Bearings’ Vibration Fault Detection via Physics-Informed Deep Learning

Lucas Xavier Teixeira de Souza[1], Leonardo Streck Raupp[1], Caio Souto Maior[1,2], Isis Didier Lins[1], Thiago Cavalcanti[1], Márcio José das Chagas Moura[1], Gustavo de Novaes Pires Leite[1]

[1] *Center for Risk Analysis, Reliability Engineering and Environmental Modeling (CEERMA), Federal University of Pernambuco, Brazil*

[2] *The B. John Garrick Institute for the Risk Sciences, University of California - Los Angeles, USA*

ABSTRACT

The reliability of rotating machinery is essential in industrial environments, where early fault detection can prevent significant losses. In this scenario, Condition-Based Maintenance strategies benefit from the combined use of signal processing and machine learning techniques. Although Deep Learning based models present good results in automatic fault classification, their exclusive dependence on data can generate inconsistent predictions and non-compliance with physics. This study presents a Physics-Informed Deep Learning (PIDL) approach for fault detection in rolling element bearings based on vibration signal analysis. Experimental data were collected from a controlled test bench that simulates real operational conditions, using a self-aligning double row ball bearing (NSK 1205K C3) under three health conditions: healthy, light and severe damage. Vibration signals were acquired using piezoelectric accelerometers and pre-processed through envelope analysis to enhance fault-related features in the frequency domain. The classification task was reformulated into a binary problem - healthy vs. unhealthy - aiming to improve the generalization of the model and its practical utility. A convolutional neural network (CNN) was developed for binary classification, which combines convolutional, clustering, and dense layers. A custom loss function was proposed to incorporate physical knowledge, penalizing false negatives and false positives differently, according to the criticality of each type of error in industrial contexts. The experimental results showed that the PIDL model achieved a balance precision of 94.77%, outperforming the traditional deep learning model (94.05%) not only in overperformance, but rather reducting false negatives, which are critical for preventive maintenance. The findings demonstrate that incorporating domain knowledge into the training process leads to safer and more reliable fault detection, making the PIDL approach particularly suitable for industrial applications where early and accurate detection is crucial to avoid unplanned downtime and equipment failure.

1. INTRODUCTION

Reliability is a key factor in ensuring the success and competitiveness of modern industries. With increasingly complex systems and the growing need for higher availability while reducing costs, the importance of maintaining the integrity of industrial assets cannot be overstated. To address these challenges, Condition-Based Maintenance (CBM) strategies have gained prominence, focusing on the management of equipment according to their health status (Jardine, Lin, & Banjevic, 2006). This approach is often integrated with Prognostics and Health Management (PHM) programs (Pecht, 2009), which enable continuous monitoring of system health and allow interventions before critical failures occur (Maior, Moura, & Lins, 2019).

Among the many components that contribute to the reliability of mechanical systems, bearings play a crucial role, particularly in rotating machinery such as motors, compressors, and gearboxes. Bearings are often responsible for a significant proportion of system failures—up to 55% of all failures in rotating machines, according to studies by (Yang, Yu, & Cheng, 2007). The ability to detect faults in bearings early on is, therefore, vital to minimizing unexpected downtimes, reducing maintenance costs, and ensuring operational safety.

One of the most effective methods for detecting bearing faults is vibration analysis. Techniques such as the Hilbert-Huang Transform, envelope analysis, and frequency spectrum analysis have proven useful in extracting relevant features from vibration signals, enabling early detection of anomalies (Soualhi, Medjaher, & Zerhouni, 2015). The Hilbert Transform, in particular, is known for its sensitivity to damage in rotating components, as it highlights small minor variations in the time domain that are indicative of fault progression.

At the same time, advancements in data storage, the availability of affordable sensors, and enhanced computational capabilities have fueled the rise of data-driven methods, including Machine Learning (ML) and Deep Learning (DL). These techniques are particularly compelling in tackling complex systems, extracting patterns from large datasets, and applying them to a variety of tasks, such as failure mode classification (Yuan et al., 2020), Remaining Useful Life (RUL) prediction (Maior et al., 2016; Sikorska, Hodkiewicz, & Ma, 2011), and anomaly detection (Bakdi & Kouadri, 2017; Zhou & Paffenroth, 2017). However, relying solely on data can introduce limitations, such as false alarms or low confidence in predictions when the data fails to fully represent the nuances of specific equipment or conditions (S. Shen et al., 2021).

Despite their success, one of the main drawbacks of these data-driven models is their lack of integration with physical principles, which are often essential for understanding the underlying causes of system behavior. This gap becomes particularly important when dealing with equipment such as bearings in the rolled element, such as those found in pumps, where bearings account for a significant proportion of system failures. This issue raises critical questions about how to combine the strengths of data-driven approaches with a deeper understanding of physical systems to improve fault detection and prediction reliability. To overcome these limitations, a new hybrid approach called Physics-Informed Deep Learning (PIDL) has emerged. This methodology incorporates physical knowledge directly into the training process of neural networks, typically by modifying the loss function with additional terms that represent constraints based on physical laws (S. Shen et al., 2021; Thuerey et al., 2021).

This study proposes the application of a PIDL model for fault detection in bearings based on vibration analysis. Signals were acquired from an experimental test bench with different fault conditions in self-aligning ball bearings. Data preprocessing includes Hilbert Transform application and segmentation into 512-sample temporal windows. The convolutional neural network model is trained with a modified loss function that penalizes inconsistent predictions according to amplitude thresholds defined from bearing physical characteristics.

The main contribution of this study is the explicit incorporation of physical knowledge into the machine learning process, enhancing the robustness and reliability of the fault detection system. Comparative experiments between PIDL and traditional DL models demonstrate the superiority of the former in industrially sensitive scenarios where early fault detection is critical for equipment safety and continuous operation.

2. THEORETICAL BACKGROUND

2.1. Machine Learning, Neural Networks, and Deep Learning

ML can be defined as the development of algorithms and statistical models that enable computational systems to improve performance on specific tasks based on data, without relying exclusively on explicit instructions from programmers. According to (Maior, Moura, Santana, & Lins, 2020), machine learning is the improvement of a computational program's performance on a given task, as measured by a criterion, based on acquired experience.

Artificial Neural Networks (ANNs), introduced by (McCulloch & Pitts, 1943), represent a class of models within ML inspired by the human brain's structure and function for learning and pattern recognition. These networks consist of computational units called neurons, organized into layers: input, hidden, and output. Each neuron applies an activation function to its input signals, producing outputs that propagate through the network. Synaptic weights, adjustable during training, determine the relative influence of each input.

During training, ANNs adjust weights to minimize the error between predicted outputs and desired targets via a procedure called backpropagation. This technique is associated with optimization methods like gradient descent to minimize predefined cost functions.

DL is a subfield of ANNs distinguished by the use of multiple hidden layers, known as Deep Neural Networks (DNNs). This structural depth allows hierarchical feature extraction, facilitating the modeling of complex relationships within the data. Unlike traditional ANNs, DL employs various layer types for specific purposes. Convolutional layers, for example, are effective in processing grid-structured data like images and time series signals, due to their ability to capture spatial hierarchies of features.

The success of DL owes to factors such as large data availability, advances in computational resources, and optimization algorithm improvements. This approach has been effective in solving complex problems across domains, including computer vision, signal processing, and predictive maintenance.

2.2. Physics-Informed Deep Learning (PIDL)

Despite DL models' remarkable performance in various applications, many of these methods are purely data-driven, which may lead to predictions inconsistent with underlying physical laws. This limitation is critical in domains where safety, robustness, and interpretability are essential. Models that ignore physical principles can make incorrect classifications, violating expected system behavior and compromising reliability.

Physics-Informed Deep Learning (PIDL) addresses this by integrating physical knowledge directly into DL training. The primary strategy involves modifying the model's loss function to include terms that penalize deviations from known physical laws. Thus, the loss function reflects not only predictive error but also compliance with domain-specific physical principles.

During training, the PIDL model optimizes to minimize discrepancies between predictions and real data while respecting constraints imposed by physical equations or system properties. This approach is particularly advantageous in scenarios with scarce or noisy data, where incorporating prior knowledge about system behavior improves model generalization and stability.

2.3. Vibration Monitoring in Bearings

Vibration monitoring is a widely used and effective technique for assessing bearing condition, which is crucial for the reliable operation of rotating mechanical systems. Bearings support radial and axial loads, allowing smooth rotation of shafts and gears. Self-aligning bearings, in particular, accommodate angular misalignments between shaft and housing, ideal for applications with alignment variations caused by vibrations, thermal expansion, or wear.

Accurate fault characterization requires detailed experiments using vibration sensors, such as accelerometers or velocity transducers (Safizadeh & Latifi, 2014), capable of capturing vibrational signals at strategic points on the structure.

3. METHODOLOGY: PIDL MODEL IMPLEMENTATION

The physics-informed deep learning approach adopted in this study is based on the work of (C. Shen et al., 2020) and (Streck et al., 2024). The method involves creating a threshold model that reflects the expected physical behavior of faults. The output of this model is integrated into the DL system's loss function, allowing training to be guided not only by statistical patterns but also by relevant physical knowledge. The process follows typical AI modeling steps:

1. Data features undergo pre-processing including segmentation, transformation, variable selection, train-test splitting, labeling, and shuffling.
2. Model development, which includes:
   - The deep neural network
   - A customized loss function incorporating the threshold model
3. Classification is performed, and results are analyzed.

3.1. Dataset Description

Data were collected from a vibration test bench designed to simulate real operational conditions of bearings in rotating systems. Located at the Center for Risk Analysis, Reliability Engineering and Environmental Modeling (CEERMA) at the Federal University of Pernambuco (UFPE), the bench comprises a transmission system driven by an adjustable-speed induction motor controlled via a frequency inverter, as illustrated in Figure 1.

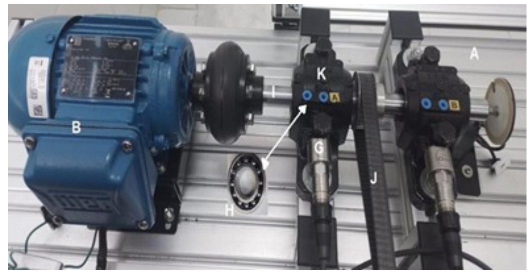

Figure 1. Vibration Bench Components

Table 1. Bearing and Accelerometer Specifications

Bearing Specifications	Accelerometer Specifications
Inner diameter (d), mm: 25	Sensitivity: 10.2 mV/(m/s²)
Outer diameter (D), mm: 52	Measur. Range: ± 490 m/s²
Pitch diameter (Pd), mm: 38.5	Freq. Range: 0.5 - 10K Hz
Ball diameter (Bd), mm: 7.14	Resonant Frequency: 25 Hz
No. rolling elements (Nb): 12	Broadband Res: 3434 μm/sec²
	Non-Linearity: ± 1%
	Transverse Sensitivity: $\leq 7\%$

The bearing used was the NSK 1205K C3, a self-aligning double-row ball bearing. This bearing type is especially suitable for applications with angular misalignment and shaft deflections, exhibiting low friction coefficient and high radial load capacity. Its unsealed design facilitates direct failure analysis and damage propagation. Main geometric and technical specifications of the bearing and sensors used are presented in Table 1.

Vibration monitoring was performed using two piezoelectric accelerometers (model 603C01), magnetically attached to the split bearing housings. Vibration signals were amplified 45 times before digitization by a Labjack U12 acquisition device, with a sampling rate of 2048 Hz and scan rate of 4096 Hz, configured via LJscope V1.09 software.

During tests, the motor frequency was fixed at 15 Hz, with measurements taken under three bearing conditions:

- Healthy: bearing without damage
- Light Fault: artificial 1 mm damage on the bearing outer race
- Heavy Fault: artificial 3 mm damage on the bearing outer race

Figure 2 shows the fault states considered. Standardizing these conditions enabled replicable experiments and facilitated comparative analyses.

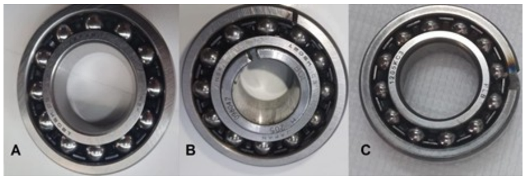

Figure 2. Bearing Conditions: (A) Healthy; (B) Light damage; (C) Heavy damage

Table 2. Fault Characteristic Frequencies

Characteristic Frequency	Value (Hz)
Ball Pass Frequency, Outer Race (BPFO)	73.305
Ball Pass Frequency Inner (BPFI)	106.695
Ball Spin Frequency (BSF)	78.08
Fundamental Train Frequency (FTF)	6.10875

3.2. Data Pre-processing

Following data acquisition, the raw vibration signals were segmented into overlapping windows of 512 data points, a size empirically determined to balance resolution and computational efficiency. Each window represented a time series sample corresponding to a specific condition of the bearing. Next, an envelope analysis was conducted to transform the time-domain signal into the frequency domain, a process particularly effective in highlighting amplitude modulations caused by structural defects. This approach is supported by previous studies such as (Lessmeier et al., 2016) which emphasize the relevance of frequency-domain features in fault detection.

The Hilbert Transform was applied to the signal to extract the envelope, enhancing the fault-related frequency components in the frequency domain. This step is crucial to better highlight characteristic fault frequencies such as Ball Pass Frequency Outer race (BPFO), Ball Pass Frequency Inner race (BPFI), Ball Spin Frequency (BSF), and Fundamental Train Frequency (FTF). These frequencies were calculated based on the bearing specifications and the operating condition of 15 Hz, and are summarized in Table 2.

The envelope signal was then segmented into windows of 512 samples each, resulting in time frames of approximately 0.25 seconds at a sampling frequency of 2048 Hz. Each segment was labeled according to the bearing condition (healthy or faulty) for supervised training.

This work considered, in addition to each calculated characteristic frequency, up to three of its harmonics. To account for noise and minor variations, a tolerance band of ±5% was applied around each frequency and its corresponding harmonics. Furthermore, the dataset was split into training and testing sets using an 80%-20% ratio, followed by a random shuffling of the data to ensure unbiased model evaluation.

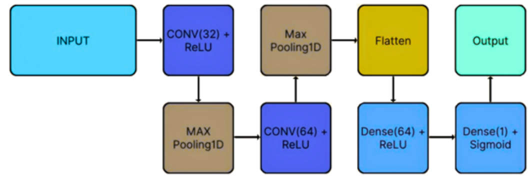

Figure 3. Network Architecture

Here, for simplification purpose, the original problem—initially framed as a multi-class classification with three states (healthy, light damage, and heavy damage)—was reformulated into a binary classification problem, defined as follows:

- Class 0 (Healthy): samples corresponding to undamaged bearings
- Class 1 (Unhealthy): samples containing any level of damage, whether light or heavy

This binary reformulation allows the model to focus on identifying the presence of anomalies, regardless of their severity. In practical industrial scenarios, this approach is valuable, as early identification of any abnormality can trigger maintenance inspections and prevent unscheduled downtimes or catastrophic failures.

3.3. Model Architecture

To compare the performance of the PIDL model with that of a conventional DL model using vibration data collected from our bearing vibration test bench, we adopted an architecture inspired by the model proposed by (S. Shen et al., 2021), which we modified to better suit the specific requirements of our binary classification problem. In particular, the kernel size and stride were empirically adjusted based on the characteristics of our dataset. The final architecture is illustrated in Figure 3.

The architecture consists of:

- **1D Convolutional Layer** with filters, kernel size 5, and ReLU activation: This convolutional layer is responsible for extracting temporal features from vibration data. The convolutional filter acts locally on the data, applying a convolution operation to identify specific patterns, such as peaks and variations in vibration waves, which are indicative of bearing faults. The ReLU (Rectified Linear Unit) activation introduces non-linearities to the model, allowing it to capture more complex patterns.
- **MaxPooling1D Layer** with pool size 2: Pooling is an important operation that reduces the dimensionality of the data by applying an aggregation function (typically the maximum value) over each sub-region of the input. The goal is to reduce the number of parameters while retaining the most important information, such as the most prominent features of bearing faults.

- **Flatten Layer**: This layer converts the multidimensional output from the previous layers into a one-dimensional vector, which is required for input into the subsequent dense layer. The goal of the Flatten layer is to organize the representations extracted by the convolutional and pooling layers for the network's final decision-making process.
- **Dense Layer** with neurons and ReLU activation: Dense layers, which are fully connected, learn more complex patterns in the representations provided by the previous layers. The ReLU activation here also helps introduce non-linearities, allowing the network to learn more sophisticated representations of the vibration data.
- **Output Layer** with 1 neuron and sigmoid activation: The final layer of the network provides a probability between 0 and 1, indicating the classification of the bearing as healthy (0) or faulty (1). The sigmoid activation function is used to ensure that the output is within the range [0, 1], which is suitable for a binary classification task.

3.4. Custom Loss Function (PIDL Loss)

The loss function plays a key role in guiding the model's training. In the PIDL approach, the standard loss is modified to include penalties for misclassifications, based on physical knowledge about bearing failures. The loss function is:

$$\text{Loss} = \text{Std\_Loss} + \alpha \times \text{False\_Pos} + \beta \times \text{False\_Neg} \quad (1)$$

Where:

- Standard_Loss is typically binary cross-entropy
- α and β are weights to penalize false positives and false negatives, respectively

In the study, $\alpha = \beta$ was used to balance both types of errors. However:

- $\alpha > \beta$: penalizes false positives more (useful for reducing false alarms)
- $\beta > \alpha$: penalizes false negatives more (useful for detecting more faults and avoiding missed failures)

Choosing $\alpha = \beta$ ensures balanced performance, but adjusting these values can significantly shift the model's behavior depending on the application priority.

4. RESULTS

The experimental results showed that both the traditional DL model and the PIDL model achieved high balanced accuracy rates. However, there are significant differences in the nature of the errors made by each model. While the traditional DL model tended to classify faulty bearings as healthy (false negatives), the PIDL model showed a slight tendency to classify

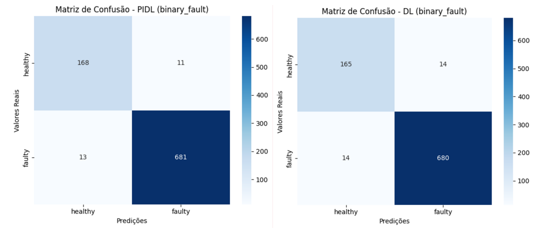

Figure 4. Confusion matrices: (a) PIDL: 94.776% accuracy balanced; (b) DL: 94.050% balanced accuracy

Table 3. Classification metrics for PIDL model

Class	Precision	Recall	F1-score
Healthy	0.9300	0.9140	0.9206
Non-healthy	0.9780	0.9815	0.9797
Accuracy		0.9677	
Macro avg	0.9540	0.9478	0.9502
Weighted avg	0.9678	0.9677	0.9676

healthy bearings as faulty (false positives). This behavior was unexpected since the weights α and β were set equal, but false positives are less harmful to industry than false negatives.

Both models (traditional DL and PIDL) were trained using the same data and hyperparameters (50 epochs, batch size = 64, Adam optimizer with learning rate = 0.001). The average balanced accuracy results were: 94.776% for PIDL and 94.050% for traditional DL.

Beyond the slight improvement in metric, the most relevant difference lies in the nature of the errors:

- Traditional DL: tends to classify faulty bearings as healthy (false negatives)
- PIDL: tends to classify some healthy bearings as faulty (false positives)

The PIDL approach proved to be more robust and reliable for implementation in industrial scenarios where minimizing the occurrence of undetected failures is a priority. Tables 3 and 4 present the detailed classification metrics for each model.

The PIDL model shows superior performance in recall for healthy bearings (0.9140 vs 0.8983) while maintaining comparable precision. Both models achieve high accuracy, with PIDL showing a slight advantage (96.77% vs 96.54%). The most significant improvement comes in the F1-score for healthy bearings (0.9206 vs 0.9147), suggesting PIDL provides better balance between precision and recall for this critical class.

5. CONCLUSION

This study introduced a PIDL approach for the detection of bearing faults based on vibration signal analysis. By incorporating domain-specific physical insights into the training pro-

Table 4. Classification metrics for DL model

Class	Precision	Recall	F1-score
Healthy	0.9326	0.8983	0.9147
Non-healthy	0.9740	0.9827	0.9783
Accuracy		0.9654	
Macro avg	0.9533	0.9405	0.9465
Weighted avg	0.9655	0.9654	0.9653

cess—specifically through a custom loss function that penalizes critical misclassifications—the resulting model became not only accurate but also safer and more suitable for real industrial applications.

Results show a slightly increased accuracy when comparing the traditional and the physics-informed DL models: 94.776% and 94.050%, respectively. The main contributions include:

- The proposal and implementation of a PIDL loss function that integrates knowledge about the criticality of false negatives and false positives
- A comparative analysis demonstrating that even a small modification to the learning objective can result in more reliable predictions in safety-sensitive scenarios

This approach is particularly promising for industrial environments, where predictive maintenance relies heavily on reliable early detection of mechanical faults to prevent failures and optimize operational costs.

ACKNOWLEDGMENT

The authors would like to thank the Center for Studies and Testing in Risk and Environmental Modeling (CEERMA) at the Federal University of Pernambuco for providing the experimental facilities and technical support.

REFERENCES

Bakdi, A., & Kouadri, A. (2017). A new adaptive pca based thresholding scheme for fault detection in complex systems. *Chemometrics and Intelligent Laboratory Systems*, *162*, 83–93. doi: 10.1016/j.chemolab.2017.01.013

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510. doi: 10.1016/j.ymssp.2005.09.012

Lessmeier, C., et al. (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. *Proceedings of the European Conference of the PHM Society*, *3*, Article 1. doi: 10.36001/phme.2016.v3i1.1577

Maior, C. B. S., Moura, M. D. C., & Lins, I. D. (2019). Particle swarm-optimized support vector machines and preprocessing techniques for remaining useful life estimation of bearings. *Eksploatacja i Niezawodnosc*, *21*(4), 610–619. doi: 10.17531/ein.2019.4.10

Maior, C. B. S., Moura, M. J. d. C., Santana, J. M. M., & Lins, I. D. (2020). Real-time classification for autonomous drowsiness detection using eye aspect ratio. *Expert Systems with Applications*, *158*. doi: 10.1016/j.eswa.2020.113505

Maior, C. B. S., et al. (2016). Remaining useful life estimation by empirical mode decomposition and support vector machine. *IEEE Latin America Transactions*, *14*(11), 4603–4610.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133. doi: 10.1007/BF02478259

Pecht, M. G. (2009). *Prognostics and health management of electronics*. Encyclopedia of Structural Health Monitoring.

Safizadeh, M. S., & Latifi, S. K. (2014). Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. *Information Fusion*, *18*, 1–8. doi: 10.1016/j.inffus.2013.10.002

Shen, C., et al. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*, *56*(5), e2019WR026731.

Shen, S., et al. (2021). A physics-informed deep learning approach for bearing fault detection. *Engineering Applications of Artificial Intelligence*, *103*, 104295. doi: 10.1016/j.engappai.2021.104295

Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, *25*(5), 1803–1836. doi: 10.1016/j.ymssp.2010.11.018

Soualhi, A., Medjaher, K., & Zerhouni, N. (2015). Bearing health monitoring based on hilbert huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, *64*(1), 52–62. doi: 10.1109/TIM.2014.2330494

Streck, L., Lins, I. D., Maior, C., Cavalcanti, T., Moura, C., & Leite, G. d. N. P. (2024). Bearings' vibration fault detection via physics-informed deep learning. *Sendai International Center*.

Thuerey, N., et al. (2021). Physics-informed deep learning for fluid mechanics. *Annual Review of Fluid Mechanics*, *53*, 477–508.

Yang, Y., Yu, D., & Cheng, J. (2007). A fault diagnosis approach for roller bearing based on imf envelope spectrum and svm. *Measurement*, *40*(9-10), 943–950. doi: 10.1016/j.measurement.2006.10.010

Yuan, L., et al. (2020). Rolling bearing fault diagnosis based on convolutional neural network and support vector machine. *IEEE Access*, *8*, 137395–137406. doi: 10.1109/ACCESS.2020.3012053

Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 665–674). doi: 10.1145/3097983.3098052