

Sistema de Múltiplos Classificadores para Detecção de Defeitos em Dormentes de Aço

Leonardo Pessoa Freitas e Silva * Glauco Ferreira Gazel Yared **
Agnaldo José da Rocha Reis ***

* Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto (UFOP) e Instituto Tecnológico Vale (ITV), Ouro Preto, MG (e-mail: leonardopessoa.eng@gmail.com).

** Departamento de Eng. Elétrica da Universidade Federal de Ouro Preto (UFOP), Rua 36, 115 - Loanda, 35931-008, João Monlevade, MG (e-mail: glauco@ufop.edu.br)

*** Depto de Eng. de Controle e Automação da Universidade Federal de Ouro Preto (UFOP) - Campus Morro do Cruzeiro s/n, 35400-000, Ouro Preto, MG (e-mail: reis@ufop.edu.br)

Abstract: The rail system is essential for commercial activities and transport of people in several countries, playing an important role in improving economic indicators. In order to ensure the reliability and safety of rail transport, it is becoming increasingly important to monitor the conditions of the railway and to execute planned maintenance. The defect in the sleepers can cause an overload on adjacent sleepers, accelerating the structure fatigue of such components, contributing to the occurrence of new defects and finally affecting the track gauge. In this context, one proposes a new method for detecting defects in steel sleepers from the permanent way geometric signals, based on signal processing and machine learning. Three classifiers with different learning characteristics were trained: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and AdaBoost. In addition, a multiple classifier system was implemented to improve system accuracy.

Resumo: O sistema ferroviário é essencial para as atividades comerciais e transporte de pessoa em diversos países, desempenhando um papel importante na melhoria dos indicadores econômicos. Para garantir a confiabilidade e a segurança do transporte ferroviário, torna-se cada vez mais importante monitorar as condições da ferrovia e executar a manutenção planejada. Os defeitos nos dormentes podem provocar uma sobrecarga nos dormentes adjacentes, acelerando a fadiga da estrutura desses componentes e contribuindo para a ocorrência de novos defeitos, afetando finalmente a bitola. Neste contexto, propõe-se um novo método para detecção de defeitos em dormentes de aço a partir de sinais geométricos da via permanente, com base em técnicas de processamento de sinais e aprendizado de máquina. Foram treinados três classificadores com diferentes características de aprendizado: Redes Neurais Artificiais (RNA), máquinas de vetores de suporte (SVM, do inglês *Support Vector Machine*) e o *AdaBoost* (ADB). Além disso, implementou-se um sistema de múltiplos classificadores (*ensemble*) para melhorar a acurácia do sistema.

Keywords: Machine learning; steel sleepers; defect detection; pattern analysis; railway safety.

Palavras-chaves: Aprendizado de máquina; dormentes de aço; detecção de defeitos; análise de padrões; segurança ferroviária.

1. INTRODUÇÃO

O sistema ferroviário é essencial para as atividades comerciais e transporte de pessoa em diversos países, desempenhando um papel importante na melhoria dos indicadores econômicos. Essas atividades estão associadas a custos

* O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001; do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ); da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG); da Universidade Federal de Ouro Preto (UFOP) e da Vale SA.

operacionais mais baixos em comparação com atividades rodoviárias e aéreas, além de serem mais eficazes em termos de utilização de energia e emissão de carbono (Khan et al., 2018). A demanda cada vez maior no transporte de cargas é desejada, embora contribua severamente para a degradação da infraestrutura ferroviária. Este problema trouxe novos desafios relacionados às atividades de manutenção, buscando aumento de eficiência, redução de custos, ganho de produção e questões de segurança.

As tecnologias para analisar as condições estruturais dos componentes ferroviários têm atraído muita atenção da

academia nos últimos anos. Dois métodos de aprendizado de máquina foram introduzidos para detecção de defeitos em rodas de trens (Krummenacher et al., 2017). Outro trabalho (Ng et al., 2019) apresentou um sistema para verificar as relações entre os defeitos da superfície do trilho (RSDs, do inglês *Rail Surface Defects*) e seus correspondentes sinais de aceleração da caixa de eixo (ABA, do inglês *Axle Box Acceleration*). Além disso, foi investigada a influência da distância entre os dormentes no crescimento da corrosão ferroviária (Ng et al., 2018).

A via permanente é responsável por absorver e dissipar o impacto da carga do material rodante. Especificamente no que diz respeito aos dormentes, eles devem suportar os dispositivos de fixação dos trilhos e a capacidade estrutural de transmitir as forças dos trilhos ao lastro. Os danos ocorridos na forma de fissuras nos dormentes podem introduzir situações perigosas dependendo da carga diária de tráfego e do tipo de fissura. Rachaduras não visíveis podem piorar e se tornar uma fratura completa. Qualquer ruptura de um determinado dormente provoca uma sobrecarga nos dormentes adjacentes, acelerando a fadiga da estrutura desses componentes e contribuindo para a ocorrência de novos defeitos, afetando finalmente a bitola.

Convencionalmente, os dormentes das ferrovias são inspecionados manualmente, realizados por inspetores que percorrem a via permanente observando aspectos visuais da ferrovia, em busca de possíveis problemas. No entanto, essas atividades são uma tarefa de manutenção extremamente demorada e trabalhosa, além de expor a saúde e segurança dos inspetores a riscos. Alguns exemplos dessas situações são: caminhar longas distâncias na chuva ou sob temperaturas extremas; o risco de cair ao caminhar em superfícies irregulares; inspeção em áreas de risco ou de difícil acesso. Além desses problemas, a avaliação visual é subjetiva, dependendo da percepção de cada inspetor. Por fim, em algumas áreas, os dormentes podem estar em situações não visíveis, cobertos por lastro, dificultando a inspeção.

Para superar os problemas intrínsecos de inspeção mencionados acima, apresenta-se neste trabalho um novo método para detecção de dormentes de aço defeituosos. Os métodos de classificação utilizados visam aumentar a confiabilidade da infraestrutura ferroviária, garantindo maior segurança e reduzindo custos de manutenção. Para atingir este objetivo, propõe-se o uso de técnicas de processamento de sinal e aprendizado de máquina que permitem avaliar os dados existentes e extrair as informações sobre dormentes de aço defeituosos. Os dados utilizados nesta pesquisa estão associados à geometria da via permanente e foram adquiridos por um Carro Controle (CC) de uma ferrovia existente, localizada no Brasil, a Estrada de Ferro Vitória-Minas (EFVM).

2. TRABALHOS RELACIONADOS

Diferentes abordagens foram propostas na literatura com relação ao problema de detecção e classificação de dormentes defeituosos, com diferentes casos de uso. Os métodos e técnicas utilizados nos trabalhos apresentados geralmente contêm métodos automatizados baseados em processamento de imagens ou reconhecimento de padrões ou híbridos. Embora existam diversas pesquisas sobre mé-

todos de detecção de danos em dormentes de concreto e madeira, até o momento pouco se sabe sobre os métodos para detecção de defeitos em dormentes de aço.

No trabalho desenvolvido por Franca and Vassallo (2020) foi possível demonstrar a capacidade do sistema de trabalhar com imagens de ferrovias, classificando os dormentes em dois tipos (madeira ou aço), além de detectar defeitos nos dormentes de madeira. A transformada de Haar e a “imagem integral” (do inglês *integral image*) foram utilizadas, assim como outras técnicas de processamento de imagens, como detecção de bordas e cálculo de entropia, juntamente com aspectos da topologia ferroviária.

Para os dormentes de concreto, Delforouzi et al. (2017) introduziram um método baseado em visão computacional para detecção de trincas usando um sistema integrado incluindo solução de *hardware* e *software*. *Template Matching* e métodos para encontrar deslocamentos de imagem são as principais abordagens usadas para detecção de dormentes, além de algumas técnicas de limiarização binária. Outro trabalho relacionado (Clark et al., 2017) defende o uso de sensores de emissão acústica para detectar problemas estruturais em dormentes de concreto. Os autores apresentaram investigações experimentais para detectar trincas centradas no centro. Os testes (flexão de três pontos) foram realizados em laboratório com quatro dormentes de concreto. Os resultados demonstraram que a tecnologia de sensoriamento de emissão acústica é eficaz na detecção de eventos iniciais de trinca.

Com relação aos dormentes de aço, foi apresentado um método para detecção de trincas baseado na medição de vibração (Yared et al., 2019). Os sinais de vibração dos dormentes de aço foram adquiridos após a aplicação de um impacto impulsivo com uma marreta. Em seguida, foi aplicado um algoritmo de pré-processamento e técnicas de reconhecimento de padrões para diagnosticar o estado de saúde dos dormentes de aço, que apresentou uma taxa de acerto de 85%.

3. SISTEMA DE MEDIÇÃO E CARACTERÍSTICAS DE DEFEITOS

3.1 Método de Aquisição dos Dados

As inspeções dos elementos da via permanente podem ocorrer com auxílio de um CC (Plasser, 2021). Esses veículos são capazes de coletar diversas informações, parte das quais normalmente estão relacionadas às características geométricas da via permanente ou mesmo imagens para verificar o estado de trilhos, dormentes e dispositivos de fixação. O CC (modelo EM80) utilizado para coleta de dados desta pesquisa trafega na ferrovia a uma velocidade nominal de 80 km/h, fazendo medições a cada 25 cm. Os dados utilizados são fornecidos por diferentes dispositivos instalados neste veículo, como leitor de bitola, medidores de aceleração vertical e horizontal e uma unidade de medida inercial (IMU, do inglês *Inertial Measurement Unit*). Este último, consiste basicamente em dois tipos de sensores: acelerômetros e giroscópios. O primeiro é usado para medir a aceleração inercial, enquanto o segundo para medir a rotação angular. A partir destes dispositivos foram extraídos sete sinais geométricos para buscar estabelecer uma correspondência com as condições do dormente de

aço: alinhamento esquerdo e direito, nivelamento longitudinal esquerdo e direito, bitola, empeno e superelevação.

Atualmente, esses dados são analisados considerando alguns limiares que definem as condições de normalidade dos elementos da via, de acordo com as normas de segurança. Nesse sentido, não é uma tarefa simples correlacionar as análises de forma isolada dos sinais a um problema em um elemento específico. Dessa forma, considerando que essa verificação ainda é feita de forma pontual pela equipe responsável, busca-se desenvolver um novo processo para a análise dos defeitos que seja sistematizado e mais exato.

3.2 Características de Defeito em Dormentes de Aço

Os diferentes tipos de defeitos associados aos dormentes de aço podem estar relacionados a diversos sintomas, comprometendo o próprio dormente ou mesmo as condições da via permanente. Dentre esses defeitos, destaca-se a perda de capacidade estrutural do dormente, caracterizada pelos sintomas de corrosão e trinca. Estes sintomas muitas vezes não são visíveis devido ao local onde aparecem e porque o lastro da via permanente costuma escondê-los. Outro defeito, em nível mais crítico, é a fratura, que é facilmente percebida pois a aba do dormente fica elevada. Considera-se ruptura quando existe uma fratura completa no dormente. Nesta pesquisa, são consideradas trincas, fraturas e rupturas como modo de defeito nos dormentes de aço. As figuras 1 e 2 mostram, respectivamente, o problema de trincas e fraturas em dormentes de aço.



Figura 1. Evidência de trinca em um dormente de aço.



Figura 2. Evidência de fratura em um dormente de aço.

4. O MÉTODO PROPOSTO

Os dormentes de aço são empregados em vários continentes, mas aparecem em números mais expressivos na Oceania, Europa e América do Sul. No entanto, a grande maioria dos dormentes utilizados no mundo são feitos de concreto ou madeira (Ferdous and Manalo, 2014). É provavelmente por isso que nenhuma solução industrial para avaliar a condição estrutural dos dormentes de aço foi apresentada até o momento. Portanto, neste artigo apresenta-se um novo método para detecção de defeitos em dormentes de aço baseado em processamento de sinais e aprendizado de máquina. Para tanto, três classificadores (RNA, SVM e ADB) com diferentes características de aprendizado foram treinados e avaliados isoladamente e sob a forma de *ensembles*. Eles serão apresentados e descritos em detalhes nas seções que se seguem.

Este trabalho está dividido em duas etapas principais. A primeira consiste em aplicar a metodologia de um sistema de reconhecimento de padrões que, essencialmente, envolve os seguintes três aspectos: aquisição e pré-processamento dos dados, representação dos dados e classificação (Jain et al., 2000). As técnicas de cada etapa, desde o pré-processamento de dados até os sistemas de classificação, são apresentadas nas subseções a seguir. A segunda etapa utiliza os melhores modelos gerados na primeira etapa para a construção do *ensemble* de classificadores. Com esta nova abordagem é possível melhorar a acurácia na detecção dos dormentes de aço defeituosos, com base na apresentação dos resultados de forma estratificada.

4.1 Aquisição de Dados e Pré-Processamento

Os dados utilizados nesta pesquisa foram coletados de três diferentes supervisões da EFVM: Conselheiro Pena (CP), Governador Valadares (GV) e Mário Carvalho (MR), de acordo com o método de aquisição de dados explicado na Seção 3.1. Esses dados foram armazenados em arquivos divididos de acordo com algumas características, como o tipo de elemento (curva ou tangente), a supervisão (CP, GV ou MR) e a linha férrea em que se localiza o elemento (a EFVM possui duas linhas férreas nos trechos investigados neste trabalho).

O papel do módulo de pré-processamento é aprimorar a qualidade dos dados coletados. Frequentemente, os dados apresentam diversos problemas tais como dados ruidosos (valores incorretos para os atributos), grande desproporção entre o número de exemplos de cada classe, grande quantidade de valores desconhecidos, entre outros; os quais podem impactar de forma negativa no desempenho do sistema de classificação. No caso deste trabalho, instabilidades e erros inesperados durante a aquisição de dados do CC podem produzir inconsistências, como dados não numéricos ou mesmo amplitudes incompatíveis com o restante dos dados em um determinado momento.

4.2 Extração de Características

Os sinais espaciais associados à geometria da via permanente possuem informações intrínsecas que, a princípio, caracterizam o estado de funcionamento dos componentes da infra e superestrutura ferroviária de forma interligada.

Assim, técnicas de extração e seleção de características buscam extrair informações capazes de revelar aspectos de interesse, como padrões de comportamento relacionados a danos estruturais.

Neste trabalho, optou-se pela utilização dos seguintes atributos espaciais e estatísticos comumente utilizados: energia, variância, taxa de cruzamento por zero, diferença entre as amplitudes máxima e mínima, primeira e segunda derivada e função de autocorrelação a partir de valores de medidas geométricas da via permanente. Por convenção, considerando a característica espacial dos sinais estudados, essas medidas serão tratadas nesta pesquisa como atributos espaciais.

Com o objetivo de localizar trechos da via permanente que contenham dormentes de aço danificados e a fim de reduzir a natureza não estacionária dos dados coletados ao longo da ferrovia, utilizou-se o processo de janelamento dos dados. Foi escolhida uma janela retangular de tamanho 128, ou seja, no interior de cada janela estará presente um total de 128 amostras. A janela espacial é aplicada para determinar um vetor de parâmetro na etapa de extração de características. O n -ésimo vetor de atributos corresponde ao n -ésimo deslocamento da janela de dados, considerando uma janela deslizante com incremento unitário. Para a escolha do tamanho da janela tomou-se como base o tamanho do menor elemento da base de dados. Vale lembrar que como a amostragem é realizada a cada 25 cm, uma janela com 128 amostras corresponde a um trecho analisado de 32 m.

4.3 Seleção de Características

No processo de representação dos dados, é importante se preocupar com a relação entre o tamanho da amostra e o número de atributos, pois estão diretamente relacionados ao desempenho do classificador. Embora o aumento do número de atributos possa levar a uma melhoria no desempenho, na prática, além de um certo ponto, a adição de novas características pode, na verdade, levar a uma redução no desempenho do sistema de classificação. Este fenômeno é denominado como “maldição da dimensionalidade” (Bishop et al., 1995), o que leva ao “fenômeno do pico” (Jain and Chandrasekaran, 1982) no projeto de classificadores. Se o número de amostras de treinamento usadas para projetar o classificador for pequeno em relação ao número de atributos, o desempenho do classificador pode realmente degradar. Nesta pesquisa, há uma preocupação com a quantidade de amostras de treinamento, uma vez que há uma pequena quantidade de dados referentes a dormentes de aço defeituosos se ou quando comparado ao total de amostras.

Para lidar com o problema citado, é necessário utilizar alguma técnica de redução de dimensionalidade. Nesta pesquisa, utilizou-se a razão discriminante de Fisher (FDR, do inglês *Fisher's Discriminant Ratio*). Essa técnica atua na seleção de características e, conseqüentemente, reduz a dimensionalidade do problema.

O Discriminante Linear de Fisher (FLD) é uma das abordagens mais eficientes para redução de dimensão no reconhecimento estatístico de padrões (Webb, 2003). O objetivo da análise de Fisher é realizar a redução de

dimensionalidade preservando ao máximo a informação discriminatória da classe (Sharma et al., 2016). A abordagem adotada por Fisher foi encontrar uma combinação linear das variáveis que separa as duas classes. Esse critério proposto ficou conhecido como FDR. Dessa forma, é possível quantificar a contribuição de cada parâmetro para a discriminação de classes, sendo que aquela com maior capacidade possui os maiores valores de FDR. Decidiu-se variar o número de atributos selecionados de 2 a 24 (2, 4, 6, ..., 24), considerando que nenhuma melhora de desempenho foi observada para valores maiores.

4.4 Classificação

Existem muitos classificadores que podem ser construídos a partir de um determinado conjunto de dados. Nesta pesquisa, utilizou-se as Redes Neurais Artificiais, a máquina de vetores de suporte e o *AdaBoost*. Em relação à complexidade de cada classificador, não houve uma análise sistemática para encontrar a melhor configuração. Nesse sentido, para todos os classificadores a complexidade foi empiricamente selecionada.

Redes Neurais Artificiais O tipo de rede neural utilizado para o desenvolvimento de modelos baseados em RNA foi o *perceptron* multicamadas (MLP, do inglês *multilayer perceptron*). Para o sistema MLP desenvolvido, utilizou-se uma única camada oculta, mas variou-se o número de neurônios em cada iteração (20, 25, 30, 35 e 40) e um único neurônio na camada de saída. Na primeira parte do treinamento foi utilizado o algoritmo *Backpropagation* de gradiente decrescente com *momentum* e taxa adaptativa. Foi utilizada a tangente hiperbólica como função de ativação para a camada oculta e a *sigmóide* para a camada de saída. O treinamento é interrompido quando se alcança o número máximo de épocas (definido como 1000) ou o desempenho é minimizado para o objetivo ($1e^{-4}$). Após algumas iterações e redução de erros, o processo continua com uma função de treinamento de rede que atualiza os valores de peso e *bias* de acordo com a otimização de *Levenberg-Marquardt*.

Máquinas de Vetores de Suporte É um algoritmo de aprendizado de máquina supervisionado que tem sido utilizado como uma ferramenta para resolver problemas práticos de classificação binária (Sun et al., 2005), (Bhowmik et al., 2009). Segundo Haykin (2007), as SVMs podem ser consideradas como algoritmos de aprendizado supervisionado, baseado no princípio de minimização de risco estrutural, advindo das teorias de aprendizado estatístico. A SVM utiliza diferentes *kernels* que podem efetivamente modelar funções de decisão não lineares. Nesta pesquisa, o treinamento realizado para construir os modelos é baseado em duas variações em relação à função *kernel*: linear e polinomial. O *kernel* polinomial é definido como $K(X_i, X_j) = (\gamma X_i \cdot X_j + C)^p, \gamma > 0$. Foi utilizado $p = [2, 3]$, $C = 1$ e γ é sintonizado usando um procedimento heurístico. Esse procedimento heurístico usa subamostragem, de modo que as estimativas podem variar de uma chamada para outra. Portanto, para reproduzir os resultados, definiu-se uma semente de número aleatório antes do treinamento.

AdaBoost O algoritmo de *boosting* utilizado, o *AdaBoost*, foi implementado utilizando classificadores baseado

em árvores de decisão como aprendiz fraco para fins de treinamento. Também chamado de *Adaptive Boosting*, o *AdaBoost* atribui um peso maior a instâncias classificadas incorretamente em cada iteração, buscando assim uma maior otimização do algoritmo final. Neste trabalho, variou-se a complexidade do classificador (número de ciclos de aprendizagem do *ensemble* a serem realizados) de 20 a 100 (20, 40, 60, 80 e 100).

5. MODELOS E BASE DE DADOS

Um único conjunto de dados foi montado para treinar os três classificadores escolhidos e avaliar o desempenho das diferentes configurações analisadas. Assim como mencionado, foram utilizados os sinais geométricos da via permanente coletados pelo CC (Seção 3.1). Embora o defeito estrutural no dormente de aço apareça de maneiras diferentes (Seção 3.2), consideramos apenas os defeitos de trincas, fraturas e rupturas. Portanto, considerando um modelo de classificação binária, os dormentes caracterizados com esses três tipos de defeito são rotulados como defeituosos e, caso contrário, são considerados saudáveis.

Experimentos iniciais demonstraram que, quando dados dos elementos de diferentes supervisões são submetidos ao treinamento em conjunto, os classificadores tiveram desempenhos ruins. O que possivelmente explica esse baixo desempenho é o fato de que cada supervisão possui uma característica particular, seja estrutural, do solo ou mesmo climática, dificultando o aprendizado dos classificadores. Nesse sentido, decidiu-se realizar os experimentos separadamente para cada supervisão (CP, GV e MR), linha (1 ou 2) e tipo de elemento (curvas e tangentes), o que resulta em 12 configurações diferentes. Portanto, foram utilizados modelos diferentes para cada uma dessas configurações.

5.1 Modelos e Aplicação

Inicialmente, o conjunto de dados selecionado foi utilizado para avaliar o desempenho individual de cada classificador. Dessa forma, seguindo-se as etapas de um sistema de reconhecimento de padrões, inicialmente foi realizado o processo de extração de características dos dados considerando a janela de tamanho 128. Em seguida, antes do treinamento dos classificadores, selecionou-se de forma iterativa um número de atributos da FDR (vide Seção 4.3). Além disso, como mencionado na Seção 4.4, variou-se também a complexidade de cada classificador. Desse modo, a partir de cada variação, avaliava-se o modelo gerado quanto as métricas analisadas (Taxa de Acerto na localização de dormentes defeituosos e Taxa de Esforço Desnecessário, que serão apresentadas no Capítulo 6).

Para o *ensemble*, a estratégia foi construir uma matriz em que cada coluna seja representada pela saída individual de cada classificador (0 para o diagnóstico dos dormentes saudáveis e 1 para o diagnóstico dos dormentes defeituosos). A última coluna da matriz recebe a soma da previsão de cada classificador individual. Nesse sentido, o menor valor que poderia aparecer nesta coluna é 0 (quando o diagnóstico de todos os classificadores indica a ausência de defeito) e o maior valor seria 3 (quando o diagnóstico de todos os classificadores indica a presença de defeito). Valores intermediários indicam o conflito entre classificadores. Cada

grupo é nomeado de acordo com a pontuação na última coluna da matriz. Portanto, as previsões podem pertencer ao Grupo 0, 3 ou grupos intermediários. Ao final, é possível verificar o desempenho do sistema para cada um.

5.2 Base de Dados

Foi selecionado um conjunto de dados contendo 225 elementos das três supervisões escolhidas (CP, GV e MR), sendo 132 curvas e 93 tangentes, totalizando uma extensão de 78.151 metros. Dentro desse trecho analisado, há um total de 128.404 dormentes instalados, dos quais 980 apresentam defeitos (0,76%). Nesse sentido, fica claro que a relação entre dados saudáveis e defeituosos é desequilibrada. O desequilíbrio entre os dados refere-se à distribuição desigual de classes dentro de um conjunto de dados, ou seja, há muito menos eventos em uma classe em comparação com as outras. Para contornar o problema de desequilíbrio de classe, as linhas nos dados de treinamento são re-amostradas. O conceito básico é alterar as proporções das classes (distribuição a priori) dos dados de treinamento para obter um classificador que possa prever efetivamente a classe minoritária (os dormentes de aço com defeito). Para resolver esse problema, reequilibrou-se as proporções de classe dos conjuntos de dados sub-amostrando-se aleatoriamente a classe maior. Portanto, 50% das medições estão relacionadas com os dormentes saudáveis e os outros 50% com os defeituosos. As tabelas 1 e 2 mostram o número de elementos em cada configuração e o respectivo número de amostras.

Tabela 1. Base de dados (curvas).

Supervisão/Linha	Números de elementos	Números de amostras
CP/Linha1	11	1772
CP/Linha2	24	9078
GV/Linha1	17	1654
GV/Linha2	20	2630
MR/Linha1	32	24104
MR/Linha2	28	3786

Tabela 2. Base de dados (tangentes).

Supervisão/Linha	Números de elementos	Números de amostras
CP/Linha1	12	2562
CP/Linha2	8	3700
GV/Linha1	19	1238
GV/Linha2	18	500
MR/Linha1	24	600
MR/Linha2	12	5548

6. RESULTADOS EXPERIMENTAIS

Para avaliação de desempenho dos métodos computou-se duas métricas: Taxa de Acerto (TA) e Taxa de Esforço Desnecessário (TED). A TA, também conhecida como Taxa de Verdadeiros Positivos está relacionada com a capacidade de identificação dos dormentes de aço defeituosos e pode ser calculada por

$$TA = \frac{NDDIC}{NTDD} \quad (1)$$

em que NDDIC é o Número de Dormentes Defeituosos Identificados Corretamente e NTDD é o Número Total de Dormentes Defeituosos. Por outro lado, a TED busca medir o esforço desnecessário empregado durante a inspeção da ferrovia e é formulada como

$$TED = \frac{NDSII}{NTD} \quad (2)$$

onde NDSII é o Número de Dormentes Saudáveis Identificados Incorretamente e NTD é o Número Total de Dormentes. Deve-se ressaltar que a TED, apesar de ter sido calculado de forma diferente da métrica conhecida de Taxa de Falsos Positivos, se aproxima bastante, uma vez que a proporção existente entre dormentes saudáveis e defeituosos dos dados de teste são de aproximadamente 99% para 1%, respectivamente.

6.1 Seleção e Avaliação dos Modelos

Para todos os experimentos nesta seção, o desempenho é calculado a partir de um conjunto de teste que não foi usado para treinamento e seleção do modelo/parâmetro. Para evitar problemas de *overfitting* e *underfitting*, repetiu-se cada experimento 10 vezes em novas divisões aleatórias de treinamento e teste, e relatou-se a média e o intervalo de confiança com um nível significância de 95%. O motivo de não utilizar a validação cruzada no processo de verificação do desempenho dos modelos foi pelo fato de que a sequência temporal dos dados (a ordem das inspeções) deve ser levada em consideração para o processo de treinamento/validação e teste. Portanto, a aleatoriedade na seleção dos conjuntos de dados pode não ser a melhor escolha em alguns casos práticos.

O CC realiza inspeções bimestralmente ao longo do ano, totalizando 6 inspeções. Portanto, escolheu-se 4 delas para as divisões entre treinamento e teste do conjunto de dados. Uma dessas inspeções (aproximadamente 25% dos dados) foi selecionada para teste enquanto as outras três (aproximadamente 75% dos dados) para treinamento/seleção dos modelos. Como as proporções de classe para ambos os conjuntos de dados não são balanceadas (vide Seção 5.2), reequilibrou-se as proporções de classe dos conjuntos de dados sub-amostrando aleatoriamente a classe maior. Treinar e avaliar os classificadores sem esse procedimento poderia fazer com que houvesse uma maior probabilidade de classificação para a classe sobre-representada.

6.2 Resultados RNA

Em um primeiro momento, desenvolveu-se um sistema para verificar o desempenho individual de cada classificador para o conjunto de dados, considerando as mesmas configurações. Os valores nas tabelas 3 e 4 indicam a média e o intervalo de confiança com nível de significância de 95% para cada configuração, utilizando o classificador RNA. A cada geração de um novo modelo (considerando as variações dos atributos analisados) comparava-se a TA e TED do modelo atual com os desempenhos do anterior. O método de escolha do modelo consistiu-se em fixar um valor aceitável para a TED (40%) e assim selecionar o maior valor de TA. Ao final, o melhor modelo de cada configuração foi salvo para ser utilizado na implementação

do *ensemble*. O mesmo procedimento foi realizado para os outros dois classificadores apresentados nas seções seguir.

Tabela 3. Desempenho RNA (curvas).

Supervisão/Linha	TA (%)	TED (%)
CP/1	100 ± 0	6 ± 1
CP/2	74 ± 6	33 ± 5
GV/1	94 ± 8	28 ± 3
GV/2	81 ± 5	21 ± 4
MR/1	81 ± 1	14 ± 2
MR/2	78 ± 10	34 ± 7

Tabela 4. Desempenho RNA (tangentes).

Supervisão/Linha	TA (%)	TED (%)
CP/1	98 ± 2	35 ± 5
CP/2	86 ± 4	32 ± 5
GV/1	72 ± 1	17 ± 2
GV/2	95 ± 8	40 ± 6
MR/1	97 ± 8	25 ± 2
MR/2	70 ± 3	32 ± 5

Como pode ser visto, utilizando o classificador RNA, o menor desempenho geral para as curvas foi para a supervisão de CP/Linha2 com uma média de 74% para a TA e 33% para a TED. Por outro lado, a supervisão de CP/Linha1 alcançou um percentual médio de 100% para a TA e 6% para a TED. No caso das tangentes, o menor desempenho geral para as curvas foi para a supervisão de MR/Linha2 com uma média de 70% para a TA e 32% para a TED. Por outro lado, a supervisão de CP/Linha1 alcançou um percentual médio de 98% para a TA e 35% para a TED.

6.3 Resultados SVM

Assim como para a RNA, os valores nas tabelas 5 e 6 indicam os resultados para o classificador SVM.

Tabela 5. Desempenho SVM (curvas).

Supervisão/Linha	TA (%)	TED (%)
CP/1	100 ± 0	2 ± 1
CP/2	82 ± 1	37 ± 1
GV/1	99 ± 1	10 ± 1
GV/2	79 ± 1	19 ± 1
MR/1	81 ± 1	12 ± 1
MR/2	94 ± 1	40 ± 1

Tabela 6. Desempenho SVM (tangentes).

Supervisão/Linha	TA (%)	TED (%)
CP/1	100 ± 0	24 ± 1
CP/2	79 ± 1	26 ± 1
GV/1	86 ± 1	39 ± 1
GV/2	99 ± 1	12 ± 1
MR/1	100 ± 0	10 ± 1
MR/2	87 ± 1	39 ± 1

Como resultado utilizando o classificador SVM, o menor desempenho geral para as curvas foi para a supervisão de

GV/Linha2 com uma média de 79% para a TA e 19% para a TED. Por outro lado, a supervisão de CP/Linha1 alcançou um percentual médio de 100% para a TA e 2% para a TED. No caso das tangentes, o menor desempenho geral para as curvas foi para a supervisão de CP/Linha2 com uma média de 79% para a TA e 26% para a TED. Por outro lado, a supervisão de CP/Linha1 alcançou um percentual médio de 100% para a TA e 24% para a TED.

6.4 Resultados ADB

Da mesma forma, os valores nas tabelas 7 e 8 indicam os resultados para o classificador ADB.

Tabela 7. Desempenho ADB (curvas).

Supervisão/Linha	TA (%)	TED (%)
CP/1	100 ± 0	4 ± 1
CP/2	79 ± 4	35 ± 3
GV/1	93 ± 5	13 ± 2
GV/2	89 ± 2	39 ± 1
MR/1	86 ± 1	31 ± 1
MR/2	72 ± 40	24 ± 2

Tabela 8. Desempenho ADB (tangentes).

Supervisão/Linha	TA (%)	TED (%)
CP/1	95 ± 2	20 ± 1
CP/2	91 ± 1	38 ± 1
GV/1	70 ± 5	45 ± 4
GV/2	100 ± 0	14 ± 2
MR/1	100 ± 0	39 ± 6
MR/2	78 ± 2	29 ± 2

Como resultado utilizando o classificador ADB, o menor desempenho geral para as curvas foi para a supervisão de MR/Linha2 com uma média de 72% para a TA e 24% para a TED. Por outro lado, a supervisão de CP/Linha1 alcançou um percentual médio de 100% para a TA e 4% para a TED. No caso das tangentes, o menor desempenho geral para as curvas foi para a supervisão de GV/Linha1 com uma média de 70% para a TA e 45% para a TED. Por outro lado, a supervisão de GV/Linha2 alcançou um percentual médio de 100% para a TA e 14% para a TED.

6.5 Resultados Ensemble

Ao analisar as tabelas de resultados das seções 6.2, 6.3 e 6.4, é possível observar que, para uma mesma configuração (supervisão/linha/tipo do elemento), os três classificadores apresentam desempenhos semelhantes. Essa verificação é importante para garantir que nenhum classificador influencie negativamente o sistema *ensemble*.

Depois de verificar as performances individuais, o melhor modelo de cada configuração é utilizado para a implementação do *ensemble*. Como resultado, a TA e a TED podem ser calculados para cada um dos 4 grupos definidos pela combinação dos classificadores. Esses grupos são analisados individualmente para as medidas de desempenho, de forma que seja possível entender o potencial de cada um deles na detecção de defeitos nos dormentes de aço. Os

Tabela 9. Performance do *ensemble* para cada configuração e grupo (curvas).

Supervisão/Linha	Métrica	G3 (%)	G2 (%)	G1 (%)	G0 (%)
CP/Linha1	TA	100,00	0	0	0
	TED	0,19	0,57	2,91	93,95
CP/Linha2	TA	71,6	18,52	4,32	5,56
	TED	16,24	21,65	21,39	35,65
GV/Linha1	TA	100	0	0	0
	TED	3,03	5,7	12,97	76,89
GV/Linha2	TA	83,33	12,5	4,17	0
	TED	11,75	7,35	24,57	54,44
MR/Linha1	TA	78,18	9,7	6,06	6,06
	TED	11,93	10,25	26,81	42,47
MR/Linha2	TA	78,67	29,33	6,67	5,33
	TED	11,49	20,45	20,09	46,10

Tabela 10. Performance do *ensemble* para cada configuração e grupo (tangentes).

Supervisão/Linha	Métrica	G3 (%)	G2 (%)	G1 (%)	G0 (%)
CP/Linha1	TA	100	0,00	0,00	0,00
	TED	6,13	10,72	24,7	54,92
CP/Linha2	TA	87,30	1,59	3,17	7,94
	TED	24,54	6,03	9,52	55,26
GV/Linha1	TA	68,18	27,27	4,55	0
	TED	7,71	15,71	37,38	38,82
GV/Linha2	TA	100	0	0	0
	TED	1,67	4,13	21,95	71,98
MR/Linha1	TA	100	0	0	0
	TED	1,43	5,06	23,41	69,74
MR/Linha2	TA	67,57	24,32	0	8,11
	TED	21,28	17,01	16,23	41,37

resultados para os 4 grupos das curvas e tangentes de cada supervisão/linha são apresentados nas tabelas 9 e 10.

Esses resultados evidenciam as diferentes contribuições de TA e TED para cada grupo. Nesse sentido, é possível notar que o Grupo 3, em todos os casos, apresentou a maior TA entre todos os grupos e, além disso, uma baixa TED. Os outros grupos não apresentaram um comportamento comum em todos os casos, mas seguiram uma lógica de que, na maioria dos casos, o desempenho diminui à medida que parte-se do Grupo 3 para o Grupo 0.

É importante estar claro que, individualmente, o Grupo 3 não irá superar o resultado obtido pelos classificadores individuais. Entretanto, pode ser que, junto a outro grupo, esse desempenho se aproxime dos outros resultados quanto a TA. Neste último caso, mesmo não alcançando o melhor resultado com relação à TA, a TED cai significativamente no Grupo 3, trazendo uma grande contribuição para o desempenho. Além disso, a aplicação dessa metodologia permitiu que os resultados fossem estratificados e, consequentemente, que fosse possível a criação de um sistema de prioridades de acordo com o compromisso entre o esforço empregado na inspeção e a taxa de acerto na localização dos dormentes danificados. Em outras palavras, agora é possível procurar por trechos da via permanente contendo dormentes defeituosos com uma maior probabilidade de localizá-los e, ao mesmo tempo, com uma menor chance de realizar um esforço desnecessário.

7. CONCLUSÃO

Foram apresentados diferentes métodos de aprendizado de máquina para detecção de defeitos em dormentes de aço. Os métodos se valeram de sete sinais geométricos (alinhamento esquerdo e direito, nivelamento longitudinal esquerdo e direito, bitola, superelevação e empeno) da via permanente fornecidos por dispositivos instalados em um Carro Controle. Os resultados encontrados evidenciaram o potencial das técnicas utilizadas em cada etapa, assim como a capacidade dos sinais geométricos na identificação de dormentes de aço defeituosos. Os primeiros experimentos mostraram o desempenho médio individual de cada classificador, sendo que o menor desempenho apresentado foi de 70% na detecção dos dormentes defeituosos, enquanto que em outros casos esse desempenho foi próximo ou igual a 100%. Com relação ao sistema *ensemble*, pode-se dizer que os resultados encontrados foram muito promissores devido à sua característica de apresentação. A maneira com que o método foi implementado possibilitou a criação de um sistema de prioridades de acordo com o compromisso desejado entre o esforço empregado na inspeção e a TA na localização dos dormentes danificados. Adicionalmente, a combinação entre os grupos foi capaz de superar o desempenho dos resultados individuais quanto a TA em alguns casos, além de que a TED reduz significativamente nos grupos mais significativos. A partir dos resultados apresentados, fica evidente a contribuição industrial do trabalho para o aumento da confiabilidade da infraestrutura ferroviária, reduzindo custos de manutenção. Além disso, existe uma contribuição importante de saúde e segurança, uma vez que possibilita a redução significativa das horas de exposição dos inspetores aos riscos físicos e condições climáticas intensas.

Um direcionamento que poderia ser levado em consideração como trabalho futuro, seria uma abordagem ao problema baseado em técnicas de detecção de anomalias, em que se modela apenas a classe dos dados normais, visto que o conjunto de dados é bastante desbalanceada. Além disso, com relação à seleção dos dados de treinamento para a geração dos modelos, é possível utilizar uma estratégia de sobreamostragem dos dados de defeito, considerando-se que neste trabalho foi utilizado um processo de subamostragem dos dados normais.

AGRADECIMENTOS

O presente trabalho, o qual se encontra inserido na *Cátedra Under Rail*, foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), da Universidade Federal de Ouro Preto (UFOP) e da Vale SA.

REFERÊNCIAS

Bhowmik, T.K., Ghanty, P., Roy, A., and Parui, S.K. (2009). Svm-based hierarchical architectures for hand-written bangla character recognition. *International Journal on Document Analysis and Recognition (IJ-DAR)*, 12(2), 97–108.

- Bishop, C.M. et al. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Clark, A., Kaewunruen, S., Janeliukstis, R., and M., P. (2017). Damage detection in railway prestressed concrete sleepers using acoustic emission. *IOP Conference Series: Materials Science and Engineering*.
- Delforouzi, A., Tabatabaei, A.H., Khan, M.H., and Grzegorzec, M. (2017). A vision-based method for automatic crack detection in railway sleepers. In *International Conference on Computer Recognition Systems*, 130–139. Springer.
- Ferdous, W. and Manalo, A. (2014). Failures of mainline railway sleepers and suggested remedies—review of current practice. *Engineering Failure Analysis*, 44, 17–35.
- Franca, A.S. and Vassallo, R.F. (2020). A method of classifying railway sleepers and surface defects in real environment. *IEEE Sensors Journal*.
- Haykin, S. (2007). *Redes Neurais: Princípios e Prática*. Artmed.
- Jain, A., Duin, R., and Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22, 4–37.
- Jain, A.K. and Chandrasekaran, B. (1982). 39 dimensionality and sample size considerations in pattern recognition practice. *Handbook of statistics*, 2, 835–855.
- Khan, H.U.R., Siddique, M., Zaman, K., Yousaf, S.U., Shoukry, A.M., Gani, S., Khan, A., Hishan, S.S., Saleem, H., et al. (2018). The impact of air transportation, railways transportation, and port container traffic on energy demand, customs duty, and economic growth: Evidence from a panel of low-, middle-, and high-income countries. *Journal of Air Transport Management*.
- Krummenacher, G., Ong, C.S., Koller, S., Kobayashi, S., and Buhmann, J.M. (2017). Wheel defect detection with machine learning. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1176–1187.
- Ng, A.K., Martua, L., and Sun, G. (2018). Influence of sleeper distance on rail corrugation growth. In *2018 International Conference on Intelligent Rail Transportation (ICIRT)*, 1–5. IEEE.
- Ng, A.K., Martua, L., and Sun, G. (2019). Dynamic modelling and acceleration signal analysis of rail surface defects for enhanced rail condition monitoring and diagnosis. In *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*. IEEE.
- Plasser (2021). Carro controle eletrônico auto propelido em80h. URL <https://www.plasser.com.br/pt/maquinas-sistemas/em80h.html>.
- Sharma, L., Yadav, D.K., and Singh, A. (2016). Fisher's linear discriminant ratio based threshold for moving human detection in thermal video. *Infrared Physics & Technology*, 78, 118–128.
- Sun, B.Y., Huang, D.S., and Fang, H.T. (2005). Lidar signal denoising using least-squares support vector machine. *IEEE Signal processing letters*, 12(2), 101–104.
- Webb, A.R. (2003). *Statistical pattern recognition*. John Wiley & Sons.
- Yared, G.F.G., Barbosa, C.H.N.R., Leite, S.N.C., Barbosa, V.M., Andrade, M.P., Almeida, A.B.M., Silva, D.A., Vasconcelos, R.L., Oliveira, L., and Rodrigues, R.F. (2019). Vibration analysis for crack detection in railroad steel sleepers. *International Heavy Haul STS Conference (IHHA 2019)*.