

# Análise Comparativa da Aplicação de Técnicas de Clusterização no Monitoramento da Integridade Estrutural

Raimundo M. B. Neto

Universidade Federal do Sul e Sudeste do Pará  
Instituto de Geociências e Engenharias  
Faculdade de Computação e Engenharia Elétrica  
Marabá, Pará 68505-080  
Email: netobarros@unifesspa.edu.br

Luan Silva

Universidade Federal do Sul e Sudeste do Pará  
Instituto de Geociências e Engenharias  
Faculdade de Computação e Engenharia Elétrica  
Marabá, Pará 68505-080  
Email: luansilvatec@unifesspa.edu.br

Adam Santos

Universidade Federal do Sul e Sudeste do Pará  
Instituto de Geociências e Engenharias  
Faculdade de Computação e Engenharia Elétrica  
Marabá, Pará 68505-080  
Email: adamdreyton@unifesspa.edu.br

João Cordeiro

Universidade Federal do Sul e Sudeste do Pará  
Instituto de Geociências e Engenharias  
Faculdade de Computação e Engenharia Elétrica  
Marabá, Pará 68505-080  
Email: joao.cordeiro@unifesspa.edu.br

Victor Souza

Universidade Federal do Sul e Sudeste do Pará  
Instituto de Geociências e Engenharias  
Faculdade de Computação e Engenharia Elétrica  
Marabá, Pará 68505-080  
Email: victoor.souza@unifesspa.edu.br

**Resumo**—O paradigma de monitoramento de integridade estrutural (SHM) consiste em um processo que fornece informações precisas sobre a condição e o desempenho estrutural. As informações obtidas pela SHM são usadas para planejar e projetar atividades de manutenção, aumentar a segurança, verificar hipóteses, reduzir incertezas e aumentar o conhecimento sobre a estrutura. No contexto da aplicação de algoritmos de aprendizado de máquina na detecção de dano estrutural, os algoritmos de clusterização se destacam por serem uma abordagem capaz de revelar as principais condições estruturais quando da operação normal de uma estrutura. Neste trabalho são comparadas três abordagens de detecção de dano baseadas nos algoritmos K-Means, DBSCAN e MeanShift. Os resultados obtidos demonstraram que a abordagem baseada no DBSCAN obteve a melhor performance em termos de erros tipo I e II.

**Keywords**—Monitoramento de integridade estrutural, Clusterização, Detecção de dano.

## I. INTRODUÇÃO

A preocupação com a integridade de estruturas civis é de suma importância, pois são elementos essenciais e onipresentes em todas as sociedades. É completamente impossível imaginar uma sociedade sem edifícios, estradas, ferrovias, pontes, túneis, barragens e usinas de energia. Entretanto, o mau funcionamento de estruturas civis muitas vezes tem consequências graves [1].

Essas estruturas estão suscetíveis ao envelhecimento decorrente da passagem do tempo, bem como a degradação provocada por fenômenos naturais (terremotos, por exemplo). O dano nas estruturas pode ser classificado como dano global, quando está em toda a estrutura, e dano local, quando encontra-se em um local específico [2]. Eles causam prejuízos a estrutura e conseqüentemente ao seu desempenho. A identificação de dano é realizada através de inspeções visuais ou por técnicas destrutivas ou não destrutivas, mas esses procedimentos necessitam de profissionais com bastante experiência [3]. Tendo em conta que, na maioria dos casos, a substituição de estruturas é economicamente inviável, torna-se necessária a implementação de sistemas capazes de estender o tempo de utilização estrutural, de modo que possam continuar a ser utilizadas com segurança, sem oferecer prejuízos econômicos e, o mais importante, a vida humana [4].

Nesse contexto surge o paradigma de monitoramento de integridade estrutural (SHM), o qual consiste em um processo que proporciona informações precisas e em tempo real sobre a condição e desempenho estrutural [5]. Trata-se de uma gravação contínua permanente ou periodicamente contínua de parâmetros estruturais em um curto ou longo prazo. As informações obtidas a partir da SHM são usadas para planejar e projetar atividades de manutenção, aumentar a segurança, verificar hipóteses, reduzir incer-

tezas e para ampliar o conhecimento sobre a estrutura que está sendo monitorada [1].

Um grande desafio para a aplicação da tecnologia SHM é a presença de variabilidades ambientais e operacionais (temperatura e carga operacional, por exemplo) que alteram os dados de medição de resposta estrutural, provocando alterações na correlação entre a magnitude das características sensíveis à dano e o nível de dano [6]. Note que essas variabilidades são efeitos normais aos quais uma estrutura pode estar sujeita.

Se essas variabilidades normais não forem levadas em conta no processo de detecção de dano, tais mudanças nas características de resposta dinâmica podem resultar em falsas indicações de dano [4]. Existem várias abordagens utilizadas para separar mudanças nas medidas de resposta dinâmica causadas por variações ambientais e operacionais das mudanças causadas por dano. A abordagem utilizada neste estudo consiste na aplicação de algoritmos de aprendizado de máquina, que tem o intuito de desenvolver modelos capazes de representar o estado normal da estrutura, mesmo estando sob efeitos ambientais ou operacionais extremos. A utilização desses algoritmos dispensa a medida direta das fontes de variabilidades, possibilitando a utilização desses métodos em sistemas SHM, independente da complexidade da estrutura [7].

No contexto de aplicação de algoritmos de aprendizado de máquina na detecção de dano, destacam-se os algoritmos de clusterização, por constituírem uma abordagem capaz de descobrir automaticamente as condições normais e estáveis de um sistema estrutural. O objetivo principal em abordagens de clusterização é agrupar de forma lógica as observações referentes as medições de resposta estrutural, relacionando esses agrupamentos a estados físicos da estrutura, incluindo estados que correspondem a influências de variabilidades ambientais e operacionais.

Neste trabalho, três abordagens de detecção de dano são consideradas, sendo baseadas no algoritmos de clusterização K-Means, MeanShift e DBSCAN. Essas abordagens combinam um algoritmo de clusterização para determinar as condições normais da estrutura, com um algoritmo de detecção de anomalias para detectar e quantificar o dano estrutural. As abordagens foram testadas na base de dados da Ponte Z-24, a qual apresenta dados das condições normal e anormal, com vários cenários de dano. A avaliação dos resultados obtidos é realizada através de erros do tipo I (indicação de dano falsa-positiva) e II (indicação de dano falsa-negativa).

As próximas seções deste artigo estão organizadas da seguinte forma. Na seção II são apresentados os trabalhos correlatos. Na seção III são descritas as abordagens de detecção de dano baseadas em técnicas de clusterização. A seção IV apresenta uma explanação mais específica sobre a base de dados utilizada. Na seção V são apresentadas as metodologias e técnicas utilizada nos testes e avaliação dos resultados. A seção VI mostra e discute os resultados obtidos com os experimentos, apontando a abordagem que

obteve o melhor desempenho. E por fim, na seção VII são feitas as considerações finais deste estudo.

## II. TRABALHOS CORRELATOS

O emprego de algoritmos de aprendizado de máquina na detecção de danos tem sido explorado em diversos trabalhos ao longo dos anos. Em Figueiredo et al. (2011) [8] busca-se detectar dano estrutural com variações ambientais e operacionais em uma estrutura de laboratório. Quatro algoritmos de aprendizado de máquina foram utilizados para tal: *Singular Value Decomposition* (SVD), *Factor Analysis* (FA), *Auto-Associative Neural Network* (AANN), e *Mahalanobis Squared-Distance* (MSD). A partir dos resultados que foram obtidos, os algoritmos FA e SVD são mais apropriados quando se quer minimizar indicações de dano falsas-positivas, enquanto que os algoritmos AANN e MSD são preferidos quando se quer minimizar indicações de dano falsas-negativas.

Seguindo a mesma linha, no trabalho [6] utiliza-se quatro algoritmos não supervisionados de aprendizado de máquina baseado em kernel, esses são adaptados para realizar a normalização de dados e para detecção de dano, tais algoritmos são: *One-class Support Vector Machine* (*One-Class SVM*), *Support Vector Data Description* (SVDD), *Kernel Principal Component Analysis* (KPCA) e *Greedy Kernel Principal Component Analysis* (GKPCA). O conjunto de dados no qual os algoritmos foram aplicados é o mesmo citado anteriormente. Esses dados foram transformados a partir de um modelo autoregressivo. Nos resultados obtidos os algoritmos KPCA e GKPCA obtiveram os melhores resultados em termos de minimização de erros relacionados a classificação de estados normais e anormais, se comparado aos demais.

Considerando os dois trabalhos anteriores, é importante destacar a necessidade de validação dos algoritmos em estruturas do mundo real, uma vez que análises laboratoriais normalmente estão em condições controladas, que nem sempre simulam efeitos ambientais e operacionais em sua plenitude.

No trabalho de Diez et al. (2016) [9], é proposto uma abordagem baseada em *clusters* para SHM. As técnicas de SHM foram aplicadas na ponte Sydney Harbour, onde a partir dos dados da ponte (vibrações), foram realizados 4 etapas: extração de características, remoção de *outliers*, processamento de sinais e caracterização do comportamento. O algoritmo de clusterização k-means foi utilizado em diversas situações, em uma pequena parte da ponte (juntas) até ela por completo. Os resultados da clusterização indicaram semelhança entre juntas localizadas em diferentes locais e áreas da ponte, ajudando a agrupar comportamentos semelhantes. Também foi possível identificar uma junta danificada e uma junta com um sensor com defeito, a partir da clusterização dos resultados, demonstrando a robustez da abordagem ao detectar um dano estrutural ou um problema no sensor.

Em Nick et al. (2015) [10], no âmbito de detecção de dano em aeronaves, foram utilizadas técnicas de aprendizado não supervisionado: k-means e mapas auto-organizáveis. Técnicas de aprendizado supervisionado também foram aplicadas: máquinas de vetores de suporte, Naive Bayes e redes neurais. O principal objetivo do trabalho foi explorar essas técnicas e observar suas características para que combinações delas possam ser usadas adequadamente em várias circunstâncias. Nos resultados obtidos notou-se que para as técnicas supervisionadas e não supervisionadas, existem compensações entre precisão e tempo de execução para encontrar uma combinação de técnicas apropriadas para uma determinada situação.

### III. ALGORITMOS PARA DETECÇÃO DE DANO

Existem várias formas de trabalhar com grandes quantidades de dados atualmente, independente do tipo de dado. Uma dessas formas, a que é utilizada nesta pesquisa, é a classificação ou formação de grupos (*clusters*) em determinadas categorias. Essa forma de análise dos dados facilita seu entendimento por conta de sua clareza e visualização.

A análise dos dados através da clusterização pode ser dividida em duas partes [11]: as análises exploratórias ou confirmatórias, mediante os modelos apropriados para a fonte de dados. As duas abordagens levam em consideração a classificação de dados feita a partir das seguintes formas:

- Boa adequação a um modelo postulado;
- Grupos naturais descobertos através de análise.

Um exemplo de clusterização é mostrado na Figura 1. Cada cor (verde, amarelo e roxo) representa um grupo diferente, assim como cada ponto no plano pertence a um grupo determinado, correspondente a sua cor. Um grupo ou *cluster* pode ser resumido por uma média dos elementos pertencentes a esse grupo, normalmente chamada de centróide do grupo.

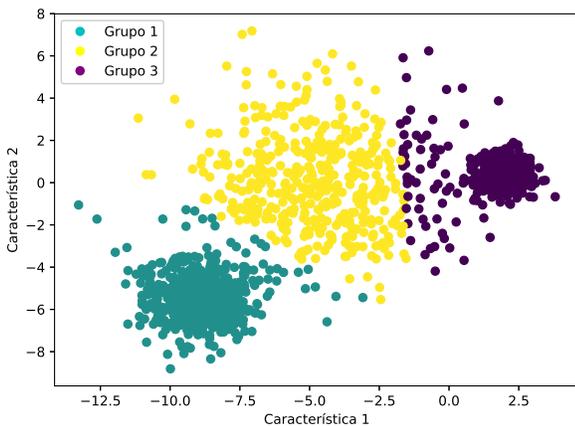


Figura 1: Exemplo de Clusterização [12]

Cada algoritmo de clusterização aplica sua técnica de classificação, de acordo com o problema proposto, porém todas essas técnicas seguem uma similaridade no momento

da classificação. Tal semelhança será melhor exemplificada pelos três algoritmos utilizados neste trabalho.

A seguir os princípios de funcionamento dos algoritmos de clusterização e da abordagem de detecção de anomalias são apresentados.

#### A. Algoritmo K-Means

O algoritmo é baseado na minimização de medida de custo e distância mediante padrões de agrupamento [13]. Essa minimização do algoritmo busca encontrar um mínimo local para a função objetivo, o qual dependerá do ponto ou centróide inicial do algoritmo.

O método utilizado pelo k-means consiste em dividir as  $m$  observações, que estão em  $d$  dimensões, em  $k$  grupos ou *clusters*, onde a soma dos quadrados dentro do *cluster* seja minimizada [14]. De acordo com [14], não é comumente prático exigir que uma solução que seja uma soma mínima de quadrados nas partições, exceto quando  $m$  e  $n$  são pequenos e  $k = 2$ . Com tudo, o objetivo é buscar por ótimos locais, onde nenhum movimento de um ponto para outro ponto reduzirá a soma dos quadrados dentro do *cluster* [14].

O algoritmo utiliza os centróides como protótipos representativos dos grupos, onde evidentemente, o centróide representa o centro de um grupo, calculado pela média de todas as observações do grupo.

#### B. Algoritmo MeanShift

O MeanShift é um algoritmo baseado também nos centróides, que funcionam atualizando os candidatos a centróides, para que os centróides continuem sendo a média das observações dentro de uma partição ou determinada região. Os candidatos são filtrados para formar o conjunto final de centróides [15].

Esse método é não supervisionado e não paramétrico, visando estimar o gradiente de uma função de probabilidade [16]. Sua aplicação tem por objetivo a localização dos máximos locais da função de probabilidade, de forma a estimar o gradiente dessa função [16]. Então, para se obter os máximos locais basta calcular as raízes do gradiente da função de probabilidade. Os máximos da função de probabilidade são regiões do espaço de características onde encontra-se alta densidade [17].

#### C. Algoritmo DBSCAN

Para divisão dos clusters, o algoritmo inicia um ponto arbitrário  $p$  e recupera os pontos acessíveis por densidade a partir de  $p$ . Se  $p$  é um ponto central, esse procedimento gera um *cluster* [18]. Em outras palavras, o algoritmo baseia-se na densidade espacial dos dados. De acordo com [19], o DBSCAN busca por valores globais, logo, dois *clusters*, mesmo com densidades diferentes, podem ser mesclados em um grupo, se ambos estiverem próximos.

O algoritmo é uma forma não paramétrica que não se atenta a identificação inicial do número de *clusters* e pode identificar agrupamentos de diferentes formas, sendo capaz também de lidar com ruídos nos dados.

#### D. Detecção de anomalia estrutural

Uma vez identificados os agrupamentos, que são representações das condições normais da estrutura sob variações operacionais e ambientais, pode-se computar indicadores de dano/anomalia através da menor distância (por exemplo, distância Euclidiana) entre uma nova observação e o centróide de cada agrupamento descoberto. Os indicadores de anomalias representam uma quantificação da severidade do dano na estrutura. Dessa forma, quando uma estrutura opera em seu estado normal, o indicador de dano (*damage indicator* - DI) deve ter um valor pequeno, enquanto que uma estrutura que opera em estado anormal apresenta valores abruptos no que diz respeito ao indicador de anomalias.

É importante ressaltar que um limiar de confiança deve ser computado para classificar observações com ou sem dano. Normalmente, esse limiar considera 95% de confiança nos dados de treinamento dos algoritmos, o que implica na possibilidade de erro de 5%.

#### IV. CONJUNTO DE DADOS

A Ponte Z-24 foi uma estrutura que se localiza na Suíça. A ponte era de concreto e composta por um vão principal de 30m, com duas vigas laterais pré-tencionadas de 14m de acordo [20]. Os testes com a ponte foram elaborados antes de sua completa demolição, com o objetivo de fornecer dados confiáveis para a utilização em SHM.

O período de tempo que durou a elaboração dos testes foi de aproximadamente um ano, com início em 11 de novembro de 1997 até 10 de setembro de 1998 [20]. Através do monitoramento, visava-se quantificar as variabilidades operacionais e ambientais presentes na ponte, assim como o dano inserido artificialmente de forma controlada, já no último mês de operação da ponte.

De acordo com [21], foram elaborados dois tipos de monitoramento na ponte, que são eles: Monitoramento Intermitente e o Monitoramento Contínuo. Durante o monitoramento intermitente, onde ocorreram restes de dano progressivo, foi removido um dos pilares para a instalação de um sistema de liquidação [21]. Toda ponte foi medida em nove instalações de 33 acelerômetros. Três sensores de referência são comuns a todas as configurações.

No monitoramento contínuo o objetivo foi fornecer dados ambientais e vibracionais [21]. A cada hora, quantidades ambientais foram medidas, com o auxílio de 49 sensores capturando parâmetros como temperatura do ar, características de vento, umidade, expansão da ponte, temperatura do solo e temperaturas do concreto [21]. Além disso, 8 acelerômetros capturavam, a cada hora durante 11 minutos, vibrações da ponte.

#### V. METODOLOGIA

O vigente artigo baseou-se na análise de três algoritmos inteligentes, como já foi dito anteriormente: K-Means, MeanShift e DBSCAN. A clusterização tem por objetivo, identificar e agrupar registros similares de uma base de

dados, formando assim, *clusters* (grupos) [11], como foi abordado na Seção V.

É importante que fique claro o método de avaliação dos erros, utilizados pelos algoritmos, no tratamento dos dados. Esses métodos, que são modelos estatísticos, são usados para minimizar indicações de dano. Esse método se divide em duas categorias:

- Falso-positivo (Erro Tipo I): indicação de dano quando não há dano na estrutura;
- Falso-negativo (Erro Tipo II): sem indicação de dano quando há a existência de dano na estrutura.

É evidente que os erros Tipo I são indesejáveis, pois acarretam tempo de inatividade desnecessário e consequentemente perda de receita [22], como também, a perda da confiança no sistema, que é um fator importante de avaliação de um algoritmo. Contudo, é importante se atentar para os erros do Tipo II, pois, esses são dados referentes a identificação de dano na estrutura. Os erros do Tipo I são encontrados acima do limiar e os do Tipo II são encontrados abaixo do limiar. Essa divisão é baseada no limiar de 95% de confiança de avaliação dos dados de treino, onde 5% é o máximo de possibilidades aceitas para erros do algoritmo.

De acordo com do conjunto de observações, o total de 3476 estão relacionadas a condição normal da estrutura e o restantes são pertinentes a condição anormal da ponte. Os algoritmos propostos são empregados neste trabalho em modo não supervisionado, uma vez que apenas os dados da condição sem dano são utilizados na modelagem estatística.

A matriz dos vetores de características foi dividida em matrizes de treinamento e teste. A matriz de treinamento permite que o algoritmo aprenda a condição normal da estrutura e incorpore ao classificador os efeitos operacionais e ambientais de variabilidades. Matriz esta, que por sua vez possui dimensões de  $T^{3476 \times 4}$ . Já a matriz de teste possui dimensões de  $Z^{3932 \times 4}$ , composta por todos os dados, sejam eles considerados normais ou anormais.

O processo para detecção de dano será aqui conduzido em duas etapas. A primeira etapa onde um algoritmo de clusterização é empregado para separar dados da matriz de treinamento  $T$  em um número  $k$  de partições. Essa etapa é importante, pois captura a essência das alterações causadas por variações ambientais e operacionais representadas por possíveis deslocamentos quantitativos das frequências naturais. Nesta etapa de normalização (filtragem de efeitos de variabilidades), o algoritmo deve identificar diferentes *clusters*. Os modelos gerados servem para classificação de toda a base  $Z$  com intenção de indicar corretamente dados discrepantes que devem corresponder às condições da ponte no período em que a estrutura continha dano [4].

Por conseguinte, a segunda etapa do processo está propriamente relacionada à detecção de dano. A determinação de observações que apresentam discrepância é feita por

meio da distância euclidiana, onde a distância da amostra ao centro do *clusters* mais próximo representa seu DI.

Atualmente há duas abordagens utilizadas para separação e identificação de danos estruturais. A primeira categoria para Danos Operacionais e a segunda para Danos Ambientais. A primeira abordagem consiste em medir as variações operacionais em ambientais e a segunda abordagem consiste em aplicar algoritmos inteligentes que avaliem a influência de variabilidade operacional e ambiental em ambientes sensíveis a danos [18].

## VI. RESULTADOS

Para a etapa de execução dos algoritmos foram definidos alguns parâmetros para serem aplicados aos mesmos. Esses parâmetros são de suma importância, pois podem variar de acordo com o tamanho da base de dados. A Tabela I mostra os valores e parâmetros de cada algoritmo executado.

Tabela I: Parâmetros dos algoritmos.

Algoritmo	n_clusters	quantile	eps	m_samples
K-Means	4	X	X	X
DBSCAN	X	X	0,101	18
MeanShift	X	0,04	X	X

Após a definição dos parâmetros, a etapa de treino dos algoritmos foi realizada. Essa etapa fornece a divisão de grupos ou *clusters* feita por cada algoritmo. A divisão de grupos é fundamental, para que na etapa de teste o algoritmo possa diferenciar qual dado está em condição normal e qual não está.

O algoritmo k-means identificou exatamente o número de clusters que foi definido no seu parâmetro, conforme é mostrado na Figura 2, ou seja, 4 (quatro) *clusters*. Os dois primeiros (Amarelo e Preto), podem estar relacionados com as condições operacionais da estrutura, no estado normal, que ocorreram no período de coleta dos dados. O terceiro e quarto (Azul e Vermelho), são pertinentes as variações de baixa temperatura da ponte.

A divisão de clusters gerada no DBSCAN identificou três *clusters* (ver Figura 3). O primeiro *clusters* (Vermelho) está fortemente relacionado as variações operacionais da ponte. O Segundo e terceiro (Preto e Azul) ao período de pré-congelamento e congelamento total do asfalto em baixas temperaturas do ano.

Dentre os algoritmos executados o MeanShift apresentou o menor número de *clusters*, apenas 2 *clusters* conforme ilustra a Figura 4. Esse resultado demonstra o bom desempenho do algoritmo referente a classificação dos dados.

Após a etapa de treinamento, os resultados em termos de erros tipo I e II gerados através dos testes serão mostrados nas figuras a seguir. Ambos os algoritmos coincidentemente mostraram a mesma quantidade de erros tipo I, exatamente a porcentagem esperada, ou seja, 5% (ver Tabela II).

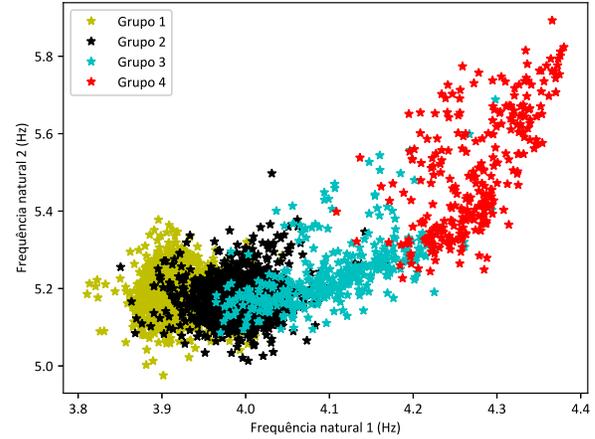


Figura 2: Divisão de clusters - K-Means

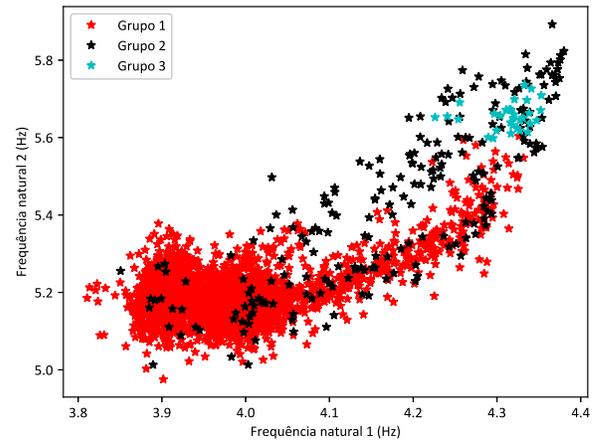


Figura 3: Divisão de clusters - DBSCAN

Tabela II: Resultados dos Erros tipo I e II.

Algoritmos	Erros tipo I	Erros tipo II	Percentual total de erros
K-Means	174 (5,0%)	11 (2,41%)	4,70%
DBSCAN	174 (4,50%)	3 (0,65%)	4,50%
MeanShift	174 (5,0%)	14 (3,07%)	4,78%

A Figura 5 mostra os resultados obtidos no k-means. Em erros do tipo I, o algoritmo apresentou os 5% esperado. Já em erros do tipo II o algoritmo exibiu 11 erros. É importante que se atende ao números de erros do Tipo II apresentaram.

O algoritmo DBSCAN identificou um total de 177 erros (4.50%), e diferente do demais resultados, o algoritmo conseguiu obter um excelente resultado referente aos erros do Tipo II como mostra a Figura 6. Indicando a robustez do algoritmo para análise do problema proposto.

Já o MeanShift identificou um total de 188 erros, tendo

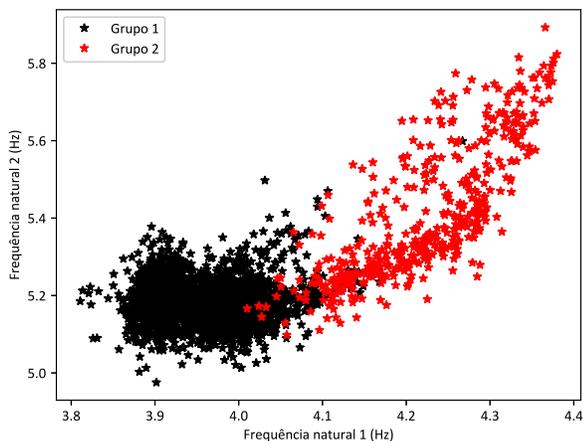


Figura 4: Divisão de clusters - MeanShift

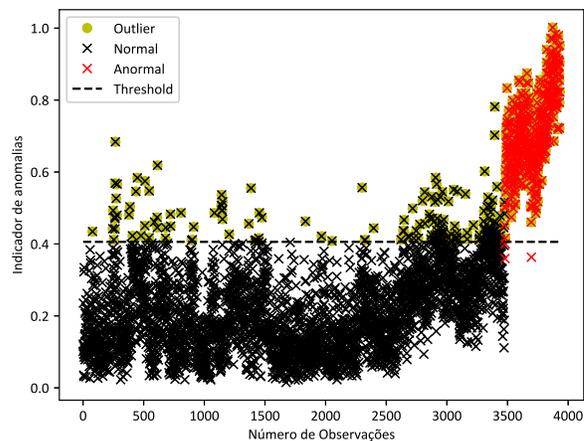


Figura 6: Resultado - DBSCAN

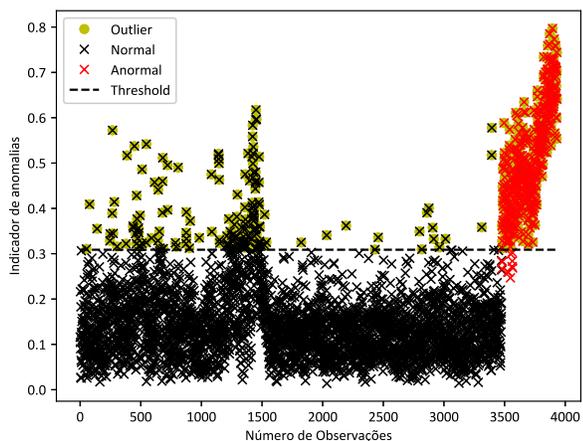


Figura 5: Resultado - K-Means

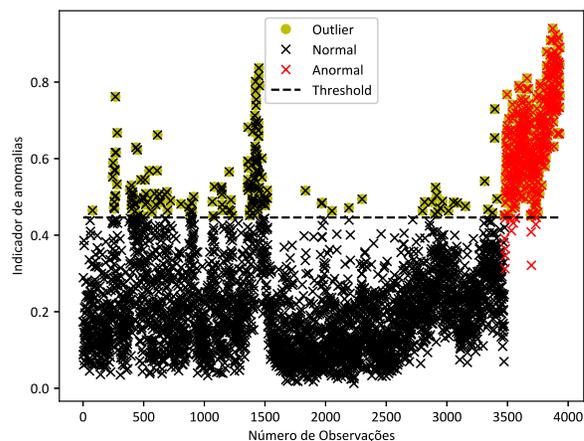


Figura 7: Resultado - MeanShift

um percentual de erros do tipo II um pouco acima se comparado ao k-means e o DBSCAN, conforme a Figura 7 ilustra. Através deste resultado é possível verificar que o algoritmo não conseguiu desempenhar corretamente seu método de classificação, devido a complexidade do problema.

É importante ressaltar que a discrepância da distribuição dos dados ilustrados nas Figuras 5, 6 e 7 sobre os resultados dos algoritmos, porém mais evidenciados nas Figuras 5 e 7, estão ligadas ao período de pré-congelamento e congelamento total da estrutural da ponte, que ocorria em determinados períodos do ano, devido ao clima. Através desse aspecto é possível verificar mais uma vez o melhor desempenho do algoritmo DBSCAN se comparado aos demais, ao qual conseguiu atenuar essa dados de acordo com sua distribuição no plano. Evidenciando o bom desempenho do algoritmo em analisar os dados propostos.

No cenário de testes, ambos os algoritmos mostraram resultados satisfatórios, tendo em vista que no geral a quantidade de erros tipo II foi pequena. Quanto menor a quantidade dos erros tipo II, mais satisfatórios são os resultados, pois, significa que o algoritmo errou pouco em relação a condição danificada da estrutura, demonstrando um bom desempenho mediante a análise do problema. Esse aspecto é importantíssimo, já que, a pesquisa lida não só com a integridade estrutural de construções civis (prédios, edifícios ou pontes), mas com vidas de pessoas que diariamente transitam nessas estruturas.

Se tratando da literatura que visam essa problemática, exemplificados neste artigo na Seção II, comparando seus resultados, os algoritmos desempenharam boas avaliações do problema em que cada um foi proposto. No entanto, o algoritmo DBSCAN se comparado aos algoritmos, que visavam a mesma problemática, como o SVD, AANN, MSD, dentre outros, obteve um melhor resultado referente

a diminuição de erros (Tipo I e Tipo II) identificados, demonstrando a eficácia do algoritmo e ratificando a importância da busca do algoritmo em extinguir os erros.

É importante frisar que, os desempenhos dos algoritmos varia de acordo com a problemática (classificação, diminuição de ruído, compressão de dados, etc.) em que ele é empregado, pois cada algoritmo tem suas características e formas de tratamentos dos dados.

### CONCLUSÃO

Este trabalho apresentou o desempenho de três algoritmos de clusterização desenvolvidos para detecção de danos estruturais. Como foi mostrado anteriormente, todos os algoritmos foram submetidos a testes sobre o conjunto de dados extraídos da ponte Z-24. Logo, o intuito do trabalho foi verificar seus desempenhos na verificação dos danos, levando em consideração todos os efeitos estocásticos quanto os efeitos operacionais que acometiam a ponte.

Os testes com os algoritmos mostraram resultados bem satisfatórios com as identificações dos danos na estrutura. É importante frisar mais uma vez que, os algoritmos seguiram a restrição de confiança baseada no limiar de 5% de erros que um algoritmo pode ter nos testes de detecção, que podem ser verificados no tópico anterior. Todos os algoritmos foram capazes de fazer agrupamentos coerentes ao estado da estrutura, tanto para os dados em condição normal, quanto para os danos na condição danificada da base de dados. Essa coerência indica o bom desempenho dos algoritmos em saber diferenciar mudanças que são resultados de variabilidade ambiental ou operacional.

É importante frisar que todos os algoritmos foram submetidos as mesmas condições de teste, para avaliação da base de dados. Com tudo, o algoritmo que apresentou o melhor resultado nos testes foi o DBSCAN, pois o mesmo apresentou a menor taxa de erros do tipo II, dentre os algoritmos, e como foi dito nos tópicos anteriores, os erros do tipo II são erros que refletem a condição da estrutura que estão com possíveis danos. O fato do DBSCAN ser bem mais robusto, se comparado com o k-means (algoritmo com o segundo melhor resultado), proporciona uma análise mais profunda e detalhada sobre o problema proposto, tornando a identificação de padrões mais suscetível ao algoritmo. Com isso, o algoritmo se mostrou muito eficiente para medição e avaliação dos dados propostos.

### REFERÊNCIAS

- [1] A. D. F. dos Santos, M. F. M. da Silva, C. d. S. de Sales Junior, M. J. de Sousa, C. S. Fernandes, and J. C. W. A. Costa, "Data management system for structural health monitoring," *IEEE Latin America Transactions*, vol. 13, no. 4, pp. 1090–1097, 2015.
- [2] V. B. Falcão *et al.*, "Aplicação de metodologias de reconhecimento de padrões para detecção de dano em vigas de concreto," 2017.
- [3] C. G. Cardoso, "Identificação de danos em estruturas de viga por meio de suas propriedades," *Science*, vol. 20, no. 2, pp. 93–100, 2005.
- [4] A. P. Araujo, "Aplicação de algoritmo de clusterização na detecção de danos estruturais," 2017.
- [5] L. V. Palomino *et al.*, "Técnicas de inteligência artificial aplicadas ao método de monitoramento de integridade estrutural baseado na impedância eletromecânica para monitoramento de danos em estruturas aeronáuticas," 2012.
- [6] A. Santos, E. Figueiredo, M. Silva, C. Sales, and J. Costa, "Machine learning algorithms for damage detection: Kernel-based approaches," *Journal of Sound and Vibration*, vol. 363, pp. 584–599, 2016.
- [7] A. Santos, E. Figueiredo, and J. Costa, "Clustering studies for damage detection in bridges: A comparison study," *Structural Health Monitoring 2015*, 2015.
- [8] E. Figueiredo, G. Park, C. R. Farrar, K. Worden, and J. Figueiras, "Machine learning algorithms for damage detection under operational and environmental variability," *Structural Health Monitoring*, vol. 10, no. 6, pp. 559–572, 2011.
- [9] A. Diez, N. L. D. Khoa, M. M. Alamdari, Y. Wang, F. Chen, and P. Runcie, "A clustering approach for structural health monitoring on bridges," *Journal of Civil Structural Health Monitoring*, vol. 6, no. 3, pp. 429–445, 2016.
- [10] W. Nick, K. Asamene, G. Bullock, A. Esterline, and M. Sundaresan, "A study of machine learning techniques for detecting and classifying structural damage," *International Journal of Machine Learning and Computing*, vol. 5, no. 4, p. 313, 2015.
- [11] N. Lopes Cavalcanti Junior, "Clusterização baseada em algoritmos fuzzy," Master's thesis, Universidade Federal de Pernambuco, 2006.
- [12] (2019, aug) Scikit learn: Clustering. [Online]. Available: <https://scikit-learn.org/stable/modules/clustering>
- [13] A. A. M. de Castro and P. P. L. do Prado, "Algoritmos para reconhecimento de padrões," *Revista Ciências Exatas*, vol. 8, no. 2002, 2002.
- [14] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [15] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 5, pp. 603–619, 2002.
- [16] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 8, pp. 790–799, 1995.
- [17] G. G. Schardong, L. J. Silva, A. T. Winck, and C. T. Pozzer, "Agrupamento de dados baseado em mean shift aplicado a legendas de séries televisivas," in *Proceedings of the Symposium on Knowledge Discovery, Mining and Learning. Sao Carlos, SP, Brazil*, 2013.
- [18] G. A. Bauer, "Geração de conhecimento através de dados da plataforma lattes com o uso de técnicas de mineração de dados," 2016.
- [19] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.
- [20] A. Santos, R. Santos, M. Silva, E. Figueiredo, C. Sales, and J. C. Costa, "A global expectation-maximization approach based on memetic algorithm for vibration-based structural damage detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 661–670, 2017.
- [21] B. Peeters, J. Maeck, and G. De Roeck, "Vibration-based damage detection in civil engineering: excitation sources and temperature effects," *Smart materials and Structures*, vol. 10, no. 3, p. 518, 2001.
- [22] C. R. Farrar, H. Sohn, and G. Park, "A statistical pattern recognition paradigm for structural health monitoring," in *9th ASCE Specialty Conference on Probabilistic Mechanics*, 2004.