

Enable Transferability of Interpretable ML Models for Bearing Diagnostic in Changing Operating Conditions

Julian Schauer^{1,2}, Houssam El Moutaouakil^{1,2}, Andreas Schütze²

¹*Centre for Mechatronics and Automations Technology, Saarbrücken, Germany*

²*Lab for Measurement Technology, Saarland University, Saarbrücken, Germany*

(julian.schauer),(h.moutaouakil)@zema.de, schuetze@lmt.uni-saarland.de

Abstract

In industrial machine learning applications, changing environmental and operating conditions significantly reduce the robustness and predictive quality of machine learning models, leading to performance drops under domain shifts. Interpretable machine learning algorithms based on physically meaningful features allow us to build models that are more robust to data distribution shifts. However, for significant domain shifts, the model should be able to adapt to the novel domain to avoid a performance decrease. Especially in industrial settings, the influences of external sources are hard to control. This study utilizes a novel approach to represent interpretable machine learning algorithms as deep neural networks, enabling transfer learning to adapt to novel domains. In this study, an interpretable machine learning algorithm comprising feature extraction, feature selection, and classification was developed to predict the condition of a roller bearing using accelerometer data. The novel approach demonstrates the ability to adapt the model via transfer learning to a new, unseen rotational speed with only 32 new transfer samples. While keeping information from the source domain out, the novel approach increases the model's accuracy from 77.6% to 100% in the unseen operating condition.

Keywords: Smart Sensors, Interpretable Machine Learning, Transfer Learning, Domain Shifts

1. Introduction

Reliable bearing diagnostics are essential for ensuring the safety, efficiency, and longevity of rotating machinery in a wide range of industrial applications. The degradation of roller bearings directly affects overall system performance and maintenance costs [1]. The introduction of Predictive Maintenance (PM) and Condition Monitoring (CM) based on Machine Learning (ML), along with appropriate sensors, enables efficient and reliable diagnosis of bearing conditions [2]. However, data-driven approaches like ML are sensitive to variations in the underlying data, which may result from changing operational or environmental conditions [3]. This is particularly relevant for roller bearings, which often operate under varying loads, speeds, and unpredictable environmental conditions. Developing diagnostic models that can adapt to these new domains with minimal additional data is therefore a key challenge for achieving robust, scalable PM. An approach to building robust and interpretable ML models is to use Feature Extraction (FE), Feature Selection (FS), and Classification (C) algorithms (FESC) [4]. These algorithms are based on physically interpretable features and have proven to achieve robust predictions in several applications [3]. Especially

on unseen data, the FESC model shows reliable performance. Nevertheless, when FESC models face significant domain shifts, they lack transferability to new, unseen domains, without building a new model from scratch with data from the novel domain. This results in a high-effort collection of new datasets from the target domain to build a novel trained model.

In the field of Deep Neural Networks (DNNs), Transfer Learning (TL) has emerged as a powerful strategy to address this challenge [5]. TL enables knowledge learned from an initial domain to be effectively reused in a related target domain by fine-tuning the initial model. However, in industrial environments interpretable and trustworthy ML models are required to avoid unwanted false predictions. The “Black-Box” nature of DNNs with very limited interpretability and their limited robustness often prevent their use in industrial environments [6].

For this purpose, a novel approach to represent the inference of interpretable FESC models as DNN (IDNNRep) was recently introduced [7]. Primarily, this approach enables the implementation of FESC inference after offline training on generic AI accelerators, thereby reducing inference time and leading to a more energy-efficient implementation [8]. Furthermore, this

approach allows combining the interpretable and robust characteristics of FESC models with recently available TL techniques from the field of DNNs to adapt models to new and unseen domains with a small number of transfer samples.

This study investigates the retraining of specific parts of an FESC model using IDNNRep to adapt to new, unseen operating conditions of a roller bearing, while preserving the interpretable nature and physical meaning of the engineered features. For this purpose, data from a roller bearing testbed are investigated by varying the testbed's rotational speed to create domain shifts.

2. Methods

2.1 Data

The paper investigates a roller bearing testbed designed to assess the robustness of ML models. Figure 1 shows the testbed to collect the dataset. The dataset [9], [10] contains data from a triaxial accelerometer collected from damaged and undamaged roller bearings at a sampling frequency of 20 kHz. Three roller bearings were tested at four different mounting positions within three separate runs. Each run contains four force levels, each operated at four distinct operational speeds. Thus, the complete dataset contains 1151 measurement files, each containing a one-minute measurement from the three-axis accelerometer for a specific combination of covariates. In this study, only the accelerometer axis vertical to the shaft is used. The covariate variations are shown in Table 1.

Table 1: Covariates for the testbed included in the dataset

Parameter	Quantity	Values
Bearing	3	B10, B20, B30
Damage State	2	No Damage, Small Damage
Run	3	R1, R2, R3
Position	4	PA, PB, PC, PD
Force Level [N]	4	0, 1600, 2500, 3000
Rotational Speed [rpm]	4	85, 392, 706, 969

Table 1 also highlights the importance of dataset variability, when investigating ML models in the context of domain shifts in industrial applications. By trying to capture most of the variability of a complex system, the dataset and the experimental design become significantly larger, often resulting in incomplete coverage of that variability. This study focuses on the changing rotational speed, which represents one possible change in the operating condition. The introduced dataset allows keeping variations as small as possible by changing only the rotational speed. This preserves the influence of different variations without disturbing the model, allowing it to focus on predicting damage at the defined operating rotational speed.

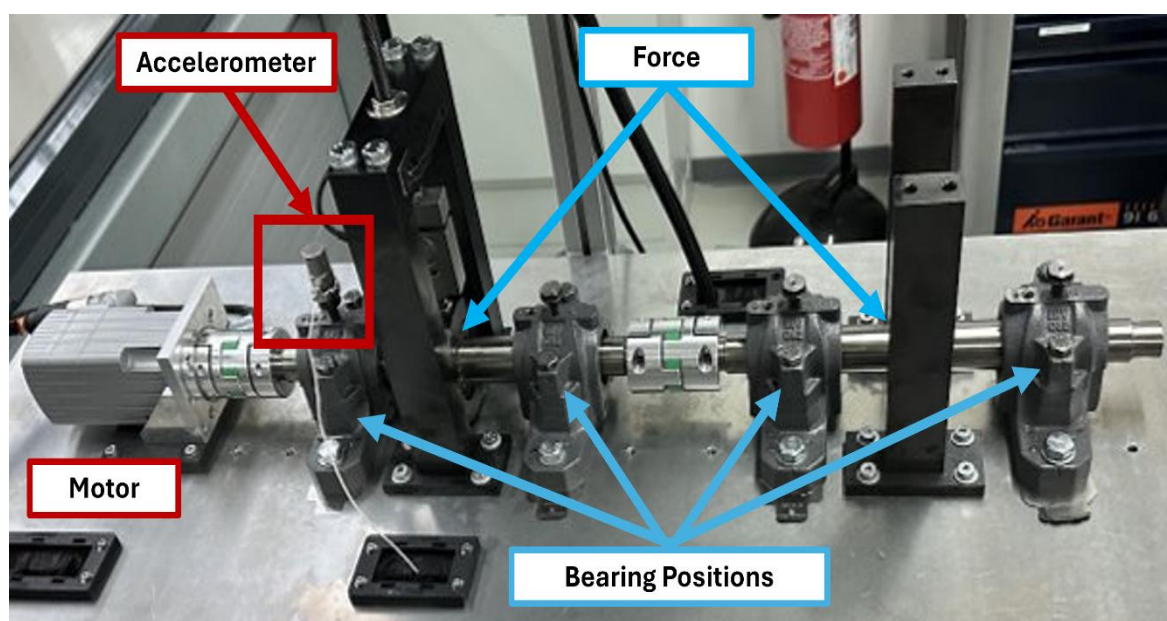


Figure 1: Roller bearing testbed with triaxial accelerometer

2.2 IDNNRep

The IDNNRep represents the inference equation of a trained FESC algorithm using DNN layers. For that purpose, the inference equation of the FESC algorithm is broken down into basic mathematical operations, such as addition, subtraction, multiplication, and division, as well as DNN-specific operations, such as convolution and pooling. The FESC algorithm, which is used to predict the roller bearing condition based on the data of an accelerometer, consists of the following parts and is included in the LMT Toolbox [11]:

- **Best Fourier Coefficient (BFC):** The BFC algorithm calculates the amplitudes and phases of the spectrum for each input time series and returns the 10% of amplitudes with the highest average absolute value and their corresponding phases. The 10% represent a suitable trade-off between a low approximation error and a low number of features [12].
- **Pearson Correlation:** The Pearson correlation algorithm calculates the linear relationship between the extracted features

and the target. The number of features to keep was defined to 20[4].

- **Linear Discriminant Analysis and Mahalanobis distance classifier (LDAMahal):** The LDAMahal algorithm first performs a linear projection of the feature space into the lower-dimensional LDA [13] subspace. After performing the LDA, the Mahalanobis distance [14] to each class center is calculated. The algorithm then classifies the input data based on the smallest calculated Mahalanobis distance.

Each part of the described FESC algorithm can be represented as a DNN; the detailed description is provided in [7]. The final FESC algorithm is then simply a concatenation of the layer graphs for each part. As shown in Figure 2, the final IDNNRep can be represented with the basic mathematical layers. This IDNNRep lays the foundation for the following TL to adapt the model to a novel domain.

2.3 Transfer Learning

Especially in industrial applications, environmental and operating conditions can change,

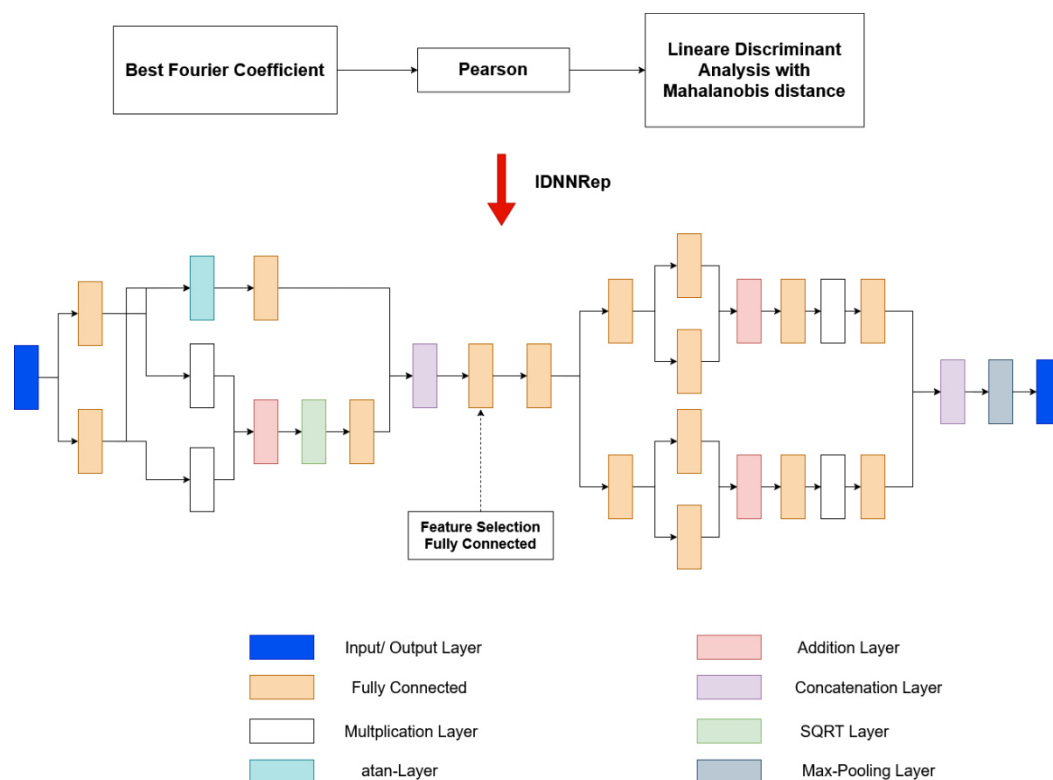


Figure 2: IDNNRep of the FESC algorithm

resulting in significant domain shifts. To avoid a decrease in model performance, it is necessary to retrain the underlying ML model to the novel domain. Besides data-based transfer techniques, especially TL, have shown significant benefits when domain shifts occur. This study focuses on the change in the operating condition of a roller bearing. When the rotational speed changes, the ML model that predicts the roller bearing's damage state experiences a significant domain shift. The goal is to retrain the IDNNRep of the FESR model using a small number of samples from the new domain to adapt to it, as shown in Figure 3.

For this purpose, an initial model was trained on accelerometer data at 706 rpm and 969 rpm and subsequently tested on 50% of the 396 rpm data, which were held out as validation data. To minimize data variation, only one roller bearing was used. Also, the other covariates, run (R1), position (PA), and the force level (0 N), were kept constant. After training the initial model, it is converted to an IDNNRep, and only the Feature Selection IDNNRep (one fully connected layer) part is retrained on a data subset from the target domain at 396 rpm. This allows us to retain information from the initial domain and fine-tune only a small number of parameters for the new domain.

3. Results and Discussion

To demonstrate the significance of the domain shift, a short investigation of the data is shown.

For that purpose, a Principal Component Analysis (PCA) is calculated based on the extracted features of the BFC. Figure 4 demonstrates the clear separation between the three rotational speeds. It also shows a clear separation between the damaged and undamaged data. As this study focuses on transferring an interpretable ML model to a new domain, this data shows good characteristics for such a problem.

To test the retraining of the IDNNRep, 50% of the data at 396 rpm is left out during training. These data are used to test the initial model and compare this with the retrained IDNNRep on the new domain. For retraining in the new domain, the amount of transfer samples was subsequently increased. The number of transfer samples investigated in this study was 0 (initial model), 8, 16, and 32. For retraining, all learnable parameters of the IDNNRep were frozen, except the weights and biases of the fully connected layer, which represent the FS. This layer was trained with an initial learning rate of 10^{-3} and a learning rate drop of 0.9 for 10 epochs. Table 2 lists the testing results of the initial model and the TL of the IDNNRep, with N denoting the number of samples used for finetuning. The initial model still achieves acceptable performance on an unseen operating speed with an accuracy of 77.6%, again demonstrating the robustness of FESC models under domain shifts. However, in most industrial applications an accuracy of 77.6% would be too low for reliable bearing diagnostic.

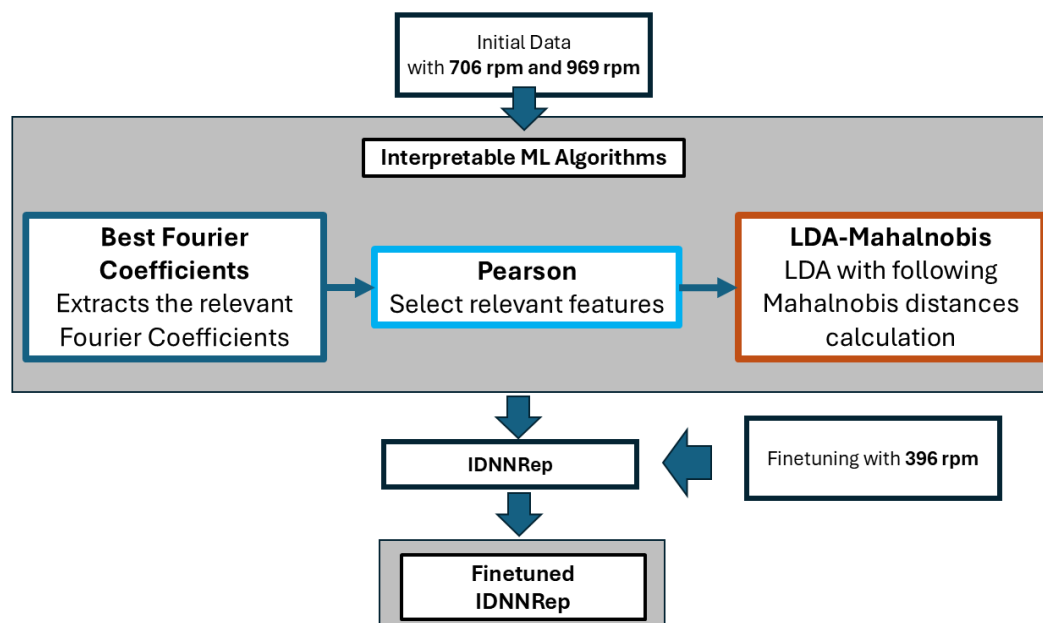


Figure 3: Design of Experiment for retraining the IDNNRep

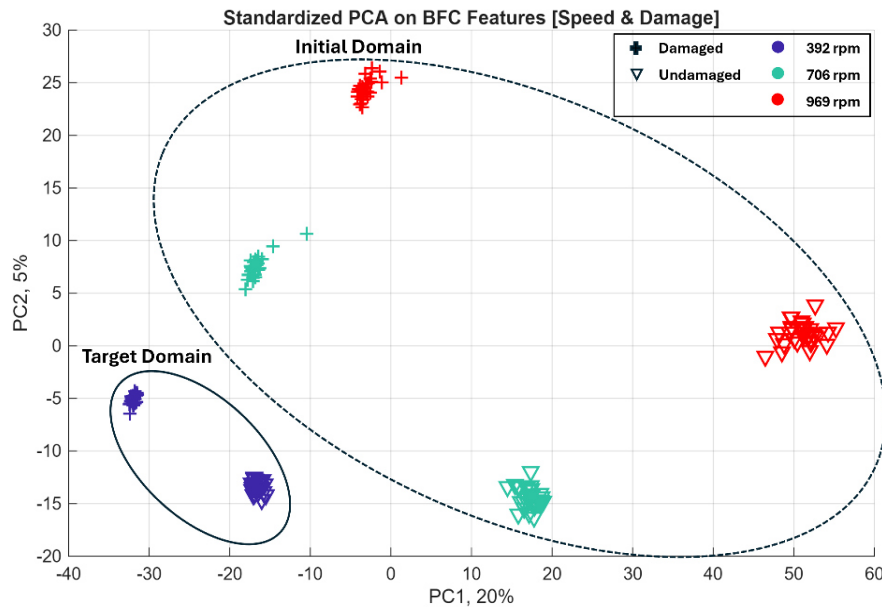


Figure 4: Standardized PCA of BFC Features

Retraining the initial model as an IDNNRep with 8 transfer samples already increases accuracy by 15.7% to 93.3%. This demonstrates a significant improvement compared to the initial model. Increasing from 8 transfer samples to 16 achieves an accuracy to 95.4%, an increase of only 2.1% compared to the retrained IDNNRep with 8 transfer samples. Using 32 transfer samples, the IDNNRep achieves 100% accuracy and yields the best possible result under the tested conditions.

Table 2: Validation Accuracy

	Accuracy (Testing)
Initial Model (N = 0)	77.6 %
TL (N=8)	93.3 %
TL (N=16)	95.4 %
TL (N=32)	100 %

4. Conclusion and Future Work

This study demonstrates a novel approach for model-based transferability for interpretable ML models. With the introduction of the IDNNRep, it is possible to retrain an initial interpretable ML after a significant domain shift. Even with only 8 transfer samples, TL increases the initial model's accuracy from 77.6% to 93.3%. By including 32 transfer samples, the model achieves 100%

accuracy, demonstrating the significant improvement of the new approach.

This demonstrates that the presented approach can be used to adapt interpretable ML models to novel domains, which presents an important challenge in the field of Machine Learning. Since this study focuses on a relatively simple domain shift, future investigations should include more complex domain shifts, also including additional covariates from the dataset. Additionally, more effective TL strategies and additional hyperparameter tuning offer the potential for further improvements.

Acknowledgement

This paper was written in the context of the VProSaar project which is funded by the Ministry of Economic Affairs, Innovation, Digital and Energy (MWIDE) and the European Funds for Regional Development (EFRE). The project is located in the federal state of Saarland and is executed by ZeMA – Centre for Mechatronics and Automation Technology gGmbH.

References

- [1] R. V. Daniel, S. A. Siddhappa, S. B. Gajanan, S. V. Philip, and P. S. Paul, "Effect of Bearings on Vibration in Rotating Machinery," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 225, no. 1, p. 012264, Aug. 2017, doi: 10.1088/1757-899X/225/1/012264.
- [2] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2,

- pp. 485–520, Feb. 2011, doi: 10.1016/j.ymssp.2010.07.017.
- [3] P. Goodarzi, A. Schütze, and T. Schneider, “Domain shifts in industrial condition monitoring: a comparative analysis of automated machine learning models,” *J. Sens. Sens. Syst.*, no. 14, pp. 119–132, 2025.
 - [4] T. Schneider, N. Helwig, and A. Schütze, “Automatic feature extraction and selection for classification of cyclical time series data,” *tm-Technisches Messen*, vol. 84, no. 3, pp. 198–206, 2017, doi: 10.1515/teme-2016-0072.
 - [5] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J Big Data*, vol. 3, no. 1, p. 9, May 2016, doi: 10.1186/s40537-016-0043-6.
 - [6] V. Buhrmester, D. Münch, and M. Arens, “Analysis of explainers of black box deep neural networks for computer vision: A survey,” *Machine Learning and Knowledge Extraction*, vol. 3, no. 4, pp. 966–989, 2021, doi: 10.3390/make3040048.
 - [7] J. Schauer, P. Goodarzi, and A. Schütze, “Efficient Hardware Implementation of Interpretable Machine Learning Based on Deep Neural Network Representations for Sensor Data Processing,” *Journal of Sensors and Sensor Systems*, 2025, doi: 10.5162/sensoren2024/A1.4.
 - [8] J. Schauer, P. Goodarzi, J. Morsch, and A. Schütze, “A Performance Study of Deep Neural Network Representations of Interpretable ML on Edge Devices with AI Accelerators,” *Sensors*, vol. 25, no. 18, p. 5681, 2025, doi: 10.3390/s2518568.
 - [9] C. Schnur, P. Goodarzi, Y. Robin, J. Schauer, and A. Schütze, “A Machine Learning Dataset of Artificial Inner Ring Damage on Cylindrical Roller Bearings Measured Under Varying Cross-Influences,” *Data*, vol. 10, no. 5, p. 77, May 2025, doi: 10.3390/data10050077.
 - [10] C. Schnur *et al.*, “A Machine Learning Dataset of Artificial Inner Ring Damages on Cylindrical Roller Bearings Measured Under Varying Cross-Influences.” Zenodo, Apr. 15, 2025. doi: 10.5281/zenodo.11108503.
 - [11] ZeMA-gGmbH, “LMT-ML-Toolbox,” *GitHub repository*. GitHub, 2017. [Online]. Available: <https://github.com/ZEMA-gGmbH/LMT-ML-Toolbox>
 - [12] F. Mörchén, “Time series feature extraction for data mining using DWT and DFT.” Citeseer, 2003.
 - [13] R. H. Riffenburgh, “Linear discriminant analysis,” PhD Thesis, Virginia Polytechnic Institute, 1957. [Online]. Available: <http://hdl.handle.net/10919/80187>
 - [14] G. J. McLachlan, “Mahalanobis distance,” *Resonance*, vol. 4, no. 6, pp. 20–26, 1999, doi: 10.1007/BF02834632.