

*Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference*  
 Edited by Eirik BJORHEIM ABRAHAMSEN, Terje Aven, Frederic Boudier, Roger Flage, Marja Ylönen  
 ©2025 ESREL SRA-E 2025 Organizers. Published by Research Publishing, Singapore.  
 doi: 10.3850/978-981-94-3281-3\_ESREL-SRA-E2025-P8201-cd

## A Study on the Use of Simulation Data for Data-Driven Fault Diagnosis of Various Rolling Bearings Using Transfer Learning

Marcel Braig, Fabian Mauthe, and Peter Zeiler

*Institute for Technical Reliability and Prognostics (IZP), Esslingen University of Applied Sciences, Germany.*  
*E-mail: marcel.braig@hs-esslingen.de, fabian.mauthe@hs-esslingen.de, peter.zeiler@hs-esslingen.de*

Rolling bearings are key components of numerous engineering systems and are subject to wear due to the mechanical contacts. Consequently, bearing fault diagnosis is imperative for the reliability and efficiency of these systems, such as rotating machinery. This paper explores the utilization of simulation data for training data-driven fault diagnosis methods. To this end, a self-developed bearing simulation and self-collected measurement data from test rigs are employed, considering varied operating conditions and bearing types. The study evaluates the effectiveness of simulation data in improving the diagnosis performance of real bearing faults. In particular, transfer learning methods are examined, encompassing both inductive and transductive transfer learning approaches, implemented with three types of neural networks. The findings demonstrate the effectiveness of the developed simulation model in generating data that is conducive to fault diagnosis. Already the training with simulation data alone indicates the potential benefits of incorporating simulation data. The study further demonstrates that inductive transfer learning exhibits superior performance in comparison to training with real measurement data alone. However, no improvements are achieved through transductive transfer learning.

*Keywords:* PHM, prognostics and health management, fault diagnosis, data-driven methods, simulation data, measurement data, transfer learning, rolling bearing, similar systems, operating conditions.

### 1. Introduction

Data-driven methods are among the most commonly utilized and empirically validated methods to diagnose the condition of engineering systems. However, training these methods requires a substantial amount of degradation data (Zio, 2022). Collecting data from all types of faults and levels of damage severity is a time-consuming and often intractable task. Furthermore, the problem is exacerbated when engineering systems operate under many different operating conditions or when there are a large number of similar systems that share technical characteristics but differ significantly in some aspects (Braig and Zeiler, 2023). In numerous industrial applications, it is not economically feasible to generate data from such a wide range of different conditions by damaging systems.

An approach to increasing the amount of data without damaging real systems is to generate synthetic data using simulations. Based on the physical equations and models that underpin the simulation, it is possible to generate degradation data from various similar systems with different faults

and under different operating conditions. However, these simulations are typically based on simplifications and assumptions that are only applicable to a limited extent in practice. Consequently, a discrepancy exists between the simulation data and the actual measured data. This discrepancy can be addressed using transfer learning (TL).

Therefore, the objective of this work is to study whether information from simulation data can be used to improve the fault diagnosis of rolling bearings. The study is based on measured condition data from rolling bearings of different types and sizes, recorded under different operating conditions on two different test rigs. A self-developed simulation model for rolling bearing condition data is used to generate the corresponding simulation data. Two concepts of TL are investigated: parameter transfer with fine-tuning and feature alignment by adversarial training. Different neural network types are evaluated, including the multilayer perceptron (MLP), the 1D convolutional neural network (CNN), and the temporal convolutional network (TCN).

In the following, Section 2 introduces related work. Section 3 describes the two TL methods applied and the data sets used. The study conducted is presented in Section 4, and Section 5 concludes the results of this work.

## 2. Related Work

The aim of TL is to transfer information between so-called domains. According to Braig and Zeiler (2023), a domain is defined by the set  $D = \{\mathcal{X}, P(X)\}$ , where  $\mathcal{X}$  represents the feature space and  $P(X)$  denotes the marginal distribution. The sample  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_m\} \in \mathcal{X}$  consists of  $m$  elements, each represented by a feature vector  $\mathbf{x}_i$ . Within a domain, a learning task is typically characterized by the set  $T = \{\mathcal{Y}, P(Y|X)\}$ , which includes the label space  $\mathcal{Y}$  and the prediction function  $P(Y|X)$ , interpreted as a conditional distribution. The set of labels  $Y = \{y_1, \dots, y_m\} \in \mathcal{Y}$  corresponds to the  $m$  feature vectors in the sample  $X$ .  $P(Y|X)$  is learned in a supervised manner using the pairs  $\{\mathbf{x}_i, y_i\}$ . TL is applied when there is a discrepancy between the source domain  $D_s$  and the target domain  $D_t$ , or if the source learning task  $T_s$  and the target learning task  $T_t$  differ. (Braig and Zeiler, 2023)

Rolling bearings are one of the most important machine elements in the literature on fault diagnosis of engineering systems (Lei et al., 2020; Raj et al., 2024). This also applies to TL applications for fault diagnosis (Braig and Zeiler, 2023). For example, Chen and Xiao (2024) and Wang et al. (2023) deal with the fault diagnosis of rolling bearings using TL. Among other things, the focus is on bearings under different operating conditions and a diagnosis using small sample sizes. A current overview of further TL approaches with the aim of fault diagnosis of rolling bearings can be found in (Hakim et al., 2023) and (Chen et al., 2023).

The generation of bearing degradation data is time-consuming. In the literature, damage to the bearings is either artificially added or the bearings wear out over time—often at an accelerated rate. Another possibility, also used in the literature, is the artificial generation of bearing fault data using simulations. For example, Peng et al. (2022)

provide an overview of the types of simulation and prognostics and health management techniques currently in use.

Simulations are usually based on assumptions and simplifications, resulting in deviations from the real measurement data. This difference must be overcome, and TL is often used for this purpose. Nguyen et al. (2024) use simulation data in conjunction with unlabeled measurement data to diagnose real bearing faults using TL. Wang et al. (2024), Liu et al. (2023), and Liu and Gryllias (2022) in addition use labeled measurement data. In (Xie et al., 2024), (Hou et al., 2023), and (Liu et al., 2023) training is carried out with simulation data and, in addition, with measurement data from the healthy bearing state. Xu et al. (2024), Ai et al. (2023), and Zhu et al. (2022) limit the training exclusively to simulation data. The main TL methods used include domain adaptation, either using maximum mean discrepancy or adversarial training, and parameter transfer. However, many publications are limited to a single TL method or a single network type. Furthermore, frequently only one operating condition or only one bearing type is considered.

## 3. Transfer Learning Methods and Data

In this section, a description of the two TL methods employed is given in Section 3.1, and the data sets used are presented in Section 3.2 and 3.3.

### 3.1. Employed transfer learning

This paper examines two types of TL: inductive and transductive TL. In inductive TL, labeled source data  $\{\mathbf{X}, \mathbf{Y}\}_S$  and labeled target data  $\{\mathbf{X}, \mathbf{Y}\}_T$  are available for training. In transductive TL, the training is based only on labeled source data  $\{\mathbf{X}, \mathbf{Y}\}_S$  and unlabeled target data  $\{\mathbf{X}\}_T$ . For inductive TL, parameter transfer with fine-tuning is implemented, and domain-adversarial training is used for transductive TL.

The used **parameter transfer with fine-tuning** consists of three steps. First, a pre-training of the network parameters with the labeled source data is conducted. During this training, source data are also used as validation data. Next, the deeper layers of the network are fine-tuned with labeled

target data while the remaining layers are frozen. Finally, the whole network is retrained with the same labeled target data (Li et al., 2022). During both steps, target data are used as validation data.

**Domain-adversarial training** is used to find a feature representation minimizing the difference in the data distribution between the source and target domains. In this work, a domain-adversarial neural network (DANN) architecture is implemented (Ajakan et al., 2015; Ganin and Lempitsky, 2015). It consists of a feature extractor, a label predictor, and a domain discriminator. The feature representation generated by the feature extractor is used by both the label predictor and the domain discriminator. It is chosen in such a way that the task of label classification is simplified, but at the same time the distributions of the instances of the domains are as identical as possible.

**3.2. Measurement data**

The measurement data serve as target data in the TL methods. These data are collected on two rolling bearing test rigs and under different operating conditions—speed and radial load. Bearings of different types with different defects are mounted in the test rigs, and the resulting vibration signal and speed are measured. The signals are sampled for one second at 15.625 kHz, which also applies to the simulation data. In the following, the term “recording” refers to such a one-second logging. The two test rigs differ in bearing support; otherwise, they are identical. Figure 1 shows the test rigs and the bearing types used. The configurations used to record the data can be found in Table 1. Ten recordings are made for each of the possible combinations. The data include four distinct classes: healthy bearing, inner race fault, outer race fault, and rolling element fault.

**3.3. Simulation data**

The simulation data serve as source data in the TL methods. These data are generated by a developed simulation model presented in more detail in (Mauthe et al., 2025)<sup>a</sup>. When a mechanical contact occurs within a bearing that impacts the

<sup>a</sup>A uniform damage severity level is used for all fault classes, whereby it should be noted that the model could also be used to simulate an increasing severity of damage over time.

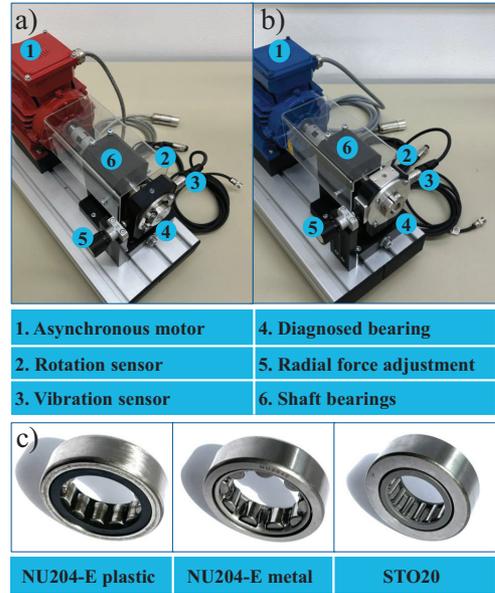


Fig. 1. The setup used to generate the measurement data, including two rolling bearing test rigs (a) and (b), and different types of rolling bearings used (c). The shaft position and vibration acceleration signals are recorded, along with the measuring time. The speed is calculated from the position and time values.

fault location, an impulse is generated that excites structural resonances in the bearing and its support (McFadden and Smith, 1984). This impulse repeats as the bearing rotates, and the time intervals between them are determined by the speed, fault type, and bearing geometry. These impulses are modeled in the simulation, combined with incorporating ball sliding theory and a random sliding effect, resulting in signals more representative of real bearing faults (Antoni, 2007). To emulate the impulses, the entire bearing and its support are modeled as a single-degree-of-freedom (SDOF) oscillating system (D’Elia et al., 2018; Ai et al., 2023). The time-domain impulse response can be described as (Mauthe et al., 2025):

$$x_{SDOF}(t) = \frac{J}{m\omega_d} e^{-\zeta\omega_n t} \sin(\omega_d t). \quad (1)$$

$\zeta = \frac{c}{2\sqrt{mk}}$  is the damping ratio,  $\omega_n = \sqrt{k/m}$  the natural frequency,  $\omega_d = \omega_n \sqrt{1 - \zeta^2}$  the damped natural frequency, and  $J$  the impulse. Here,  $m$  is the mass,  $k$  the stiffness, and  $c$  the damping coefficient of the modeled SDOF oscillating system.

Table 1. Configurations used to record the measurement rolling bearing data set. For each combination of bearing types, faults, speeds, and loads, ten recordings over one second are logged.

Bearing type <sup>(1)</sup>	Fault <sup>(2)</sup>	Speed / min <sup>(3)</sup>	Load <sup>(4)</sup>
NU204-E (a, plastic, 12, cylinder, 7.5, 20, 47)	OK / OR / IR / RE	500 / 1000	1 / 2
NU204-E (b, plastic, 12, cylinder, 7.5, 20, 47)	OK / OR / IR / RE	500 / 1000	1 / 2
NU204-E (b, metal, 9, cylinder, 7.5, 20, 47)	OK / IR / RE	500 / 1000	1 / 2
STO20 (b, metal, 14, needle, 3.5, 20, 47)	OK / OR / IR / RE	500 / 1000	1 / 2

<sup>(1)</sup>(test rig in Fig. 1, cage material, number of rolling elements, rolling element type, diameter of rolling elements / mm, inner diameter / mm, outer diameter / mm); <sup>(2)</sup> OK=healthy, OR=outer race fault, IR=inner race fault, RE=rolling element fault; <sup>(3)</sup> Controlled set speed; <sup>(4)</sup> Specified in the virtual unit "load level". The load is set using a micrometer screw and a rubber spring element.

To adapt the model to the test rigs, the parameters in Eq. (1) must be selected accordingly. To this end, the equation is differentiated twice with respect to time in order to obtain the acceleration signal. This is followed by an approximation to the real measured acceleration signal based on the impulse response when the outer race fault is rolled over. In this work, during data preprocessing, the amplitudes of the simulation recordings are scaled per load level with the maximum amplitude over all measurement recordings. Therefore,  $J/m\omega_a$  is replaced by a parameter  $A$  during numerical optimization, representing the initial amplitude of the impulse response. In order to enable a temporal alignment, the phase shift  $\phi$  is added to the sine argument. The parameters  $m$ ,  $k$ , and  $c$ , as well as  $A$  and  $\phi$ , are chosen based on a nonlinear least squares algorithm so that the equation reflects the measured acceleration signal as closely as possible. As with the measurement data, ten recordings are simulated for each of the configuration combinations in Table 1. In addition, outer race faults are also simulated for the third bearing type. In order to obtain variations between the simulation data of the same combination, stochastic effects such as speed fluctuations around the set speed, slippage on the bearing, and noise are simulated.

#### 4. Study

In this section, an overview of the procedure within the study is given in Section 4.1, and in the subsequent Section 4.2, the evaluation results are presented and discussed.

#### 4.1. Procedure

The following provides an overview of the network structures, hyperparameter optimization, features, and evaluation process.

*Network structures:* Three network types are used for the evaluations: MLP, CNN, and TCN with activated skip connection. In the context of parameter transfer, these network types form the initial layers, with dense layers added at the end. For DANN, the MLP, CNN, and TCN serve as feature extractors, while the rest of the network is implemented as MLP. A dropout layer is implemented for each dense layer and a max pooling layer after each CNN layer and TCN block. The stride of the CNN and TCN layers is set to one, and for the max pooling, it is set as the pooling size. The output activation function for the label prediction is softmax, and for the DANN, the sigmoid function is selected for the domain prediction. Remaining activation functions are ReLU, and cross-entropy is used as loss function. As a benchmark, alongside the TL methods, the three fundamental network types are trained using only source data and only target data. The implementations are based on TensorFlow.

*Hyperparameter optimization:* For hyperparameter optimization, the tree-structured Parzen estimator is used with up to 150 steps and early stopping after 20 attempts without improvement. The present study focuses on a comparison of the utilization of TL with that of non-use, rather than on a comparison of the individual network types.

Consequently, the specific settings and constraints for hyperparameter optimization are of less importance, provided they are consistent in both cases.

*Features:* The mean absolute value, the standard deviation, the peak-to-peak value, and the impulse factor are used as features. With MLP, these features are calculated over the entire vibration signal, whereas a moving window is used for the CNN and TCN, resulting in time series of length 30. The corresponding time value is assigned to each entry in the time series. Load and speed are included as additional features for all network types. The features are z-score normalized, where the normalization parameters are derived from the training data (both source and target). It is important to note that the generation of frequency features, as well as the targeted examination of fault frequencies, is explicitly excluded to avoid physics-based feature generation.

*Evaluation process:* The aim of the evaluation considered in this work is to classify the types of bearing faults. The information transfer from the simulation data to the real measurements is evaluated for each bearing type. In addition, each case is repeated ten times with a random distribution in training (50 %), validation (20 %), and test data (30 %). Within these data, operating conditions and fault types are evenly distributed. The number of training epochs is set to 200 for pre-training and training of the DANN, 100 for fine-tuning, and 50 for retraining. Premature termination occurs after ten epochs without improvements. Adam is used as optimizer. For the DANN, the adversarial training is conducted exclusively between instances where the source domain shares the same operating conditions as the target domain. Various classification metrics (Grandini et al., 2020) listed in Table 2 are used for the evaluation, and the mean value and standard deviation across all bearing types and repetitions are specified.

The performance ratio achieved between methods that use source data and those that do not use source data is expected to increase as the amount of target data decreases. In order to obtain as unbiased an evaluation as possible, the identical amount of data was used in the source and target domains, as described in Section 3.2 and 3.3.

## 4.2. Results

The results of the study are shown in Table 2. The mean values and standard deviations of the evaluation metrics across all bearing types and repetitions are given for all network types examined. Performance on the target test data, i.e., measurement data, is shown. The results of the parameter transfer with fine-tuning are marked with (*ST*), and the results of the domain-adversarial training are marked with DANN. (*S*) and (*T*) indicate the networks that are exclusively trained with source data and target data, respectively.

The approaches (*T*) and (*ST*), which use labeled target data, perform significantly better for all network types and all evaluation metrics than approaches that do not use labeled target data. This is reasonable since labeled target data for training are very valuable for performance in the target domain. Metric values of over 85 % up to 93 % are achieved with the (*T*) and (*ST*) approaches. The additional use of source data in approach (*ST*) leads to an improvement for all network types and all evaluation metrics despite the high performance already achieved in approach (*T*). The strong performance across all metrics and network types indicates that the resulting models are balanced and do not favor one class over another, ensuring reliable predictions independent of the respective network type. Pre-training with labeled source data seems to help find a better local optimum in the target domain. This is remarkable as all fault types and operating conditions are already present in the target training data. One reason for this could be that the “ideal” behavior of the respective bearing fault type is mapped in the simulation data, only superimposed by some additional stochastic effects. The algorithms therefore learn these ideal, physically meaningful relations between vibration and fault type during pre-training. It is also worth noting that the (*ST*) approach demonstrates superior performance in terms of accuracy when evaluated using source test data, as shown in Table 3. This observation holds true for all network types. This indicates that despite fine-tuning and retraining, part of the information from the source domain is still present in the networks.

Table 2. Values of the metrics when evaluating on target test data.

Approach	Accuracy		f1-score		Precision	
	Mean	Std	Mean	Std	Mean	Std
MLP( <i>T</i> )	0.914	0.081	0.908	0.094	0.916	0.093
MLP( <i>S</i> )	0.492	0.151	0.445	0.135	0.492	0.150
MLP( <i>ST</i> )	<b>0.923</b>	<b>0.073</b>	<b>0.920</b>	<b>0.076</b>	<b>0.928</b>	<b>0.070</b>
MLP-DANN	0.494	0.141	0.445	0.147	0.471	0.160
CNN( <i>T</i> )	0.920	0.139	0.909	0.169	0.912	0.175
CNN( <i>S</i> )	0.597	0.185	0.550	0.182	0.610	0.195
CNN( <i>ST</i> )	<b>0.928</b>	<b>0.130</b>	<b>0.921</b>	<b>0.153</b>	<b>0.935</b>	<b>0.149</b>
CNN-DANN	0.567	0.210	0.530	0.225	0.571	0.231
TCN( <i>T</i> )	0.864	0.189	0.844	0.230	0.855	0.237
TCN( <i>S</i> )	0.554	0.204	0.503	0.214	0.552	0.248
TCN( <i>ST</i> )	<b>0.884</b>	<b>0.167</b>	<b>0.869</b>	<b>0.203</b>	<b>0.879</b>	<b>0.212</b>
TCN-DANN	0.525	0.224	0.482	0.237	0.502	0.250

(*T*)=trained with labeled target data, (*S*)=trained with labeled source data, (*ST*)=pre-training with labeled source data and fine-tuning with labeled target data, best results per network type are shown in bold

Looking at the approach (*S*) and the DANN-based approach in Table 2, both of which do not use labeled target training data, metric values in the order of 50 % are achieved. Since the bearing fault diagnosis under consideration is a four-class classification problem, an accuracy of 25 % can be expected with random guessing. The performance values achieved are significantly higher. This indicates that information from the source domain can be successfully applied in the target domain to a certain extent.

However, the results do not show any noticeable advantage in using unlabeled target data in a DANN to achieve domain alignment. In fact, pure training with labeled source data tends to perform slightly better. It therefore seems to be difficult to find a feature representation in which the distribution difference between source and target data is minimal and which at the same time still allows a good classification of fault types. In order to check the latter, Table 3 can be analyzed, which shows the accuracies of the evaluation based on the source test data. In the case of the convolution-based networks, it can be seen that the (*S*) approach performs better than the DANN-based networks. This indicates that the choice of feature space towards domain invariance worsens the classification performance. An exception is the MLP, which is consistent with the fact that in this

network type the DANN has better accuracy on the target data than the (*S*) approach.

Table 3. Mean values of the accuracy when evaluating on source test data.

Approach	Mean accuracy		
	MLP	CNN	TCN
( <i>T</i> )	0.360	0.433	0.428
( <i>S</i> )	0.851	<b>0.973</b>	<b>0.929</b>
( <i>ST</i> )	0.429	0.553	0.569
DANN	<b>0.863</b>	0.912	0.836

(.) as for Table 2

Another reason for the poorer performance of the DANN-based networks could be that the domain alignment is unsupervised, i.e., without considering the fault labels of the data. This would only be possible if labeled target data were available for training of the DANN, which was intentionally omitted in this work in order to check whether unlabeled target data are already useful for TL. However, a possible future extension could be the use of so-called pseudo-labels. The target training data are assigned the labels that the label predictor outputs in the current training status. Subsequently, only an alignment between the conditional distributions of the source and target data with the same labels is performed. The

efficacy of this procedure, therefore, also depends on how well the label predictor trained with the labeled source data classifies the target data. However, since the performance of the approach ( $S$ ) on the target test data is better than pure guessing, it makes sense to work with pseudo-labels.

An analysis of the standard deviations in Table 2 reveals analogous trends in the performance of the approaches. The larger the mean value of the metrics, the smaller the standard deviation usually is. Small standard deviations indicate high robustness and generalizability. Here too, the best results across all network types and metrics are achieved with the ( $ST$ ) approach.

#### Additional evaluation driven by the results:

With regard to the comparison of approaches ( $T$ ) and ( $ST$ ), the results of the above evaluation raised the question of how the results change if not all operating conditions occur in the target training data. This is briefly examined in the following. The number of target data is reduced from ten per configuration combination, as described in Section 3.2, to now only two. A random, operating-condition-independent division into training, validation, and test data no longer ensures that all operating conditions occur in the target training data. Consequently, a decreasing performance is to be expected for both the approach ( $T$ ) and ( $ST$ ). However, an even more significant improvement in the ( $ST$ ) approach compared to the ( $T$ ) approach could also be expected. As shown in Table 4, both apply to the MLP as well as to the CNN. The accuracies are in the range of about 70 % for the MLP and about 70 to 80 % for the CNN, compared to over 90 % in Table 2. The improvement of the ( $ST$ ) approach over the ( $T$ ) approach is now 3.4 % for the MLP, compared to 0.9 % in Table 2, and now even 7.7 % for the CNN. As expected, the accuracies for the TCN have also deteriorated, which are in the range of 65 %, compared to over 85 % in Table 2. However, the smaller number of target data also appears to have a strong negative effect on fine-tuning and retraining, as the improvement in the accuracy of TCN( $ST$ ) in relation to TCN( $S$ ) could not be improved compared to the results in Table 2.

Table 4. Mean values of the accuracy when evaluating on target test data with reduced target training data.

Approach	Mean accuracy		
	MLP	CNN	TCN
( $T$ )	0.684	0.729	0.653
( $ST$ )	<b>0.718</b>	<b>0.806</b>	<b>0.655</b>
(.) as for Table 2			

## 5. Conclusion and Outlook

This work has investigated the use of artificially generated data from a self-developed rolling bearing simulation to improve the data-driven fault diagnosis of rolling bearings. Using only simulation data, i.e., without real measurement data, an accuracy of up to almost 60 % can be achieved. It has also been shown that using inductive TL can increase the accuracy to almost 93 %. Given that simulation data can be generated in seconds with the simulation model used, while the actual introduction of faults is much more time-consuming, the enormous potential becomes evident. However, transductive TL does not show any improvement compared to training with only simulation data. As described, the use of pseudo-labels is therefore recommended for future studies.

In this work, the ideal behavior in the simulation data was distorted by stochastic effects (noise, slip, speed fluctuations). More research should be carried out in the future to investigate how the extent of these distortions affects the training results. It may even turn out that it is best to dispense with distortions, as the ideal physical behavior is then modeled. The number of simulation and measurement data, as well as the coverage of operating conditions and fault types, can also be further varied for future investigations.

#### Acknowledgement

This work was supported in part by the Research and Development Projects at Universities of Applied Sciences (Innovative Projects) by the State of Baden-Württemberg, Germany; in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant no. 514247199; and in part by the German Federal Ministry of Education and Research through the Program "Professorales Personal an Fachhochschulen" (Professorial Staff at Universities of Applied Sciences) under Grant 03FHP115.

## References

- Ai, T., Z. Liu, J. Zhang, H. Liu, Y. Jin, and M. Zuo (2023). Fully simulated-data-driven transfer-learning method for rolling-bearing-fault diagnosis. *IEEE Trans. Instrum. Meas.* 72, 1–11.
- Ajakan, H., P. Germain, H. Larochelle, F. Laviolette, and M. Marchand (2015). Domain-adversarial neural networks. *arXiv*.
- Antoni, J. (2007). Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *J. Sound Vib.* 304(3-5), 497–529.
- Braig, M. and P. Zeiler (2023). Using data from similar systems for data-driven condition diagnosis and prognosis of engineering systems: A review and an outline of future research challenges. *IEEE Access* 11, 1506–1554.
- Chen, X., R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang (2023). Deep transfer learning for bearing fault diagnosis: A systematic review since 2016. *IEEE Trans. Instrum. Meas.* 72, 1–21.
- Chen, Y. and L. Xiao (2024). A multisource–multitarget domain adaptation method for rolling bearing fault diagnosis. *IEEE Sensors J.* 24(3), 3406–3419.
- D’Elia, G., M. Cocconcelli, and E. Mucchi (2018). An algorithm for the simulation of faulted bearings in non-stationary conditions. *Meccanica* 53(4-5), 1147–1166.
- Ganin, Y. and V. Lempitsky (2015). Unsupervised domain adaptation by backpropagation. In F. Bach and D. Blei (Eds.), *Proc. 32nd Int. Conf. Mach. Learn.*, pp. 1180–1189. JMLR.
- Grandini, M., E. Bagli, and G. Visani (2020). Metrics for multi-class classification: an overview. *arXiv*.
- Hakim, M., A. A. B. Omran, A. N. Ahmed, M. Al-Waily, and A. Abdellatif (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 Eng. J.* 14(4), 101945.
- Hou, W., C. Zhang, Y. Jiang, K. Cai, Y. Wang, and N. Li (2023). A new bearing fault diagnosis method via simulation data driving transfer learning without target fault data. *Measurement* 215, 112879.
- Lei, Y., B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process* 138, 106587.
- Li, W., R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, and K. Gryllias (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mech. Syst. Signal Process* 167, 108487.
- Liu, C. and K. Gryllias (2022). Simulation-driven domain adaptation for rolling element bearing fault diagnosis. *IEEE Trans. Ind. Inf.* 18(9), 5760–5770.
- Liu, J., H. Cao, S. Su, and X. Chen (2023). Simulation-driven subdomain adaptation network for bearing fault diagnosis with missing samples. *Eng. Appl. Artif. Intell.* 123, 106201.
- Liu, X., S. Liu, J. Xiang, and R. Sun (2023). A transfer learning strategy based on numerical simulation driving 1d cycle-gan for bearing fault diagnosis. *Information Sciences* 642, 119175.
- Mauthe, F., S. Hagemeyer, and P. Zeiler (2025). Holistic simulation model of the temporal degradation of rolling bearings. In E. B. Abrahamsen, T. Aven, F. Boudier, R. Flage, and M. Yloenen (Eds.), *Proc. 35th Eur. Saf. Reliab. Conf.* Research Publishing.
- McFadden, P. D. and J. D. Smith (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *J. Sound Vib.* 96(1), 69–82.
- Nguyen, T. H., V. V. Hung, D. D. Thinh, T. T. Tran, and H. S. Hong (2024). Generalized simulation-based domain adaptation approach for intelligent bearing fault diagnosis. *Arab. J. Sci. Eng.* 49(12), 16941–16957.
- Peng, F., L. Zheng, Y. Peng, C. Fang, and X. Meng (2022). Digital twin for rolling bearings: A review of current simulation and phm techniques. *Measurement* 201, 111728.
- Raj, K. K., S. Kumar, and R. R. Kumar (2024). Systematic review of bearing component failure: Strategies for diagnosis and prognosis in rotating machinery. *Arab. J. Sci. Eng.*, 1–23.
- Wang, C., G. Zhu, T. Liu, Y. Xie, and Di Zhang (2023). A sub-domain adaptive transfer learning base on residual network for bearing fault diagnosis. *J. Vib. Control.* 29(1-2), 105–117.
- Wang, H., P. Wang, S. Wang, and D. Li (2024). Rolling bearing fault diagnosis method based on dynamic simulated model from source domain to target domain with improved alternating transfer learning. *Nonlinear Dynamics*, 1–26.
- Xie, J., S. Wang, Y. Li, T. Wang, J. Yang, and T. Pan (2024). A simulated-to-real transfer fault diagnosis method based on prototype clustering subdomain adversarial adaptation network for hst bogie bearing. *IEEE Trans. Instrum. Meas.* 73, 1–13.
- Xu, Z., G. Tang, and B. Pang (2024). Simulation-driven unsupervised fault diagnosis of rolling bearing under time-varying speeds. *Struct. Health Monit.*, 1–17.
- Zhu, P., S. Dong, X. Pan, X. Hu, and S. Zhu (2022). A simulation-data-driven subdomain adaptation adversarial transfer learning network for rolling element bearing fault diagnosis. *Meas. Sci. Technol.* 33(7), 075101.
- Zio, E. (2022). Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice. *Reliab. Eng. Syst. Saf.* 218, 108119.