

Diagnóstico de Falhas em Rolamentos usando Redes Convolucionais: Otimização da Representação de Sinais e uma Nova Metodologia de Avaliação

Rodrigo Kobashikawa Rosa, Vicente Knobel Borges, Danilo de Souza Braga e Danilo Silva

Resumo— Este artigo trata do problema do diagnóstico automatizado de falhas de rolamento em máquinas rotativas usando dados de vibração, o qual é importante para a operação eficiente e contínua na indústria. Um trabalho recente mostrou que a tradicional divisão em treino e teste do conjunto mais amplamente utilizado em pesquisas apresentava vazamento de dados, levando a resultados excessivamente otimistas em modelos de estado da arte baseados em redes neurais profundas. Neste artigo, é proposta uma nova metodologia de divisão do conjunto de dados que, ao mesmo tempo em que evita problemas de vazamento, também proporciona maior diversidade ao conjunto de treino, permitindo uma maior robustez à mudança de domínio. Além disso, diversas representações 1D e 2D de sinais são investigadas como entrada para redes convolucionais, tendo o melhor modelo alcançado um desempenho que supera com folga o atual estado da arte.

Palavras-Chave— Falha de rolamento, manutenção preditiva, CWRU, aprendizado profundo

Abstract— This paper addresses the problem of automated diagnosis of rolling bearing faults in rotating machinery using vibration data, which is crucial for efficient and continuous operation in the industry. A recent study has revealed that the traditional train-test split widely used in research suffered from data leakage, resulting in overly optimistic outcomes for state-of-the-art models based on deep neural networks. In this article, a new data splitting methodology is proposed that not only mitigates leakage issues but also provides greater diversity to the training set, enabling enhanced robustness to domain variation. Additionally, various 1D and 2D signal representations are explored as input to convolutional networks, with the best-performing model significantly outperforming the current state of the art.

Keywords— Bearing failure, predictive maintenance, CWRU, deep learning

I. INTRODUÇÃO

Falhas em máquinas industriais representam desafios para a operação eficiente e contínua da produção, exigindo detecção precoce e estratégias de manutenção preditiva a fim de evitar danos graves, reduzir custos e minimizar o tempo de parada. Em máquinas rotativas, os rolamentos são componentes críticos, sendo frequentemente a principal fonte de falhas mecânicas [1]. Nesse contexto, o uso de técnicas automatizadas

para detecção e diagnóstico de falhas em rolamentos a partir de dados coletados por sensores de vibração tem se mostrado uma ferramenta promissora para assegurar a eficiência operacional e a confiabilidade das máquinas.

Inúmeras técnicas tem sido propostas, desde métodos clássicos físicos [2]–[4] a métodos *data-driven* utilizando aprendizado de máquina [5]. Nos últimos anos, o uso de redes neurais profundas [6]–[8] a partir dos sinais brutos de vibração tem se popularizado nesta área, permitindo alcançar resultados superiores aos métodos de aprendizado de máquina clássicos baseados na extração de atributos [5]. Em particular, as redes neurais convolucionais (*convolutional neural networks*, CNN) representam hoje o estado da arte. Algumas arquiteturas, como a *Deep Convolutional Neural Networks with Wide First-layer Kernels* (WDCNN) [7] e a TICNN [8], as quais utilizam como entrada uma representação 1D de sinais, foram desenvolvidas especialmente para esta tarefa; outra abordagem é utilizar uma arquitetura convencional para reconhecimento de imagens, como a ResNet [9], e alimentá-la com uma representação 2D do sinal, tal como um espectrograma. Embora o aprendizado profundo permita teoricamente dispensar o uso de engenharia atributos, na prática observa-se que a representação do sinal de entrada tem impacto significativo no desempenho do modelo.

Para comparação entre as diversas técnicas propostas, é indispensável o uso de um conjunto de dados público e representativo de condições realistas, sendo o conjunto de falhas de rolamento da *Case Western Reserve University* (CWRU) [10] o maior e mais utilizado na literatura. Tradicionalmente, o CWRU é usado sendo dividido em subconjuntos com base nas cargas de trabalho, supostamente um cenário representativo de condições reais de operação, a fim de testar a robustez dos modelos. Porém, em [11] é apontada uma falha nessa divisão por vazamento de dados de treinamento para o teste e uma nova divisão sem vazamentos é proposta. Não surpreendentemente, nesta nova divisão, modelos de estado da arte que antes alcançavam 95% de acurácia apresentaram um desempenho significativamente inferior (53%).

Além dos resultados inferiores—embora provavelmente muito mais realistas—um novo problema surge com a proposta em [11]. O artigo não descreve a metodologia de otimização de hiperparâmetros utilizada, gerando a suspeita de que pode ter sido realizada no conjunto de teste, provocando viés nos resultados. Por outro lado, se parte do conjunto de treino fosse usada como conjunto de validação, o mesmo vazamento seria observado, comprometendo uma otimização de hiperparâmetros eficaz e robusta.

Rodrigo Kobashikawa Rosa, Danilo Silva, Departamento de Engenharia Elétrica e Eletrônica, UFSC, Florianópolis-SC, e-mails: rodrigo.k@posgrad.ufsc.br, danilo.silva@ufsc.br. Vicente Knobel Borges, Departamento de Automação e Sistemas, UFSC, Florianópolis-SC, e-mail: vicentekborges@gmail.com. Danilo de Souza Braga, Dynamox, Florianópolis-SC, e-mail: danilo@dynamox.net. Este trabalho foi parcialmente financiado pelo CNPq (164299/2021-1).

Neste artigo, propomos uma nova metodologia de divisão dos conjuntos de treino, validação e teste que, por um lado, permite uma otimização de hiperparâmetros mais robusta e, por outro, fornece mais diversidade ao conjunto de treino, permitindo alcançar desempenho superior ao de [11].

Além disso, foram investigadas e otimizadas diversas representações 1D e 2D de sinais como entrada para redes convolucionais. Os melhores resultados foram obtidos por um *ensemble* de duas representações 2D, superando com folga o atual estado da arte.

II. METODOLOGIA

A. Conjunto de dados CWRU

O conjunto de dados [10] consiste em sinais de vibração medidos de um motor elétrico usando dois acelerômetros localizados próximos aos rolamentos do motor (*drive end* e *fan end*). Além de rolamentos em condições normais, falhas nos rolamentos foram introduzidas com diâmetros de 7, 14, 21 e 28 mils nas regiões interna (*inner*), externa (*outer*) e nos elementos rolantes dos rolamentos (*ball*). Em todos os experimentos, foram realizadas medições em quatro condições de carga operacional do motor, variando de 0 a 3 HP. A maioria dos experimentos foram conduzidos com taxa de amostragem de 12 kHz e alguns em 48 kHz, no entanto, apenas os dados em 12 kHz são utilizados no presente trabalho.

B. Descrição dos subconjuntos de treinamento

Os subconjuntos de dados adotados foram organizados conforme ilustrado na Figura 1. Foram utilizadas falhas em duas localizações de acelerômetros para os três tipos de falha e três cargas distintas, totalizando 19 sinais por subconjunto (18 com falhas e 1 sem falha). O único rolamento físico que se repete nos três subconjuntos é o rolamento normal, uma vez que existe apenas um exemplar desse tipo no conjunto de dados. Por essa razão, a divisão dos dados foi realizada atribuindo-se uma carga diferente de normalidade para cada subconjunto.

As classes utilizadas para a classificação pelos modelos de predição foram divididas de acordo com o tipo de falha, podendo ser *inner race* (IR), *outer race* (OR) ou *ball*, e a localização, podendo ser *drive end* (DE) ou *fan end* (FE), além da classe de normalidade. Portanto, há um total de sete classes: DE IR, DE OR, DE BALL, FE IR, FE OR, FE BALL e Normal. Essa divisão das classes, proposta em [11], também difere da metodologia tradicional adotada em outros artigos, na qual existem dez classes para uma única localização do acelerômetro, com distinção de classe baseada no tamanho da falha. Na divisão adotada neste artigo, o foco é identificar qual é a falha e de qual rolamento ela provém.

C. Nova proposta de divisão de conjuntos de treino e teste

Em [11], Hendriks *et al.* argumentam que, ao contrário da abordagem tradicionalmente utilizada, a divisão do conjunto de dados em treino e teste deve ser feita por rolamentos físicos, de forma que não apareçam no conjunto de teste sinais de um mesmo rolamento que apareça no conjunto de treinamento,

uma vez que esta situação não seria realista.¹ Assim, Hendriks *et al.* propõe uma metodologia de avaliação em que cada um dos conjuntos D, E e F é treinado individualmente e testados nos outros dois.

Visando proporcionar uma maior diversidade ao conjunto de treinamento, mas mantendo a abordagem de utilizar rolamentos distintos para treinamento e teste, estamos propondo uma modificação que combina dois subconjuntos para o treinamento e utiliza o subconjunto restante para o teste. Essa abordagem implica em reduzir pela metade os testes entre diferentes domínios, tornando a metodologia de avaliação idêntica a uma validação cruzada K -Fold com $K = 3$. No entanto, a expectativa é que haja um ganho em relação à adaptação ao tamanho da falha, devido ao aumento da diversidade presente no conjunto de treinamento, o que deverá contribuir para uma melhor generalização dos modelos.

D. Metodologia de otimização de hiperparâmetros

Para a busca de hiperparâmetros, é necessária a divisão em conjuntos de treinamento, validação e teste. Para separarmos um conjunto de validação e realizar a otimização, foi utilizada uma estratégia de busca fazendo validação cruzada K -Fold com $K = 2$ para os subconjuntos D, E e F, após isolar o conjunto de teste. Note que esta estratégia corresponde ao laço interno de um processo de validação cruzada aninhada [12]. Desta forma, para encontrar os melhores hiperparâmetros de cada modelo, foi analisada a curva de validação média do treinamento individual de cada um dos subconjuntos. Por exemplo, para a escolha dos hiperparâmetros do treinamento do subconjunto DE e teste no subconjunto F, é feito o treinamento em D, validando em E e logo após, o treinamento em E e validação em D. A partir das curvas de validação, é feita a média de ambas para avaliar quais curvas desempenharam melhor.

Para cada combinação dos *folds* de treinamento, validação e teste, foram explorados 8 valores distintos para o ajuste das taxas de aprendizado do otimizador, variando de 10^{-6} a 10^{-2} . Além disso, foram executadas 5 iterações para obter a média de desempenho e o desvio padrão entre diferentes realizações, empregando uma inicialização aleatória dos pesos do modelo. Como resultado, foram realizados 120 treinamentos do modelo para cada tipo de entrada distinta.

E. Técnicas de representação de sinais

Foram utilizadas representações 1D e 2D como entradas dos modelos, baseando-se nas técnicas utilizadas na literatura para a detecção e diagnóstico de falhas.

1) *Representações 1D*: Para as representações 1D, foi utilizado o próprio sinal bruto no tempo, o espectro do sinal através da transformada discreta de Fourier e o cepstrum de potência [4].

O cepstrum de potência é uma técnica que envolve o cálculo do espectro de potência do logaritmo do espectro do sinal

¹Uma analogia pode ser feita com um problema de diagnóstico médico, no qual, para evitar vazamento de dados, é necessário fazer a divisão do conjunto de dados de forma que todos os exames pertencentes ao mesmo paciente se encontrem apenas em um dos subconjuntos de treino ou teste.

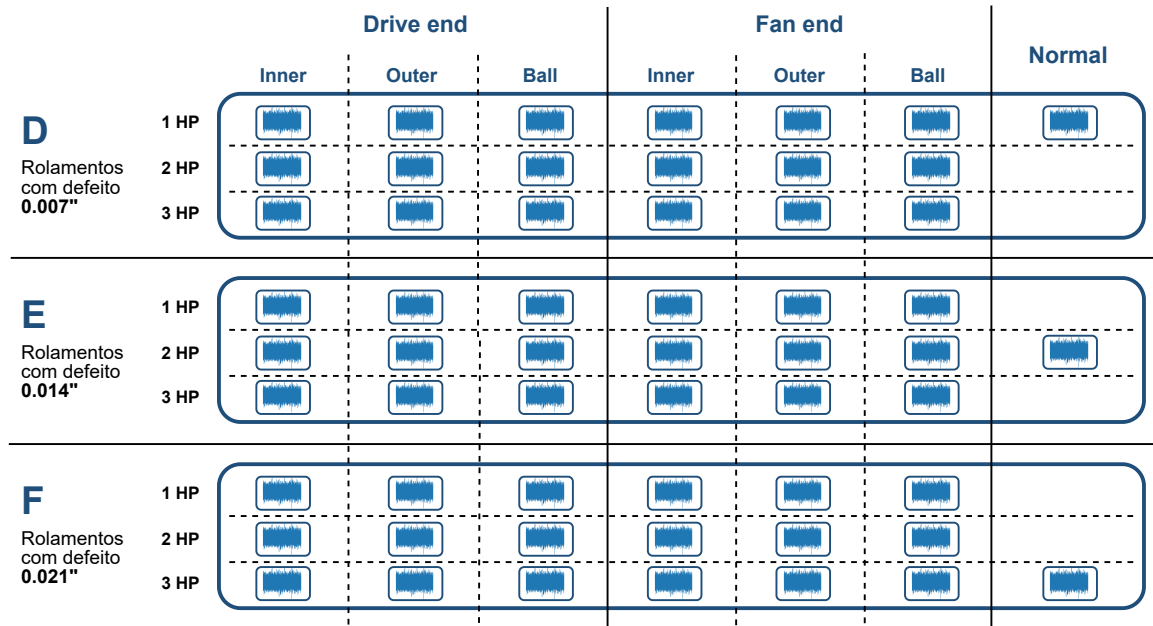


Fig. 1: Organização dos dados em subconjuntos.

original e é um método útil para detectar choques periódicos no sinal, como é o caso muitas vezes nas falhas de rolamentos [4]. Um problema da técnica é que ela produz muito picos próximos ao ponto zero, dificultando a interpretação do sinal.

2) *Representações 2D*: Para as representações 2D, usamos o espectrograma do sinal processado pela *Short-Time Fourier Transform* (STFT) [2] e o escalograma processado pela transformada de *wavelet* [13], fazendo o uso da *wavelet* FBSP para a convolução. Além disso, também foram experimentadas duas representações pouco utilizadas na literatura de predição de falhas, o cepstrograma (utilizado originalmente para análise de fala [14]) e o espectrograma de modulação (também utilizado para aplicações com sinais biomédicos [15], [16]).

O cepstrograma é uma representação em tempo x quefrência do cepstrum enquanto que o espectrograma de modulação é uma representação em frequência de modulação x frequência, o que torna interessante ao considerar que as falhas de rolamento são sinais modulados.

F. Pré-processamento dos sinais

A etapa de segmentação e pré-processamento seguiu os passos descritos em [11] e consistiu na divisão do sinal de 10 segundos (aproximadamente 120.000 pontos de medição, amostrados a 12 kHz) de cada experimento em segmentos menores, com 2048 pontos para os dados no domínio do tempo, 4096 pontos para os dados de espectro e cepstrum e 11500 pontos para as representações 2D. Para aumentar a quantidade de amostras de treinamento, foi aplicado um *overlap* de 97% entre os segmentos.

G. Redes convolucionais utilizadas

Duas redes convolucionais do estado da arte, comumente citadas em artigos, foram adotadas para nossos experimentos.

Para os dados 1D, usamos a WDCNN [7] que foi desenvolvida com o objetivo de possibilitar o funcionamento direto com dados brutos e apresentar bom desempenho de mudança de domínio. Sua estratégia consiste em utilizar uma camada convolucional com *kernels* largos para extrair características, seguida por várias camadas com *kernels* menores.

Para os dados 2D, usamos a ResNet18 com pesos inicializados aleatoriamente. A ResNet [9] é uma arquitetura famosa de redes convolucionais profundas que alcançou desempenho de ponta em tarefas de visão computacional e inspirou o desenvolvimento de outras arquiteturas baseadas em conexões residuais, sendo aplicada em diversas áreas.

Para garantir uma comparação justa com a WDCNN, cuja entrada é 1D, algumas modificações foram feitas no modelo ResNet. Primeiramente, optou-se por utilizar apenas um canal do acelerômetro em vez dos dois disponíveis, a fim de manter a consistência. Isso significa que todos os experimentos foram conduzidos utilizando apenas o acelerômetro do DE para classificar as 7 classes de falhas.

Ambos os modelos foram treinados com o otimizador Adam e a taxa de aprendizado (TA) utilizada foi escolhida a partir da otimização de hiperparâmetros.

H. Otimização dos hiperparâmetros de representação

Para cada representação 2D, foram realizadas otimizações nos parâmetros para encontrar os melhores tamanhos de segmentos da STFT e o *overlap* entre os segmentos. Portanto, configurações diferentes foram utilizadas para os espectrogramas nos nossos experimentos em comparação com os em [11], que utilizou um tamanho de janela e *overlap* capazes de gerar imagens com a mesma largura e altura das imagens originalmente utilizadas pela ResNet (224x224 pixels).

A partir da otimização dos parâmetros do espectrograma foi utilizado 192 pontos para o tamanho de janela e 104

TABELA I: Resultados dos experimentos usando a divisão proposta em [11].

Modelo	Entrada	TA	Épocas	Acurácia e Desvio Padrão (%)						MÉDIA
				D→E	D→F	E→D	E→F	F→D	F→E	
WDCNN (em [11])	Tempo	5E-4	4	29,7	51,5	21,5	32,5	44,9	37,1	36,2
	Espectro	5E-4	4	38,1	51,2	57,7	54,1	46,8	52,2	50,0
ResNet18 (em [11])	Espectrograma	5E-4	4	53,2	62,6	55,9	52,3	53,2	41,4	53,1
WDCNN (Nosso)	Tempo	1E-4	4	15,5±3,0	42,5±6,0	28,6±3,0	30,7±6,4	50,7±8,6	35,8±6,8	34,0±12,1
	Espectro	5E-3	4	22,0±3,9	27,5±5,3	33,4±4,6	33,2±4,9	46,8±7,3	24,4±5,0	31,2±8,9
	Cepstrum de potência	1E-3	4	38,1±6,7	37,0±8,0	53,9±10,5	55,1±6,7	36,6±14,5	49,5±5,7	45,0±8,8
ResNet18 (Nosso)	Espectrograma	5E-4	5	30,9±12,2	39,6±6,2	39,0±2,9	48,6±4,2	49,3±8,1	43,7±7,7	41,9±6,9
	Wavelet Contínua	1E-4	6	16,4±1,0	44,4±5,6	20,0±11,9	43,0±4,2	58,6±11,9	38,7±16,0	36,8±16,0
	Cepstrograma	1E-3	6	47,0±12,3	62,0±6,6	61,0±10,0	63,4±18,0	62,3±6,0	52,1±10,0	57,9±6,8
	Espectrograma de modulação	1E-6	4	56,0±3,7	56,7±10,3	53,8±7,0	70,6±2,9	54,2±4,8	49,6±4,2	56,8±7,2

pontos de *overlap*, resultando em uma imagem de 96x129. Para o cepstrograma, o tamanho de janela utilizado foi de 1024 pontos e com 940 pontos de *overlap*, resultando em uma imagem de 256x125. E finalmente, para o espectrograma de modulação, o tamanho de janela utilizado foi de 256 pontos com 176 pontos de *overlap*, resultando em uma imagem de 129x71.

Os escalogramas obtidos pela transformada de *wavelet* são dados por 200 frequências interpolando de 6000 a 10 Hz em uma curva exponencial inversa. A imagem de 11500x200 é então redimensionada para 224x224, para ser agilmente manipulada pelo algoritmo de treinamento e fazer melhor uso dos *kernels* de convolução quadrados da ResNet.

III. RESULTADOS

Inicialmente, foram reproduzidos os experimentos descritos em [11]. No entanto, observa-se na Tabela I que, para alguns tipos de entrada, não foi possível obter o mesmo desempenho de acurácia apresentado no artigo. No caso do modelo de dados 1D usando a WDCNN, os resultados com entrada no domínio do tempo apresentaram uma acurácia média muito próxima à mencionada no artigo, mas os resultados utilizando o espectro não conseguiram alcançar o mesmo desempenho. O mesmo pode ser observado para a ResNet utilizando espectrogramas, porém é importante ressaltar que neste caso, nossos resultados inferiores são coerentes devido às alterações feitas no treinamento, como utilizar apenas um canal de entrada e não ter informações detalhadas sobre a versão específica da ResNet utilizada.

Os experimentos com o cepstrum de potência, cepstrograma e espectrograma de modulação foram os que apresentaram os melhores resultados. Obtendo médias de acurácia superiores àquelas demonstradas no artigo em [11].

Na Tabela I, são apresentados os resultados da otimização da taxa de aprendizado e a quantidade de épocas, que obtiveram o melhor desempenho para cada tipo de entrada diferente.

A. Resultados da nova proposta de divisão

Seguindo a metodologia de otimização de hiperparâmetros em que separamos um dos experimentos em conjuntos de

treinamento e validação, consideramos que cada divisão de treinamento e teste do modelo pode ser tratada de forma independente. Dessa forma, ao termos um conjunto de validação para cada divisão, podemos otimizar a taxa de aprendizado e a quantidade de épocas para cada conjunto. As taxas de aprendizado e a quantidade de épocas podem ser vistas na Tabela II.

TABELA II: Resultados da otimização de hiperparâmetros.

Entrada	EF → D		DF → E		DE → F	
	TA	Época	TA	Época	TA	Época
Tempo	1E-3	2	1E-4	3	1E-5	6
Espectro	5E-3	3	1E-3	3	5E-3	2
Cepstrum de potência	1E-3	3	5E-3	2	1E-3	3
Espectrograma	5E-4	5	1E-3	6	1E-3	4
Wavelet Contínua	1E-3	6	1E-5	8	1E-3	10
Cepstrograma	5E-4	4	1E-3	3	1E-3	6
Espectrograma de modulação	5E-5	4	1E-6	3	1E-6	3

A proposta de combinar os conjuntos de dados para o treinamento, aumentando a diversidade de falhas, resultou em uma melhora no desempenho médio de todos os experimentos, conforme pode ser observado nas Tabelas III e IV. É importante observar na Tabela III, que o desempenho do cepstrum de potência foi o melhor entre as entradas 1D, utilizando o modelo WDCNN, com uma acurácia média de 57,34%. Já o cepstrograma e o espectrograma de modulação apresentaram os melhores resultados entre as entradas 2D, utilizando o modelo ResNet treinado do zero, com acurácias médias próximas a 70%.

Após a realização dos experimentos com as diferentes entradas, um último experimento foi conduzido com o objetivo de realizar um *ensemble*, com estratégia *soft-voting*, dos modelos de cepstrograma e espectrograma de modulação, os quais apresentaram as maiores acurácias médias. Os resultados desse experimento estão apresentados na Tabela IV como entrada *Ensemble*, revelando um aumento de aproximadamente 3% na acurácia média em relação ao melhor modelo anteriormente obtido.

TABELA III: Resultados dos experimentos usando a nova divisão proposta com o WDCNN.

Entrada	Acurácia e Desvio Padrão (%)			
	EF → D	DF → E	DE → F	MÉDIA
Tempo	48,4±5,6	35,1±6,3	48,3±4,3	43,9±7,7
Espectro	50,2±5,9	26,9±4,5	62,8± 6,0	46,7±18,2
Cepstrum de potência	59,5±7,4	55,9±5,0	56,6±4,7	57,3±1,9

TABELA IV: Resultados dos experimentos usando a nova divisão proposta com a ResNet18.

Entrada	Acurácia e Desvio Padrão (%)			
	EF → D	DF → E	DE → F	MÉDIA
Espectrograma	65,3±7,7	34,5±7,1	62,1±9,1	53,9±16,9
Wavelet Contínua	43,5±16,9	19,3±2,3	59,8±12,1	40,9±20,4
Cepstrograma	63,9±12,3	66,1±3,0	74,2±6,1	68,1±5,4
Espectrograma de modulação	56,4±1,7	60,3±5,1	79,1±4,4	65,2±12,1
Ensemble	72,6±2,4	69,2±1,6	73,9±5,1	71,9±2,4

B. Teste de ganho de diversidade

Outro experimento foi conduzido para verificar se o aumento no desempenho dos modelos foi de fato causado pelo aumento da diversidade e não simplesmente pelo aumento no número de amostras de treinamento. O experimento consistiu em utilizar o modelo que apresentou o melhor desempenho médio de acurácia, que foi o cepstrograma com a ResNet, dividindo os sinais de 10 segundos dos experimentos do CWRU pela metade, resultando em segmentos de 5 segundos. Dessa forma, foi possível utilizar o mesmo número de amostras de treinamento da proposta de divisão em [11], porém com a nova abordagem com maior diversidade de tamanhos de falha.

Os resultados na Tabela V mostram que o desempenho médio do treinamento com metade dos sinais foi próximo ao do treinamento com sinais completos, apresentando uma diferença de aproximadamente 2,5% a menos na média de acurácia. Essa diferença ainda é consideravelmente superior ao desempenho observado na proposta de divisão em [11]. Portanto, com base nestes resultados, fica evidente que há um ganho de desempenho real ao aumentar a diversidade de tamanhos de falha por meio da combinação de conjuntos de dados no treinamento.

TABELA V: Resultados dos experimentos testando o ganho de diversidade.

Entrada	Acurácia e Desvio Padrão (%)			
	EF → D	DF → E	DE → F	MÉDIA
Cepstrograma	63,9±12,3	66,1±3,0	74,2±6,1	68,1±5,4
Cepstrograma (metade do sinal)	60,8±10,8	65,2±6,5	70,7±5,9	65,6±5,0

IV. CONCLUSÕES

Os resultados obtidos indicam que as mudanças no pré-processamento dos dados de entrada para os modelos de

previsão de falhas de rolamento e uma nova metodologia de otimização de hiperparâmetros resultaram em um aumento próximo de 20% na acurácia média em relação ao desempenho obtido em relação à maior acurácia relatada em [11].

Ao aprimorar a metodologia de avaliação, adicionando repetições dos experimentos para maior diversidade estatística e propondo uma metodologia de separação de um conjunto de validação para a nova proposta de divisão de dados, tornou-se possível realizar uma avaliação mais robusta de novos pré-processamentos de dados e arquiteturas de modelos para a tarefa de predição de classes de falha de rolamentos.

As descobertas deste estudo são importantes para o desenvolvimento de conjuntos de dados mais robustos e realistas, visando melhorar continuamente os sistemas de predição de falhas de máquinas, um tópico de grande relevância na pesquisa.

REFERÊNCIAS

- [1] J. Zarei, M. A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," *Mechatronics*, vol. 24, no. 2, pp. 151–157, 2014.
- [2] R. B. Randall, *Vibration-based condition monitoring: industrial, automotive and aerospace applications*. John Wiley & Sons, 2021.
- [3] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical systems and signal processing*, vol. 64, pp. 100–131, 2015.
- [4] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *Journal of Failure Analysis and Prevention*, vol. 16, no. 2, pp. 271–284, 2016.
- [5] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.
- [6] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical systems and signal processing*, vol. 72, pp. 303–315, 2016.
- [7] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [8] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical systems and signal processing*, vol. 100, pp. 439–453, 2018.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [10] "Case western reserve university bearing data center website." <https://engineering.case.edu/bearingdatacenter>. Acessado em: 10 de Maio de 2023.
- [11] J. Hendriks, P. Dumond, and D. Knox, "Towards better benchmarking using the CWRU bearing fault dataset," *Mechanical Systems and Signal Processing*, vol. 169, p. 108732, 2022.
- [12] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv preprint arXiv:1811.12808*, 2018.
- [13] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal processing*, vol. 96, pp. 1–15, 2014.
- [14] C. Che and Q. Lin, "Cepstrogram and its application to speech analysis," *The Journal of the Acoustical Society of America*, vol. 96, no. 5, pp. 3351–3351, 1994.
- [15] R. Cassani and T. H. Falk, "Spectrotemporal modeling of biomedical signals: Theoretical foundation and applications," *Reference Module in Biomedical Sciences*, 2018.
- [16] R. Cassani, I. Albuquerque, J. Monteiro, and T. H. Falk, "AMA: An Open-source Amplitude Modulation Analysis Toolkit for Signal Processing Applications," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1–4, 2019.