento. Agradeço pelas experiências enriquecedoras e pelos desafios que me ajudaram a amadurecer profissionalmente. Além disso, sou grato pela equipe maravilhosa com a qual tive a honra de fazer parte.

Agradeço também a todos os professores que me guiaram e incentivaram ao longo do caminho. Suas orientações e apoio foram inestimáveis. Em especial agradeço ao meu orientador, Prof. Dr. Luis Gustavo Giacon Villani por toda a paciência em me guiar na construção desse trabalho, atenção dedicada durante essa jornada e confiança no meu potencial. Obrigado por me instigar a alcançar meu melhor.

Por fim, agradeço a instituição Universidade Federal do Espírito Santo pelas experiências e oportunidades proporcionadas. Foram incontáveis noites mal dormidas ou viradas, mas nada comparado ao crescimento pessoal e profissional adquirido. Serei eternamente grato!

Resumo

O monitoramento da integridade estrutural (SHM) visa garantir a segurança e a confiabilidade de vários sistemas de infraestrutura. É uma abordagem proativa que melhora a segurança num contexto geral por meio de monitoramento e análise contínua da integridade de estruturas. O SHM fornece detecção precoce de possíveis danos ou deterioração, permitindo manutenção e reparo oportunos que refletem no ganho de desempenho da estrutura, ajuda a minimizar o tempo de inatividade e os custos de manutenção. Entretanto, abordagens tradicionais para diagnóstico de danos em estruturas civis dependem principalmente de inspeções visuais, que podem ser demoradas, necessitam de mão de obra especializada e pode não detectar danos ocultos. Estudos recentes na literatura propõem o uso de modelos supervisionados de aprendizado profundo para diagnóstico de condição de estruturas baseado na resposta à vibração. Tais modelos apresentam maior capacidade de generalização profunda dos dados. No entanto, são mais complexos e requerem maior recurso computacional em comparação com modelos tradicionais. Em virtude disso, o presente estudo estima modelos supervisionados de aprendizado de máquina baseado em árvore de decisão para detecção e localização de danos estruturais com base na resposta à vibração. Ao longo do trabalho é desenvolvida uma metodologia que faz uso do modelo classificador de floresta aleatória e utiliza de características estatísticas em domínio do tempo e da frequência, bem como características de coeficientes autorregressivos. Os modelos são treinados, testados e validados sob dados provenientes de uma bancada de simulações na *Qatar University* (QU). Durante os testes dos modelos, identifica-se a ineficácia das características estatísticas em treiná-los para prever o dano estrutural. Em contrapartida, as características de coeficientes autorregressivos demonstraram eficácia em treinar os modelos para a prever o dano estrutural. Por conta disso, o resultado é um conjunto de modelos supervisionados especificados para a estrutura estudada, capazes de diagnosticar com clareza a existência e localização de dano por meio de características de coeficientes autorregressivo, as acurácias dos modelos durante a etapa de teste variam entre 100% e 84%. Por fim, ao comparar os modelos desenvolvidos com os modelos de aprendizado profundo da literatura, comprova-se a viabilidade da metodologia escrita nesse trabalho, demonstra-se ser possível obter resultados promissores utilizando de abordagens tradicionais, eliminando a necessidade das abordagens de aprendizado profundo para diagnóstico de dano estrutural.

Palavras-chaves: 1. Monitoramento da Integridade Estrutural (SHM). 2. Aprendizado de máquina. 3. Classificador de floresta aleatória.

Abstract

Structural Health Monitoring (SHM) aims to ensure the safety and reliability of various infrastructure systems. It is a proactive approach that improves security in an overall context through continuous monitoring and analysis of the integrity of structures. SHM provides early detection of potential damage or deterioration, allowing for continuous maintenance and repair that reflects on structure performance gain, helps to minimize downtime and maintenance costs. However, traditional approaches to diagnosing damage to civil structures rely primarily on visual inspections, which can be time-consuming, require skilled labor, and may not detect hidden damage. Recent studies in the literature propose the use of supervised deep learning models for condition diagnosis of structures based on vibration response. Such models have greater capacity for deep generalization of the data. However, they are more complex and require more computational resources compared to traditional models. As a result, the present study estimates supervised decision tree-based machine learning models for detecting and localizing structural damage based on vibration response. Throughout the work, a methodology is developed that makes use of the random forest classifier model and uses statistical characteristics in the time and frequency domain, as well as characteristics of autoregressive coefficients. The models are trained, tested and validated against data from a simulation bench at Qatar University (QU). During the tests of the models, the ineffectiveness of the statistical characteristics in training them to predict the structural damage is identified. In contrast, the characteristics of autoregressive coefficients demonstrated effectiveness in training the models to predict structural damage. Because of this, the result is a set of supervised models specified for the studied structure, capable of clearly diagnosing the existence and location of damage through characteristics of autoregressive coefficients, the accuracies of the models during the test stage vary between 100% and 84%. Finally, when comparing the models developed with the deep learning models in the literature, the viability of the methodology written in this work is proven, it is possible to obtain promising results using traditional approaches, eliminating the need for deep learning approaches to diagnosis of structural damage.

Keywords: 1. Structural health monitoring (SHM). 2. Machine Learning. 3. Random Forest Classifier.

Lista de ilustrações

Figura 1 – Algoritmo genérico para criação de um modelo de aprendizado de máquina.	16
Figura 2 – Árvore de decisão baseada no ganho de informação gerada a partir de um banco de dados de melancias.	24
Figura 3 – Fluxograma das etapas realizadas para criação do modelo de aprendizado de máquina.	26
Figura 4 – Projeto estrutural da arquibancada.	28
Figura 5 – Estrutura principal da bancada de simulações em <i>QU</i> com indicação das posições dos sensores.	29
Figura 6 – Os acelerômetros modelo (a) PCB 393B04. (b) B&K 8344.	29
Figura 7 – Excitador TMS 2100E11 acoplado à estrutura.	30
Figura 8 – Da direita para a esquerda: Os dois dispositivos de aquisição de dados (DT9857E-16), o amplificador de energia (“SmartAmp” 2100E21-400) e o notebook para registro de dados.	30
Figura 9 – Juntas (a) Com parafusos apertados. (b) Com parafusos frouxos. . . .	31
Figura 10 – Fluxograma de criação do banco de dados de características de um sensor qualquer k a partir do subconjunto de dados desse sensor.	33
Figura 11 – Estrutura do banco de dados representativo das amostras.	35
Figura 12 – Coeficientes das correlações de <i>Pearson</i> entre características estatísticas no domínio do tempo extraídas.	39
Figura 13 – Coeficientes das correlações de <i>Pearson</i> entre características estatísticas no domínio da frequência.	40
Figura 14 – Coeficientes das correlações de <i>Pearson</i> entre características de valor estatísticos.	41
Figura 15 – Correlações de <i>Pearson</i> entre características de coeficientes autorregressivos.	42
Figura 16 – Representatividade dos valores da variável objetivo.	43
Figura 17 – Probabilidade de dano no cenário 0, sem dano existente, na simulação do banco de dados B.	46
Figura 18 – Probabilidade de dano no cenário 1 na simulação do banco de dados B.	47
Figura 19 – Probabilidade de dano no cenário 17 na simulação do banco de dados B.	47
Figura 20 – Probabilidade de dano no cenário 8 na simulação do banco de dados B.	48
Figura 21 – Probabilidade de dano no cenário 10 na simulação do banco de dados B.	49
Figura 22 – Comparativo de condição da estrutura no cenário 0 entre modelos. . . .	50
Figura 23 – Comparativo de condição da estrutura no cenário 9 entre modelos. . . .	51
Figura 24 – Comparativo de condição da estrutura no cenário 1 entre modelos. . . .	52
Figura 25 – Comparativo de condição da estrutura no cenário 8 entre modelos. . . .	53

Figura 26 – Probabilidade de dano nos cenários referencia e 1 do banco de dados B.	61
Figura 27 – Probabilidade de dano nos cenários 2, 3 e 4 do banco de dados B. . . .	62
Figura 28 – Probabilidade de dano nos cenários 5, 6 e 7 do banco de dados B. . . .	63
Figura 29 – Probabilidade de dano nos cenários 8, 9 e 10 do banco de dados B. . .	64
Figura 30 – Probabilidade de dano nos cenários 11, 12 e 13 do banco de dados B. .	65
Figura 31 – Probabilidade de dano nos cenários 14, 15 e 16 do banco de dados B. .	66
Figura 32 – Probabilidade de dano nos cenários 17, 18 e 19 do banco de dados B. .	67
Figura 33 – Probabilidade de dano nos cenários 20, 21 e 22 do banco de dados B. .	68
Figura 34 – Probabilidade de dano nos cenários 23, 24 e 25 do banco de dados B. .	69
Figura 35 – Probabilidade de dano nos cenários 26, 27 e 28 do banco de dados B. .	70
Figura 36 – Probabilidade de dano nos cenários 29 e 30 do banco de dados B. . . .	71

Lista de tabelas

Tabela 1 – Acurácia dos modelos otimizados na previsão da variável objetivo no banco de dados A.	44
Tabela 2 – Hiper-parâmetros obtidos nas otimizações dos modelos.	60

Lista de abreviaturas e siglas

SHM	Monitoramento da integridade estrutural
QU	<i>Qatar University</i>
EDA	Análise exploratória de dados
FFT	Transformada Rápida de Fourier
PSD	Densidade espectral de potência
AR	Modelos de autorregressão
PoD	Probabilidade de dano

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
2	UMA VISÃO GERAL SOBRE APRENDIZADO DE MÁQUINAS	16
2.1	Variável objetivo	17
2.2	Classificação dos métodos de aprendizado de máquina	17
2.3	Análise exploratória de dados	18
2.4	Engenharia de variáveis	18
2.4.1	Métricas estatísticas no domínio do tempo	19
2.4.2	Métricas estatísticas no domínio da frequência	21
2.4.3	Coeficientes autorregressivos	21
2.4.4	Correlação de <i>Pearson</i>	22
2.5	Divisão de dados para treino e teste	22
2.6	Modelo de floresta aleatória	23
2.7	Modelos de aprendizado profundo	25
3	METODOLOGIA	26
3.1	Bancada de simulação em <i>Qatar University</i>	27
3.1.1	Estrutura física da arquibancada	27
3.1.2	Instrumentação	28
3.1.3	Tipo de dano simulado	30
3.1.4	Descrição do banco de dados	31
3.2	Engenharia de Variáveis	32
3.3	Pré-processamento de dados	33
3.4	Modelo de aprendizado de máquina	35
3.4.1	Treinamento e teste dos modelos	35
3.4.2	Validação dos modelos	36
4	APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS	38
4.1	Vetor de características de valores estatísticos	38
4.2	Vetor de características de coeficientes autorregressivos	41
4.3	Cardinalidade	42
4.4	Desempenho em amostras de teste	44
4.4.1	Modelos treinados com vetor de características de valores estatísticos	45

4.4.2	Modelos treinados com vetor de características de coeficientes autorregressivos	45
4.5	Desempenho em validação com banco de dados B	45
4.6	Comparação de desempenho com modelos da literatura	49
5	CONSIDERAÇÕES FINAIS	55
5.1	Conclusão	55
5.2	Trabalhos futuros	56
	REFERÊNCIAS	57
	APÊNDICES	59
	APÊNDICE A – HIPER-PARÂMETROS OBTIDOS NAS OTIMI- ZAÇÕES DOS MODELOS	60
	APÊNDICE B – PROBABILIDADE DE DANO EM TODOS OS CENÁRIOS	61

1 Introdução

Monitoramento de integridade estrutural, referenciado na literatura como SHM, é um processo crucial no campo da engenharia que visa garantir a segurança e a confiabilidade de vários sistemas de infraestrutura. [Farrar e Worden \(2007\)](#) descrevem que o SHM envolve a implementação de uma estratégia de identificação de danos para detectar mudanças no material, propriedades geométricas, condições de contorno e conectividade do sistema dessas estruturas. Ao monitorar e analisar continuamente a integridade desses sistemas, o SHM fornece detecção precoce de possíveis danos ou deterioração, permitindo manutenção e reparo oportunos. Essa abordagem proativa não apenas melhora a segurança geral e o desempenho da infraestrutura, mas também ajuda a minimizar o tempo de inatividade e os custos de manutenção. Os autores enfatizam a importância do SHM na melhoria da longevidade e resiliência da infraestrutura de engenharia aeroespacial, civil e mecânica.

No contexto do SHM, as abordagens tradicionais para diagnóstico de danos em estruturas civis dependem principalmente de inspeções visuais. Essas inspeções envolvem o exame físico da estrutura para identificar quaisquer sinais visíveis de dano ou deterioração. No entanto, [Avci et al. \(2021\)](#) destacam vários desafios associados a esses métodos tradicionais. Em primeiro lugar, o tamanho das estruturas civis pode tornar as inspeções visuais regulares trabalhosas e demoradas. Em segundo lugar, como essas abordagens dependem do julgamento humano, elas exigem mão de obra qualificada e altamente treinada para interpretar pistas visuais e fazer julgamentos subjetivos sobre a saúde estrutural. Por último, muitas vezes é um desafio inspecionar visualmente os membros estruturais de suporte de carga, pois eles são normalmente cobertos por componentes não estruturais e coberturas decorativas. Essas limitações dificultam a aplicação prática de abordagens tradicionais para diagnóstico de danos.

Além disso, [Avci et al. \(2021\)](#) enfatizam o uso da abordagem baseada em vibração como uma alternativa interessante aos métodos tradicionais. Essa abordagem envolve o registro e análise da resposta em vibração ao processar sinais medidos por uma rede de acelerômetros. Os benefícios em relação aos métodos tradicionais são os de superar os desafios descritos anteriormente, elimina a subjetividade dos dados por promover uma análise orientada por dados, pode substituir a mão de obra qualificada por algoritmos e técnicas de aprendizado de máquina, ademais em caso de grandes estruturas basta posicionar mais sensores e, uma vez posicionados, não há necessidade de acesso à estrutura para efetuar diagnóstico. Ao utilizar técnicas baseadas em vibração, os pesquisadores visam melhorar a eficiência e a precisão do diagnóstico de danos em estruturas civis, tornando-se uma abordagem atraente no campo do monitoramento de saúde estrutural. [Park et al. \(2006\)](#) dizem que o objetivo do SHM é o desenvolvimento de um sistema de monitoramento

autossustentável e internalizado na própria estrutura que monitora com o mínimo de mão de obra.

Vale apontar que a análise dos sinais de vibração usando algoritmos de aprendizado de máquina desempenha um papel crucial no monitoramento de integridade estrutural. Os modelos de aprendizado de máquina tornaram-se ferramentas valiosas para, dentre outras aplicações, analisar dados de vibração. Esses algoritmos oferecem a capacidade de extrair informações significativas desses sinais de vibração medidos, permitindo a detecção precisa e eficiente de danos ([ABDELJABER et al., 2017](#)).

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo geral do presente trabalho é propor um algoritmo de aprendizado de máquina supervisionado que identifique e localize dano na bancada de simulação em *Qatar University*, que é uma estrutura tipo treliça, utilizando dados de aceleração medidos por uma rede de acelerômetros.

1.1.2 Objetivos Específicos

Como objetivos específicos do trabalho, cita-se:

- Estudar a extração de métricas estatísticas clássicas calculadas no domínio do tempo e da frequência como característica sensível ao dano;
- Estudar a extração de coeficientes de modelos autorregressivos como característica sensível ao dano;
- Estimar um modelo supervisionado de aprendizado de máquina baseado em árvore de decisões, capaz de identificar e localizar a presença de dano na estrutura da bancada de simulação em *Qatar University*;
- Comparar o desempenho do modelo supervisionado baseado em estatísticas clássicas com o modelo supervisionado baseado em coeficientes autorregressivos;
- Comparar desempenho do modelo classificador de floresta aleatória desenvolvido nesse estudo com desempenho dos modelos de aprendizado profundo da literatura ([ABDELJABER et al., 2017](#)).

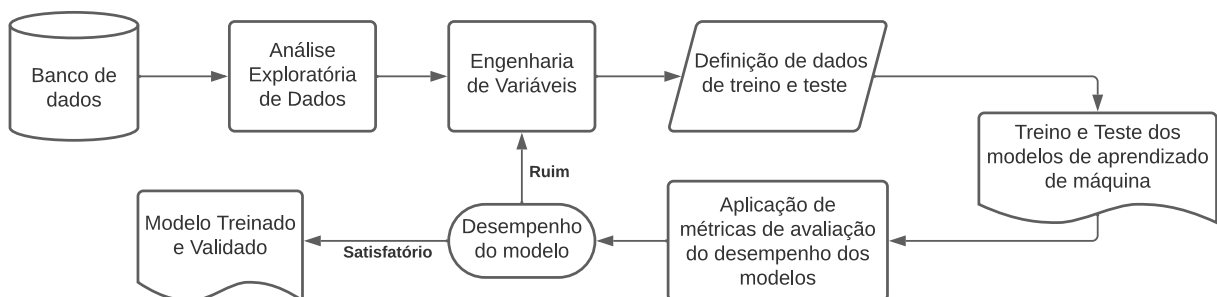
2 Uma Visão Geral Sobre Aprendizado de Máquinas

O aprendizado de máquinas é um subcampo da inteligência artificial e tem como objetivo desenvolver algoritmos capazes de aprender a partir de dados, aprimorando seu desempenho ao longo do tempo, sem a necessidade de programação explícita. Para isso, o aprendizado baseia-se em fundamentos teóricos provenientes de disciplinas matemáticas amplamente estudadas, como estatística, álgebra linear, cálculo e otimização. Como consequência, os algoritmos de aprendizado de máquina têm a capacidade de identificar padrões e relações nos dados que não seriam facilmente perceptíveis pelos seres humanos, o que os torna uma poderosa ferramenta para solucionar problemas complexos e aprimorar a eficiência em diversos setores. Dessa forma, tornou-se um dispositivo indispensável para empresas e organizações, permitindo extrair informações valiosas de grandes volumes de dados e embasar decisões em fundamentos sólidos. Vale ressaltar que algoritmos de aprendizado de máquina já são aplicados em diversas áreas, incluindo reconhecimento de imagem e fala, análise preditiva e recomendação personalizada (SOMVANSHI et al., 2016).

Russell (2020) explica que a adoção de modelos de aprendizado de máquina se dá principalmente por dois motivos. O primeiro decorre da incapacidade de antecipar todas as possíveis situações futuras. Além disso, programar um sistema para lidar com todos os cenários possíveis pode ser impraticável ou mesmo impossível. O segundo motivo envolve a carência de conhecimento ou experiência para calcular, analiticamente, o conjunto de regras que regem o cenário.

Um projeto típico de criação de modelos de aprendizado de máquina está apresentado na Figura 1 onde cada uma das etapas expostas estão descritas nas subseções subsequentes.

Figura 1 – Algoritmo genérico para criação de um modelo de aprendizado de máquina.



Fonte: De autoria própria.

2.1 Variável objetivo

No contexto do aprendizado supervisionado, a variável objetivo é normalmente a variável que o modelo é treinado para prever com base nos dados de entrada. Infere-se de [Zhou e Liu \(2021\)](#) que para as variáveis objetivo têm-se dois tipo, categórica ou contínua.

Variável categórica, também conhecida como variável qualitativa, representa dados que se enquadram em categorias ou classes específicas. Os modelos de aprendizado de máquina treinados para prever esse tipo de variável objetivo são chamados de algoritmos de classificação ([PATEL, 2019](#)). Alguns exemplos de variáveis categóricas incluem variáveis binárias (com duas categorias, como “sim” ou “não”) e variáveis multiclasse (com mais de duas categorias, como “vermelho”, “verde” ou “azul”).

Variável contínua, também conhecida como variável quantitativa, representa dados que podem assumir qualquer valor numérico em um determinado intervalo. O termo é usado quando a variável de destino é uma medida numérica ou uma quantidade de valor real. Os modelos de aprendizado de máquina treinados para prever esse tipo de variável objetivo são chamados de algoritmos de regressão ([PATEL, 2019](#)). Exemplos de variáveis contínuas incluem variáveis como temperatura ou altura.

A variável objetivo que o presente estudo propõe-se a prever é uma variável categórica binária (dano ou saudável). Dessa forma, o algoritmo desenvolvido é um algoritmo de classificação. A identificação a priori da variável objetivo é de vital importância para selecionar o modelo de aprendizado de máquina apropriado e determinar as métricas para avaliar o desempenho do modelo.

2.2 Classificação dos métodos de aprendizado de máquina

[Géron \(2017\)](#) classifica os métodos de aprendizado de máquina em 4 categorias principais que se diferenciam quanto a quantidade de supervisão recebida durante a etapa de treino. Essas categorias são:

- Aprendizado supervisionado;
- Aprendizado não supervisionado;
- Aprendizado semi-supervisionado;
- Aprendizado por reforço.

Aprendizado supervisionado é o método que envolve, no processo de treinamento, fornecer ao modelo de aprendizado de máquina um conjunto de dados rotulado, onde cada dado consiste em um objeto de entrada e uma variável objetivo correspondente. Alguns dos

modelos que utilizam essa abordagem de aprendizado são modelos baseados em árvores de decisão e redes neurais (SOMVANSHI et al., 2016).

Aprendizado não supervisionado é o método que lida com banco de dados o sem uso de rótulos, ele tenta entender a estrutura subjacente dos dados sozinho, fazendo agrupamentos de forma que os dados em um dos grupos sejam semelhantes entre si, mas diferentes dos dados em outro grupo (PATEL, 2019). Zhou e Liu (2021) trazem como um exemplo de modelo não supervisionado o modelo de misturas de gaussiana (GMM).

Aprendizado semi-supervisionado envolve uma abordagem que combina dados com e sem rótulos durante a etapa de treinamento do modelo. A ideia por trás disso é a de que os dados sem rótulos possuem informação valiosa sobre padrões ocultos enquanto que os dados com rótulos são excelentes para guiar o processo. Um modelo que representa essa categoria é o modelo regenerativo (ZHOU; LIU, 2021).

Aprendizado por reforço é outro método que se diferencia na etapa de treinamento do modelo. Para aprender, o modelo se recompensa ou se penaliza de acordo com ações tomadas observando o ambiente. Dessa forma ele aprende sozinho qual a política a ser adotada que maximiza a recompensa (GÉRON, 2017).

Apesar de haver diversos métodos de aprendizado de máquinas, o presente trabalho faz uso apenas da metodologia de aprendizado supervisionado, haja vista que há presença de dados rotulados. O presente estudo não visa comparar metodologias diferentes, mas sim, obter o desempenho de uma metodologia escolhida.

2.3 Análise exploratória de dados

A Análise exploratória de dados (EDA) é um processo na ciência de dados que envolve a exploração interativa de um conjunto de dados desconhecido. Essa análise é uma etapa inicial que permite aos cientistas de dados obterem uma compreensão mais profunda inerente à estrutura do banco de dados que estão avaliando. Isso é realizado por meio de ferramentas de visualização dos dados que mostram, por exemplo, a distribuição dos valores e a falta de dados. Por meio da EDA, identifica-se padrões e tendências que podem fornecer *insights* valiosos e ajudar na formulação hipóteses (MILO; SOMECH, 2020).

2.4 Engenharia de variáveis

A etapa de engenharia de variáveis ocorre logo após a EDA (Seção 2.3) e envolve o processo de extração e seleção das características usadas como de entrada para o modelo posteriormente adotado.

A extração de características é um conceito fundamental no campo de aprendizado

de máquina e análise de dados. Ele envolve extrair um subconjunto de características relevantes de um conjunto maior de variáveis ou atributos de entrada. Seu objetivo é transformar os dados originais em uma representação mais compacta que capture as informações essenciais necessárias para uma tarefa específica, reduzindo assim a dimensão dos dados. Ao fazer isso, pode melhorar a eficiência computacional e reduzir o risco de *overfitting*, fenômeno que descreve modelos que só desempenham bem nos cenários em que foram treinados. Esse processo de extração requer um aprofundamento cuidadoso no domínio do problema e das características dos dados, por conta disso, é necessário que seja feito após uma análise exploratória de dados. Várias técnicas e algoritmos podem ser empregados, variando de medidas estatísticas simples a métodos mais complexos, como análise de componentes principais (PCA) (ARAÚJO; COELHO, 2009). Em última análise, o sucesso da extração de características reside em sua capacidade de descobrir padrões e relacionamentos significativos nos dados, permitindo análises e previsões mais precisas e eficientes (ISABELLE; ELISSEEFF, 2006).

Executar a seleção de características após a extração é uma etapa essencial na preparação das variáveis de entrada para modelos de aprendizado de máquina. As características extraídas podem conter informações redundantes entre si ou irrelevantes para o sistema, portanto, conduzir a seleção delas ajuda a identificar e reter as informações mais relevantes e discriminativas para parametrização do modelo. Guyon e Elisseeff (2003) discorre sobre como o processo ajuda a reduzir dimensionalidade das variáveis de entrada e aprimorar o desempenho do modelo enfatizando que ao incluir variáveis irrelevantes ou redundantes pode levar ao aumento da complexidade computacional, redução interpretabilidade e *overfitting*. Ao combinar cuidadosamente a extração e a seleção de características, pode-se otimizar efetivamente a representação dos dados, levando a uma maior precisão e eficiência nos modelos de aprendizado de máquina.

A seguir apresentam-se as características utilizadas nesse trabalho, extraídas de sinais de vibração medidos.

2.4.1 Métricas estatísticas no domínio do tempo

As métricas estatísticas no domínio do tempo são ferramentas essenciais na análise de dados pois fornecem uma compreensão das características de um conjunto de dados, permitindo uma análise mais precisa e uma tomada de decisão embasada em evidências. A seguir estão algumas métricas estatísticas retiradas de Afonso e Nunes (2019) para um vetor \vec{x} de k amostras definido pela Equação 2.1.

$$\vec{x} = [x(1), x(2), x(3), \dots, x(k-2), x(k-1), x(k)] \quad (2.1)$$

A média é uma medida de tendência central que representa o valor médio de um conjunto de dados. A Equação 2.2 descreve a média (\bar{x}) de um agrupamento de n amostras

do vetor \vec{x} .

$$\bar{x} = \sum_{i=1}^n \frac{x(i)}{n} \quad (2.2)$$

A mediana é o valor que divide o conjunto de dados em duas partes iguais, onde metade dos valores está abaixo e metade está acima desse valor. Ela é especialmente útil quando há valores extremos ou uma distribuição assimétrica. A Equação 2.3 descreve a mediana (M) de um agrupamento de n amostras do vetor \vec{x} .

$$M = \frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right)}{2} \quad (2.3)$$

O desvio padrão é uma medida de dispersão que indica o quanto os valores de um conjunto de dados se afastam da média. A Equação 2.4 descreve o desvio padrão (σ^2) de um agrupamento de n amostras do vetor \vec{x} .

$$\sigma^2 = \sqrt{\sum_{i=1}^n \frac{(x(i) - \bar{x})^2}{n}} \quad (2.4)$$

A assimetria indica o grau de desvio de uma distribuição em relação à simetria. A Equação 2.5 descreve a Assimetria de um agrupamento de n amostras do vetor \vec{x} , o termo σ^3 representa o quarto momento central (AFONSO; NUNES, 2019).

$$Assimetria = \frac{\sum_{i=1}^n (x(i) - \bar{x})^3}{(n-1)\sigma^3} \quad (2.5)$$

A amplitude é uma medida de dispersão que indica a diferença entre o maior e o menor valor de um conjunto de dados. Por conta disso, ela fornece uma ideia da variação total dos dados. A Equação 2.6 descreve amplitude de um agrupamento de n amostras do vetor \vec{x} .

$$Amplitude = \max(\vec{x}) - \min(\vec{x}) \quad (2.6)$$

A curtose é uma medida de achatamento que indica o grau de achatamento ou picosidade de uma distribuição em relação à distribuição normal, pode ser positiva, negativa ou zero. A Equação 2.7 descreve a Curtose amostral (k_a) de um agrupamento de n amostras do vetor \vec{x} , o termo m_4 representa o quarto momento amostral e o termo s^4 representa o quarto momento central amostral elevado à quarta potência (AFONSO; NUNES, 2019).

$$k_a = \frac{n^2(n+1)m_4}{(n-1)(n-2)(n-3)s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (2.7)$$

Para os propósitos deste trabalho, as definições de Assimetria e Curtose são as mesmas utilizadas pela biblioteca *Pandas*¹, implementada na linguagem de programação *Python*. Uma vez estimadas as métricas estatísticas no domínio do tempo, elas podem ser utilizadas como entrada para algoritmos de aprendizado de máquinas.

¹ <<https://pandas.pydata.org/>>

2.4.2 Métricas estatísticas no domínio da frequência

Para entender as métricas estatísticas no domínio da frequência, é importante compreender o conceito de Transformada Rápida de Fourier (FFT). De acordo com Butz (2006), a FFT decompõe um sinal f no domínio do tempo t em componentes no domínio da frequência, calculando a amplitude g das componentes ω de frequência, como mostra a Equação 2.8 onde i representa a unidade imaginária $\sqrt{-1}$.

$$g(\omega) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{i\omega t} dt \quad (2.8)$$

Ao aplicar a FFT, analisam-se essas amplitudes bem como as fases das componentes de frequência do sinal, fornecendo informações valiosas sobre o conteúdo espectral de um sinal. Além disso, permite-se determinar a densidade de potência ou energia em diferentes frequências, identificar frequências dominantes e observar como a potência do sinal varia em relação à frequência (BUTZ, 2006).

O espectro de amplitude é uma característica que mostra a distribuição das amplitudes em função da frequência. A densidade espectral de potência (PSD) obtida a partir da FFT é uma medida de distribuição das potências dos sinais em função da frequência (BUTZ, 2006).

A partir da amplitude das frequências e da PSD calculadas por meio da FFT, calculam-se as mesmas métricas estatísticas do sinal no domínio do tempo, agora no domínio da frequência. Dessa forma, elas também podem ser utilizadas como entrada para algoritmos de aprendizado de máquinas.

2.4.3 Coeficientes autorregressivos

Modelos de autorregressão (AR) são um tipo método estatístico que utilizam de observações passadas para prever comportamentos futuros, modelando esse comportamento por meio de uma série temporal (MORETTIN; TOLOI, 2018). Essa característica de dependência temporal das observações passadas demonstra-se valiosa para treinar modelos de aprendizado de máquina. De acordo com Morettin e Toloi (2018), um modelo autorregressivo de ordem p para o comportamento $Z(t)$ é denotado pela Equação 2.9, onde Φ_t são os coeficientes autorregressivos, Z_{p-t} são valores das observações passadas e α_t é chamado de ruído branco.

$$Z(t) = \sum_{t=0}^p \Phi_t Z_{p-t} + \alpha_t \quad (2.9)$$

Uma vez estimados para uma sequência temporal, os coeficientes Φ_t podem ser utilizados como entrada para algoritmos de aprendizado de máquinas.

2.4.4 Correlação de *Pearson*

[Guyon e Elisseeff \(2003\)](#) descrevem métodos utilizados para determinação de correlação entre variáveis, dentre eles destaca-se a correlação de *Pearson*. Essa correlação promove um coeficiente entre -1 e +1 que detecta relações lineares entre as variáveis em foco, coeficiente -1 representa uma relação inversamente proporcional perfeita e +1 uma relação diretamente proporcional perfeita, 0 indica não haver correlações lineares entre as variáveis. O autor apresenta que o coeficiente da correlação de *Pearson* $R(i)$ entre as variáveis representadas pelos vetores \bar{x} e \bar{y} pode ser encontrado pela Equação 2.10, onde a anotação em barra define a média dos agrupamentos de m valores.

$$R(i) = \frac{\sum_{k=1}^m (x_{k,i} - \bar{x}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^m (x_{k,i} - \bar{x}_i)^2 \sum_{k=1}^m (y_k - \bar{y})^2}} \quad (2.10)$$

É comum em aplicações de aprendizado de máquina adotar valores de coeficientes de correlações entre variáveis como critério para seleção de características, ou seja, variáveis de entrada com alta correlação são descartadas, pois representam a mesma informação e isso pode levar ao *overfitting*.

2.5 Divisão de dados para treino e teste

[Zhou e Liu \(2021\)](#) discutem a importância de dividir o conjunto de dados original fornecido em um subconjunto de treinamento e um subconjunto de teste para, respectivamente, treinar o modelo de aprendizado de máquina e avaliar seu desempenho em dados não vistos durante o treino. O equilíbrio adequado entre os tamanhos dos subconjuntos é necessário porque caso o conjunto de teste seja muito pequeno, a avaliação pode ser menos confiável, enquanto que um conjunto de teste maior pode levar a uma diferença substancial entre o modelo treinado no conjunto de treinamento e o modelo treinado em todo o conjunto de dados.

Para encontrar o equilíbrio descrito anteriormente, [Zhou e Liu \(2021\)](#) sugerem utilizar cerca de 2/3 a 4/5 do conjunto de dados original para treinamento e o restante para teste. Além disso, mencionam que um conjunto de teste deve conter idealmente pelo menos 30 amostras para uma avaliação confiável. Essas diretrizes ajudam a garantir que o modelo treinado possa generalizar bem para dados novos e fornecer previsões precisas.

Uma abordagem comum de ser utilizada em estudos envolvendo aprendizado de máquina é a aplicação de um regime de validações cruzadas, nessa abordagem divide-se o conjunto de dados em k subconjuntos chamados de *folds*, sendo k um valor escolhido com base na quantidade de amostras disponíveis. O modelo então é treinado com $k-1$ subconjuntos e testado com o *fold* restante. Isso é feito uma quantidade de k vezes, com cada *fold* atuando como subconjunto de teste uma vez de modo que a cada iteração realizada

registra-se a métrica desejada e altera o *fold* atuante. Zhou e Liu (2021) argumentam que a abordagem de validações cruzadas melhora o uso das amostras disponíveis por garantir que todas sejam usadas tanto para treino quanto para teste. Dessa forma, aumenta-se a confiabilidade do desempenho e generalização de modelos e reduz a variâncias.

Outro fator importante que influencia o desempenho do modelo é a representatividade e imparcialidade dos subconjuntos em relação ao conjunto original. Para isso, Brownlee (2018) comenta sobre o uso do recurso de embaralhamento dos dados antes da divisão dos subconjuntos. Por meio desse recurso, garante-se que não haja viés no modelo em encontrar padrões específicos na ordem dos dados, pois cada subconjunto conterá uma representação aleatória e diversa deles.

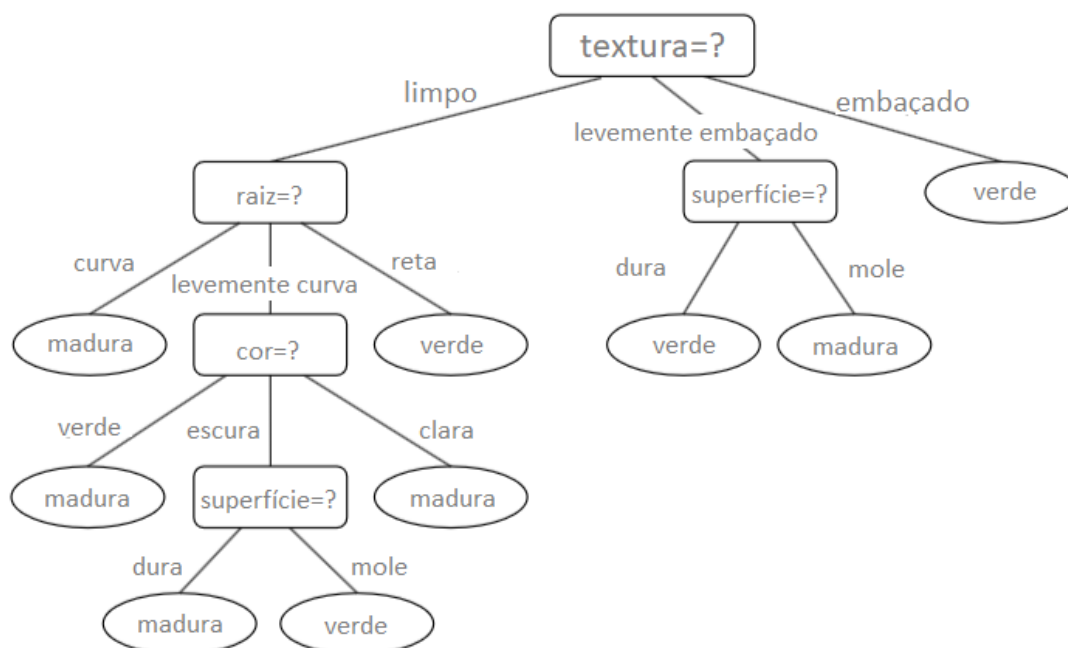
2.6 Modelo de floresta aleatória

O presente estudo faz uso do modelo classificador de floresta aleatória, modelo esse baseado em árvores de decisão. A sua escolha é dada porque as florestas aleatórias são úteis em dados de alta dimensionalidade por poderem lidar com um grande número de características com menor probabilidade de *overfitting*. Em dados de alta dimensionalidade, o número de características pode ser muito maior do que o número de amostras, o que pode levar ao baixo desempenho na generalização do modelo (ZHOU; LIU, 2021). As florestas aleatórias abordam esse problema selecionando aleatoriamente um subconjunto de características para cada árvore, o que reduz a correlação entre as árvores, melhorando sua diversidade e reduzindo a variância do modelo (ZHOU, 2014).

Uma árvore de decisão é formada por divisões recursivas do conjunto de dados com base em diferentes características. A princípio, cada divisão consiste em uma decisão com base no valor de uma ou mais características específicas dos dados. Sendo assim, a árvore se ramifica para representar diferentes resultados possíveis dessa divisão estabelecida, de forma que cada ramo resultante possui um subconjunto dos dados. Esse processo continua até que algum critério de parada seja atendido, como atingir um máximo de divisões ou um número mínimo de amostras em um subconjunto (RUSSELL, 2020). A Figura 2, retirada de Zhou e Liu (2021), é um exemplo de árvore de decisão sobre melancias verdes e maduras onde se observa a estrutura da árvore bem como as decisões tomadas.

A estrutura de uma árvore de decisão é composta por três componentes principais, sendo eles nó raiz, nó interno e nó folha. O nó raiz é o ponto inicial da árvore de decisão e está associado à(s) característica(s) usada(s) na primeira tomada de decisão. Os nós internos são pontos de decisão internos na árvore onde é avaliado uma ou mais características para a tomada de decisão, com base nos resultados dessa decisão a árvore é ramificada para diferentes nós filhos. Por fim, os nós folha são os nós finais da árvore, eles não possuem nós filhos e representam a previsão resultante da árvore de decisão (ZHOU; LIU, 2021).

Figura 2 – Árvore de decisão baseada no ganho de informação gerada a partir de um banco de dados de melancias.



Fonte: Adaptado de Zhou e Liu (2021).

Floresta aleatória é um modelo de aprendizado de máquina que combina várias árvores de decisão para fazer previsões. Cada uma dessas árvores são treinadas sob características aleatórias, de modo que a previsão final é determinada pela agregação das previsões de todas as árvores. Esses modelos oferecem precisão e robustez aprimoradas em comparação com árvores de decisão individuais (ZHOU, 2014).

O modelo classificador de floresta aleatória utilizado foi importado da biblioteca *scikit-learn* da linguagem de programação *python* e chama-se *RandomForestClassifier*². Nesse modelo fornecido pelo *scikit-learn* é possível definir configurações das métricas sobre a estrutura da floresta aleatória como quantidade de árvores, quantidade máxima de nós por ramo na árvore, entre outros. Tais métricas são chamadas de hiper-parâmetros e afetam o seu desempenho e capacidade de generalização do modelo.

A otimização de hiper-parâmetros visa encontrar a combinação ideal dessas métricas para maximizar o desempenho do modelo haja vista que diferentes configurações podem levar a resultados significativamente distintos (SRINIVAS; KATARYA, 2022). O pacote *Optuna* (AKIBA et al., 2019) da linguagem de programação *python*, apresenta um algoritmo que automatiza a busca por hiper-parâmetros que maximizam ou minimizam alguma métrica de avaliação de resultado estipulada pelo cientista de dados.

² <<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>>

É importante ressaltar também que a atenção do presente trabalho não está voltada na comparação de diferentes modelos de aprendizado de máquinas. Em vez disso, busca-se investigar diferentes conjuntos de características e suas relações de causalidade com a variável objetivo. A comparação com outros algoritmos é feita com base em resultados disponíveis na literatura para a solução do mesmo problema ([ABDELJABER et al., 2017](#)).

2.7 Modelos de aprendizado profundo

Os modelos de aprendizado profundo são um tipo de modelo supervisionado de aprendizado de máquina projetado para aprender padrões e relacionamentos complexos em dados. Eles são compostos por múltiplas camadas de nós interconectados, onde cada nó aplica uma função de ativação não linear à soma ponderada de suas entradas. As camadas são geralmente organizadas de maneira hierárquica, com cada camada aprendendo a representar os dados de entrada em um nível mais alto de abstração ([ZHOU; LIU, 2021](#)).

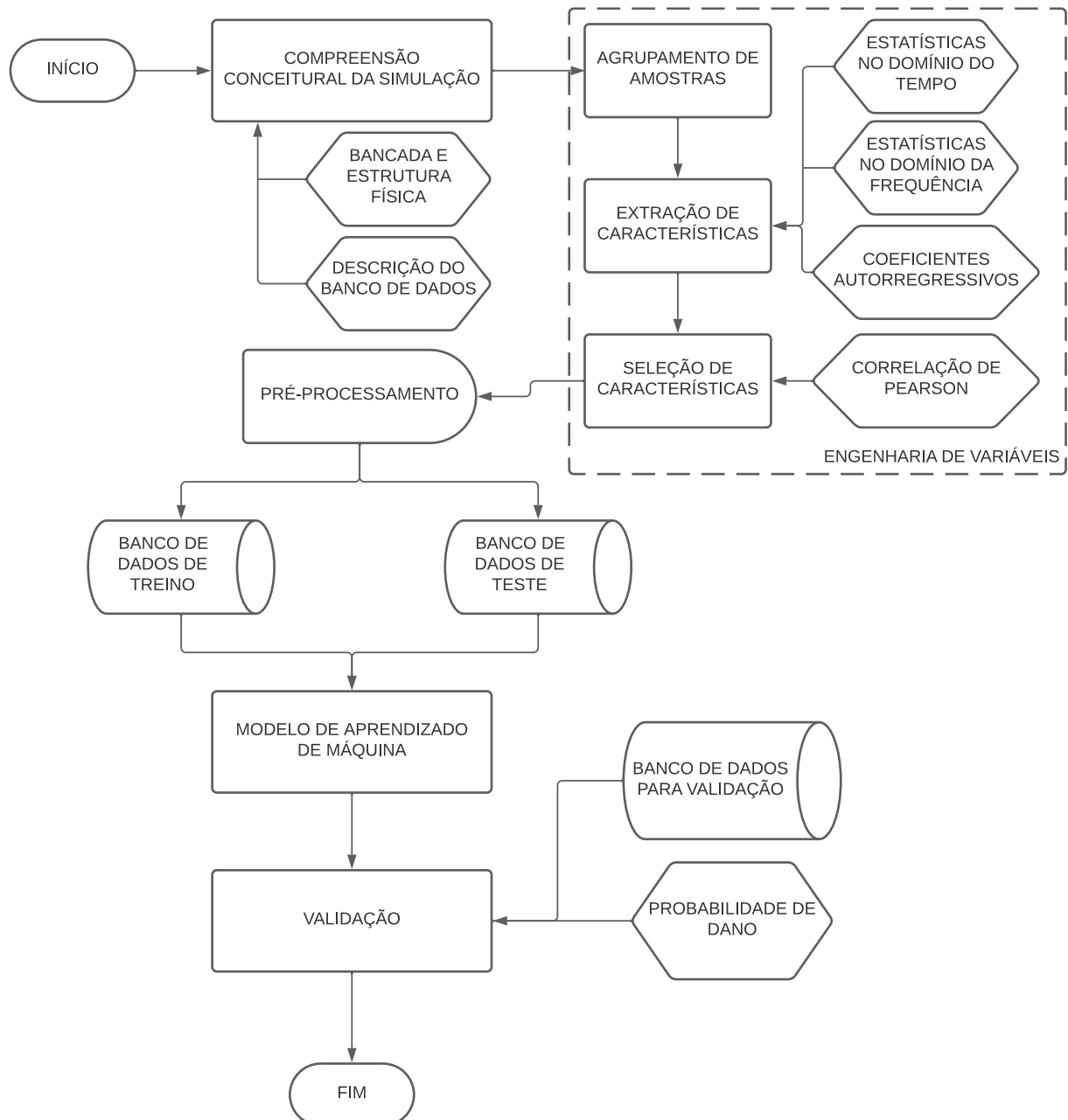
Uma das principais vantagens dos modelos de aprendizado profundo em relação aos tradicionais é a capacidade de aprender automaticamente características relevantes a partir de dados brutos, reduzindo a necessidade de uma etapa manual de engenharia de variáveis ([ABDELJABER et al., 2017](#)). Os modelos de redes neurais convolucionais de uma dimensão desenvolvidos por [Abdeljaber et al. \(2017\)](#), utilizam dessa abordagem e realizam a previsão da condição da estrutura a partir dos dados brutos, sem realização de engenharia de variáveis.

Entretanto, [Zhou e Liu \(2021\)](#) discorrem que os modelos de aprendizado profunda podem ser computacionalmente intensivos e requerem grandes quantidades de dados para treinar de forma eficaz. Eles também podem ser difíceis de interpretar e entender, o que pode dificultar o diagnóstico e a correção de erros.

3 Metodologia

A Figura 3 mostra um fluxograma que representa as etapas do trabalho realizado.

Figura 3 – Fluxograma das etapas realizadas para criação do modelo de aprendizado de máquina.



Fonte: De autoria própria.

A metodologia adotada neste estudo inicia-se com a compreensão conceitual da origem dos dados, o qual descreve a estrutura física em questão, a instrumentação de aquisição e a apresentação final dos dados, assunto tratado na seção 3.1.

Próxima etapa configura engenharia de variáveis. Nela está inclusa a extração de características dos dados, do inglês “*features*”, uma etapa crucial para modelos parametrizados, pois a característica extraída servirá de informação de entrada para o modelo que irá utilizá-la como critério para previsão. Além disso, também se faz presente a seleção de característica nessa etapa, este processo avalia a correlação entre as variáveis extraídas para realização de um filtro que evita informação redundante na hora de treinar os modelos. Tudo isso está retratado na seção 3.2.

Posteriormente é realizado um pré-processamento dos dados, nessa etapa leva-se em conta a característica estabelecida na etapa anterior para formatar os dados originais de modo a serem usados devidamente como entrada no modelo, além de checar a cardinalidade da variável objetivo no banco de dados, esse processo resulta em um novo banco de dados, pronto para ser usado no modelo. A descrição dos métodos abordados nesse processo encontra-se na seção 3.3.

Na etapa seguinte são estabelecidos parâmetros e características para o desenvolvido do modelo supervisionado de aprendizado de máquina. Este modelo é treinado e testado utilizando o banco de dados resultante da etapa anterior com intuito de avaliar seu desempenho final em indicar dano na estrutura física da treliça descrita na seção 3.1.1. Assunto devidamente descrito na seção 3.4

Finalmente, a última etapa consiste na validação do modelo desenvolvido na etapa anterior. Esse processo é feito avaliando o desempenho do modelo para indicar a probabilidade de dano nas juntas da estrutura em uma simulação diferente da vista durante as etapas de treino e teste.

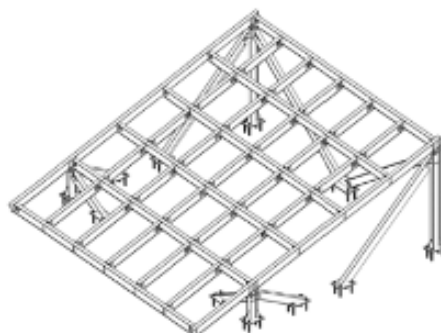
3.1 Bancada de simulação em *Qatar University*

A estrutura discutida neste trabalho é uma bancada de simulação construída na *Qatar University* (QU). Seu objetivo principal é servir como uma plataforma de teste para estudar a interação humana com estruturas e modelagem de carga induzida. A Figura 4 apresenta o projeto estrutural da arquibancada. Ao contrário de outros projetos de arquibancadas, este simulador foi projetado especificamente para testar o desempenho de vários algoritmos de detecção de danos. Para garantir a segurança e a compatibilidade com as especificações dos estádios modernos, a estrutura foi cuidadosamente projetada com considerações como conexões variáveis e membros removíveis (ABDELJABER et al., 2016).

3.1.1 Estrutura física da arquibancada

A bancada de simulação é uma estrutura de aço laminado a quente em formato de treliça, ela consiste em 8 vigas principais e 25 vigas menores complementares denominadas

Figura 4 – Projeto estrutural da arquibancada.



Fonte: Abdeljaber et al. (2016).

fillers, suportadas por 4 colunas. A plataforma possui inclinação em um ângulo de 20°. Essa bancada de simulação foi projetada para acomodar até 30 espectadores, com dimensões de 4,2m x 4,2m. No geral, esta estrutura de arquibancada fornece uma plataforma única para a realização de testes dinâmicos e pesquisas no campo da engenharia civil e arquitetônica (ABDELJABER et al., 2016).

Abdeljaber et al. (2016) ainda explicam que devido aos *fillers* serem removíveis ou substituíveis é possível simular diversos cenários de danos na estrutura desde afrouxamento de parafusos até reposição por peças danificadas.

3.1.2 Instrumentação

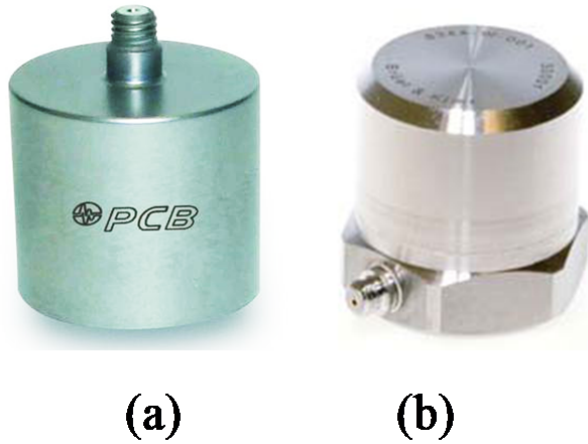
A estrutura está equipada com um total de 30 sensores acelerômetros instalados nas vigas principais da arquibancada. Os sensores foram instalados nas 30 juntas existentes entre as vigas principais e os *fillers* de forma que o sensor denominado 1 está instalado na junta 1 e assim por diante, a indicação dos sensores e suas posições estão presentes na Figura 5 e assim serão referenciados durante todo esse trabalho. São 27 acelerômetros modelo PCB 393B04 (Figura 6a) e 3 acelerômetros modelo B&K 8344 (Figura 6b) (ABDELJABER et al., 2017).

Figura 5 – Estrutura principal da bancada de simulações em QU com indicação das posições dos sensores.



Fonte: Abdeljaber et al. (2017).

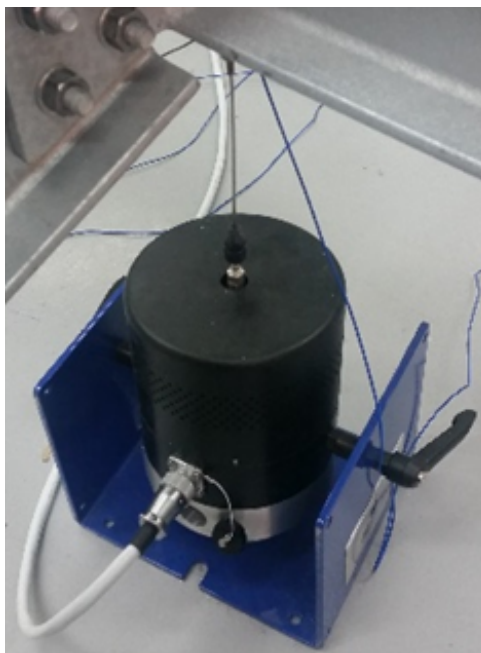
Figura 6 – Os acelerômetros modelo (a) PCB 393B04. (b) B&K 8344.



Fonte: Abdeljaber et al. (2017).

Foram utilizadas placas de montagem magnética modelo PCB 080A121 para fazer a ligação entre os acelerômetros e a estrutura. Para aplicar uma força dinâmica na estrutura foi instalado o excitador eletrodinâmico modelo 2100E11 (Figura 7) que recebe sinal de excitação pelo amplificador “SmartAmp” 2100E21-400. Finalizando a instrumentação, existem dois dispositivos de aquisição de dados de 16 canais usados para gerar a entrada pelo excitador eletrodinâmico e coletar a resposta medida pelos acelerômetros (ABDELJABER et al., 2017). A Figura 8 mostra o sistema de aquisição juntamente com o excitador.

Figura 7 – Excitador TMS 2100E11 acoplado à estrutura.



Fonte: Abdeljaber et al. (2017).

Figura 8 – Da direita para a esquerda: Os dois dispositivos de aquisição de dados (DT9857E-16), o amplificador de energia (“SmartAmp” 2100E21-400) e o notebook para registro de dados.



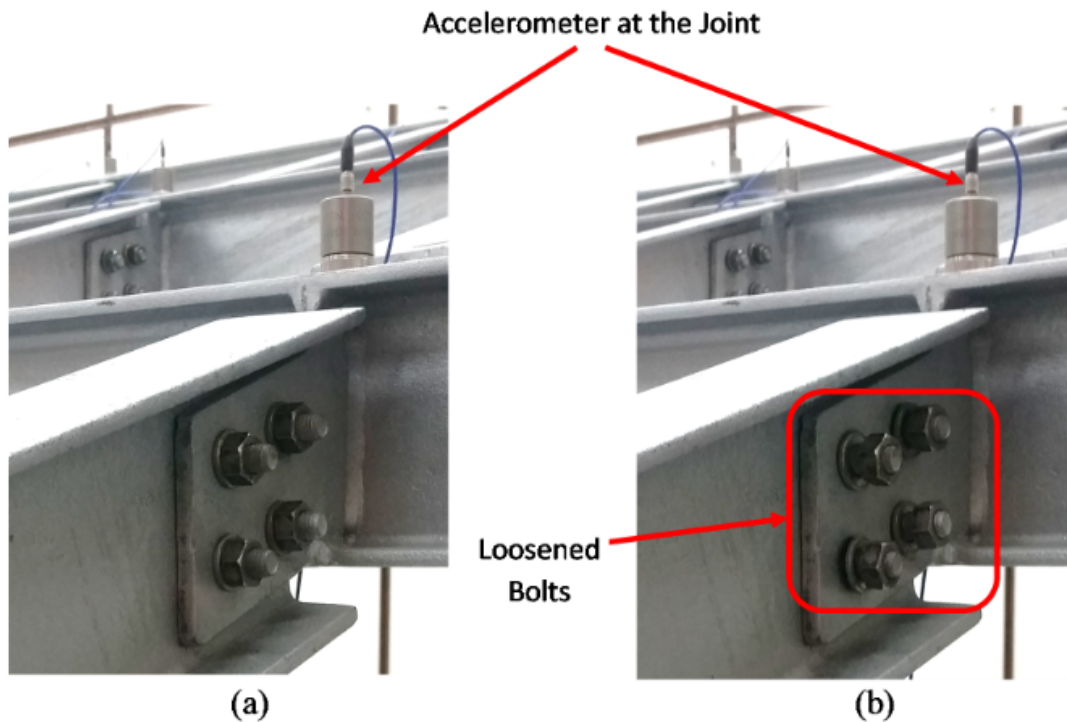
Fonte: Abdeljaber et al. (2017).

3.1.3 Tipo de dano simulado

O dano simulado consiste no afrouxamento dos parafusos das juntas que conectam as vigas e os *fillers*. A Figura 9 demonstra exemplos de junta com dano e outra sem dano, além de possibilitar observar o posicionamento do sensor tipo acelerômetro em relação à

junta. Junta sem dano corresponde as conexões em que os parafusos estão devidamente apertados (Figura 9a) e junta com dano simulado é visto na Figura 9b onde os parafusos estão afrouxados (ABDELJABER et al., 2017).

Figura 9 – Juntas (a) Com parafusos apertados. (b) Com parafusos frouxos.



Fonte: Abdeljaber et al. (2017).

3.1.4 Descrição do banco de dados

O banco de dados consiste na realização de simulações de introdução de dano à estrutura para 31 cenários diferentes. Os cenários de 1 a 30 referem-se as simulações onde há dano introduzido, respectivamente, nas juntas de 1 a 30, sendo o cenário 1 a simulação realizada na qual apenas a junta 1, onde está posicionado o sensor 1 (Figura 5), teve dano introduzido. O último cenário, denominado cenário 0, corresponde ao cenário referencial, onde a simulação não apresenta dano introduzido em nenhuma junta (ABDELJABER et al., 2017).

Todas as simulações são realizadas com a estrutura sofrendo ação do excitador. Os dados são registros do sinal de aceleração produzido pelos 30 sensores presentes na treliça. Para cada cenário simulado, esse registro é feito durante 256 segundos sob uma taxa de amostragem de 1024 Hz, totalizando 262.144 registros em domínio do tempo. O banco de dados é composto pela união dos registros de todos os 31 cenários simulados.

Cada cenário foi simulado duas vezes, dessa forma tem-se a disposição 2 bancos de dados, denominados Banco de dados A e B (ABDELJABER et al., 2017). Nesse trabalho,

o banco de dados A será utilizado na identificação do modelo classificador, e o banco de dados B nos testes de validação final.

3.2 Engenharia de Variáveis

Anteposto à etapa de extração de características, realiza-se uma seleção de dados por sensor. Para cada sensor extraem-se de um subconjunto do banco de dados com os registros referentes a esse sensor, de todos os cenários simulados. Dessa forma, o subconjunto extraído possui os dados com e sem dano simulado apenas na junta à qual o sensor refere-se. A quantidade de dados registrados presentes em um subconjunto nessa etapa é da ordem de $8,12 \times 10^6$.

Em seguida, realiza-se o agrupamento dos dados do subconjunto extraído, ou seja, os dados são divididos em janelas menores. Essa divisão é feita devido às características extraídas serem características baseadas em conjuntos de dados. Para o propósito desse trabalho, a quantidade de dados em cada um dos agrupamentos foi definido de forma arbitrária em 1024 amostras. Nesse momento, a quantidade de agrupamentos é da ordem de $7,93 \times 10^3$.

As extrações de características são realizadas nos agrupamentos. No total são três extrações realizadas e cada uma está descrita abaixo.

A seção 3.1.4 deixa explícito que o registro dos dados é feito em domínio do tempo, com isso anteposto, é elementar que as primeiras características a serem extraídas de cada agrupamento sejam características estatísticas dos dados em domínio do tempo, as quais são: média, mediana, desvio padrão, assimetria, valor mínimo, valor máximo, amplitude e curtose. Essas métricas estão detalhadas na Seção 2.4.1.

Além das estatísticas em domínio do tempo, uma segunda extração é realizada aplicando a Transformada Rápida de Fourier em cada um dos agrupamentos. Dessa forma, tanto do espectro de amplitude quanto da densidade espectral de potência (PSD) calculados pela FFT extraem-se o valor máximo, a média, a somatória e o desvio padrão, para serem usadas como características estatísticas dos dados no domínio da frequência.

Para finalizar, uma terceira extração de características é realizada. Efetua-se nos agrupamentos, a aplicação de modelagem de autorregressão de ordem 30 e a extração dos coeficientes para utilizá-los como características. A ordem da modelagem de autorregressão é definida de forma arbitrária.

As características estatísticas no domínio do tempo e no domínio da frequência foram concatenadas de maneira a formar um vetor de características de valores estatísticos. Além dele, criou-se um vetor de características de coeficientes autorregressivos para abrigar os coeficientes extraídos da autorregressão.

A seleção de características é realizada em ambos os vetores de características a fim de reduzir a dimensionalidade e redundância de informação. Para isso, aplica-se como critério de seleção a correlação de *Pearson*. O valor limite arbitrado para os coeficientes das correlações é 0,8. Dessa forma, caso duas características apresentem coeficiente de correlação entre si superior ao limite arbitrado, elimina-se uma delas.

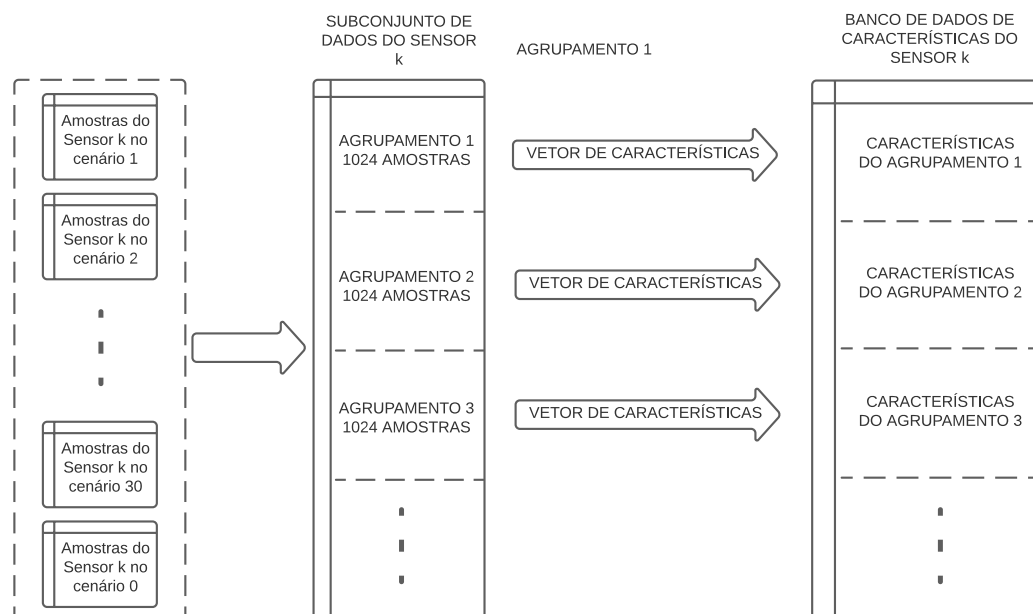
3.3 Pré-processamento de dados

O objetivo dessa etapa é servir de interface das características selecionadas com o modelo de aprendizado de máquina. Essa interface é composta de métodos que criam e moldam um banco de dados para ser usado pelos modelos tanto para treino como para teste.

O presente estudo propõe o desenvolvimento de um modelo supervisionado de aprendizado de máquina para cada um dos sensores na estrutura. Dessa forma, constrói-se um classificador binário, dano ou saudável, para cada junta da estrutura. Apesar de exigir grande quantidade de sensores, essa estratégia garante que danos simultâneos em localizações distintas possam ser detectados. Para isso, cada um dos modelos necessita de seu próprio banco de dados, com dados coletados na junta em que o sensor foi instalado.

A Figura 10 ilustra o processo de criação dos bancos de dados de características dos sensores por meio de um fluxograma. Esse banco de dados é composto pelos vetores de características obtidos a partir da seleção das características extraídas dos agrupamentos.

Figura 10 – Fluxograma de criação do banco de dados de características de um sensor qualquer k a partir do subconjunto de dados desse sensor.



Fonte: De autoria própria.

Próximo passo é adicionar a variável objetivo nos vetores de características. Essa variável é uma variável categórica binária que assume valores de 0 ou 1 dependendo do agrupamento da qual as características foram extraídas. Caso o agrupamento seja de amostras cuja junta que se referem há dano introduzido, o valor da variável objetivo é 1. Caso contrário o valor da variável objetivo é 0. É importante enfatizar que, se há outra junta com dano na medição daquele agrupamento de dados, o valor da variável objetivo é 0, pois busca-se identificar com cada modelo apenas o dano localizado na junta daquele sensor. O banco de dados resultante é um conjunto de amostras que serão usadas pelos modelos, cada amostra é composta por um vetor de características e a variável objetivo.

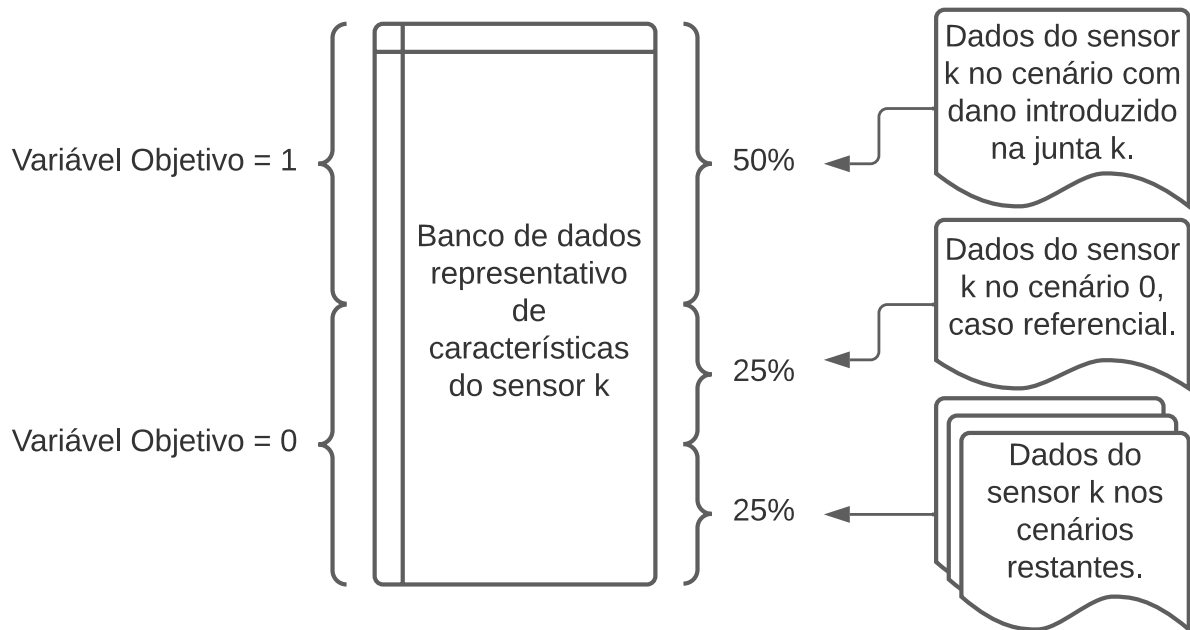
Uma etapa de verificação de cardinalidade no banco de dados de características é realizada. A finalidade desse processo é verificar se há representatividade da variável objetivo entre as amostras. Para isso, uma reestruturação deve ser realizada no banco de dados seguindo os critérios listados abaixo.

1. 50% do banco de dados deve ser composto de amostras em que a variável objetivo equivale a 1. Essas amostras são oriundas dos agrupamentos de dados coletados no cenário em que havia dano introduzido na junta relacionada.
2. 25% do banco de dados deve ser composto de amostras em que a variável objetivo equivale a 0. Essas amostras são oriundas dos agrupamentos de dados coletados no cenário 0, caso referencial.
3. 25% do banco de dados deve ser composto de amostras em que a variável objetivo equivale a 0. Essas amostras são oriundas dos agrupamentos de dados coletados no restante dos cenários que não havia dano introduzido na junta relacionada, mas sim em outras juntas da estrutura.

O banco de dados resultante é denominado banco de dados representativo de características. A Figura 11 ilustra como deve ser a estrutura do banco de dados representativo de características para cada modelo/sensor.

A última etapa do pré-processamento é a divisão do banco de dados representativo de características em dois subconjuntos complementares, para teste e para treino. Para o presente trabalho, houve duas abordagens utilizadas simultaneamente. A primeira abordagem envolve uma divisão embaralhada de amostras, definida em 70% para treino e 30% para teste. A segunda abordagem envolve validações cruzadas com 5 iterações.

Figura 11 – Estrutura do banco de dados representativo das amostras.



Fonte: De autoria própria.

3.4 Modelo de aprendizado de máquina

No trabalho em questão, foi escolhido o modelo de aprendizado de máquina classificador de floresta aleatória para aplicação e a abordagem adotada consiste em treinar um modelo para cada sensor, posicionado na sua junta de referência. Essa estratégia permitiu que o treinamento fosse adaptado conforme as características e comportamentos particulares de cada sensor, proporcionando uma modelagem mais precisa e eficiente. Ao aplicar essa técnica, é possível capturar as nuances e padrões específicos de cada sensor, além de maximizar a precisão das previsões realizadas.

Estima-se dois modelos para prever a variável objetivo estipulada. O primeiro modelo foi treinado e testado utilizando como métricas o vetor de características de valores estatísticos, não obstante o segundo modelo foi treinado e testado utilizando como métricas o vetor de características de coeficientes autorregressivos. Ao final dos testes, uma avaliação deve ser feita de modo a comparar o desempenho de ambos os modelos na proposta de predição.

3.4.1 Treinamento e teste dos modelos

A base de dados utilizada nessa etapa envolve a coleta das informações apenas do banco de dados A, referenciado na Seção 3.1.4.

A quantidade de estimadores ($n_{estimators}$) é o primeiro parâmetro estabelecido para o modelo treinado e testado com o banco de dados baseado no vetor de características

de valores estatísticos. Esse parâmetro indica o número de árvores de decisão combinadas para fazer a previsão e foi arbitrado o valor de 5000. Além dele, o número de características a ser avaliado como decisão de divisão dos nós nas árvores (*max_features*), bem como a quantidade máxima de nós na árvore de decisão (*max_depth*) também compõem os parâmetros avaliados nessa abordagem. Esses parâmetros são otimizados para um melhor desempenho dos modelos.

Assim como no modelo anterior, os mesmos parâmetros são avaliados para o modelo treinado e testado com o banco de dados baseado no vetor de características de coeficientes autorregressivo. A diferença nesse modelo envolve o valor arbitrado de quantidade de estimadores, que passa a ser de 500. A quantidade menor de estimadores, nesse caso, está relacionada a melhor performance do modelo.

A métrica para avaliação do modelo em prever a variável objetivo no banco de dados de teste é a acurácia que o modelo desempenha, definida na Equação 3.1. O termo Vp representa a quantidade de verdadeiro positivo, Vn representa a quantidade de verdadeiro negativo e Pt representa a quantidade total de previsões realizadas. Nessa métrica, 1 indica 100% de acerto nas previsões e 0 indica nenhum acerto.

$$acurácia = \frac{(Vp + Vn)}{Pt} \quad (3.1)$$

A acurácia do modelo sob o regime de validações cruzadas é definida como a média das acurácias de cada uma das interações. Para melhor descrição dos resultados, a média das acurácias é acompanhada do desvio padrão (σ_M), que é uma medida de dispersão de valores calculada pela Equação 3.2 onde n representa a quantidade de subdivisões utilizada na validação cruzada (AFONSO; NUNES, 2019).

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \quad (3.2)$$

Realiza-se a otimização por meio do pacote *Optuna* na linguagem de programação *python* que estabelece uma função objetivo para otimizar uma métrica em função de outras até um determinado limite. Para ambos os modelos nesse estudo, aplica-se a função objetivo para maximizar a acurácia do modelo, por meio da variação tanto *max_depth* de 1 à 50 quanto *max_features* de 1 a quantidade de características presente no vetor de característica, com passo de 1. O número máximo de variações foi limitado a 300.

3.4.2 Validação dos modelos

Os modelos escolhidos para a etapa de validação utilizando como base o banco de dados B são aqueles treinados e testados sob o regime de validações cruzadas com os hiper-parâmetros obtidos durante a otimização. Nessa etapa, os modelos procuram prever

a variável objetivo num banco de dados que nunca tiveram contato antes.

Para simular uma situação real onde não sabemos previamente a existência e localização do dano, a métrica que será utilizada é a de probabilidade de dano (PoD) em cada uma das juntas da estrutura. A Equação 3.3 apresenta como é calculada essa métrica para uma junta i , onde $T(i)$ é o número de amostras analisadas e $D(i)$ é a quantidade de amostras houve previsão de dano pelo sensor (ABDELJABER et al., 2017).

$$PoD(i) = \frac{D(i)}{T(i)} \times 100\% \quad (3.3)$$

Espera-se com essa métrica que os valores de probabilidade de dano sejam significativamente maiores nas juntas onde há presença de dano em relação às juntas onde não há. A PoD acima de 50% indica maior probabilidade de presença de dano na junta em relação à ausência de dano nela. Dessa forma, fornecendo uma clara indicação tanto da presença quanto da localização do dano na estrutura.

A probabilidade de falsos positivos é uma indicação das juntas onde os modelos diagnosticam dano sem que ele exista de fato. Evitar o diagnóstico de falsos positivos é essencial, pois resulta em desperdício de tempo e recursos de manutenção. Além disso, o número de falsos positivos pode ser utilizado como métrica de desempenho dos modelos.

A probabilidade de falsos negativos é uma indicação das juntas onde os modelos falham em diagnosticar dano mesmo ele se fazendo presente. A ocorrência de um falso negativo em uma estrutura representa uma ameaça significativa, afetando tanto a segurança das pessoas quanto a integridade da própria estrutura. Essa situação se deve à ausência de aplicação dos recursos de manutenção em um dano existente, mas não diagnosticado. Dessa forma, utilizar falsos negativos como métrica de desempenho dos modelos é imprescindível para evitar tais ocorrências.

4 Apresentação e discussão de resultados

Os resultados apresentados nesse estudo iniciam-se com a etapa de seleção das características extraídas seguindo o critério da metodologia, essa etapa está presente nas seções 4.1 e 4.2. Além disso, apresentam-se o resultado e a discussão da etapa de cardinalidade na seção 4.3.

A seção 4.4 mostra o desempenho dos modelos nas amostras de teste, treinados a partir de amostras tanto do vetor de valores estatísticos quanto do vetor de valores autorregressivos. Nessa seção apresentam-se ambas as acurácias dos modelos com e sem validação cruzada, com hiper-parâmetros já otimizados.

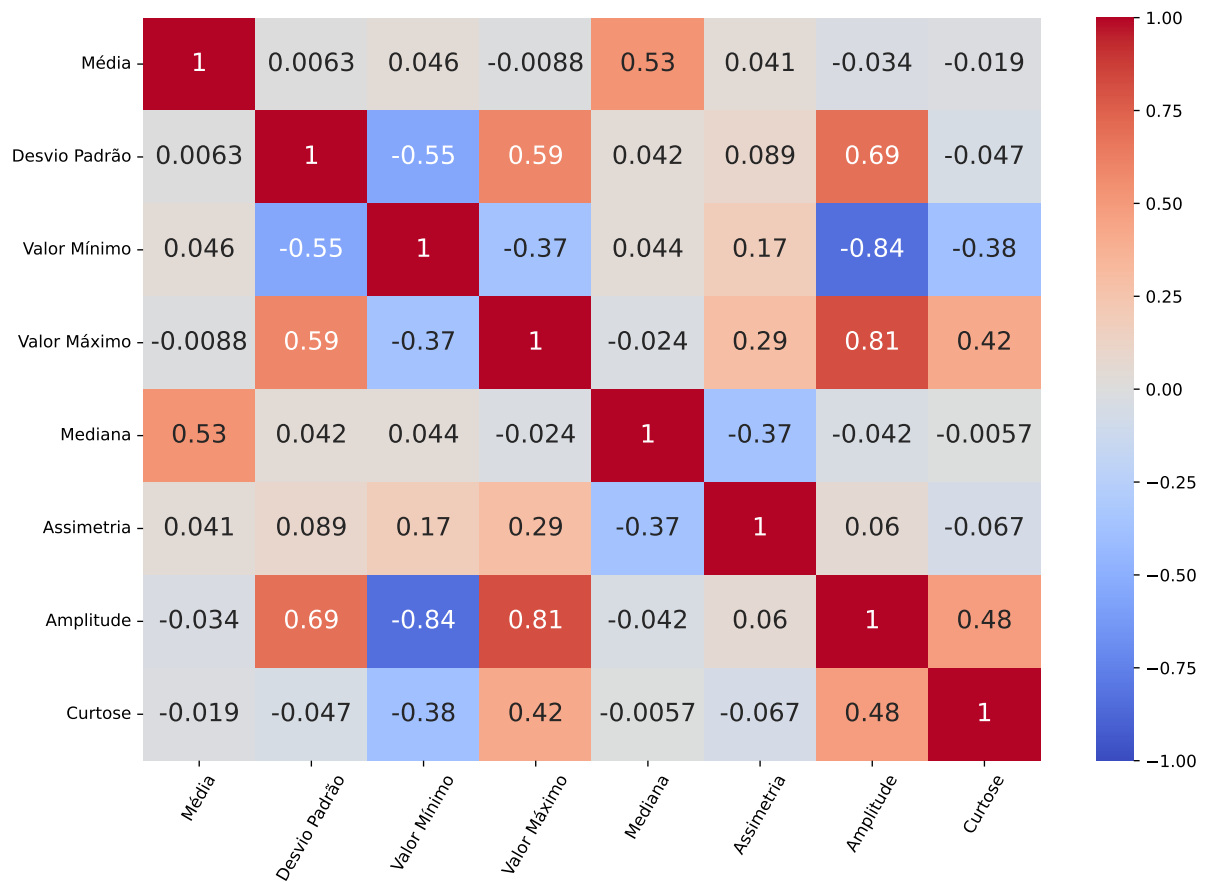
O desempenho dos modelos em prever a presença e localização de dano em amostras do banco de dados B está presente na seção 4.5. Para finalizar, a seção 4.6 compara o desempenho dos modelos desenvolvidos nesse estudo em relação aos modelos de aprendizado profundo da literatura.

4.1 Vetor de características de valores estatísticos

A Figura 12 apresenta esquematicamente os coeficientes de correlação de *Pearson* entre as características estatísticas no domínio do tempo obtidas na etapa de extração de características. Observa-se que as características de valor mínimo e valor máximo apresentam uma baixa correlação entre si, porém com uma alta correlação com a amplitude. Essa correlação pode ser justificada pela própria definição da amplitude como variável dependente do valor mínimo e do valor máximo.

Além dos coeficientes entre amplitude e os valores mínimo e máximo, não há outro coeficiente de correlação com valor superior ao critério estabelecido de 0,8 na metodologia. Portanto, a fim de reduzir a redundância de informação, eliminam-se as características de Valor Mínimo e Valor Máximo.

Figura 12 – Coeficientes das correlações de *Pearson* entre características estatísticas no domínio do tempo extraídas.

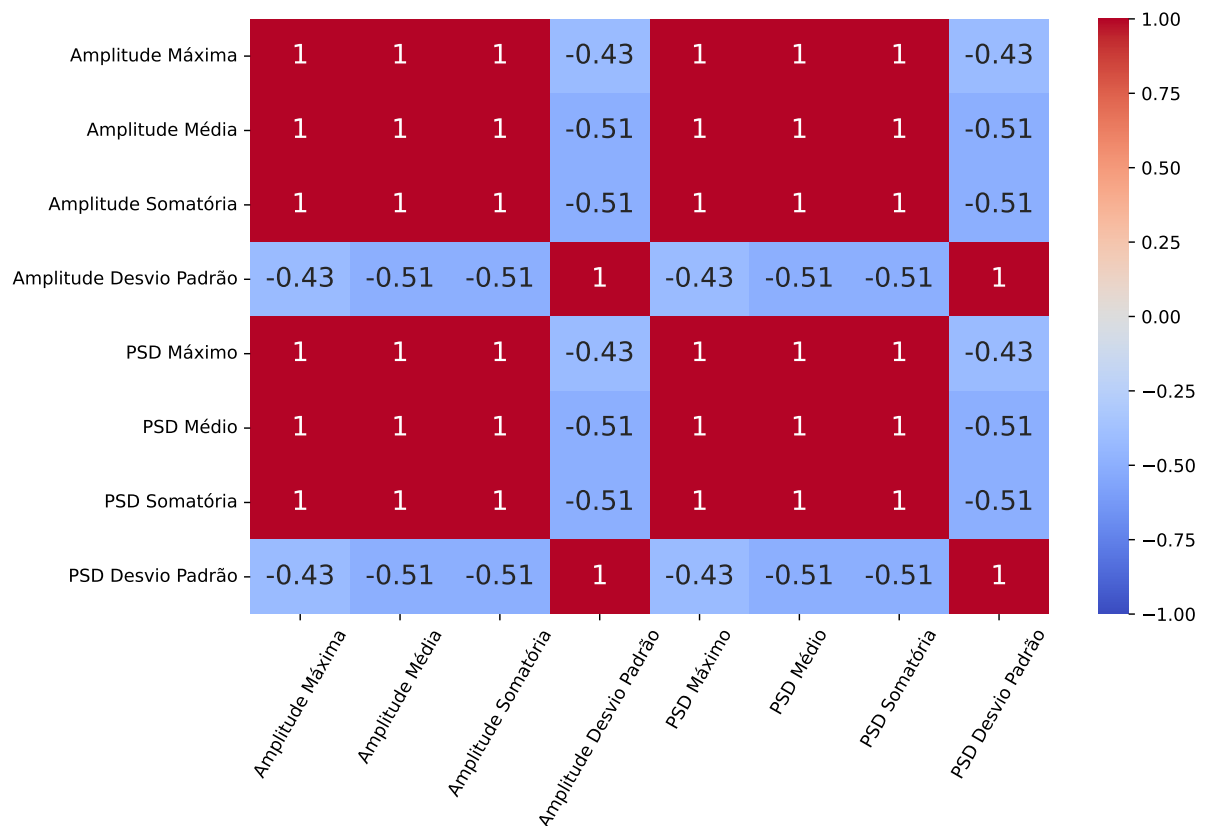


Fonte: De autoria própria.

A Figura 13 apresenta esquematicamente os coeficientes de correlação de *Pearson* entre as características estatísticas no domínio da frequência obtidas na etapa de seleção de características. De modo a evitar que picos de amplitudes em frequências específicas ditem o comportamento, as características foram correlacionadas na escala logarítmica de base 10.

Após avaliar os coeficientes da Figura 13, decide-se manter o desvio padrão do sinal espectral (Amplitude Desvio Padrão) e a máxima amplitude da PSD (PSD Máximo) como as características para representar o domínio da frequência.

Figura 13 – Coeficientes das correlações de *Pearson* entre características estatísticas no domínio da frequência.

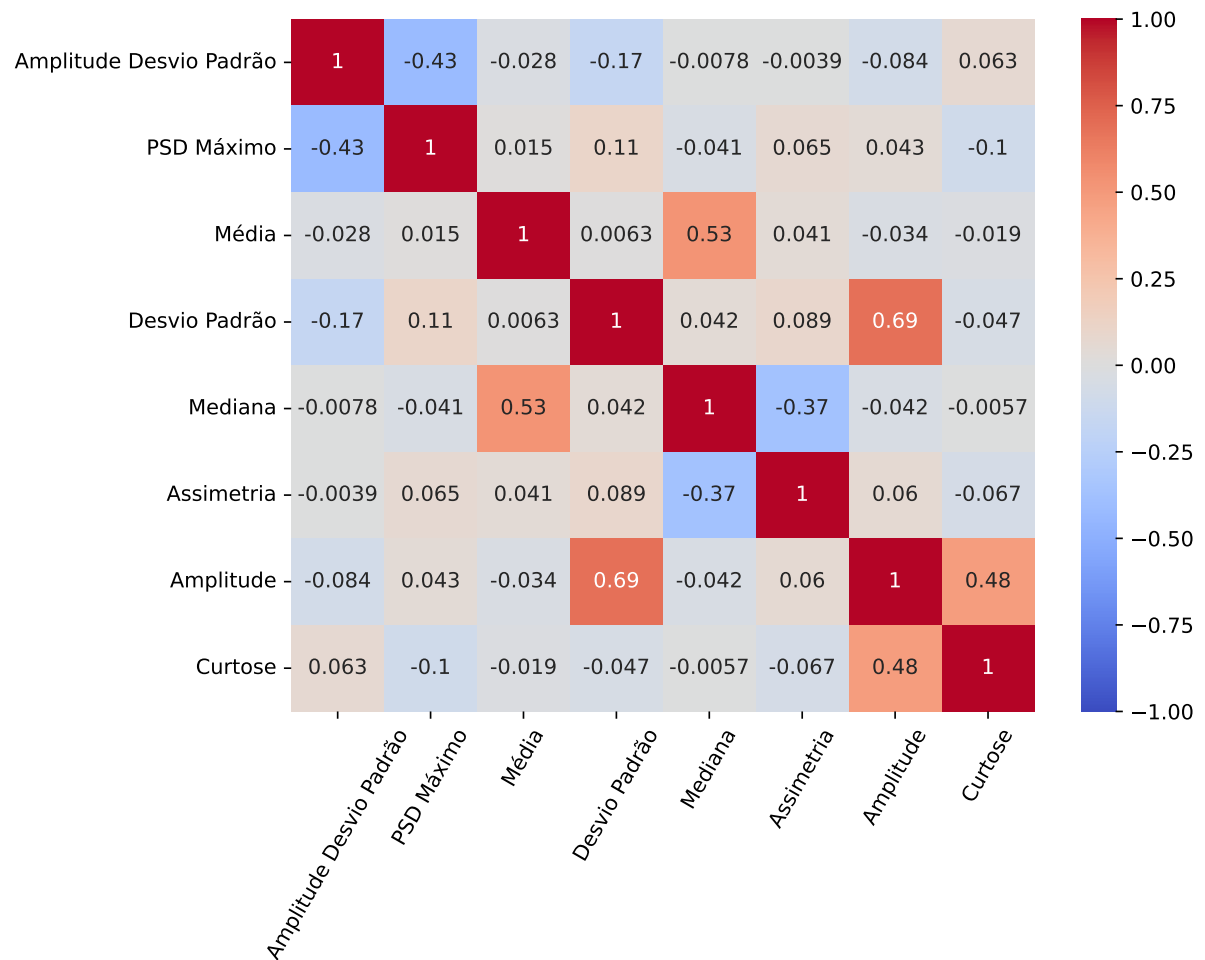


Fonte: De autoria própria.

Para finalizar, a Figura 14 apresenta os coeficientes de correlação entre todas as características selecionadas como valores estatísticos, unindo as características selecionadas no domínio do tempo com as características selecionadas no domínio da frequência.

Analisando a Figura 14, observa-se que não há coeficientes acima do critério estabelecido na metodologia para as características selecionadas. Portanto, O vetor de características de valores estatísticos apresenta as seguintes características, desvio padrão do sinal espectral (Amplitude Desvio Padrão), máxima amplitude da PSD (PSD Máximo), média, desvio padrão, mediana, assimetria, amplitude e curtose.

Figura 14 – Coeficientes das correlações de *Pearson* entre características de valor estatísticos.

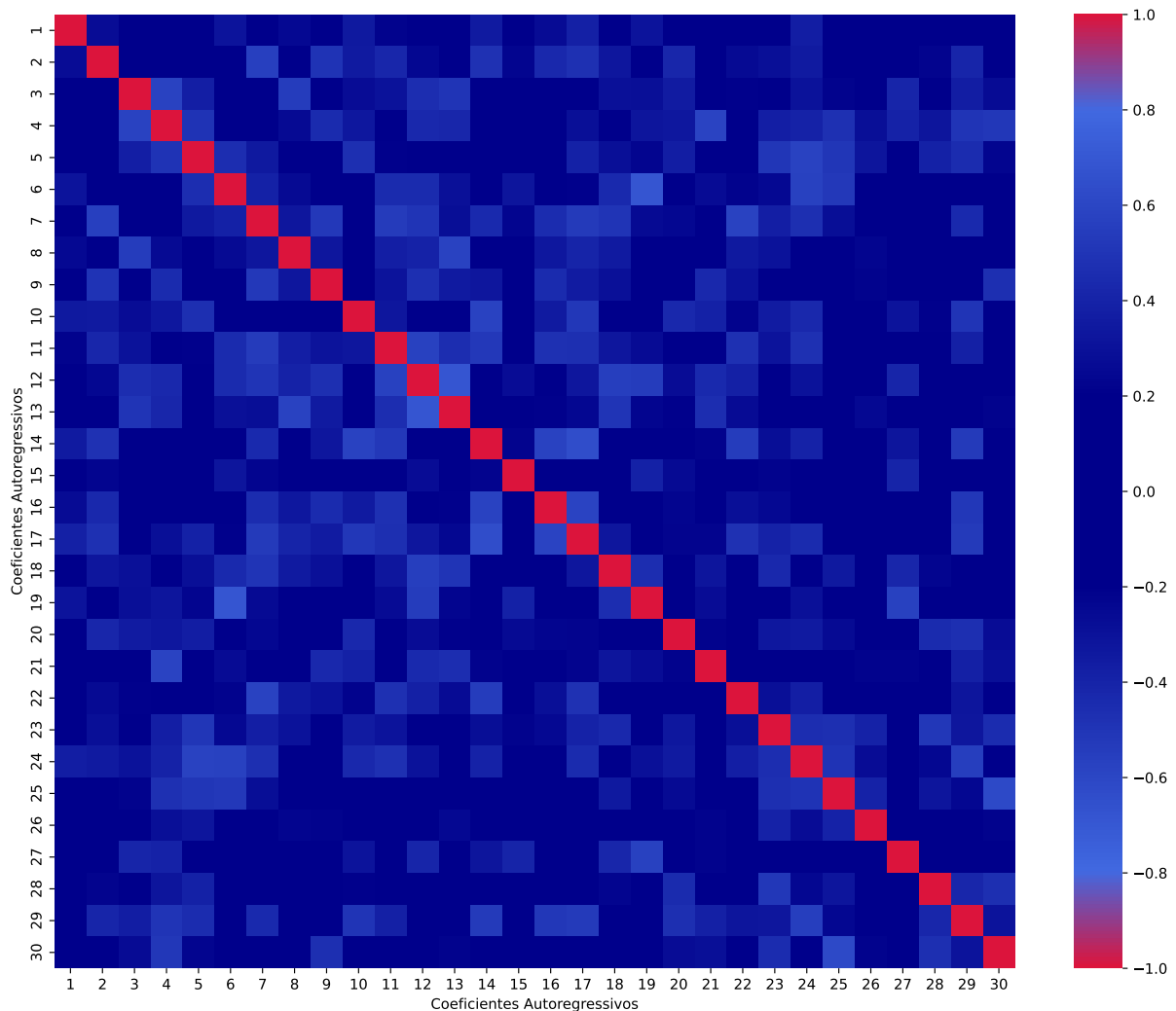


Fonte: De autoria própria.

4.2 Vetor de características de coeficientes autorregressivos

O critério de seleção de características baseado na correlação de *Pearson* também é aplicado ao vetor de características de coeficientes autorregressivos. As correlações estão presentes na Figura 15 onde indica, visualmente, se há coeficiente de correlação em que o módulo do seu valor seja superior à 0,8, critério adotado na metodologia.

Ao analisar a Figura 15, é possível observar que nenhuma correlação entre os coeficientes autorregressivos alcançou o critério estabelecido para seleção de características. Portanto, todos os coeficientes autorregressivos são mantidos no vetor de características.

Figura 15 – Correlações de *Pearson* entre características de coeficientes autorregressivos.

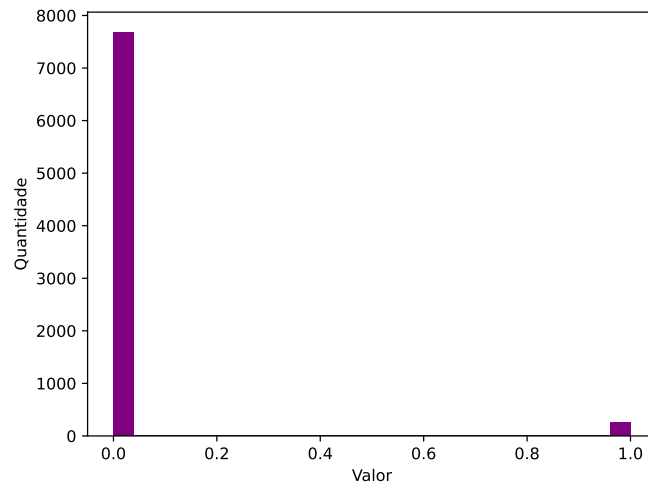
Fonte: De autoria própria.

4.3 Cardinalidade

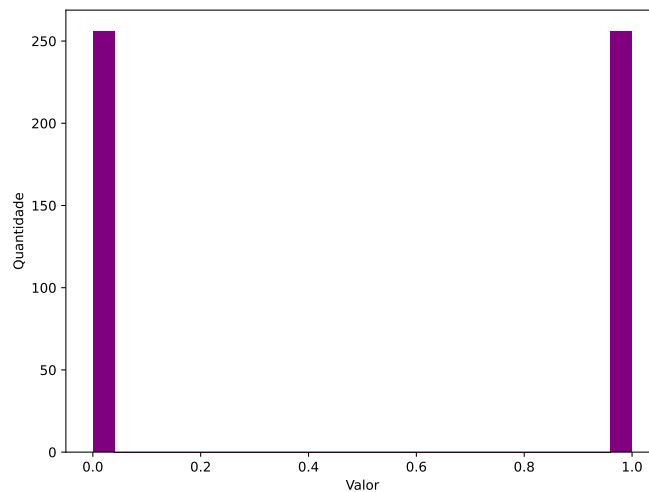
Com relação à cardinalidade no banco de dados de características, a Figura 16 mostra a disparidade entre a quantidade de amostras cuja variável objetivo tem valor 0, sem dano, e quantidade que têm valor 1, com dano. A Figura 16a mostra a cardinalidade do banco de dados antes da reestruturação proposta na metodologia e a Figura 16b mostra a cardinalidade do banco de dados representativo de características.

Ao observar a Figura 16a, apenas 3,22% do banco de dados de características correspondem a dados que fornecem ao modelo informação com variável objetivo indicando dano. Por conta disso, modelos treinados com o banco de dados sem a reestruturação da Figura 16b apresentam bom desempenho nas juntas sem danos, enquanto que o mesmo não é visto em juntas com dano aferido. Nesses casos, durante a etapa de teste, a porcentagem de acerto que os modelos desempenham está entre 96-98% em prever dano em juntas que

Figura 16 – Representatividade dos valores da variável objetivo.



(a) Cardinalidade da variável objetivo no banco de dados de características. Fonte: De autoria própria.



(b) Cardinalidade da variável objetivo no banco de dados de características após a reestruturação proposta na metodologia. Fonte: De autoria própria.

se sabe não haver dano. Entretanto, para juntas que se sabe haver dano, a porcentagem cai para 55-60% dependendo da junta. A justificativa da discrepância dos desempenhos é atribuída a baixa representatividade de dados com dano simulado, o modelo não apresenta bom desempenho para cenário que não há dano.

Por fim, adota-se a reestruturação nos bancos de dados de características de todos os sensores. Dessa forma, a Figura 16b mostra também a distribuição dos dados realmente considerada nos treinos e testes do modelo.

4.4 Desempenho em amostras de teste

A Tabela 1 apresenta as acurácias dos modelos na etapa de teste para ambos os vetores de características, com e sem validação cruzada. Os valores se referem aos modelos já otimizados e os hiper-parâmetros resultante dessa otimização estão apresentados no Apêndice A.

Tabela 1 – Acurácia dos modelos otimizados na previsão da variável objetivo no banco de dados A.

ACURÁCIA DOS MODELOS				
Sensor	Sem validação cruzada		Validação cruzada de 5 <i>folds</i>	
	Vetor estat.	Vetor autorreg.	Vetor estat.	Vetor autorreg.
1	0,6364	0,9805	$0,4730 \pm 0,0860$	$0,9922 \pm 0,0070$
2	0,5714	0,8377	$0,4786 \pm 0,0563$	$0,8831 \pm 0,0269$
3	0,8571	1	$0,2833 \pm 0,0276$	1 ± 0
4	0,6883	1	$0,8454 \pm 0,0319$	1 ± 0
5	0,5844	1	$0,6549 \pm 0,0627$	1 ± 0
6	0,7857	0,987	$0,4590 \pm 0,0805$	$0,9961 \pm 0,0035$
7	0,6818	0,9805	$0,7233 \pm 0,0770$	$0,9882 \pm 0,0064$
8	0,6169	0,9935	$0,3553 \pm 0,0261$	$0,9882 \pm 0,0064$
9	0,6623	0,9935	$0,3791 \pm 0,0715$	$0,9980 \pm 0,0017$
10	0,5325	1	$0,4378 \pm 0,0856$	$0,9980 \pm 0,0017$
11	0,6753	0,9935	$0,3713 \pm 0,0258$	$0,9961 \pm 0,0021$
12	0,6299	1	$0,4886 \pm 0,0493$	$0,9883 \pm 0,0064$
13	0,7403	0,961	$0,4906 \pm 0,0566$	$0,9628 \pm 0,0119$
14	0,6364	1	$0,6586 \pm 0,0728$	$0,9922 \pm 0,0033$
15	0,7013	0,987	$0,3767 \pm 0,0497$	$0,9883 \pm 0,0064$
16	0,7792	1	$0,6270 \pm 0,0265$	1 ± 0
17	0,8247	1	$0,8044 \pm 0,0381$	$0,9980 \pm 0,0017$
18	0,5649	0,9935	$0,8375 \pm 0,0396$	1 ± 0
19	0,5909	0,9286	$0,3945 \pm 0,0375$	$0,9610 \pm 0,0055$
20	0,6494	1	$0,4412 \pm 0,0555$	1 ± 0
21	0,7013	1	$0,5157 \pm 0,0729$	1 ± 0
22	0,7597	1	$0,5021 \pm 0,0364$	$0,9961 \pm 0,0021$
23	0,7727	1	$0,4647 \pm 0,0489$	$0,9981 \pm 0,0017$
24	0,5714	0,9935	$0,7535 \pm 0,0492$	$0,9980 \pm 0,0017$
25	0,6364	0,9935	$0,4840 \pm 0,1001$	$0,9785 \pm 0,0093$
26	0,6688	1	$0,4943 \pm 0,0350$	1 ± 0
27	0,8442	1	$0,6272 \pm 0,0609$	1 ± 0
28	0,6623	1	$0,7268 \pm 0,0847$	1 ± 0
29	0,5844	0,9416	$0,6976 \pm 0,0457$	$0,9453 \pm 0,0059$
30	0,7727	0,9805	$0,5666 \pm 0,0678$	$0,9980 \pm 0,0017$

Fonte: De autoria própria.

4.4.1 Modelos treinados com vetor de características de valores estatísticos

A partir da análise da Tabela 1, a acurácia dos modelos treinados com características de valores estatísticos é baixa mesmo com tentativas de otimização dos modelos. Além disso, a redução na acurácia das abordagens sob regime de validação cruzada indica *overfitting* dos modelos para com os subconjuntos de dados para treino.

O desempenho ruim indica que características de valores estatísticas falham em apresentar relação de causalidade com a variável objetivo de modo a ensinar modelos a realizar previsões assertivas. A hipótese provável para o baixo desempenho é a dependência das métricas estatísticas da força aplicada, ou seja, força de excitação. É provável que a variação nas métricas estatísticas esteja muito mais relacionada a variação nas estatísticas da força de entrada do que a ocorrência de dano na estrutura. Como a força de entrada não é medida, não há como separar os efeitos diretamente. A partir disso, o presente estudo não realiza a etapa de validação para os modelos mencionados nessa seção.

4.4.2 Modelos treinados com vetor de características de coeficientes autorregressivos

Logo de início, as características de coeficientes autorregressivos demonstram conter relação de causalidade com a variável objetivo. É possível observar na Tabela 1 que a maioria dos modelos apresenta desempenho satisfatório com acurácia acima de 90%. Alguns outros modelos, como é o caso do sensor 3, a acurácia apresentada chega a ser de 100%, é provável que o alto desempenho dos coeficientes esteja associado a relação existente entre eles e os parâmetros modais da estrutura, que por sua vez, estão diretamente ligados a estrutura física (distribuição de massa e rigidez) (FU; HE, 2001). Quando há variação estrutural gerada pelo dano, os parâmetros modais variam e os coeficientes acompanham essa variação. Porém, tais coeficientes não demonstram ser influenciados por variações da força de entrada, funcionando como um filtro de informação importante.

Devido ao desempenho apresentado por esses modelos, é seguro dizer que os coeficientes autorregressivos para esse estudo são mais relevantes que valores estatísticos.

4.5 Desempenho em validação com banco de dados B

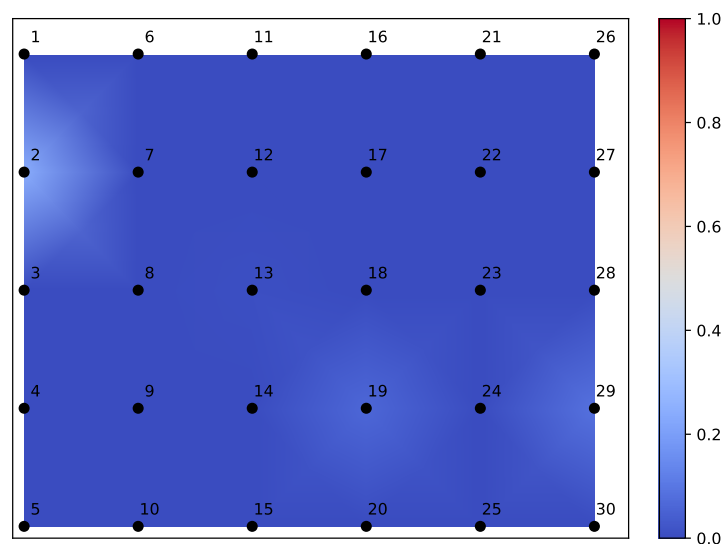
Fazendo uso dos modelos já treinados e testados, realiza-se a validação final colocando-os para prever a existência e localização de dano na simulação registrada no banco de dados B. Vale ressaltar que os modelos não tiveram contato com as amostras dessa simulação durante suas etapas de treino e teste.

Para melhor apresentação dos resultados usa-se figuras representando a estrutura e, por meio de mapas de calor, indicar a probabilidade de dano em cada junta seguindo

a nomenclatura observada na Figura 5. O Apêndice B abriga as figuras que mostram o desempenho dos modelos em todos os cenários. A seguir apresentam-se alguns desses cenários com intuito de demonstrar o desempenho dos modelos em diferentes simulações.

A Figura 17 mostra a condição da estrutura no cenário 0, sem dano. Portanto, não há possibilidade de presença de falsos negativos para esse cenário. A partir dos resultados é possível observar que os modelos indicam baixa probabilidade de presença de dano na estrutura inteira, o que é verdade. A maior PoD nesse cenário está localizada na junta 2 com 23,83%, seguido da junta 29 com 8,59%, ambos os valores representam baixa probabilidade de dano nas juntas. Além disso, a maioria das juntas restantes apresentam PoD igual a 0% e não houve presença de falsos positivos. Dessa forma, é possível afirmar o bom desempenho dos modelos em avaliar a condição da estrutura no cenário indicado.

Figura 17 – Probabilidade de dano no cenário 0, sem dano existente, na simulação do banco de dados B.

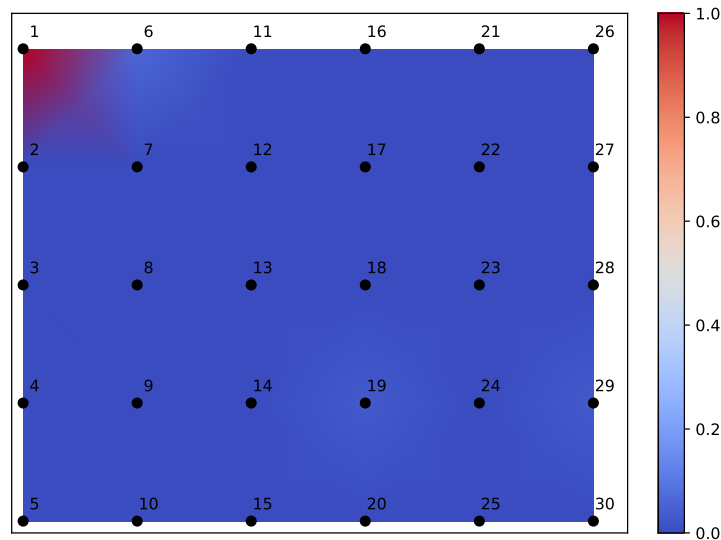


Fonte: De autoria própria.

As Figuras 18 e 19 são exemplos de cenários onde os modelos indicam com clareza a existência e localização do dano assertivamente.

A Figura 18 mostra a condição da estrutura no cenário 1, com dano existente na junta 1. A partir dos resultados observa-se que a maior PoD está localizada na junta 1 com valor de 100%, indicando corretamente que há dano e ele está localizado na junta 1. Além disso, descarta-se a probabilidade de um falso negativo, haja vista que apenas a junta 1 apresenta dano nesse cenário. Com relação às juntas restantes, a maior PoD seguinte à junta 1 está localizada na junta 6 com valor de 5,86%, indicando baixa probabilidade de falsos positivos na estrutura. Portanto, é possível afirmar o bom desempenho dos modelos em avaliar a condição da estrutura no cenário 1.

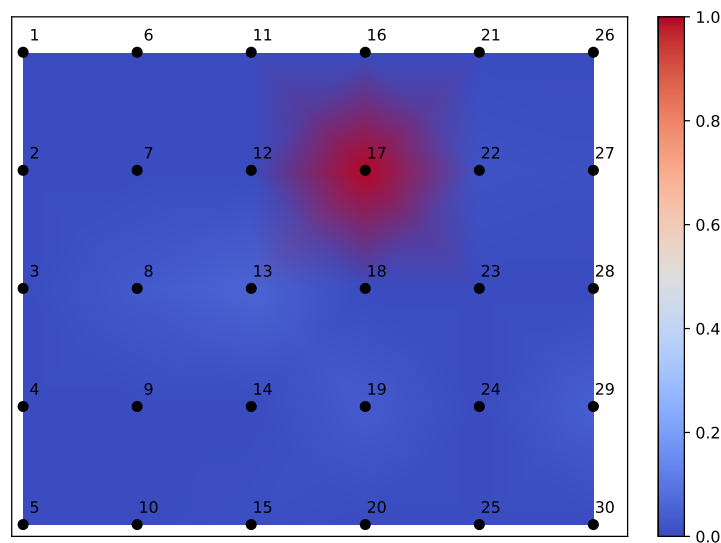
Figura 18 – Probabilidade de dano no cenário 1 na simulação do banco de dados B.



Fonte: De autoria própria.

A Figura 19 mostra a condição da estrutura no cenário 17, com dano existente na junta 17. A maior PoD observada está localizada na junta 17 com valor de 99,22%, indicando corretamente dano nessa junta e a baixa probabilidade de um falso negativo presente no cenário. Nas demais juntas, as maiores PoD seguintes à junta 17 estão localizadas nas juntas 13 e 29 com valores de 5,47% e 4,3%, respectivamente. Portanto, semelhante à análise feita para o cenário 1, indica baixa probabilidade de falsos positivos na estrutura, sendo então possível afirmar o bom desempenho dos modelos em avaliar a condição da estrutura no cenário 17.

Figura 19 – Probabilidade de dano no cenário 17 na simulação do banco de dados B.

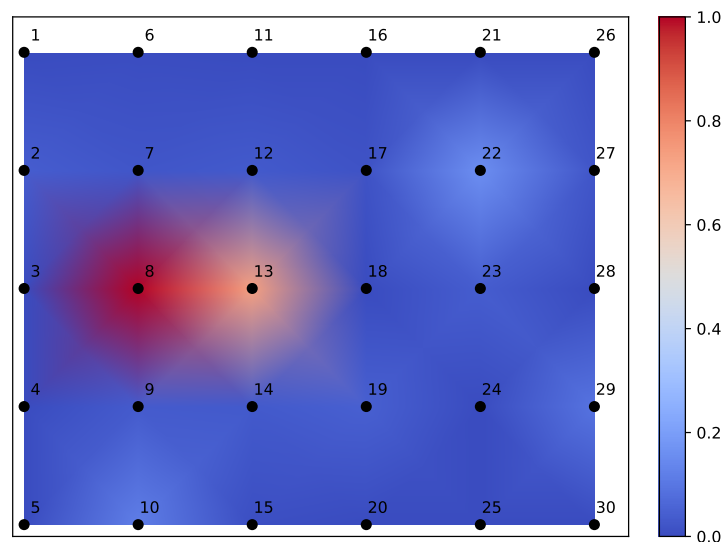


Fonte: De autoria própria.

No entanto, é importante notar que alguns cenários específicos apresentam resultados satisfatórios, o que significa que eles podem não ter atendido totalmente aos critérios ou objetivos desejados. Como exemplo disso, têm-se os cenários 8 e 10 apresentados, respectivamente, nas Figuras 20 e 21.

A Figura 20 mostra a condição da estrutura no cenário 8, com dano existente na junta 8. A maior PoD observada está localizada na junta 8 com valor de 99,61%, indicando corretamente dano na junta 8 com baixa probabilidade de falso negativo. Apesar disso, os resultados apontam 73,05% de probabilidade de dano na junta 13, o que não é verdade e indica ser um falso positivo. Nas demais juntas, as maiores PoD seguintes à junta 13 estão localizadas nas juntas 22 e 10 com valores de 15,63% e 11,72%, respectivamente. Os resultados indicam baixa probabilidade de outros falsos positivos além daquele presente na junta 13.

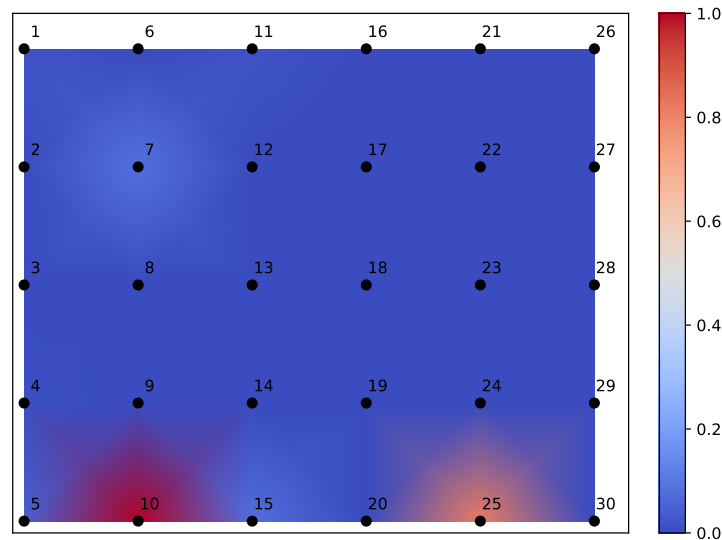
Figura 20 – Probabilidade de dano no cenário 8 na simulação do banco de dados B.



Fonte: De autoria própria.

Outro caso onde há presença de um falso positivo é o mostrado pela Figura 21, com dano existente na junta 10. A junta 25 apresenta 81,64% de probabilidade de dano, o que não é verdade, indicando ser um falso positivo. Apesar do erro, ainda sim a maior PoD está localizada na junta 10 com valor de 100%, indicando corretamente a presença de dano e descarta a probabilidade de falso negativo no cenário 10. Nas demais juntas, destacam-se as PoD nas junta 7 e 15, com valores de 8,2% e 7,81%, respectivamente.

Figura 21 – Probabilidade de dano no cenário 10 na simulação do banco de dados B.



Fonte: De autoria própria.

Acredita-se que os erros possam ser gerados por dois fatores: a influência da posição da aplicação da força e/ou os efeitos da montagem e desmontagem das juntas para as diferentes simulações de cenários.

Embora alguns cenários apresentem falsos positivos, eles ainda demonstram um nível aceitável de desempenho. Isso porque, mesmo que as juntas falsamente diagnosticadas com dano exijam tempo de inspeção, garante-se que a junta danificada também seja inspecionada. Na prática, há redução do tempo da manutenção por filtrar o excesso de locais a serem inspecionadas na estrutura.

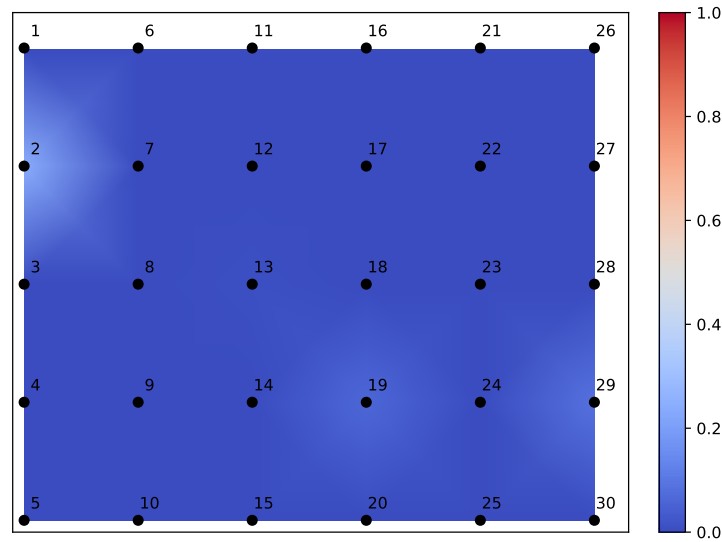
Apesar de haver cenários satisfatórios, observa-se em todas as Figuras presentes no Apêndice B que nenhum dos cenários simulados apresenta alta probabilidade de falso negativo. Em virtude disso, é possível afirmar que há eficácia dos modelos desenvolvidos por esse estudo em prever a presença e localização de dano simulado na estrutura. Demonstra-se a capacidade que os modelos tem em alcançar desempenho positivo na previsão da variável objetivo nas simulações realizadas.

4.6 Comparação de desempenho com modelos da literatura

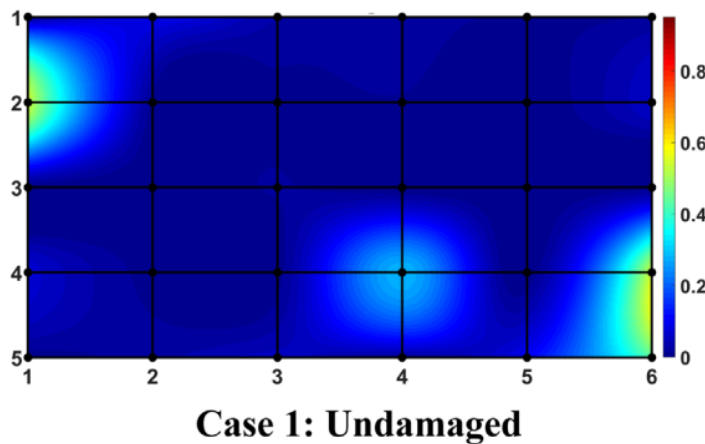
A comparação é feita entre os modelos supervisionados desenvolvidos nesse estudo e os modelos de aprendizado profundo obtidos pelo estudo feito por [Abdeljaber et al. \(2017\)](#) que, em seu artigo, implementa para a proposta de dano modelos baseados em rede neural convolucional de uma dimensão. Ambos utilizam das mesmas simulações como base de amostras e propõem-se a diagnosticar a existência e localização do dano na estrutura da simulação. Portanto, a comparação das abordagens dos modelos é válida.

A Figura 22b representa o desempenho dos modelos da literatura em prever a condição da estrutura no cenário 0, ela apresenta valores de PoD próximos, aparentemente, a 40% em 4 juntas distintas: 2, 19, 29 e 30. Enquanto isso, a Figura 22a representa para o mesmo cenário, o desempenho dos modelos supervisionados desenvolvidos nesse estudo, o resultado mostra o maior PoD localizado na junta 2 com apenas 23,83%. Apesar do desempenho do modelo da literatura na Figura 22b ainda ser satisfatório, a diferença nos resultados indica que os modelos desenvolvidos fornecem menor probabilidade de falsos positivos ao prever dano na estrutura sem dano simulado. Portanto, para o cenário 0, os modelos desenvolvidos nesse estudo desempenham melhor que os modelos da literatura.

Figura 22 – Comparativo de condição da estrutura no cenário 0 entre modelos.



(a) Modelos supervisionados desenvolvidos. Fonte: De autoria própria.

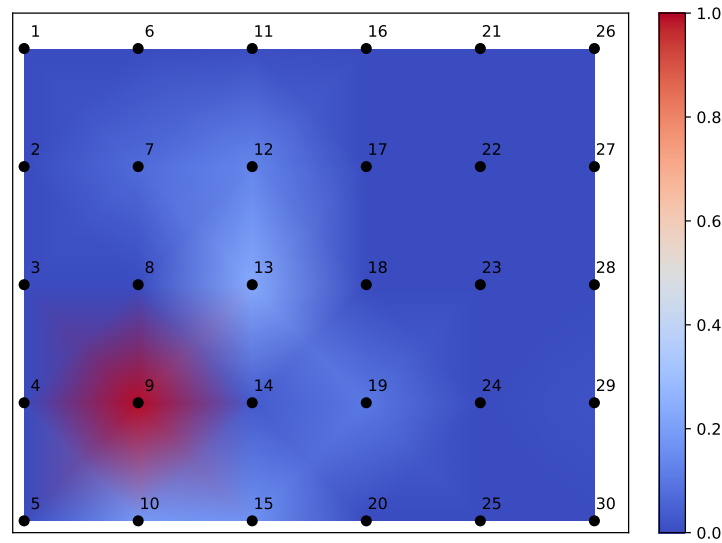


(b) Modelos de aprendizado profundo da literatura. Fonte: Abdeljaber et al. (2017).

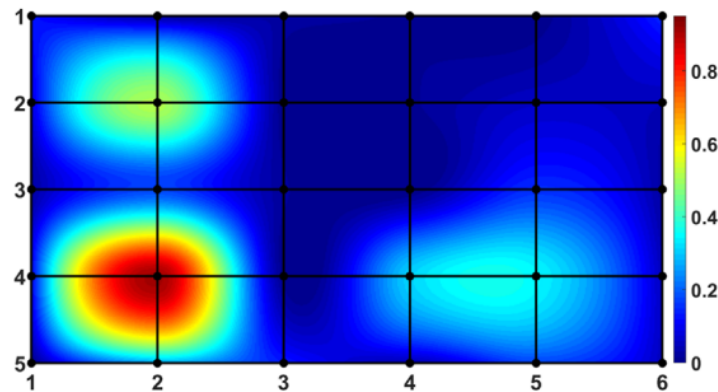
Outro exemplo de cenário em que os modelos desenvolvidos desempenham melhor é o cenário 9, representado pela Figura 23. Ambos os modelos comparados são assertivos

em diagnosticar a localização do dano, além de apresentar baixa probabilidade de falso negativo. Entretanto, o modelo da literatura representado pela Figura 23b indica focos de falsos positivos nas juntas 7, 19 e 24. Enquanto isso, o modelo supervisionado desenvolvido nesse estudo, representado pela Figura 23a, apresenta menor probabilidade de falsos positivos na estrutura, as maiores PoD localizam-se primeiramente na junta 9 com 98,83%, seguido das juntas 13 e 10 com 23,05% e 18,36%, respectivamente.

Figura 23 – Comparativo de condição da estrutura no cenário 9 entre modelos.



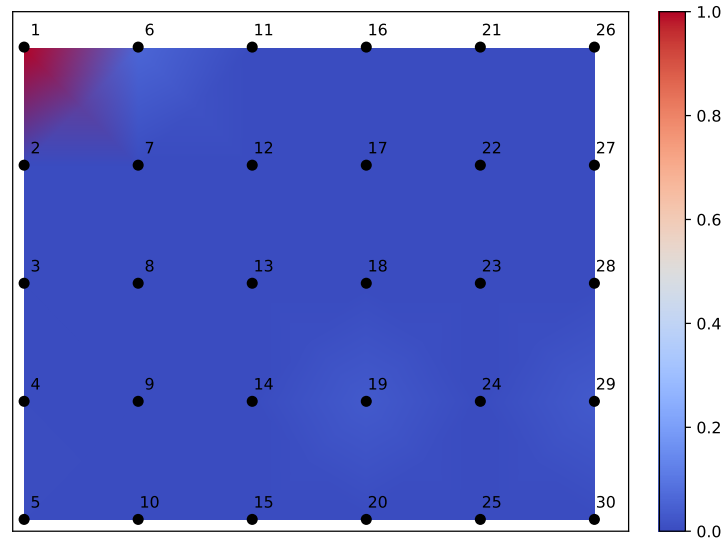
(a) Modelos supervisionados desenvolvidos. Fonte: De autoria própria.



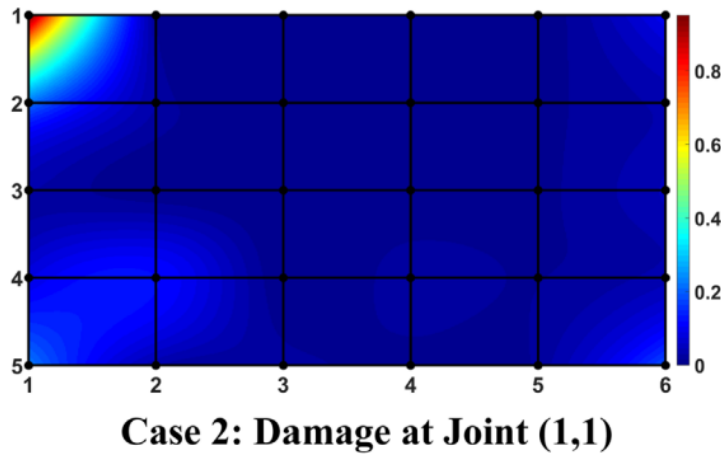
(b) Modelos de aprendizado profundo da literatura. Fonte: Abdeljaber et al. (2017).

A Figura 24 compara o desempenho dos modelos no cenário 1, com dano presente na junta 1. Avalia-se a figura citada como um exemplo em que os modelos apresentam desempenho semelhante, em ambas as Figuras 24a e 24b é possível observar um cenário assertivo, com alta probabilidade de dano localizado corretamente na junta 1 e baixa probabilidade nas demais, além de apresentarem baixa probabilidade de falso negativo na estrutura.

Figura 24 – Comparativo de condição da estrutura no cenário 1 entre modelos.



(a) Modelos supervisionados desenvolvidos. Fonte: De autoria própria.

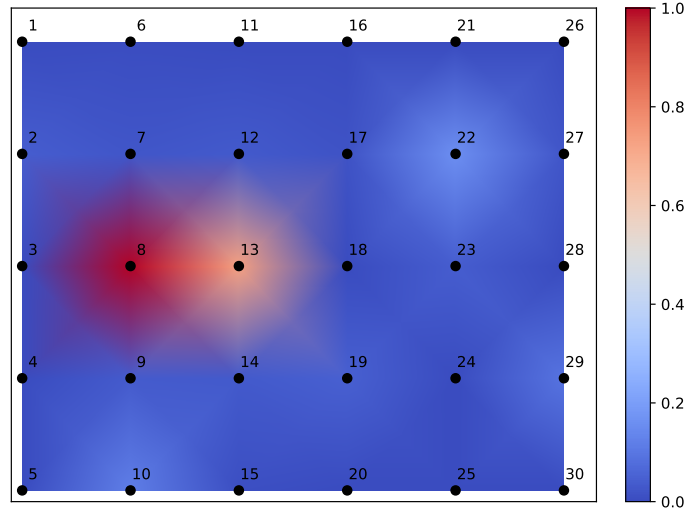


(b) Modelos de aprendizado profundo da literatura. Fonte: Abdeljaber et al. (2017).

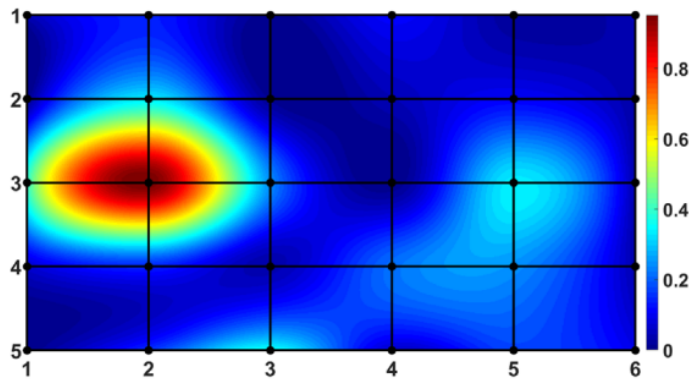
No entanto, é importante ressaltar que em alguns dos cenários avalia-se que o desempenho do modelo da literatura é superior. É o caso, por exemplo, do cenário 8. A Figura 25 ilustra a condição da estrutura com o cenário 8 simulado. O desempenho dos modelos supervisionados desenvolvidos por esse estudo é representado pela Figura 25a, onde é possível observar duas juntas com destaque em relação as demais. A junta 8, onde há realmente dano, apresenta PoD de 99,61%, já a junta 13, onde não há dano, indica um falso positivo devido à PoD de 73,05%. Com relação ao desempenho dos modelos da literatura representados pela Figura 25b, a maior PoD é observada na junta 8, onde realmente há dano. Entretanto, não há presença de um falso positivo na estrutura, a junta 3 aparenta baixa probabilidade de dano mesmo estando em destaque na Figura 25b como a segunda maior PoD no cenário. A presença de um falso positivo de alto valor de PoD

na Figura 25a indica que os modelos da literatura desempenham melhor que os modelos desenvolvidos nesse estudo no cenário 8.

Figura 25 – Comparativo de condição da estrutura no cenário 8 entre modelos.



(a) Modelos supervisionados desenvolvidos. Fonte: De autoria própria.



Case 6: Damage at Joint (2,3)

(b) Modelos de aprendizado profundo da literatura. Fonte: Abdeljaber et al. (2017).

Para finalizar, as comparações realizadas demonstram que é possível obter resultados promissores utilizando abordagens mais tradicionais. Utilizar os modelos de aprendizado de máquina supervisionado desenvolvidos no presente estudo apresenta algumas vantagens em relação a utilizar os modelos de aprendizado profundo da literatura. Em geral, o modelo desenvolvido pode oferecer benefícios em termos de simplicidade e eficiência computacional. De acordo com Zhou e Liu (2021), modelos tradicionais possuem uma menor complexidade em comparação com os modelos de aprendizado profundo, o que resulta em menor tempo de treinamento e inferência, além de exigir menos recursos computacionais. Essa vantagem pode ser especialmente relevante quando há limitações de recursos ou, como no caso desse estudo, quando o problema em questão não requer a capacidade de generalização profunda

oferecida pelos modelos mais complexos. Portanto, o modelo desenvolvido destaca-se pela capacidade de alcançar resultados comparáveis com eficiência e simplicidade, representando uma alternativa viável e eficaz para diagnosticar a existência e localização de dano em uma estrutura tipo treliça, utilizando dados de resposta à vibração.

5 Considerações Finais

Nesse capítulo apresenta-se a conclusão desse trabalho bem como a recomendação para trabalhos futuros.

5.1 Conclusão

O presente estudo desenvolveu um conjunto de modelos supervisionados de aprendizado de máquina, capazes de prever a existência e localização de dano simulado baseado na resposta à vibração da estrutura da bancada de simulações em *QU* (ABDELJABER et al., 2017). Além dos modelos, também foi desenvolvida uma metodologia detalhada para a sua criação.

O uso de características estatísticas, no domínio tanto do tempo quanto da frequência, não se demonstrou eficaz em treinar modelos de aprendizado de máquina para prever dano estrutural. Embora essas características estatísticas indiquem padrões e distribuições das amostras, elas não são suficientes para capturar a complexidade e a variabilidade dos fatores que influenciam o dano simulado.

A utilização de características de coeficientes de autorregressão para treinar modelos de aprendizado de máquina demonstrou eficácia em prever o dano com dados baseados na resposta de vibração. Os coeficientes de autorregressão capturam a dependência temporal dos dados de vibração, permitindo que o modelo aprenda padrões e tendências ao longo do tempo. Ao incorporar essas características no treinamento dos modelos, foi possível melhorar significativamente a precisão e a capacidade de prever existência e localização do dano na estrutura proposta.

O desempenho do modelo de aprendizado de máquina treinado com características de coeficientes autorregressivos foi superior ao do modelo treinado com características estatísticas, sendo que essa foi desconsiderada para desenvolvimento dos modelos finais. Os modelos de aprendizado de máquina desenvolvidos como resultado da metodologia desse estudo são modelos supervisionados classificadores de floresta aleatória, treinados e testados com características de coeficientes de autorregressão, capazes de prever a existência e localização de dano com amostras baseadas na resposta estrutural à vibração.

O desempenho em realizar diagnóstico estrutural de dano dos modelos supervisionados desenvolvidos nesse estudo foi comparado ao desempenho dos modelos disponíveis na literatura, aplicados à mesma estrutura. Como resultado da comparação, observa-se que os modelos supervisionados tiveram desempenho superior em vários cenários utilizados, indicando não haver necessidade de utilizar abordagens de aprendizado profundo neste

problema específico.

5.2 Trabalhos futuros

Embora o estudo tenha desenvolvido modelos de desempenho satisfatório, a extração de coeficientes autorregressivos foi arbitrada em 30 coeficientes. Estudos futuros podem ser realizados para otimizar a quantidade de coeficientes na autorregressão das amostras por meio de estudos sobre o “Critério de Informação de Akaike” ([AKAIKE, 1998](#)), comparando com os resultados do presente trabalho.

Sugere-se também a aplicação de diferentes modelos de aprendizados de máquinas supervisionado. Como exposto anteriormente, o presente trabalho utiliza do classificador de floresta aleatória como modelo padrão, porém a metodologia aqui descrita pode ser aplicada para diferentes classificadores. Há ainda, a possibilidade de estender a pesquisa para desenvolvimento de um único modelo capaz de generalizar as especificidades dos sensores.

Por fim, deve-se realizar uma análise de importância das variáveis de entrada na predição da saída, com o intuito de reduzir o número de entradas e otimizar o modelo. Além disso, outras técnicas de otimização de hiper-parâmetros podem ser aplicadas para melhorar os resultados obtidos.

Referências

- ABDELJABER, O. et al. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, v. 388, p. 154–170, 2017. ISSN 0022-460X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022460X16306204>>. Citado 13 vezes nas páginas 15, 25, 28, 29, 30, 31, 37, 49, 50, 51, 52, 53 e 55.
- ABDELJABER, O. et al. Dynamic testing of a laboratory stadium structure. In: _____. *Geotechnical and Structural Engineering Congress 2016*. [s.n.], 2016. cap. Site Characterization, p. 1719–1728. Disponível em: <<https://ascelibrary.org/doi/abs/10.1061/9780784479742.147>>. Citado 2 vezes nas páginas 27 e 28.
- AFONSO, A.; NUNES, C. *Probabilidades e Estatística. Aplicações e Soluções em SPSS. Versão revista e aumentada*. [S.l.]: Universidade de Évora, 2019. Citado 3 vezes nas páginas 19, 20 e 36.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike*, Springer, p. 199–213, 1998. Citado na página 56.
- AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2019. Citado na página 24.
- ARAÚJO, W. O. de; COELHO, C. J. Análise de componentes principais (pca). *University Center of Anápolis, Annapolis*, 2009. Citado na página 19.
- AVCI, O. et al. A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, v. 147, p. 107077, 2021. ISSN 0888-3270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0888327020304635>>. Citado na página 14.
- BROWNLEE, J. *Statistical methods for machine learning: Discover how to transform data into knowledge with Python*. [S.l.]: Machine Learning Mastery, 2018. Citado na página 23.
- BUTZ, T. *Fourier Transformation for Pedestrians*. 1. ed. Springer, 2006. (Fourier Series). ISBN 9783540231653, 354023165X. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=03ea20e1b0e7175fb7b669da36613372>>. Citado na página 21.
- FARRAR, C. R.; WORDEN, K. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society London, v. 365, n. 1851, p. 303–315, 2007. Citado na página 14.
- FU, Z.-F.; HE, J. *Modal analysis*. [S.l.]: Elsevier, 2001. Citado na página 45.
- GÉRON, A. Hands-on machine learning with scikit-learn and tensorflow: Concepts. *Tools, and Techniques to build intelligent systems*, O'Reilly Media, 2017. Citado 2 vezes nas páginas 17 e 18.

- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, n. Mar, p. 1157–1182, 2003. Citado 2 vezes nas páginas 19 e 22.
- ISABELLE, A. G.; ELISSEEFF. *An Introduction to Feature Extraction*. Springer Berlin Heidelberg, 2006. 1-25 p. ISBN 978-3-540-35488-8. Disponível em: <https://doi.org/10.1007/978-3-540-35488-8_1>. Citado na página 19.
- MILO, T.; SOMECH, A. Automating exploratory data analysis via machine learning: An overview. In: *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. [S.l.: s.n.], 2020. p. 2617–2622. Citado na página 18.
- MORETTIN, P. A.; TOLOI, C. M. *Análise de séries temporais: modelos lineares univariados*. [S.l.]: Editora Blucher, 2018. Citado na página 21.
- PARK, S. et al. Pzt-based active damage detection techniques for steel bridge components. *Smart Materials and Structures*, IOP Publishing, v. 15, n. 4, p. 957, 2006. Citado na página 14.
- PATEL, A. A. *Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data*. [S.l.]: O'Reilly Media, 2019. Citado 2 vezes nas páginas 17 e 18.
- RUSSELL, S. J. *Artificial intelligence a modern approach*. [S.l.]: Pearson Education, Inc., 2020. v. 4th edition. 19th chapter p. Citado 2 vezes nas páginas 16 e 23.
- SOMVANSHI, M. et al. A review of machine learning techniques using decision tree and support vector machine. In: *2016 International Conference on Computing Communication Control and automation (ICCUBE)*. [S.l.: s.n.], 2016. p. 1–7. Citado 2 vezes nas páginas 16 e 18.
- SRINIVAS, P.; KATARYA, R. hyoptxg: Optuna hyper-parameter optimization framework for predicting cardiovascular disease using xgboost. *Biomedical Signal Processing and Control*, Elsevier, v. 73, p. 103456, 2022. Citado na página 24.
- ZHOU, Z.-H. Ensemble methods. *Combining pattern classifiers*. Wiley, Hoboken, p. 186–229, 2014. Citado 2 vezes nas páginas 23 e 24.
- ZHOU, Z.-H.; LIU, S. *Machine Learning*. Singapore: Springer Nature Singapore Pte Ltd, 2021. Citado 7 vezes nas páginas 17, 18, 22, 23, 24, 25 e 53.

Apêndices

APÊNDICE A – Hiper-parâmetros obtidos nas otimizações dos modelos

A Tabela 2 apresenta os hiper-parâmetros obtidos nas otimizações dos modelos sob regime de validações cruzadas mencionados no presente estudo.

Tabela 2 – Hiper-parâmetros obtidos nas otimizações dos modelos.

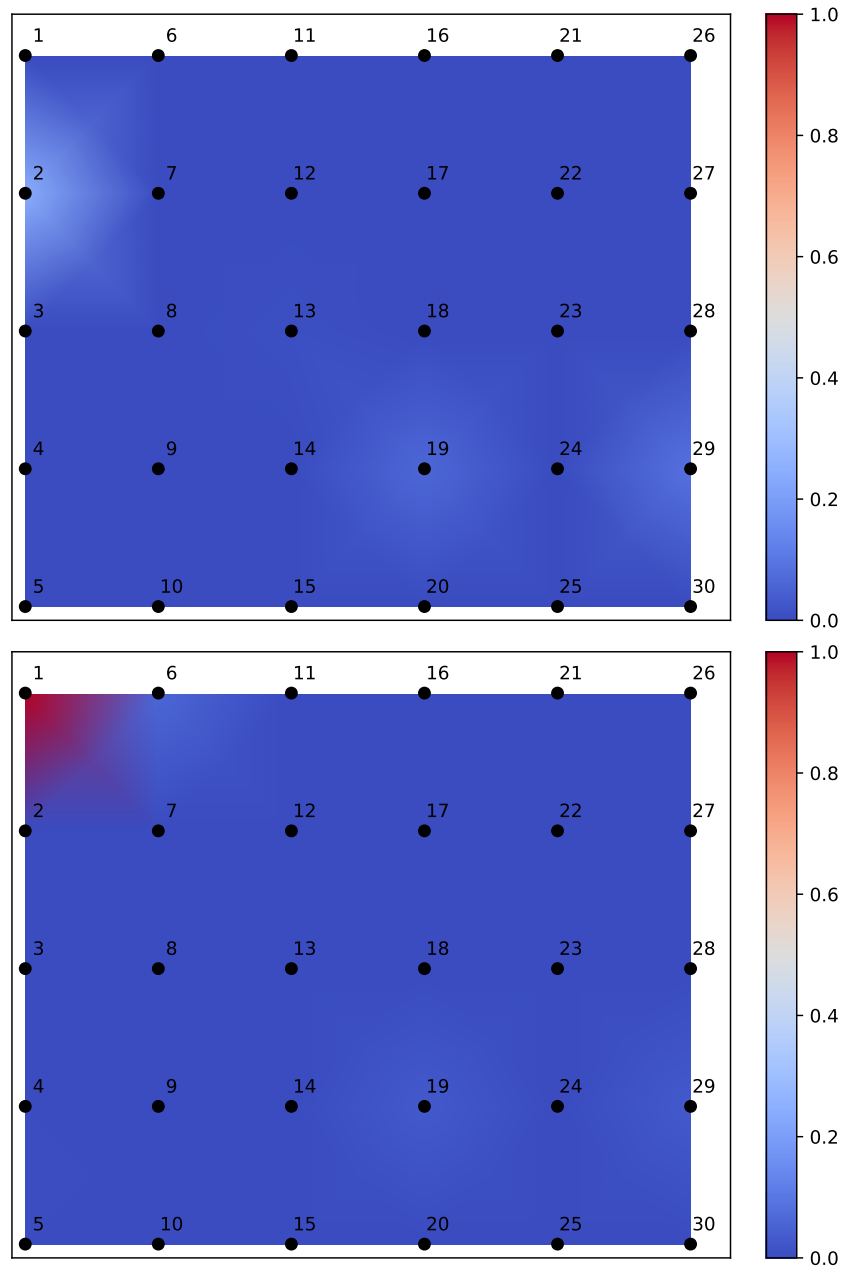
HIPER-PARÂMETROS			
Vetor estatístico		Vetor autorregressivo	
max_features	max_depth	max_features	max_depth
6	1	11	29
7	1	16	30
1	21	3	6
6	1	2	29
4	1	1	3
7	2	1	49
6	1	6	45
6	5	1	22
5	2	1	20
2	1	7	25
1	34	14	42
7	3	1	44
6	26	1	25
7	1	8	39
6	16	7	50
4	1	3	17
6	1	9	5
4	1	15	27
7	35	2	49
3	1	2	13
6	1	8	17
7	1	4	14
7	24	4	16
6	1	1	43
6	1	2	8
2	27	7	47
7	1	1	13
7	3	4	23
6	1	4	28
5	1	5	22

Fonte: De autoria própria.

APÊNDICE B – Probabilidade de dano em todos os cenários

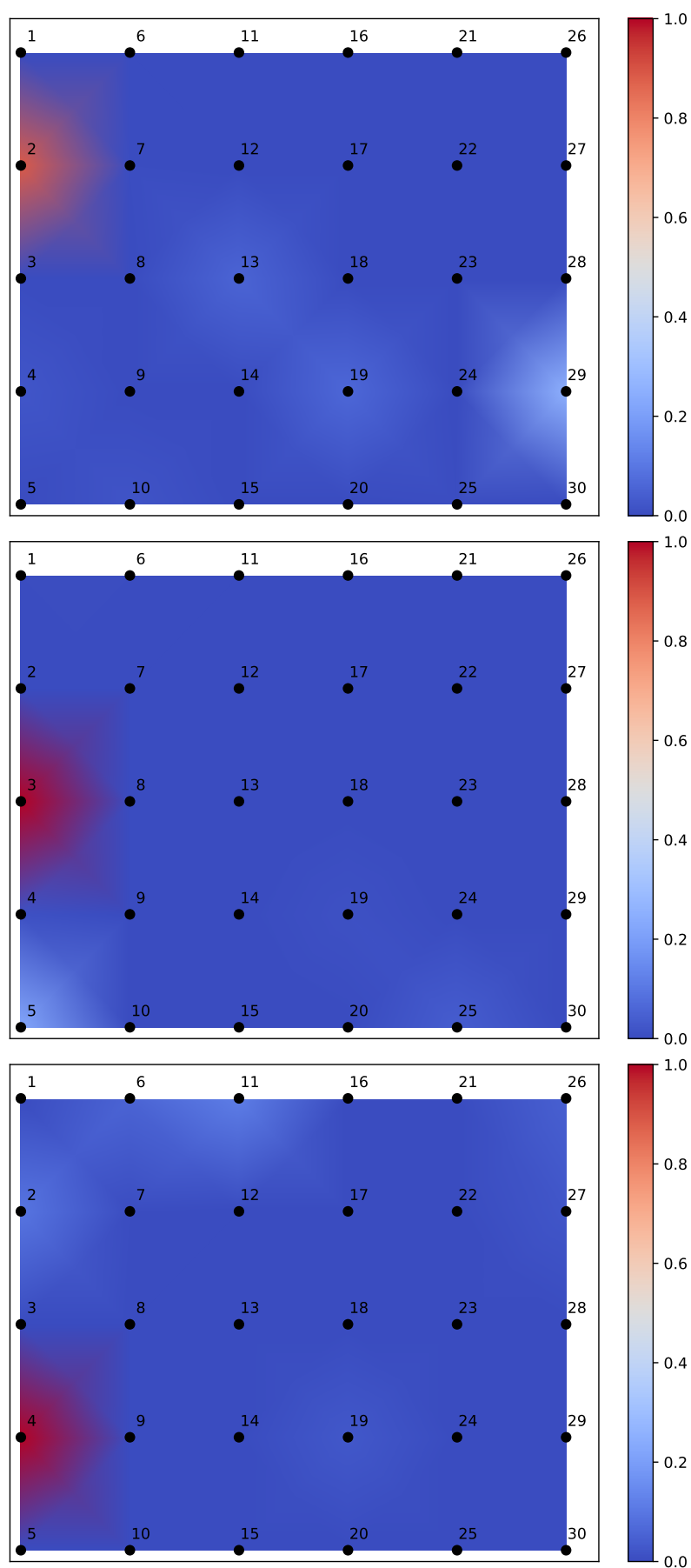
Esse apêndice abriga todas as figuras referentes à probabilidade de dano (PoD) na estrutura em cada um dos cenários.

Figura 26 – Probabilidade de dano nos cenários referencia e 1 do banco de dados B.



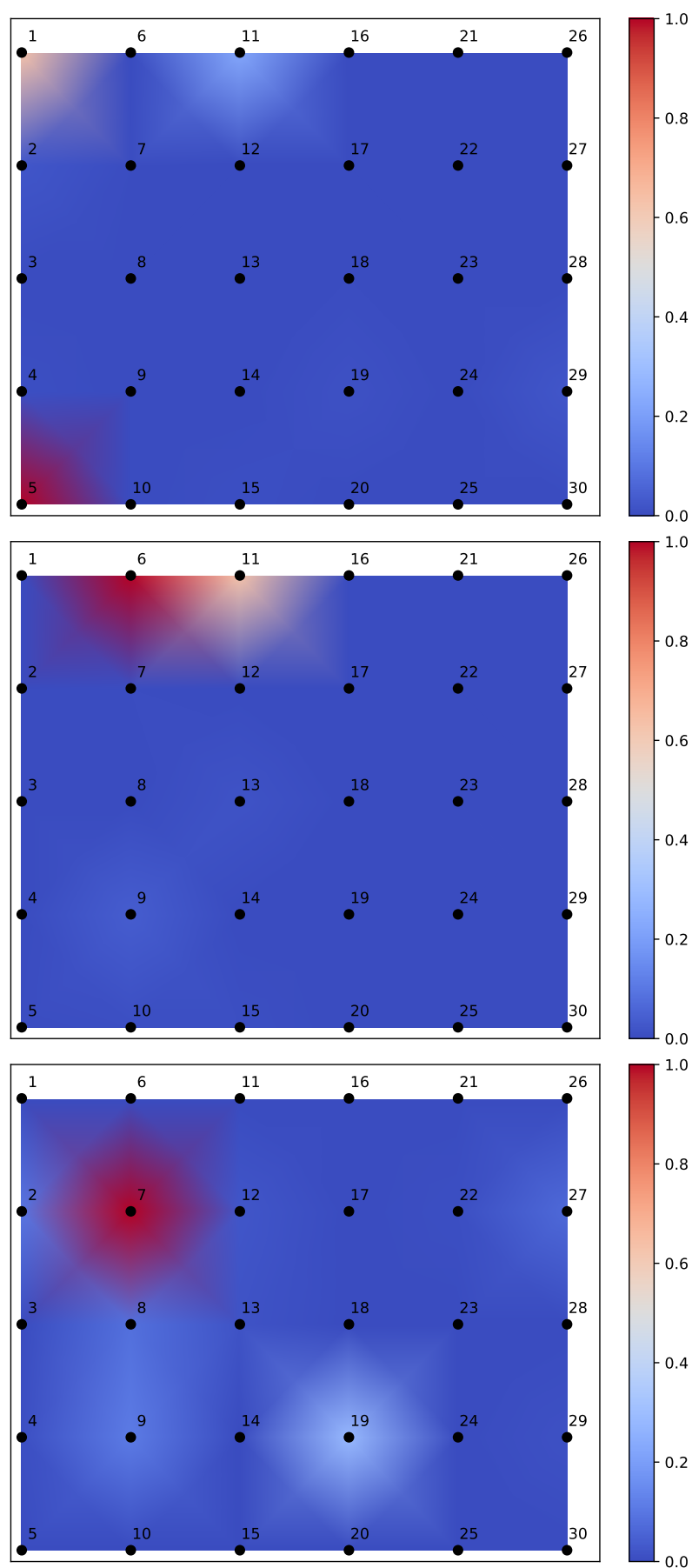
Fonte: De autoria própria.

Figura 27 – Probabilidade de dano nos cenários 2, 3 e 4 do banco de dados B.



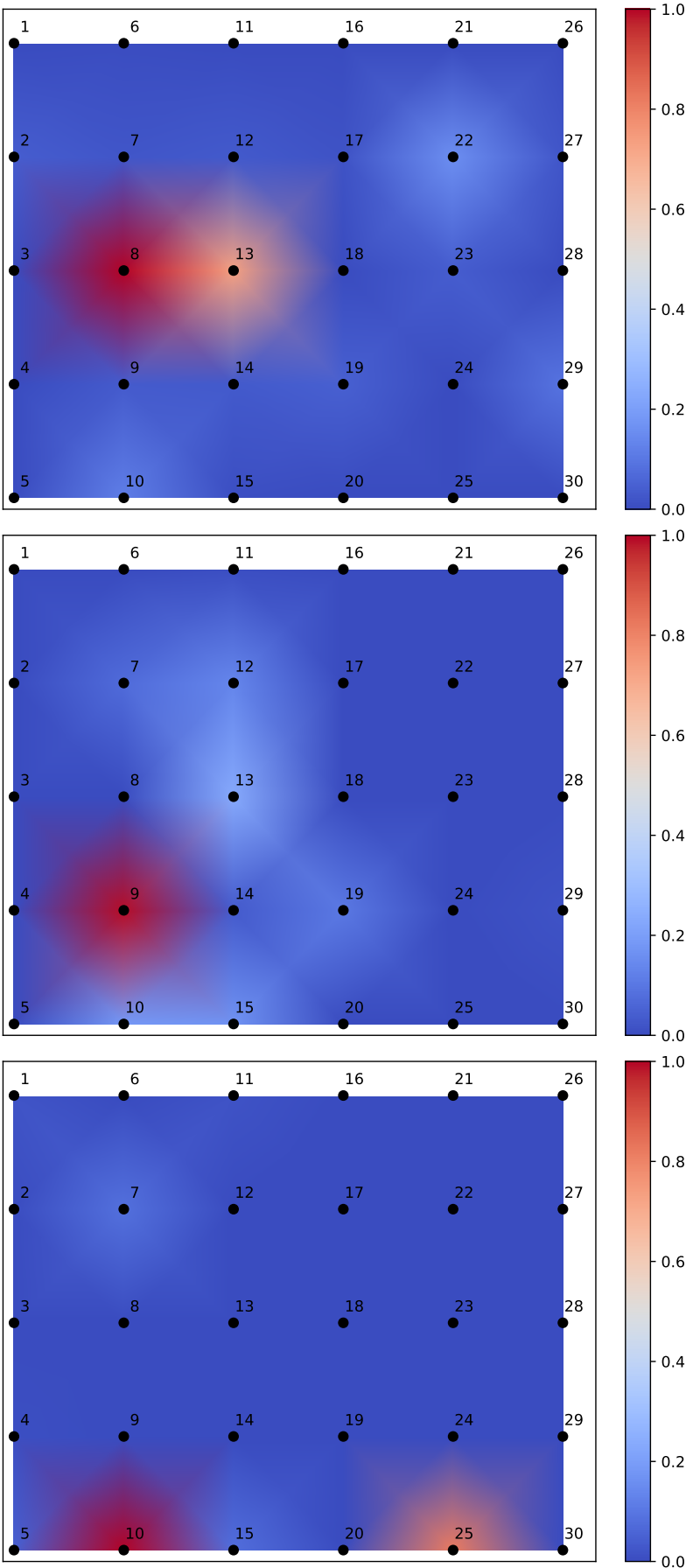
Fonte: De autoria própria.

Figura 28 – Probabilidade de dano nos cenários 5, 6 e 7 do banco de dados B.



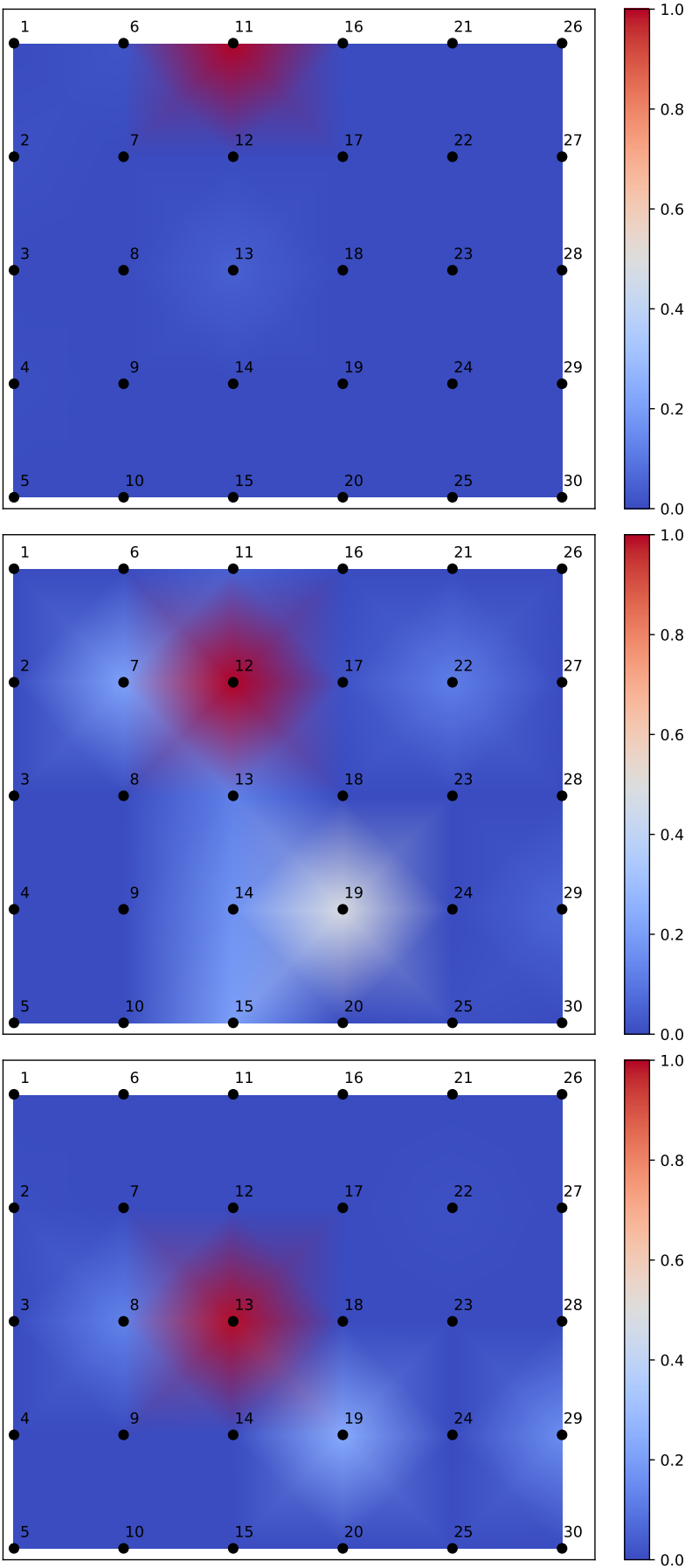
Fonte: De autoria própria.

Figura 29 – Probabilidade de dano nos cenários 8, 9 e 10 do banco de dados B.



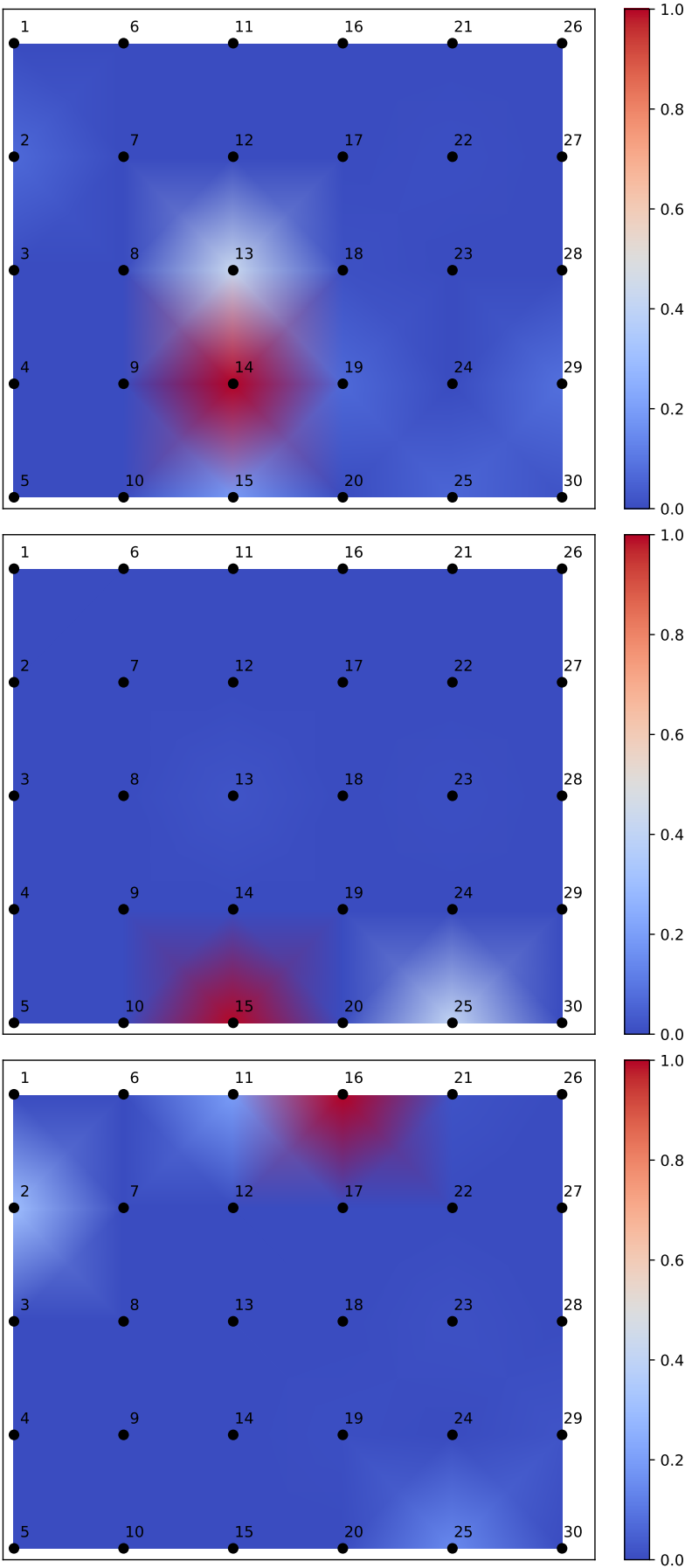
Fonte: De autoria própria.

Figura 30 – Probabilidade de dano nos cenários 11, 12 e 13 do banco de dados B.



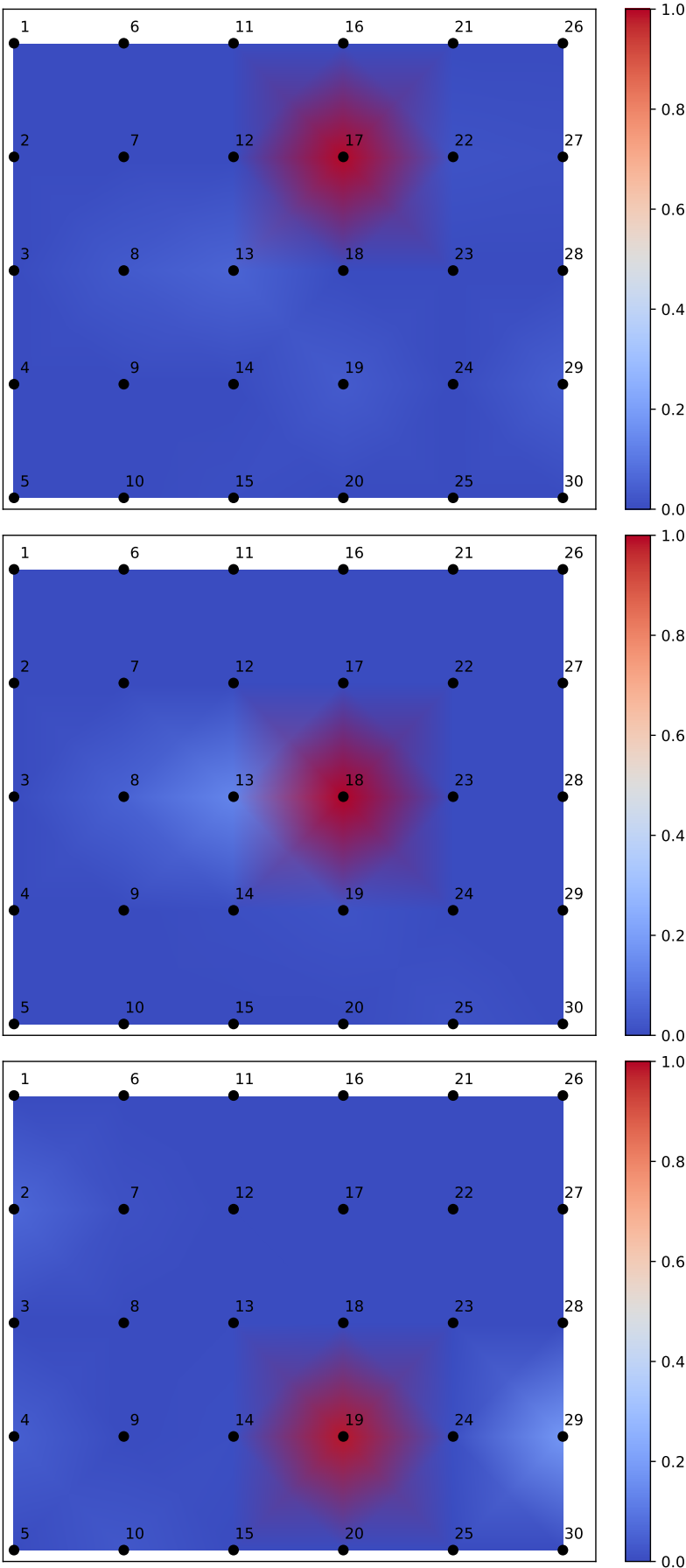
Fonte: De autoria própria.

Figura 31 – Probabilidade de dano nos cenários 14, 15 e 16 do banco de dados B.



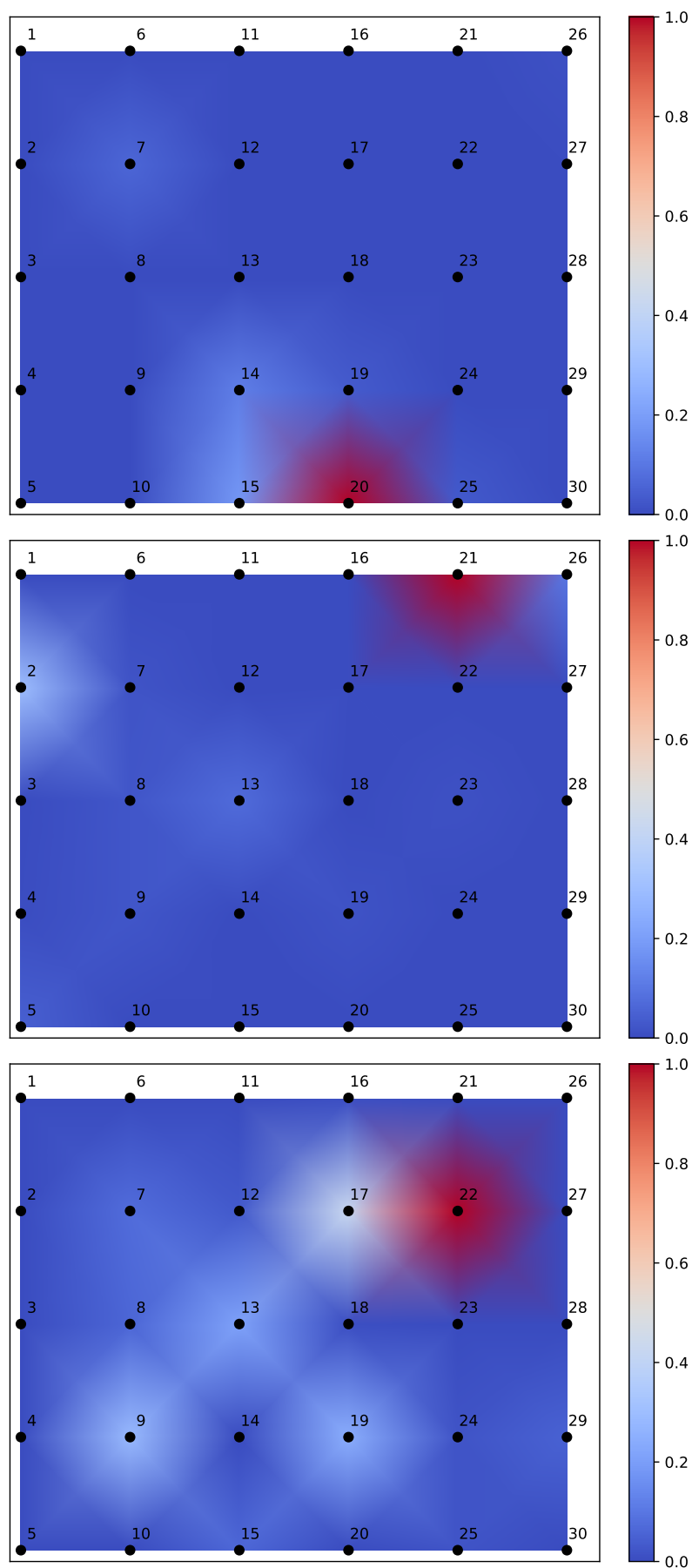
Fonte: De autoria própria.

Figura 32 – Probabilidade de dano nos cenários 17, 18 e 19 do banco de dados B.



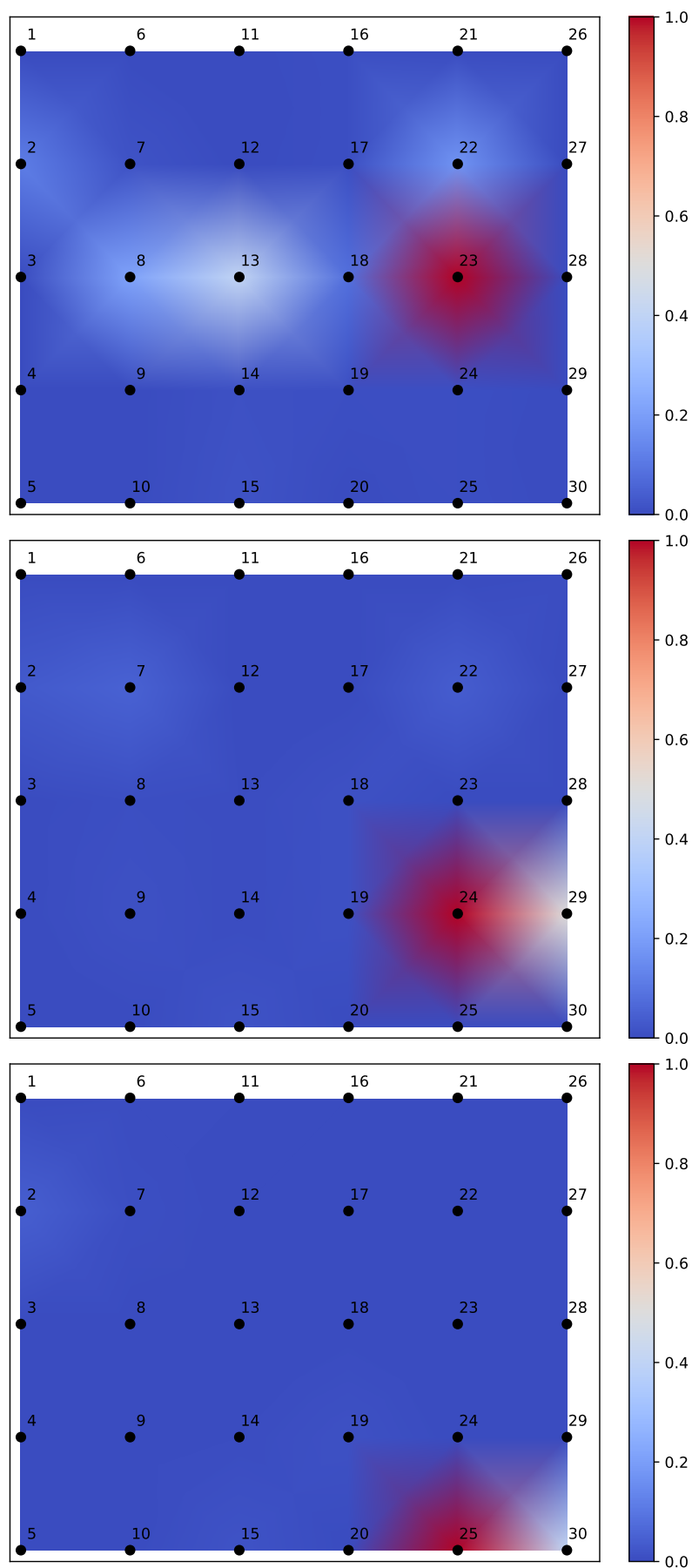
Fonte: De autoria própria.

Figura 33 – Probabilidade de dano nos cenários 20, 21 e 22 do banco de dados B.



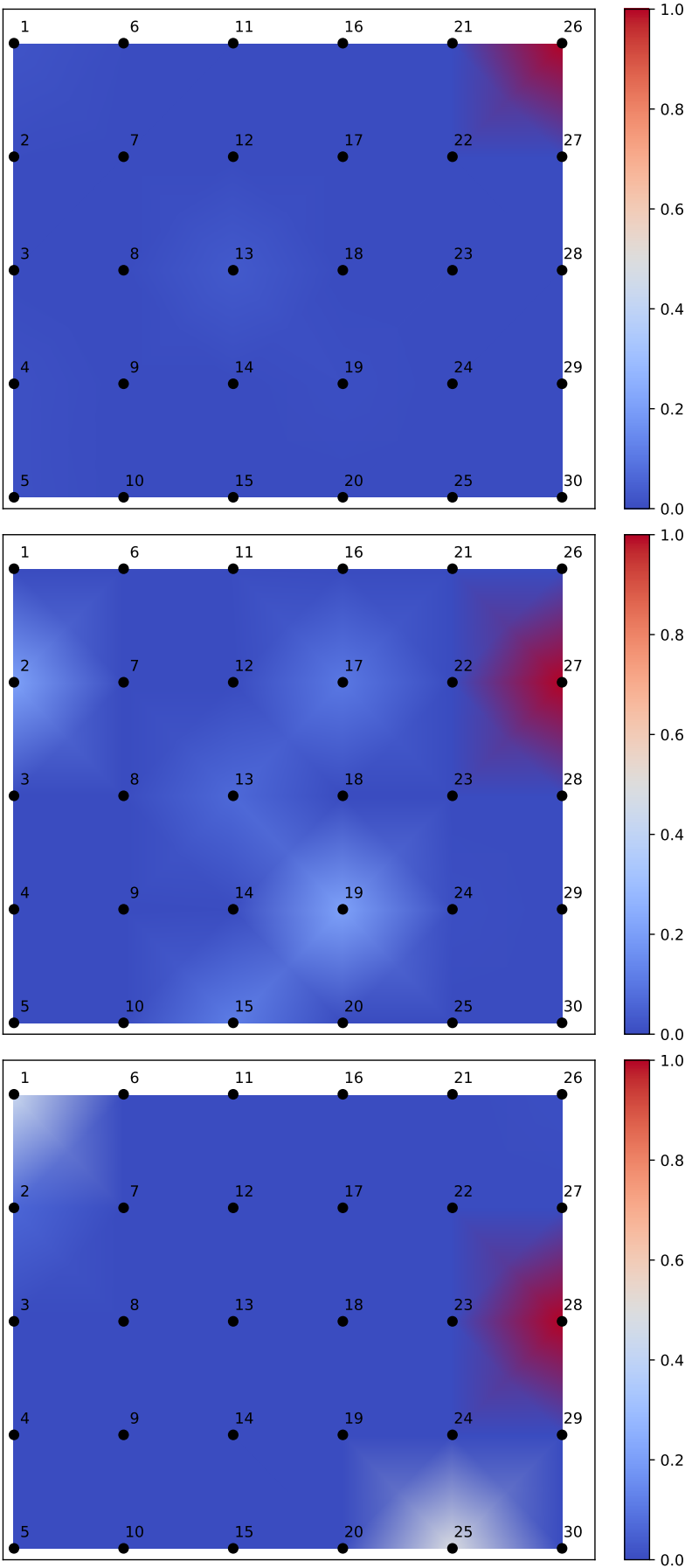
Fonte: De autoria própria.

Figura 34 – Probabilidade de dano nos cenários 23, 24 e 25 do banco de dados B.



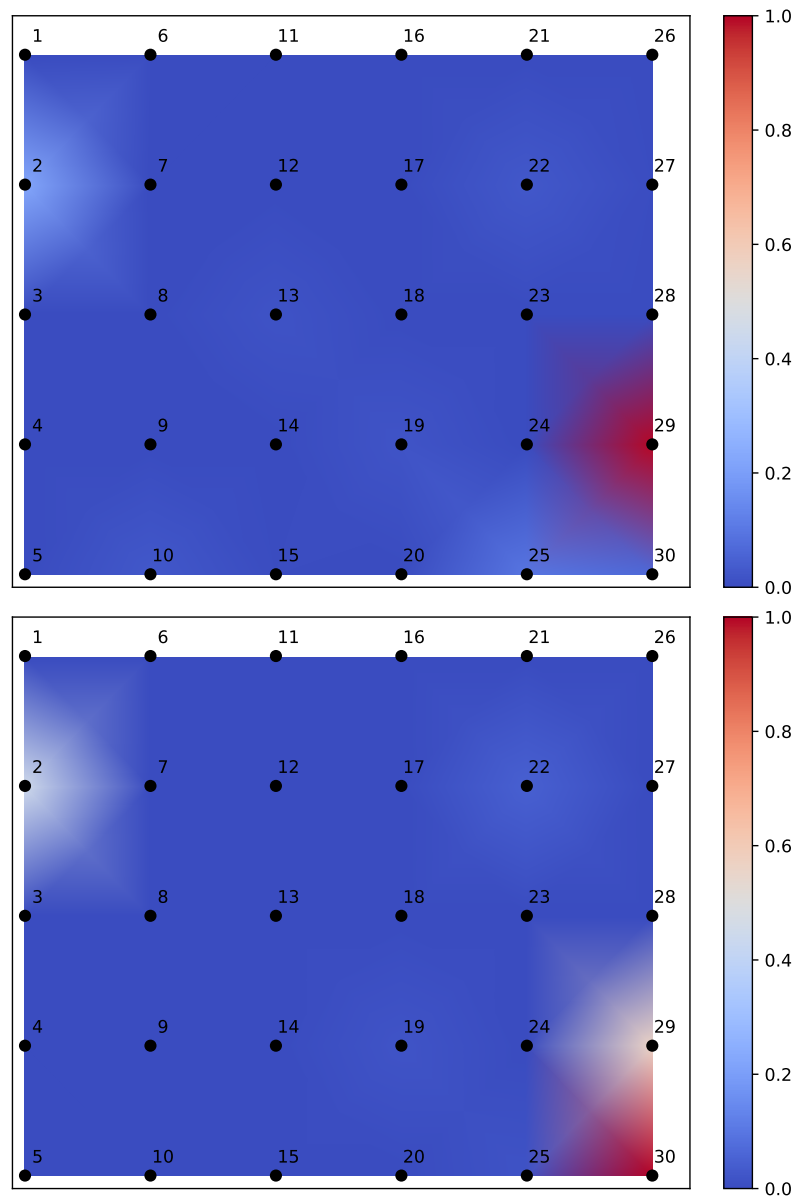
Fonte: De autoria própria.

Figura 35 – Probabilidade de dano nos cenários 26, 27 e 28 do banco de dados B.



Fonte: De autoria própria.

Figura 36 – Probabilidade de dano nos cenários 29 e 30 do banco de dados B.



Fonte: De autoria própria.