

## Efeito da seleção de métricas de desempenho em modelos de aprendizado supervisionado para detecção de amaciamento em compressores<sup>\*</sup>

Gabriel Thaler<sup>\*</sup> João V. B. Menegali<sup>\*</sup>  
Ahryman S. B. de S. Nascimento<sup>\*</sup> Antonio L. S. Pacheco<sup>\*</sup>  
Rodolfo C. C. Flesch<sup>\*</sup> João P. Z. Machado<sup>\*</sup>

<sup>\*</sup> Universidade Federal de Santa Catarina, Florianópolis, SC, (e-mails: thaler.gabriel@posgrad.ufsc.br, j.bilesimo.menegali@grad.ufsc.br, a.nascimento@labmetro.ufsc.br, pacheco@inep.ufsc.br, rodolfo.flesch@ufsc.br, joao.zomer.m@posgrad.ufsc.br).

**Abstract:** This work presents an analysis on the effect of performance metrics on the selection of supervised methods for running-in detection in reciprocating hermetic compressors. Since the running-in phenomenon takes place only in the first few hours of the lifespan of such devices, datasets extracted from the compressor operation during this period and in its normal operation are usually more representative of its post running-in behavior, and, even after balancing the dataset, an adequate choice of the performance metric might reduce the undesired effects of this situation in classification models. In order to reach the proposed goal, parameters of random forest models and of data preprocessing methods were optimized for the following performance metrics: the area under the receiver operating characteristic curve, the  $F$ -score with three different configurations, and the Matthews correlation coefficient. When applied to experimental data, the results obtained with such models showed the Matthews correlation coefficient and the  $F$ -score with more weight attributed to precision ( $\beta = 0,5$ ) as the most appropriate ones to the given case study.

**Resumo:** Este trabalho apresenta uma análise da escolha de métricas de desempenho na seleção de métodos supervisionados de detecção do estado de amaciamento em compressores herméticos alternativos para refrigeração. Dado que ensaios de amaciamento são realizáveis apenas nas primeiras horas de operação de uma unidade, e que após esse período é registrada apenas a operação normal do dispositivo, dados de ensaios geralmente contém medições majoritariamente em operação normal. O desbalanceamento entre a quantidade de amostras em cada estado de operação pode gerar efeitos negativos na classificação, que podem ser mitigados com o balanceamento artificial do conjunto e com a escolha adequada de métrica de desempenho. Foram avaliadas combinações de pré-processamento e modelos de floresta aleatória com parâmetros otimizados para as métricas da área abaixo da curva de característica de operação do receptor, o  $F$ -score para três diferentes configurações e o coeficiente de correlação de Matthews, com o conjunto de dados subamostrado aleatoriamente para balanceamento do conjunto. Ao fim, a aplicação em dados experimentais de tais modelos indicou o coeficiente de correlação de Matthews e o  $F$ -score com tendência à precisão ( $\beta = 0,5$ ) como mais apropriados ao problema.

**Keywords:** supervised machine learning; running-in; non-destructive analysis; hermetic reciprocating compressors.

**Palavras-chaves:** aprendizado de máquina supervisionado; amaciamento; análise não destrutiva; compressores herméticos alternativos.

### 1. INTRODUÇÃO

O amaciamento é um fenômeno associado ao início do contato deslizante entre duas superfícies de um mecanismo, caracterizado como um período transitório com variações na taxa de desgaste, na temperatura dos materiais, no

<sup>\*</sup> Este trabalho foi apoiado pela Nidec Global Appliance, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por meio do projeto 432116/2018-4, e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

coeficiente de atrito e em outros parâmetros correlatos que caracterizam o estado tribológico de um mecanismo. Esse fenômeno, que é um fator significativo para garantir um desempenho satisfatório e para prolongar a vida útil de sistemas, é sucedido por um regime permanente tribológico, que caracteriza a operação normal de suas superfícies componentes (Blau, 2005).

Em compressores herméticos alternativos para refrigeração, admite-se que o amaciamento ocupe uma parcela pequena de sua vida útil, referente apenas às primeiras

horas de operação de um dispositivo projetado para operar por décadas em condições usuais. Assim, quando são avaliados parâmetros de desempenho, como a potência elétrica consumida e a capacidade de refrigeração, deve-se garantir que o compressor se encontre tribologicamente em regime permanente (Schmitz, 2014). Devido aos vários mecanismos envolvidos em sua operação e à característica hermética dos dispositivos, para a análise direta dos parâmetros tribológicos representativos do amaciamento seria necessário o rompimento do selo do dispositivo e desmonte dos componentes, implicando um novo período de amaciamento após a remontagem. Na prática, isso inviabiliza o emprego de métodos baseados nos princípios clássicos de caracterização de amaciamento para compressores hermeticamente, sendo que o procedimento usual de amaciamento nesses dispositivos consiste na operação da unidade por um tempo determinado empiricamente, sem garantia do término do fenômeno.

Recentemente, um método baseado em aprendizado não supervisionado foi proposto em Thaler et al. (2021), permitindo uma estimativa do término do amaciamento em dados de ensaios já realizados. Ainda assim, a demanda por métodos de análise do amaciamento que possam detectar o término do fenômeno durante a execução de ensaios, denominada detecção on-line, ainda não foi suprida, sendo uma alternativa para tal aplicação o treinamento de ferramentas de classificação supervisionada. Porém, enquanto a operação normal de um compressor pode ser registrada facilmente em ensaios, um ensaio de amaciamento requer unidades novas, de forma que o conjunto gerado por ensaios com ambos os estados do produto normalmente possuem um desbalanceamento, com o período do amaciamento geralmente sub-representado, o que pode ser prejudicial no desempenho de modelos de aprendizado supervisionado treinados em tal conjunto.

O objetivo deste trabalho é analisar o impacto da escolha da métrica de desempenho na otimização de modelos de aprendizado supervisionado em conjuntos desbalanceados, avaliada em dados experimentais para detecção on-line do término do amaciamento para compressores hermeticamente alternativos. Tais métricas foram analisadas em um modelo de floresta aleatória, treinado com dados experimentais de compressores, cujos instantes estimados de amaciamento foram definidos utilizando o método proposto em Thaler et al. (2021).

O restante do trabalho é organizado como segue. A Seção 2 descreve o problema da identificação indireta do amaciamento em compressores e métodos de detecção presentes na literatura. A Seção 3 apresenta as principais métricas de desempenho para classificação de conjuntos desbalanceados. A Seção 4 descreve o desenvolvimento experimental do trabalho, incluindo a descrição da bancada de ensaios utilizada e dos dados nela obtidos. A Seção 5 descreve o pré-processamento dos dados e o procedimento de escolha e treinamento dos modelos de floresta aleatória empregados. A Seção 6 descreve os resultados obtidos com a aplicação dos modelos classificadores obtidos em um ensaio de teste e discorre sobre a escolha de métricas para otimização dos mesmos. Por fim, a Seção 7 apresenta a conclusão do trabalho.

## 2. IDENTIFICAÇÃO INDIRETA DO AMACIAMENTO

Esta seção descreve a motivação para este trabalho e discorre sobre trabalhos que apresentam soluções para detecção indireta do amaciamento, com ênfase em publicações que sugeriram grandezas possivelmente relacionadas ao fenômeno em máquinas rotativas.

### 2.1 Caracterização do problema

Atualmente, o método padrão para o amaciamento de compressores novos é a operação do dispositivo em condições específicas por um número de horas definido com base em observações empíricas (Martins et al., 2011). Porém, mesmo depois de aplicado tal método, resultados de parâmetros de desempenho obtidos em ensaios consecutivos de calorímetro indicam que o procedimento atual pode ser insuficiente para o amaciamento adequado dos dispositivos. A Figura 1 apresenta o valor médio e o desvio-padrão de medições da capacidade de refrigeração, um dos principais parâmetros de desempenho para compressores de refrigeração, em três ensaios consecutivos de 33 compressores de um mesmo modelo, normalizados para o intervalo  $[0;1]$ . O primeiro conjunto de medições é equivalente aos resultados típicos de ensaios de calorímetro, e a variação observada entre tal ensaio e os ensaios seguintes pode ser um indicativo dos efeitos do amaciamento.

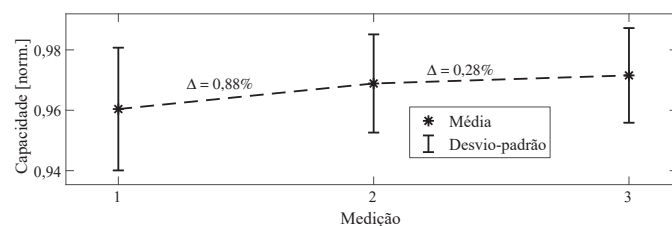


Figura 1. Média e desvio-padrão da capacidade de refrigeração escalonada em ensaios de calorímetro consecutivos.

Apesar da preparação tribológica dos compressores, realizada antes de cada um dos primeiros ensaios de calorímetro, pode-se afirmar por teste- $t$  de duas amostras, com significância 0,05, que a média da capacidade de refrigeração nos primeiros ensaios é menor do que a média dos ensaios seguintes. Tal comportamento pode ser atribuído ao fato que, para algumas unidades do modelo avaliado, o tempo de preparação previsto para o amaciamento não foi suficiente e o dispositivo continua passando pelo processo de amaciamento durante os ensaios de calorímetro, que, como indicado em Penz et al. (2012), podem levar até 4,5 h. No atual cenário, um método automático capaz de detectar o fim do período de amaciamento permitiria uma melhora na confiabilidade ao afirmar que as condições de ensaio refletem as de operação normal dos dispositivos.

O método de detecção proposto em Thaler et al. (2021) possibilita a detecção do instante de amaciamento em dados de ensaios realizados com compressores durante e após o amaciamento, utilizando aprendizado não supervisionado em dados da operação do compressor. Tal solução, apesar de apresentar um bom ponto de partida para estudos sobre o fenômeno, necessita um conjunto de dados que inclua também ensaios de referência com o

dispositivo já amaciado para análise do amaciamento em uma unidade de compressor. Dessa forma, a solução não é capaz de detecção on-line do término do amaciamento, visto que exige ensaios de referência com o compressor já amaciado, mas pode ser utilizada como especialista para rotulação de dados, permitindo, assim, o desenvolvimento de abordagens supervisionadas para detecção.

## 2.2 Possíveis grandezas relacionadas ao amaciamento

Uma revisão dos estudos presentes na literatura sobre o amaciamento é apresentada por Khonsari et al. (2021), os quais analisam o contato entre superfícies isolado da influência de outros mecanismos. Nesses estudos, o fenômeno é analisado principalmente pela sua influência na topografia e no coeficiente de atrito, além da taxa de desgaste e da análise dos detritos desbastados no processo. Fatores tribológicos como os citados são considerados os principais indicativos do processo de amaciamento (Dowson et al., 1982; Blau, 2005; Khonsari et al., 2021) e são propostos como possíveis grandezas para detecção do fenômeno em sistemas simples (Grützmacher et al., 2018; Ruggiero et al., 2020; Ghatrehsamani et al., 2022). Porém, em mecanismos com múltiplas peças móveis e diversos pontos de contato, tal análise é muito difícil, visto que fatores da operação conjunta, como o desalinhamento de peças e os acoplamentos entre subsistemas, tornam a análise muito mais complexa em relação ao caso de apenas uma peça móvel (Blau, 1991).

Já a literatura sobre detecção indireta do amaciamento, apesar de esparsa, apresenta sugestões de métodos que podem possibilitar estimativas da progressão do fenômeno em maquinário, como a detecção por análise de óleo lubrificante (Lockwood e Dalley, 1992; Neale, 1995), pela curtose da vibração (Martin, 1992), pela corrente elétrica consumida por um motor (Eissenberg e Haynes, 1992; Blau, 2009) e pela vazão mássica em compressores (Thaler et al., 2021). Empiricamente, o trabalho de Thaler et al. (2021) apontou a corrente elétrica como melhor indicativo do fenômeno.

A partir dos trabalhos levantados nessa subseção, uma bancada experimental foi implementada para medição da corrente elétrica em condições controladas de operação de compressores, utilizada também em Thaler et al. (2021) para análise não supervisionada do amaciamento. Tal bancada, bem como as rotinas de ensaio, são descritas na Seção 4.

## 3. MÉTRICAS PARA CLASSIFICAÇÃO DE CONJUNTOS DESBALANCEADOS

Dado que o fenômeno do amaciamento ocorre apenas nas primeiras horas de operação de um compressor, conjuntos de dados obtidos em ensaios com unidades durante e após o amaciamento tendem a apresentar uma distribuição desbalanceada, com menos registros da operação do compressor durante o fenômeno do que após seu término. Tal desbalanceamento pode influenciar negativamente o desempenho de um classificador mesmo após balanceamento artificial, especialmente na classe minoritária, sendo que um possível mitigador desse efeito é a escolha de uma métrica adequada para o desempenho da classificação (Guo et al., 2017).

Diversas métricas são utilizadas no contexto de modelos classificadores, de acordo com a classe de problema avaliada. Em trabalhos que utilizam conjuntos naturalmente balanceados, métricas globais como a exatidão global, que representa a proporção de casos corretamente previstos, são amplamente empregadas. Porém, em casos nos quais os conjuntos são desbalanceados, tal métrica possui uma inerente tendência à classe majoritária, dado que todo ponto classificado contribui igualmente para o resultado final (Guo et al., 2017). Para avaliar de forma mais efetiva modelos que envolvam tais conjuntos, são utilizadas métricas que ponderam a distribuição das classes na avaliação, como a área abaixo da curva de característica de operação do receptor (ROC, do inglês *receiver operating characteristics*), também chamada de ROC-AUC (do inglês *area under the ROC curve*), o *F*-score, o coeficiente de correlação de Matthews (MCC, do inglês *Matthews correlation coefficient*) e a média geométrica entre as taxas de verdadeiros positivos e de verdadeiros negativos (Guo et al., 2017). Comparações de desempenho entre algumas métricas para conjuntos desbalanceados e outras métricas globais são apresentadas por Gu et al. (2009) e por Jeni et al. (2013).

Neste trabalho, foram utilizados para avaliação dos métodos de detecção on-line do amaciamento a ROC-AUC, o *F*-score e o MCC, que são brevemente introduzidos nesta seção. Tais métricas são amplamente utilizadas na classificação de conjuntos desbalanceados, e a validade de suas aplicações em tal contexto é assunto de estudos recentes (Halimu et al., 2019; Chicco e Jurman, 2020).

As três métricas utilizadas se baseiam em proporções de acertos e erros na classificação de um conjunto. Tais acertos e erros podem ser classificados de quatro formas, com  $N_{VP}$  representando o número de verdadeiros positivos da classificação,  $N_{VN}$  representando o número de verdadeiros negativos,  $N_{FP}$  representando o número de falsos positivos e  $N_{FN}$  representando o número de falsos negativos.

### 3.1 Área abaixo da curva ROC

A partir do número de classificações em cada caso de uma matriz de confusão, obtém-se a taxa de verdadeiros positivos ( $R_{VP}$ ), também chamada de sensibilidade ou revocação, e a taxa de falsos positivos ( $R_{FP}$ ), tal que:

$$R_{VP} = \frac{N_{VP}}{N_{VP} + N_{FN}}, \text{ e } R_{FP} = \frac{N_{FP}}{N_{FP} + N_{VN}}. \quad (1)$$

Essas taxas são utilizadas na construção da curva ROC, uma ferramenta gráfica para visualização do desempenho de classificadores. A curva é utilizada para comparar a taxa de acertos e de falsos positivos para dados limiares, no caso de classificadores probabilísticos, ou do modelo em si, no caso de classificadores discretos. Como classificadores probabilísticos retornam um valor contínuo de probabilidade como estimação, cada ponto da curva corresponde a  $R_{VP}$  e  $R_{FP}$  para um dado limiar, iniciando no caso no qual o limiar é 0, e todos os pontos são classificados negativamente ( $R_{FP} = 0$ ,  $R_{VP} = 0$ ), até o ponto em que o limiar é 1, e todos os pontos são classificados positivamente ( $R_{FP} = 1$ ,  $R_{VP} = 1$ ). No caso de classificadores discretos, a estimacão é um valor binário que representa a classe estimada, sem medida probabilística de classificação. Dessa forma, a

curva possui apenas um ponto além dos pontos extremos, que representa o desempenho do modelo (Fawcett, 2006). A forma como a curva é definida também a torna pouco suscetível ao desbalanceamento do conjunto, com a taxa de acertos em ambas as classes afetando  $R_{FP}$  e  $R_{VP}$  (Jeni et al., 2013).

Para comparação entre diferentes classificadores, porém, a bidimensionalidade da curva ROC pode dificultar o processo de análise, sendo preferível uma única medida escalar. Uma métrica popular para tal é a ROC-AUC, obtida calculando a área abaixo da curva para um dado classificador, que é proporcional à probabilidade de o modelo classificar corretamente um dado ponto. Como a curva ROC é sempre contida no quadrado unitário, a ROC-AUC pode teoricamente variar no intervalo  $[0;1]$ , mas qualquer classificador avaliado com a métrica abaixo de 0,5 não é prático, já que seu desempenho médio é inferior àquele obtido com um classificador aleatório gaussiano (Fawcett, 2006).

### 3.2 F-score

O  $F$ -score, também chamado  $F_\beta$ -score, é dado pela média harmônica entre a medida de precisão  $P$ , que representa a proporção de casos positivos corretamente, e a revocação  $R$  (idêntica ao  $R_{VP}$  definido na Equação (1)). Assim, o  $F_\beta$ -score é definido tal que

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} = \frac{(\beta^2 + 1)N_{VP}}{(\beta^2 + 1)N_{VP} + \beta^2 N_{FN} + N_{FP}}, \quad (2)$$

com  $\beta$  ponderando a importância da revocação sobre a precisão, resultando em um valor dentro do intervalo  $[0;1]$ , com 0 e 1 sendo, respectivamente, o pior e o melhor desempenho. A aplicação mais tradicional dessa métrica considera  $\beta = 1$  e é chamada de  $F_1$ -score (Chicco e Jurman, 2020; Chinchor, 1992). O  $F$ -score é criticado por não considerar o número de amostras corretamente classificadas como negativas e por variar quando as classes são invertidas (Hand e Christen, 2018), porém sua capacidade de ponderação entre diferentes tipos de erro a torna uma métrica significativamente relevante no contexto de conjuntos desbalanceados.

### 3.3 Coeficiente de correlação de Matthews

O MCC leva em consideração todas as quatro possibilidades de classificação, sendo definido por

$$MCC = \frac{N_{VP}N_{VN} - N_{FP}N_{FN}}{\sqrt{(N_{VP} + N_{FN})(N_{VP} + N_{FP})(N_{VN} + N_{FP})(N_{VN} + N_{FN})}}, \quad (3)$$

resultando em um valor entre  $[-1; +1]$ , em ordem crescente de desempenho (Chicco e Jurman, 2020). Esse coeficiente é uma aplicação do coeficiente de Pearson para classificadores binários, correlacionando a classe real com a classe predita (Matthews, 1975; Chedzoy, 2006).

Apesar de bem equilibrado em geral, pode-se observar que o MCC não é definido quando a soma de alguma das linhas ou colunas da matriz de confusão é igual a zero, ou seja, quando todas as amostras do conjunto real são de apenas uma classe ou quando todas as amostras são previstas em uma única classe. Além disso, alguns contextos exigem que

maior ênfase seja dada para alguma classe específica, o que não é previsto por essa métrica (Brown, 2018).

Estudos recentes comparam o  $F_1$ -score ao MCC, e este à ROC-AUC. Em tais estudos, o MCC apresenta vantagens significativas em relação ao  $F_1$ -score (Chicco e Jurman, 2020), e grau similar de consistência à ROC-AUC (Halimu et al., 2019). Nesse último estudo, porém, também foi observado que o grau de discriminância da ROC-AUC pode se mostrar superior ao do MCC, sugerindo aquela como uma métrica mais significativa, desde que seja notado que para classificadores probabilísticos a ROC-AUC não necessariamente representa o desempenho do modelo, e sim uma média do desempenho para diversos limiares.

## 4. DESENVOLVIMENTO EXPERIMENTAL

Esta seção descreve o procedimento experimental desenvolvido de forma a obter dados com significância em relação ao amaciamento em compressores herméticos alternativos, bem como o conjunto de dados final obtido. A Subseção 4.1 descreve a estrutura física, a instrumentação e o sistema de controle da bancada experimental, enquanto a Subseção 4.2 descreve a metodologia de ensaios e a organização do conjunto de dados obtidos.

### 4.1 Bancada de ensaios

Para este trabalho, foi utilizada uma bancada experimental para ensaios em compressores como a proposta em (Thaler et al., 2021), que descreve com maior ênfase seu processo construtivo. O diagrama de tubulação e instrumentação simplificado de tal bancada é apresentado na Figura 2.

- |                                     |                                 |
|-------------------------------------|---------------------------------|
| 1 - Compressor hermético            | TT - Transdutor de temperatura  |
| 2 - Válvula de diafragma controlada | TC - Controlador de temperatura |
| 3 - Válvula agulha controlada       | PT - Transdutor de pressão      |
| 4 - Reservatório intermediário      | PC - Controlador de pressão     |
| 5 - Trocador de calor               | IT - Transdutor de corrente     |

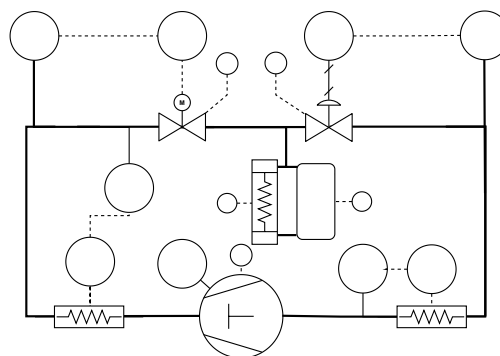


Figura 2. Diagrama simplificado da bancada automática de avaliação do amaciamento.

A bancada foi projetada para operação em ciclo quente, sem troca de fase do fluido refrigerante utilizado. Nela, a pressão de sucção do compressor (1) é controlada utilizando um controlador proporcional-integral, que atua na abertura da válvula (2) regulada com um conversor eletropneumático, e a pressão de descarga é controlada utilizando um controlador em cascata, com uma malha interna rápida para controle da posição da válvula (3) e uma malha externa para controle da pressão. Nesse sistema

de controle, a malha interna é controlada utilizando controlador proporcional, e a malha externa utilizando controlador proporcional-integral. O controle de temperatura é realizado manualmente por meio de resistores atuando sob modulação de largura de pulso, e a medição de corrente elétrica é realizada utilizando um transdutor de efeito Hall. A Figura 3 apresenta uma foto em perspectiva da bancada de amaciamento implementada.

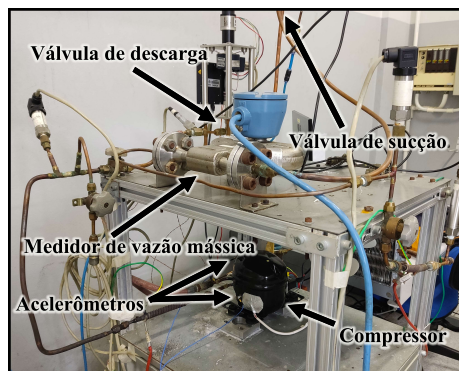


Figura 3. Bancada de amaciamento implementada.

Com a instalação da instrumentação proposta e a operação da bancada validada, foram realizados ensaios de amaciamento com duração de 20 h, tempo estabelecido em função de dados experimentais como suficiente para garantir o amaciamento do modelo. Além dos ensaios de amaciamento, também foram realizados ensaios com unidades já amaciadas, utilizados como referência no treinamento dos modelos classificadores. A descrição do conjunto de dados empregado e a definição de classes para treinamento e validação são descritas na Subseção 4.2.

#### 4.2 Dados obtidos

Para análise experimental do fenômeno do amaciamento, foi escolhido um modelo de compressor para aplicação residencial, com capacidade de refrigeração fixa de 56 W. No total, foram empregados os mesmos dados de ensaios experimentais com três unidades de um mesmo modelo de compressor hermético utilizados em Thaler et al. (2021) denominadas cronologicamente de Unidade 1 à Unidade 3, com a adição de mais uma unidade do mesmo modelo, denominada Unidade 4. Para cada unidade foi realizado um ensaio de amaciamento, com a unidade ainda não operada, e dois ensaios de referência, nos quais o dispositivo já se encontrava amaciado.

A Figura 4 apresenta o valor eficaz (RMS, do inglês *root mean square*) da corrente ao longo do ensaio de amaciamento das quatro unidades avaliadas, destacando a classificação do estado de amaciamento, realizada utilizando o método proposto em Thaler et al. (2021). É possível observar uma variação significativa entre unidades na duração do amaciamento, sendo que nas unidades 1 e 3 a transição do estado “não amaciado” para “amaciado” foi detectada em menor tempo (7 h e 6 h, respectivamente), enquanto para as unidades 2 e 4 foi necessário mais tempo até que os compressores atingissem o estado “amaciado” (14 h e 10 h para as unidades 2 e 4, respectivamente).

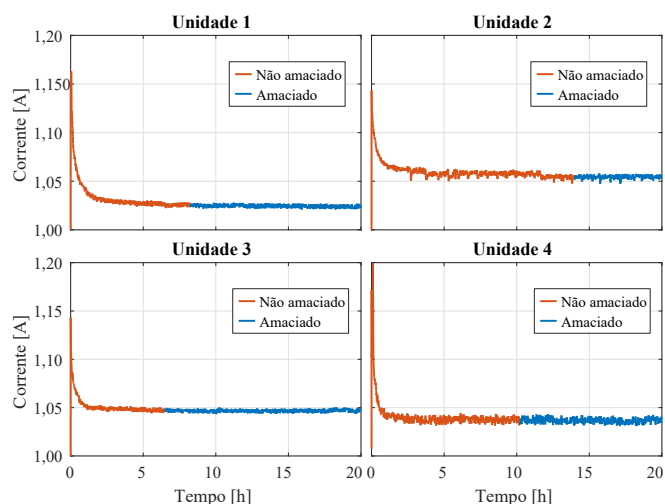


Figura 4. Corrente RMS em ensaios de amaciamento classificados utilizando o método de Thaler et al. (2021).

## 5. MÉTODO PROPOSTO

Esta seção apresenta o procedimento proposto para aplicação de aprendizado supervisionado para detecção on-line do fenômeno do amaciamento, projetado para aplicação em ensaios preparatórios de novas unidades de compressores. A Subseção 5.1 apresenta o método utilizado para processamento das séries temporais obtidas em ensaios em vetores de predição, e a Subseção 5.2 apresenta o procedimento de seleção de hiperparâmetros e a escolha de modelos de floresta aleatória para avaliação no estudo de caso proposto.

### 5.1 Processamento de dados

Para a aplicação desejada de aprendizado supervisionado, foram utilizados o valor RMS, a curtose e a variância da corrente elétrica do motor do compressor, selecionados dentre as grandezas medidas conforme descrito na Subseção 4.2. Tais dados foram extraídos uma vez a cada minuto de ensaio, porém devido ao regime transitório natural de ensaios em compressores (Penz et al., 2012), as medições referentes à primeira hora de ensaio foram descartadas.

Para suavização das séries, foi utilizado filtro de média móvel com janela  $M$ . Para construção dos vetores de observação, foi utilizada uma janela deslizante de  $N$  momentos, dentro de um espaço de atraso de  $D$  amostras. Dessa forma, sendo  $x(k)$  o valor de uma das séries avaliadas no instante  $k$  e  $\hat{x}(k)$  o seu valor filtrado naquele instante, podem-se definir:

$$\hat{x}(k) = \frac{1}{M} \sum_{i=0}^{M-1} x(k-i), \quad \text{e} \quad (4)$$

$$\hat{X}(k) = [\hat{x}(k), \hat{x}(k-D), \hat{x}(k-2D), \dots, \hat{x}(k-(N-1)D)],$$

com  $\hat{X}(k)$  sendo o vetor de observações filtradas para cada instante  $k$ . A cada vetor de observação foi também associado o instante de ensaio, de forma que o vetor final possui comprimento  $N+1$ .

As respostas associadas a cada vetor de observação foram construídas a partir dos instantes de amaciamento definidos off-line utilizando a análise por agrupamento proposta em Thaler et al. (2021), sendo que respostas anteriores ao instante de amaciamento foram definidas como “não amaciado” (classe negativa) e instantes posteriores como “amaciado” (classe positiva). Considerando que o ensaio se inicia no instante  $k = 1$ , nos primeiros  $M$  instantes, a janela de filtragem é reduzida para o maior valor condizente, e  $\hat{X}(k)$  só é definido para  $k > (N - 1)D$ , sendo que o tempo  $(N - 1)D$ , somado à hora de ensaio inicial descartada, é definido como o tempo de preparação do método.

Para análise da classificação em situação próxima à aplicação pretendida, os dados processados do ensaio de amaciamento da Unidade 4 foram selecionados como conjunto de teste, e os dados de ensaios de amaciamento e de referência das outras unidades foram empregados como conjunto de treino. Dada a natureza do fenômeno do amaciamento, descrita na Seção 1, os conjuntos de dados gerados são naturalmente desbalanceados, podendo gerar uma tendência à classe majoritária do conjunto de treino (Esposito et al., 2021). De forma a evitar tal tendência, foi aplicado no conjunto de treino subamostragem aleatória (Galar et al., 2012), de modo que o conjunto de treino final é balanceado, com o mesmo número de observações do estado “não amaciado” e “amaciado”.

### 5.2 Modelo de classificação

Como modelo classificador para análise foi escolhido o algoritmo de floresta aleatória, já utilizado em aplicações de detecção de falhas e análise de maquinário, sendo visto como uma abordagem bem estabelecida para tal classe de problemas (Janssens et al., 2016; Wang et al., 2017; Zhang et al., 2018). O modelo consiste em um conjunto de árvores de decisão treinado em seleções aleatórias independentes das observações de um conjunto original, e o resultado da classificação é a decisão majoritária entre as árvores componentes da floresta (Breiman, 2001).

A escolha dos parâmetros de pré-processamento  $M$ ,  $N$  e  $D$  foi realizada utilizando algoritmo de otimização bayesiana, buscando combinações que maximizassem o valor de cada uma das métricas analisadas, definindo limites superior de 30 amostras para  $M$ , 10 amostras para  $N$  e 30 amostras para  $D$ , definidos empiricamente de forma a manter as características do sinal e limitar o tempo máximo de computação. Para este trabalho, foram avaliadas as métricas descritas na Seção 3, sendo essas a ROC-AUC; o  $F$ -score para  $\beta=0,5$ ;  $\beta=1$  e  $\beta=2$ , representados respectivamente por  $F_{0,5}$ ,  $F_1$  e  $F_2$ ; e o MCC.

O processo de otimização foi realizado individualmente para cada uma das grandezas selecionadas, abrangendo desde a escolha de parâmetros até o treinamento do modelo de árvore aleatória para avaliação da solução, com o valor RMS, a variância e a curtose da corrente elétrica representadas respectivamente por  $I_{RMS}$ ,  $I_{Var}$  e  $I_{Kur}$ . De modo a diminuir os efeitos de sobreajuste no sistema, dois hiperparâmetros dos modelos de floresta aleatória foram incluídos no processo de otimização bayesiana com regularização L2: o número mínimo de amostras por folha e a profundidade máxima das árvores, com a última sendo representada por  $d_{max}$ . Os resultados de treinamento e teste das árvores

otimizadas, bem como uma comparação entre métricas de desempenho, são apresentados na Seção 6.

## 6. RESULTADOS E DISCUSSÃO

Buscando uma estimativa do desempenho dos modelos em relação ao conjunto de treino, foi utilizada validação cruzada por  $k$ -fold nos modelos, com dez folds. Dadas as diferenças naturais de cada métrica, o processo de otimização indicou diferentes combinações ótimas de parâmetros para cada escolha de métrica, e os modelos ótimos devidamente treinados foram também avaliados na classificação do ensaio de teste.

A Figura 5 apresenta o desempenho médio nos folds da validação cruzada do conjunto de treino e de classificação do conjunto de teste para cada grandeza, de acordo com as métricas avaliadas. Pode-se observar na figura que, com exceção dos modelos avaliados com MCC e  $F_{0,5}$ , o desempenho dos modelos para as diferentes grandezas foi muito similar, com diferença na ordem de 0,01. Ainda assim, considerando uma aplicação próxima à real, foram selecionados os melhores modelos de cada métrica, com base no desempenho de validação cruzada.

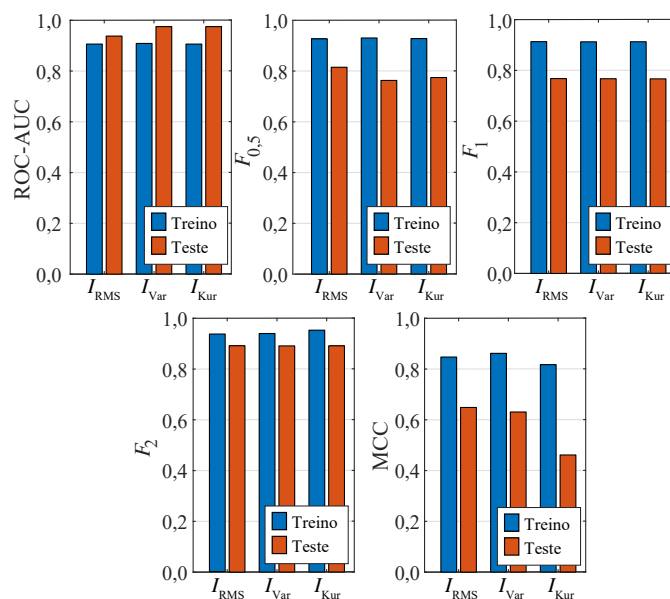


Figura 5. Desempenho na validação cruzada do conjunto de treino e na classificação do conjunto de teste em modelos classificadores selecionados para diferentes grandezas e métricas de desempenho.

De acordo com a Figura 5, também é possível observar que os únicos modelos com resultados superiores na classificação do ensaio de teste em relação à validação cruzada foram os modelos selecionados por ROC-AUC. Apesar de aparentemente contraintuitivo, tal fator pode possivelmente ser explicado pela natureza da métrica, que considera todos os possíveis limiares de probabilidade de um classificador. A Figura 6 apresenta um exemplo de curva de probabilidade de amaciamento para o modelo classificador otimizado para a variância da corrente elétrica, obtida por meio da média da decisão das árvores componentes da floresta aleatória. Na figura é possível observar que, para um limiar médio de 50%, a transição

para regime permanente seria detectada pouco antes de 5 h de ensaio, enquanto um limiar de 80% resultaria em detecção próxima a 14 h. Assim, por considerar diversas possibilidades de limiar, não é incomum tal desempenho pela ROC-AUC, e métodos de variação de limiar como o proposto em Esposito et al. (2021) podem utilizar tal propriedade de um modelo classificador para a obtenção de resultados aprimorados.

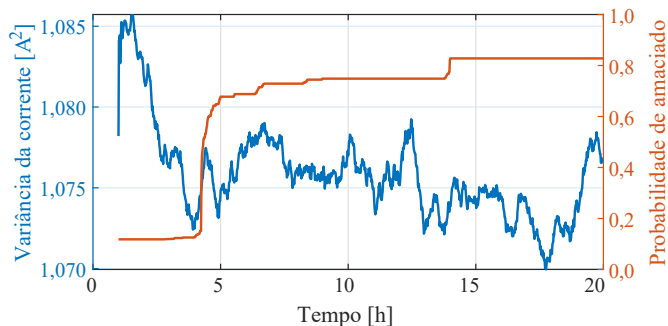


Figura 6. Variância da corrente elétrica e probabilidade de conclusão do amaciamento no ensaio de teste, para um modelo ótimo de classificação selecionado pela métrica ROC-AUC.

O Quadro 1 apresenta os parâmetros dos modelos ótimos escolhidos para cada métrica, com a exceção do número mínimo de amostras por folhas, que foi selecionado pelo algoritmo como 500 amostras para todos os modelos, e a janela  $M$  do filtro de médias móveis, selecionada como 30 para todos os modelos. Em geral, é possível observar que os modelos I e II, respectivamente selecionados pelo bom desempenho em ROC-AUC e  $F_2$ , possuem menor janela de observações, de apenas uma amostra, enquanto outros como os modelos III e IV, selecionados pelas métricas  $F_{0,5}$  e  $F_1$ , respectivamente, utilizam janelas de predição maiores. Também é possível observar que o único modelo com profundidade maior que 2 camadas foi o Modelo V, indicando maior possibilidade de sobreajuste nesse modelo.

Modelo	Métrica	Grandeza	$N$	$D$	$d_{max}$
I	ROC-AUC	$I_{Var}$	1	1	2
II	$F_{0,5}$	$I_{Var}$	4	19	2
III	$F_1$	$I_{RMS}$	2	8	2
IV	$F_2$	$I_{Cur}$	1	1	2
V	MCC	$I_{Var}$	6	1	3

Quadro 1: Modelos selecionados por otimização bayesiana, de acordo com diferentes métricas de desempenho.

Por fim, considerando a aplicação proposta, os resultados de classificação do ensaio de teste foram comparados, buscando avaliar a aplicabilidade de cada modelo. A Figura 7 apresenta tal comparação, e é possível observar que os modelos II e V, obtidos pela otimização das métricas  $F_{0,5}$  e MCC, aproximam-se mais do instante de término do amaciamento que os demais. Tais modelos apresentaram erro aproximadamente 2 h menor do que os outros, o que pode ser explicado pela ponderação mais significativa destes em relação à classe negativa do conjunto. Como no problema de amaciamento tal classe, além de minoritária, apresenta comportamento mais variável, é provável que tal ponderação com enfoque no estado “não amaciado” leve à valorização de modelos que mais bem acomodem

a parte menos representada do problema, sendo assim os mais adequados para aplicações práticas.

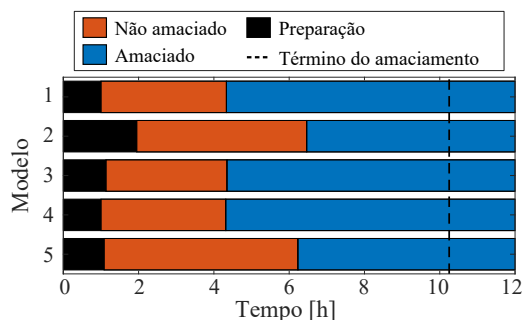


Figura 7. Resultado da classificação do conjunto de teste dos modelos avaliados.

## 7. CONCLUSÃO

Buscando avaliar o efeito da escolha de métrica de desempenho para a classificação supervisionada de conjuntos originalmente desbalanceados, mas com balanceamento artificial, foram avaliados dados experimentais de ensaios de amaciamento em compressores herméticos alternativos, buscando modelos de floresta aleatória capazes de realizar a detecção do término do fenômeno durante ensaios preparatórios de dispositivos aleatoriamente subamostrados. A partir de dados do valor RMS, curtose e variância da corrente elétrica do motor do compressor, foram selecionados modelos otimizados pelas métricas ROC-AUC; MCC e  $F$ -score com  $\beta=0,5$ ;  $\beta=1$  e  $\beta=2$ .

Dentre os modelos selecionados, os modelos otimizados pelas métricas MCC e  $F$ -score com  $\beta=0,5$  mostraram-se mais adequados para a tarefa de classificação. A avaliação dos modelos mostrou também que a seleção por ROC-AUC pode apresentar resultados superiores, contanto que sejam aplicadas técnicas de ajuste de limiar de classificação, já que a métrica é representativa do potencial de classificação do modelo para diversos limiares.

Tais resultados, além de apresentarem uma nova aplicação de aprendizado supervisionado e um avanço nos estudos de detecção indireta do amaciamento, podem também ser aplicados em outros casos de classificação desbalanceada, como o exemplo da detecção de falhas em maquinário, que pela natureza anômala dos fenômenos normalmente são sub-representadas em ensaios experimentais. Futuros trabalhos ampliarão o escopo dos experimentos para diferentes modelos de compressores, adicionando ensaios de mais unidades para obter mais informações sobre o fenômeno, e explorarão a aplicação de outras técnicas de aprendizado no problema, e o efeito de diferentes métodos de balanceamento de conjuntos nos resultados de classificação.

## REFERÊNCIAS

- Blau, P.J. (2009). *Friction science and technology: from concepts to applications*. CRC Press, Boca Raton, USA, 2 edição.
- Blau, P.J. (1991). Running-in: Art or engineering? *Journal of Materials Engineering*, 13(1), 47–53.
- Blau, P.J. (2005). On the nature of running-in. *Tribology International*, 38(11-12 SPEC. ISS.), 1007–1012.

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Brown, J.B. (2018). Classifiers and their metrics quantified. *Molecular Informatics*, 37(1), 1–11.
- Chedzoy, O.B. (2006). Phi-coefficient. In S. Kotz, N. Balakrishnan, C.B. Read, B. Vidakovic, e N.L. Johnson (eds.), *Encyclopedia of Statistical Sciences*, volume 9, 6115–6116. John Wiley & Sons, Inc., Hoboken, EUA, 2 edição.
- Chicco, D. e Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 1–13.
- Chinchor, N. (1992). Muc-4 evaluation metrics. In *Conference on Message Understanding, MUC4 '92*, 22–29. Association for Computational Linguistics, Stroudsburg, EUA.
- Dowson, D., Taylor, C., e Godet, M. (1982). The running-in process in Tribology - Proceedings of the 8th Leeds-Lyon Symposium on Tribology held in the Institut National des Sciences Appliquées de Lyon, France (September 1981). In *Tribology International*, volume 15, 264.
- Eisenberg, D. e Haynes, H. (1992). Motor-current signature analysis. In *ASM Handbook, Volume 18: Friction, Lubrication, and Wear Technology*. ASM International, Materials Park, EUA.
- Esposito, C., Landrum, G.A., Schneider, N., Stiefl, N., e Riniker, S. (2021). GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *Journal of Chemical Information and Modeling*, 61(6), 2623–2640.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., e Herrera, F. (2012). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, 42(4), 463–484.
- Ghatrehsamani, S., Akbarzadeh, S., e Khonsari, M. (2022). Relationship between subsurface stress and wear particle size in sliding contacts during running-in. *Mechanics Research Communications*, 123(March), 103891.
- Grützmacher, P.G., Rosenkranz, A., Szurdak, A., Gachot, C., Hirt, G., e Mücklich, F. (2018). Effects of multi-scale patterning on the run-in behavior of steel-alumina pairings under lubricated conditions. *Advanced Engineering Materials*, 20(1), 1700521.
- Gu, Q., Zhu, L., e Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. In Z. Cai, Z. Li, Z. Kang, e Y. Liu (eds.), *Computational Intelligence and Intelligent Systems*, 2009, 461–471. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Guo, H., Li, Y., Shang, J., Gu, M., Huang, Y., e Gong, B. (2017). Learning from class-imbalanced data: review of methods and applications. *Expert Systems with Applications*, 73, 220–239.
- Halimu, C., Kasem, A., e Newaz, S.H. (2019). Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. *ACM International Conference Proceeding Series*, 1–6.
- Hand, D. e Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, 28(3), 539–547.
- Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufer, M., Verstockt, S., Van de Walle, R., e Van Hoecke, S. (2016). Convolutional Neural Network Based Fault Detection for Rotating Machinery. *Journal of Sound and Vibration*, 377, 331–345.
- Jeni, L.A., Cohn, J.F., e De La Torre, F. (2013). Facing imbalanced data: recommendations for the use of performance metrics. In *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 245–251. IEEE, Los Alamitos, EUA.
- Khonsari, M., Ghatrehsamani, S., e Akbarzadeh, S. (2021). On the running-in nature of metallic tribo-components: A review. *Wear*, 474–475(Dec. 2020), 203871.
- Lockwood, F. e Dalley, R. (1992). Lubricant analysis. In *ASM Handbook, Volume 18: Friction, Lubrication, and Wear Technology*. ASM International, Materials Park, EUA.
- Martin, H. (1992). Vibration analysis. In *ASM Handbook, Volume 18: Friction, Lubrication, and Wear Technology*. ASM International, Materials Park, EUA.
- Martins, E.F., Flesch, C.A., Flesch, R.C., e Borges, M.R. (2011). Conception of a test bench to generate known and controlled conditions of refrigerant mass flow. *ISA Transactions*, 50(3), 513–520.
- Matthews, B. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2), 442–451.
- Neale, M.J. (ed.) (1995). *The Tribology Handbook*. Butterworth-Heinemann, Oxford, UK, 2 edição.
- Penz, C.A., Flesch, C.A., Nassar, S.M., Flesch, R.C., e de Oliveira, M.A. (2012). Fuzzy-bayesian network for refrigeration compressor performance prediction and test time reduction. *Expert Systems with Applications*, 39(4), 4268–4273.
- Ruggiero, A., Di Leo, G., Liguori, C., Russo, D., e Sommella, P. (2020). Accurate Measurement of Reciprocating Kinetic Friction Coefficient through Automatic Detection of the Running-In. *IEEE Trans. Instrum. Meas.*, 69(5), 2398–2407.
- Schmitz, F. (2014). *Desenvolvimento e validação experimental de um padrão de intercomparação para verificação de bancadas de ensaio de eficiência energética de compressores de refrigeração na indústria*. Dissertação (Mestrado), Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Santa Catarina.
- Thaler, G., Nunes, N.A., de S. Nascimento, A.S.B., Pacheco, A.L.S., e Flesch, R.C.C. (2021). Aplicação de aprendizado não supervisionado para identificação não destrutiva do amaciamento em compressores. *Proceedings do XV Simpósio Brasileiro de Automação Inteligente*, 460–466.
- Wang, Z., Zhang, Q., Xiong, J., Xiao, M., Sun, G., e He, J. (2017). Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. *IEEE Sensors Journal*, 17(17), 5581–5588.
- Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., e Si, Y. (2018). A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost. *IEEE Access*, 6, 21020–21031.