

Benchmarking domain adaptation techniques for bearing fault diagnosis

Victor Afonso Bauler¹, Júlio A. Cordioli¹, Danilo Braga³, Danilo Silva²

¹ Departamento de Engenharia Mecânica
Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 5064 – 88.035-972 – Florianópolis – SC – Brazil

² Departamento de Engenharia Elétrica e Eletrônica
Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 5064 – 88.035-972 – Florianópolis – SC – Brazil

³ Dynamox SA
Florianópolis – SC – Brazil

victor.bauler@gmail.com

julio.cordioli@ufsc.br

danilo@dynamox.net

danilo.silva@ufsc.br

Abstract. *Technological advancements have enabled the simultaneous monitoring of thousands of machines, making it possible to train machine learning algorithms for automated fault detection and diagnosis. However, one of the main challenges has been the lack of labeled data, as well as the fact that the same type of machine can be exposed to different operating conditions, altering the distribution of the acquired signal attributes. A potential solution is to use domain adaptation techniques, which can improve model performance on new types of machines, even without labeled data. However, many of the studies presented in the literature face data leakage issues, which make it impossible to compare techniques since the models performance is unrealistic. Therefore, this paper proposes a methodology for comparing domain adaptation techniques that avoids this problem. Eight domain adaptation techniques were implemented and evaluated on two datasets, resulting in significant improvements in one of the evaluation scenarios.*

Resumo. *O avanço tecnológico permitiu o monitoramento simultâneo de milhares de máquinas, possibilitando o treinamento de algoritmos de aprendizado de máquina para realizar a detecção e o diagnóstico de falhas de forma automatizada. Contudo, um dos principais desafios tem sido a falta de dados rotulados, além do fato de que um mesmo tipo de máquina pode ser exposto a diferentes condições de operação, alterando a distribuição dos atributos dos sinais adquiridos. Uma possível solução é utilizar técnicas de adaptação de domínio, que podem melhorar o desempenho dos modelos em novos tipos de máquina, mesmo sem dados rotulados. No entanto, muitos dos artigos apresentados na literatura enfrentam problemas de vazamento de dados, o que impos-*

sibilita a comparação entre técnicas, pois o desempenho dos modelos é irrealista. Diante disso, este artigo propõe uma metodologia para comparar técnicas de adaptação de domínio, evitando esse problema. Foram implementadas oito técnicas de adaptação de domínio e avaliadas em dois conjuntos de dados, resultando em ganhos significativos em um dos cenários de avaliação.

1. Introdução

A manutenção preditiva é uma estratégia de manutenção baseada em dados que visa antecipar potenciais problemas em ativos industriais, permitindo estabelecer o momento ótimo para manutenção, evitando paradas não programadas e custos desnecessários com manutenção preventiva periódica. O monitoramento de vibrações tem se destacado como uma das técnicas mais efetivas [Randall 2021], e com o advento da indústria 4.0, tem sido possível monitorar múltiplas máquinas simultaneamente, com o uso de sensores IoT e computação em nuvem. Nesse contexto, diversas técnicas para detecção automatizada de falhas em ativos industriais baseadas em dados de vibração têm sido propostas na literatura [Lei et al. 2020]. Algumas dessas técnicas foram propostas especialmente para a detecção de falhas de rolamentos [Hakim et al. 2023], que são componentes essenciais em máquinas rotativas e principal fonte de falhas mecânicas [Zhang et al. 2020].

Um dos principais desafios enfrentados é a falta de dados rotulados, fazendo com que a maioria dos estudos se concentre em bancadas experimentais de simulação de falhas [Lei et al. 2020]. Recentemente, muitos trabalhos têm dado enfoque em tentar transferir o conhecimento obtido em uma determinada condição de operação para outra [Zhang et al. 2023], utilizando técnicas de adaptação de domínio para lidar com a diferença de distribuição dos dados. No entanto, muitos dos trabalhos apresentados na literatura possuem problemas metodológicos, associados ao vazamento de dados [Kapoor and Narayanan 2023].

O vazamento de dados é um problema porque ele faz com que os modelos tenham um desempenho irrealista, tornando a comparação entre técnicas inviável. O trabalho de [Rosa et al. 2024] propõe uma metodologia para avaliar modelos no conjunto de dados da *Case Western Reserve University (CWRU¹)*, que aborda e corrige esse problema. No entanto, essa metodologia não contempla o cenário de mudança de domínio, e muitos estudos sobre adaptação de domínio ainda utilizam divisões baseadas na carga de operação [Zhang et al. 2023], uma abordagem que foi apontada como problemática por [Hendriks et al. 2022].

Este trabalho visa apresentar uma metodologia para avaliação de técnicas de adaptação de domínio de forma a evitar o vazamento de dados, garantindo assim uma avaliação mais justa de diferentes técnicas. As técnicas serão comparadas em duas condições de adaptação, primeiro com apenas uma mudança de componente dentro de uma mesma bancada, e depois com uma mudança entre bancadas e condições de operação. Para a primeira condição foi utilizado o conjunto de dados do CWRU, atualmente um dos conjuntos mais utilizados na literatura [Neupane and Seok 2020]. Para a segunda condição foi utilizado o conjunto de dados obtido pelo Laboratório de Vibração e Acústica (LVA) da Universidade Federal de Santa Catarina (UFSC).

¹ Acesso em: <https://engineering.case.edu/bearingdatacenter>

Ao total 8 técnicas de adaptação de domínio não supervisionado serão avaliadas em um problema de classificação multirrotulo com 3 classes. O objetivo das técnicas é usar dados rotulados do domínio fonte e dados não rotulados do domínio alvo para tentar aprender uma representação que obtenha um melhor desempenho no domínio alvo. Dentre elas, 3 são consideradas técnicas baseadas em alinhamento de atributos, 4 são baseadas em redes adversariais e 1 consiste em apenas uma mudança nas camadas de *Batch Normalization*. A ResNet18 será utilizada como arquitetura base e a representação de entrada será o espectrograma, seguindo os resultados apresentados por [Rosa et al. 2024].

2. Metodologia

2.1. Conjuntos de Dados

O conjunto de dados do CWRU consiste de sinais de vibração medidos de um motor elétrico usando dois acelerômetros localizados próximos aos rolamentos do motor, no lado do acionamento (LA) e no lado oposto ao acionamento (LOA). 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, com a rotação do motor variando de 1720 a 1797 RPM (rotações por minuto). A maioria dos experimentos foram conduzidos com taxa de amostragem de 12 kHz e alguns em 48 kHz, que foram reamostrados para 12 kHz, dessa forma, todos os sinais analisados neste trabalho possuem uma taxa de amostragem de 12 kHz. Todos os experimentos tiveram um tempo de aquisição de aproximadamente 10 segundos.

Com base no trabalho de [Rosa et al. 2024], serão apenas consideradas cargas que vão de 1 a 3 HP e tamanhos de falhas de 7, 14 e 21 mils. Quanto à posição da falha de trilha externa, foram utilizados preferencialmente os experimentos centrados ”@6:00”, quando estes não estavam disponíveis foram utilizados experimentos com posição ortogonal ”@3:00”. Além disso, todos os sinais utilizados neste trabalho são adquiridos no local da falha, ou seja, se a falha for em LA o sinal analisado também será coletado de LA. Os casos de normalidade, que serão considerados como casos em que ambos os rolamentos são normais, serão adicionados apenas no conjunto de teste. A Figura 1 demonstra as diferentes configurações do conjunto de dados do CWRU utilizadas neste trabalho.

Diferente do CWRU, o conjunto de dados do LVA, analisa falhas de rolamentos em 2 mancais, utilizando uma bancada de simulação de falhas fabricada pela empresa americana SpectraQuest, representada na Figura 2, especificando os mancais LA e LOA. Durante os experimentos foram utilizados 4 rolamentos defeituosos, com falhas em trilha interna, trilha externa, elemento rolante e combinada (todas as 3 falhas anteriores) e 2 rolamentos saudáveis.

Foram realizadas 5 configurações de experimentos, as quais estão descritas na Tabela 1, em todas as configurações foram realizadas medições com velocidades de 100, 1000, 2000, 3000, 4000 e 5000 RPM, exceto na configuração E, em que foram realizadas medições com 500, 1000, 2000, 3000 e 3500 RPM. A configuração D diz respeito a medições com ambos os rolamentos saudáveis e com falhas apenas no rolamento no mancal LOA, utilizando 3 níveis de desbalanceamento, na configuração E foi feita a mesma

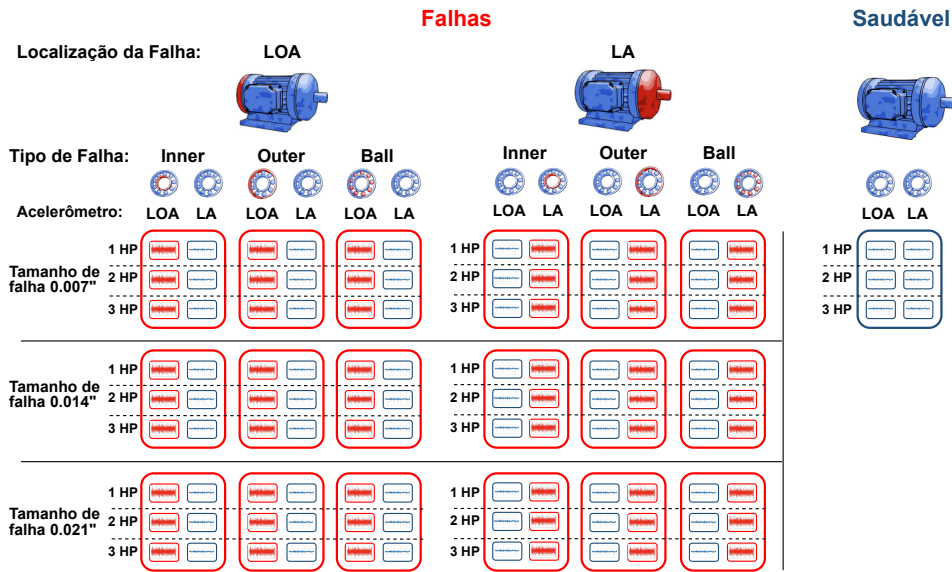


Figura 1. Divisão das aquisições do CWRU por localização, tipo e tamanho de falha e cargas distintas, em que os sinais em azul representam normalidade, enquanto os em vermelho, falha. Adaptado de [Rosa et al. 2024].

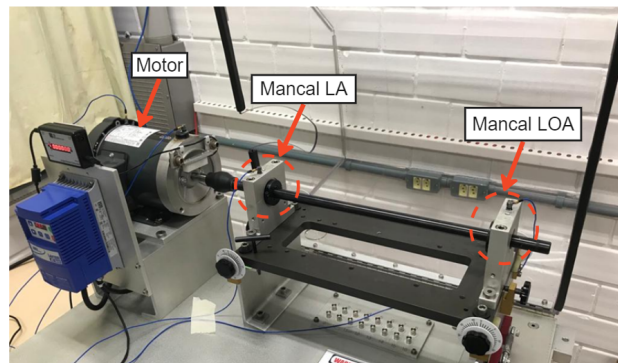


Figura 2. Bancada de simulação de falhas do LVA.

coisa, mas, ao invés de usar níveis de desbalanceamento, foi adicionada uma carga por meio de um freio magnético.

Tabela 1. Classes de configurações

Configuração	Descrição	Número de medições
A	Configuração saudável	6
B	Defeitos de rolamento no mancal LOA	24
C	Defeitos de rolamento no mancal LA	24
D	Configurações A e B com 3 níveis de desbalanceamento	90
E	Configurações A e B adicionando carga	25

Todos os sinais do conjunto de dados do LVA foram adquiridos a uma frequência de amostragem de 25,6 kHz durante 30 segundos, para as análises realizadas neste estudo, os sinais foram reamostrados para 12 kHz e cortados para ficar com uma duração de 10 segundos, de forma similar aos dados do CWRU. Além disso, também foram apenas con-

sideradas condições em que o local da falha era igual ao local em que o sinal foi coletado e os sinais de normalidade foram considerados tanto no treinamento quanto no teste, mais detalhes sobre os experimentos podem ser encontrados em [Bauler et al. 2023].

2.2. Divisão dos dados

A divisão proposta por [Rosa et al. 2024] garante máxima diversidade no CWRU, de forma a evitar vazamento de dados, no entanto, nem sempre ela é possível num contexto industrial, em que rótulos de novos tipos de máquinas são escassos. Outros trabalhos que abordaram a temática de adaptação de domínio no CWRU propuseram divisões por carga de trabalho [Wang et al. 2019, Jiao et al. 2020, Li et al. 2024], o que sofre de vazamento de informação, conforme discutido por [Hendriks et al. 2022]. Portanto, este trabalho propõe uma divisão por localização no CWRU, o qual será considerado o primeiro cenário de adaptação de domínio. Os rolamentos presentes em LA e LOA possuem dimensões distintas, dessa forma, essa divisão consegue simular uma mudança de distribuição devido a uma alteração no componente de uma mesma máquina.

Na sequência será avaliado um segundo cenário, em que o treinamento será realizado no conjunto de dados do LVA e o teste no CWRU e vice-versa. Nos casos de treinamento e teste entre bancadas, apenas o LOA do CWRU será utilizado, para facilitar a comparação com os resultados obtidos treinando com o LA do CWRU. Como o LVA possui condições de operação e rolamentos bem distintos do CWRU, podemos considerar este segundo cenário como sendo um caso mais desafiador.

Sendo assim, serão realizadas 4 divisões distintas para avaliações dos métodos de adaptação de domínio, sendo elas: i) Treinamento no LA (CWRU) e teste no LOA (CWRU); ii) Treinamento no LOA (CWRU) e teste no LA (CWRU); iii) Treinamento no LOA (CWRU) e teste no LVA; iv) Treinamento no LVA e teste no LOA (CWRU).

2.3. Metodologia de seleção e avaliação de modelos

Para otimização de hiperparâmetros, foi utilizada uma divisão de validação *holdout* composta pelos sinais de tamanho de falha de 21 mils do LA do CWRU. Para o treinamento foram utilizados os tamanhos de falha de 7 e 14 mils também do LA. Os hiperparâmetros otimizados neste trabalho foram apenas a taxa de aprendizado e tamanho do lote. Sendo que a taxa de aprendizado do classificador é 10 vezes maior que a taxa de aprendizado do extrator de atributos, de acordo com [Ganin et al. 2016]. Foram testados 3 valores distintos da taxa de aprendizado do classificador, variando de 10^{-6} a 10^{-4} e 3 valores de tamanho de lote, sendo eles: 64, 128 e 256.

Os experimentos foram realizados utilizando Python, com a biblioteca PyTorch, com uma máquina equipada com uma placa de vídeo RTX 3090. Foram executadas 5 iterações para obter a média do desempenho de cada variação, em que a melhor combinação de hiperparâmetros foi com taxa de aprendizado de 10^{-4} , tamanho de lote de 64 e o otimizador Adam [Kingma and Ba 2014].

A arquitetura, as etapas de pré-processamento e a métrica de avaliação adotadas neste trabalho seguem as mesmas descritas em [Rosa et al. 2024]. Foram utilizados segmentos de 11.500 amostras com 97% de sobreposição, e o espectrograma foi a representação de entrada utilizada. A arquitetura empregada foi a ResNet18, uma variação da ResNet [He et al. 2016], com pesos inicializados do pré-treinamento no conjunto de

dados da ImageNet. Como métrica final, buscou-se maximizar a média macro da AUROC (*Area Under ROC Curve*).

2.4. Técnicas de adaptação de domínio

A principal motivação para a utilização de técnicas de adaptação de domínio no cenário de detecção e diagnóstico de falhas está relacionada com o fato de que dados rotulados são bem escassos e uma mesma máquina pode ser exposta a diversas condições de trabalho, o que afetaria a distribuição de atributos do sinal adquirido. Dessa forma, podemos usar as técnicas de adaptação de domínio não supervisionado para tentar aprender com dados rotulados de um tipo de máquina, chamado domínio fonte, e usar dados não rotulados de outro tipo de máquina, chamado domínio alvo, para tentar melhorar o desempenho do modelo.

Comparamos técnicas do estado da arte de adaptação de domínio não supervisionado, propostas inicialmente no contexto de visão computacional, sendo elas: *Deep Adaptation Network* (DAN) [Long et al. 2015], *Domain Adversarial Neural Network* (DANN) [Ganin et al. 2016], *Deep CORAL* (D-CORAL) [Sun and Saenko 2016], *Adversarial Discriminative Domain Adaptation* (ADDA) [Tzeng et al. 2017], *Joint Adaptation Network* (JAN) [Long et al. 2017], *Conditional Domain Adversarial Network* (CDAN) [Long et al. 2018], *Adaptive Batch Normalization* (AdaBN) [Li et al. 2018] e *Margin Disparity Discrepancy* (MDD) [Zhang et al. 2019].

Para uma comparação justa, utilizou-se a mesma arquitetura base para todas as técnicas, conforme descrito na Seção 2.3, a qual será referenciada como nosso extrator de atributos. Os atributos são extraídos da camada *flatten*, antes da camada final totalmente conectada. Como classificador, foi utilizada uma camada totalmente conectada com 1024 unidades e 3 saídas, correspondendo às 3 classes a serem preditas no problema multirrótulo.

Dessa forma, para as técnicas que visam alinhar as distribuições dos atributos (DAN, JAN, D-CORAL), a saída do extrator de atributos é usada para o cálculo da perda de cada técnica, juntamente com a perda do modelo de classificação, conforme ilustrado na Figura 3.

Já para os casos adversariais (DANN, ADDA, CDAN, MDD), será utilizado um discriminador de domínio, como uma segunda saída do modelo, obtido a partir do extrator de atributos, conforme representado na Figura 4. Com exceção da técnica MDD, que substitui o discriminador de domínio por um classificador auxiliar, com a mesma arquitetura do classificador principal (uma camada totalmente conectada com 1024 unidades e 3 saídas). O discriminador de domínio utilizado neste trabalho é composto por três camadas totalmente conectadas, conforme descrito em [Ganin et al. 2016], as primeiras duas camadas têm 1024 unidades com ativação ReLu, enquanto a última camada tem saída binária indicando se o sinal pertence ao domínio fonte ou ao domínio alvo.

3. Resultados

Os resultados obtidos para o cenário de adaptação mudando apenas o componente de uma mesma bancada estão apresentados na Seção 3.1, e os resultados de adaptação entre bancadas na Seção 3.2. Todos os experimentos foram realizados 5 vezes, e a métrica de avaliação será a média macro da AUROC.

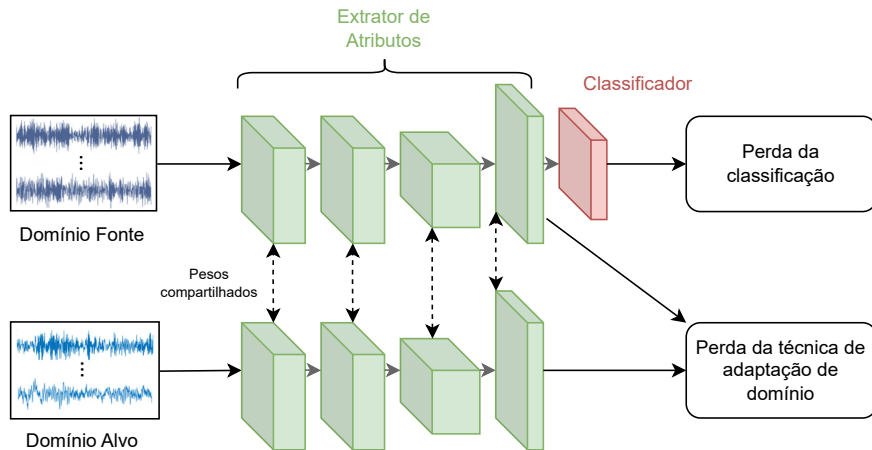


Figura 3. Exemplo de arquitetura das técnicas que visam alinhar as distribuições dos atributos.

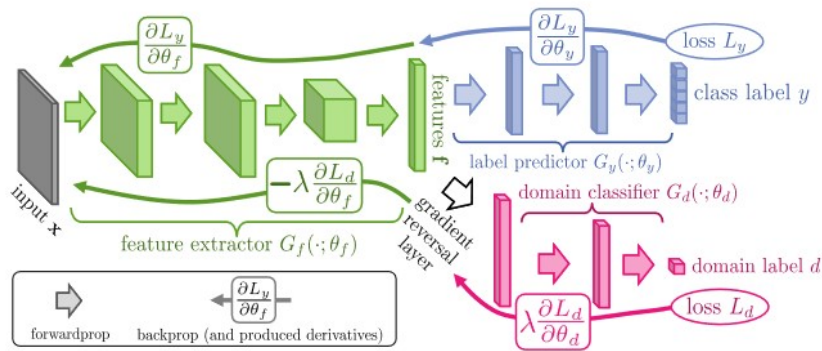


Figura 4. Exemplo de arquitetura adversarial [Ganin et al. 2016].

3.1. Mudança de componente

Conforme descrito na Seção 2.2, no conjunto de dados do CWRU, o rolamento localizado no LA apresenta dimensões diferentes do rolamento em LOA. Essa diferença dimensional implica que as frequências excitadas por uma mesma falha em ambos os rolamentos sejam distintas, o que contribui para uma mudança na distribuição entre os dados de treinamento e teste. Na Tabela 2 podemos observar que um modelo treinado apenas no domínio fonte possui um resultado pior do que os modelos treinados utilizando técnicas de adaptação de domínio. Dentre as 3 principais técnicas, duas têm como princípio aproximar as distribuições dos atributos, utilizando como perda funções que medem essa distância (D-CORAL, DAN), e uma das técnicas (ADDA) possui abordagem adversarial, em que tenta-se obter uma representação que minimize o desempenho de um discriminador de domínio.

É possível observar que os resultados obtidos realizando o treinamento no LA costumam ser mais estáveis (em termos de possuir um menor desvio padrão) do que ao realizar o treinamento no LOA. Uma análise realizada por [Smith and Randall 2015] no CWRU demonstra que alguns casos de falha de trilha interna e trilha externa em LA, assim como alguns sinais de elemento rolante em ambas as posições, possuem carac-

Tabela 2. $100 \times$ AUROC Macro (média \pm desvio padrão) das técnicas avaliadas no CWRU

Método	LA (CWRU) \rightarrow LOA (CWRU)	LOA (CWRU) \rightarrow LA (CWRU)	Média
Apenas fonte	75.27 \pm 2.33	69.35 \pm 4.44	72.31 \pm 4.57
D-CORAL	84.13 \pm 1.89	77.54 \pm 12.66	80.83 \pm 9.21
DAN	86.65 \pm 2.50	72.88 \pm 9.84	79.77 \pm 9.92
ADDA	78.60 \pm 2.91	75.49 \pm 2.74	77.05 \pm 3.13
DANN	80.19 \pm 2.92	69.96 \pm 12.74	75.08 \pm 10.24
AdaBN	77.36 \pm 2.10	68.79 \pm 3.86	73.08 \pm 5.38
MDD	67.78 \pm 10.05	67.79 \pm 8.77	67.79 \pm 8.90
JAN	70.85 \pm 7.16	60.76 \pm 13.23	65.80 \pm 11.35
CDAN	68.31 \pm 13.39	62.47 \pm 10.20	65.39 \pm 11.63

terísticas típicas de folga mecânica, o que dificultaria a detecção e diagnóstico de falhas de rolamento. Portanto, uma possível explicação para os resultados observados, é que esses sinais mais ruidosos em LA teriam um maior efeito negativo durante o teste, causando tamanha variação.

3.2. Mudança de bancada

A bancada do LVA apresenta falhas nos rolamentos de mancais em um eixo isolado, enquanto a bancada do CWRU exibe falhas nos rolamentos de um eixo dentro de um motor elétrico. Além disso, as condições de operação de ambas são distintas, de forma que o conjunto do LVA possui uma maior diversidade de velocidades de rotação e de elementos acoplados a bancada. Na Tabela 3 podemos observar que usar apenas dados do domínio fonte superou na média todas as técnicas de adaptação de domínio avaliadas, e que isso se deu principalmente devido ao desempenho apresentado ao realizar o treinamento no LVA e teste no LOA (CWRU). Todavia, as técnicas ADDA e AdaBN tiveram destaque na divisão oposta, apresentando superioridade em relação ao treinamento apenas com a fonte.

Tabela 3. $100 \times$ AUROC Macro (média \pm desvio padrão) das técnicas avaliadas entre bancadas

Método	LOA (CWRU) \rightarrow LVA	LVA \rightarrow LOA (CWRU)	Média
Apenas fonte	56.5 \pm 4.21	64.75 \pm 3.44	60.62 \pm 5.66
ADDA	57.2 \pm 2.55	59.41 \pm 2.94	58.30 \pm 2.84
AdaBN	58.31 \pm 3.78	56.16 \pm 1.51	57.24 \pm 2.94
JAN	54.81 \pm 3.26	56.26 \pm 3.92	55.54 \pm 3.48
DAN	55.92 \pm 3.66	51.78 \pm 1.63	53.85 \pm 3.45
CDAN	49.24 \pm 7.61	49.85 \pm 13.88	49.55 \pm 10.56
DANN	50.71 \pm 3.72	45.03 \pm 3.81	47.87 \pm 4.64
D-CORAL	53.89 \pm 1.84	41.62 \pm 3.25	47.76 \pm 6.93
MDD	45.35 \pm 5.76	47.86 \pm 13.63	46.61 \pm 9.95

Uma possível explicação para o desempenho pior das técnicas de adaptação de domínio neste cenário, diz respeito ao tipo de mudança de distribuição entre o domínio

fonte e alvo, causada pela mudança entre bancadas. No geral técnicas de adaptação de domínio não supervisionadas apenas conseguem lidar com mudanças nas distribuições dos atributos, também chamado de *covariate shift*.

No entanto, neste caso, pode haver não apenas um *covariate shift*, mas também um *concept shift*. *Concept shift* ocorre quando a relação entre os atributos e a variável alvo muda entre os domínios. Em outras palavras, mesmo que as distribuições dos atributos sejam semelhantes, a maneira como esses atributos influenciam ou determinam a variável alvo pode ser diferente. Isso torna a adaptação mais complexa, pois as técnicas não supervisionadas de adaptação de domínio não são projetadas para capturar essas mudanças conceituais. Para lidar com *concept shift*, geralmente é necessário o uso de métodos que requerem dados rotulados no domínio alvo, permitindo ajustar os modelos às novas relações entre atributos e variável alvo.

4. Conclusão

Neste artigo foi proposta uma nova metodologia de avaliação para técnicas de adaptação de domínio não supervisionadas, evitando vazamento de informação, com enfoque na detecção e diagnóstico de falhas de rolamentos utilizando o conjuntos de dados do CWRU e do LVA. Foram avaliados dois cenários de mudança de distribuição: i) mudança de um componente de uma mesma bancada; ii) mudança entre bancadas. Dentre as 8 técnicas de adaptação de domínio avaliadas, 3 superaram o treinamento apenas com o domínio fonte no cenário i) e nenhuma técnica superou o treinamento no domínio fonte no cenário ii).

No cenário i) foi possível observar uma diferença de desempenho entre as duas possibilidades de divisão, o que poderia ser explicado por sinais ruidosos no CWRU, conforme apontado por [Smith and Randall 2015]. Já no cenário ii), o desempenho dos modelos se aproximou de um modelo aleatório, o que pode nos indicar que existe uma mudança de distribuição muito grande entre as bancadas, indo além do *covariate shift*, quebrando hipóteses adotadas por muitas das técnicas de adaptação de domínio não supervisionadas. Em trabalhos futuros, pretendemos explorar a implementação de uma metodologia de otimização de hiperparâmetros para as técnicas de adaptação de domínio, o que acreditamos ser fundamental para melhorar o desempenho em ambos os cenários.

Referências

- Bauler, V., Cordioli, J., Braga, D., and Silva, D. (2023). Comparison of traditional vibration analysis techniques and machine learning models for bearing fault detection. In *Proceedings of the 27th International Congress of Mechanical Engineering*. ABCM.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35.
- Hakim, M., Omran, A. A. B., Ahmed, A. N., Al-Waily, M., and Abdellatif, A. (2023). A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Engineering Journal*, 14(4):101945.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

- Hendriks, J., Dumond, P., and Knox, D. A. (2022). Towards better benchmarking using the cwru bearing fault dataset. *Mechanical Systems and Signal Processing*, 169:108732.
- Jiao, J., Zhao, M., Lin, J., and Liang, K. (2020). Residual joint adaptation adversarial network for intelligent transfer fault diagnosis. *Mechanical Systems and Signal Processing*, 145:106962.
- Kapoor, S. and Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138:106587.
- Li, S., Bu, R., Li, S., Liu, C. H., and Huang, K. (2024). Principal properties attention matching for partial domain adaptation in fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 73:1–12.
- Li, Y., Wang, N., Shi, J., Hou, X., and Liu, J. (2018). Adaptive batch normalization for practical domain adaptation. *Pattern Recognition*, 80:109–117.
- Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd International Conference on Machine Learning*, page 97–105. PMLR.
- Long, M., CAO, Z., Wang, J., and Jordan, M. I. (2018). Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.
- Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2017). Deep transfer learning with joint adaptation networks. In *Proceedings of the 34th International Conference on Machine Learning*, page 2208–2217. PMLR.
- Neupane, D. and Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8:93155–93178.
- Randall, R. B. (2021). *Vibration-Based Condition Monitoring*. Wiley, 2 edition.
- Rosa, R. K., Braga, D., and Silva, D. (2024). Benchmarking deep learning models for bearing fault diagnosis using the cwru dataset: A multi-label approach. (arXiv:2407.14625). arXiv:2407.14625 [eess].
- Smith, W. A. and Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64–65:100–131.
- Sun, B. and Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In Hua, G. and Jégou, H., editors, *Computer Vision – ECCV 2016 Workshops*, page 443–450, Cham. Springer International Publishing.

- Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). Adversarial discriminative domain adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 2962–2971.
- Wang, Q., Michau, G., and Fink, O. (2019). Domain adaptive transfer learning for fault diagnosis. In *2019 Prognostics and System Health Management Conference (PHM-Paris)*, page 279–285.
- Zhang, S., Su, L., Gu, J., Li, K., Zhou, L., and Pecht, M. (2023). Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. *Chinese Journal of Aeronautics*, 36(1):45–74.
- Zhang, S., Zhang, S., Wang, B., and Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8:29857–29881.
- Zhang, Y., Liu, T., Long, M., and Jordan, M. (2019). Bridging theory and algorithm for domain adaptation. In *Proceedings of the 36th International Conference on Machine Learning*, page 7404–7413. PMLR.