UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIAS E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

LEONARDO STRECK RAUPP

PHYSICS-INFORMED DEEP LEARNING FRAMEWORK FOR BEARING FAILURE DIAGNOSTICS USING VIBRATION SIGNAL

RECIFE
2025

LEONARDO STRECK RAUPP

PHYSICS-INFORMED DEEP LEARNING FRAMEWORK FOR BEARING FAILURE DIAGNOSTICS USING VIBRATION SIGNAL

Master's dissertation presented to Production Engineering Department of Federal University of Pernambuco, as partial requirement for obtaining the title of Master in Production Engineering.

Research field: Operational Research.
Advisor: Profa. Dra. Isis Didier Lins.

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Raupp, Leonardo Streck.
   Physics-Informed Deep Learning Framework for Bearing Failure Diagnostics Using Vibration Signal / Leonardo Streck Raupp. - Recife, 2025.
   60f.: il.

   Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Tecnologias e Geociências, Programa de Pós-Graduação em Engenharia de Produção, 2025.
   Orientação: Isis Didier Lins.
   Inclui referências.

   1. Physics-Informed Deep Learning; 2. Bearing Vibration; 3. Fault Classification; 4. Vibration Analysis. I. Lins, Isis Didier. II. Título.

UFPE-Biblioteca Central

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PARECER DA COMISSÃO EXAMINADORA

DE DEFESA DE DISSERTAÇÃO DE

MESTRADO ACADÊMICO DE

LEONARDO STRECK RAUPP

"PHYSICS-INFORMED DEEP LEARNING FRAMEWORK FOR BEARING FAILURE DIAGNOSTICS USING VIBRATION SIGNAL"

ÁREA DE CONCENTRAÇÃO: PESQUISA OPERACIONAL

A comissão examinadora, composta pelas professoras abaixo, sob a presidência do(a) primeiro(a), considera o(a) candidato(a) **LEONARDO STRECK RAUPP, APROVADO(A)**.

Recife, 25 de fevereiro de 2025.

gov.br Documento assinado digitalmente
ISIS DIDIER LINS
Data: 27/02/2025 22:50:07-0300
Verifique em https://validar.iti.gov.br

Profª. ISIS DIDIER LINS, Doutora (UFPE)

gov.br Documento assinado digitalmente
MARCIO JOSE DAS CHAGAS MOURA
Data: 28/02/2025 10:53:57-0300
Verifique em https://validar.iti.gov.br

Prof. MÁRCIO JOSÉ DAS CHAGAS MOURA, Doutor (UFPE)

gov.br Documento assinado digitalmente
SAVIO SOUZA VENANCIO VIANNA
Data: 10/03/2025 17:34:26-0300
Verifique em https://validar.iti.gov.br

Prof. SÁVIO SOUZA VENÂNCIO VIANNA, PhD (UNICAMP)

ACKNOWLEDGEMENTS

First and foremost, I am immensely thankful to my family – my parents, Jair and Lourdes, and my siblings, Nathalia, Bernardo and Nicolas – for your support, encouragement, and belief in me. Your teachings and guidance provided me with strength and motivation. Thank you for always having my back.

To my friends and colleagues – you know who you are. I extend my appreciation for your friendship, stimulating discussions, and moral support. Whether through work on projects, assistance offering, or simply a listening ear, your presence made this journey way more enjoyable.

I would also like to express my deepest gratitude to my advisor, Profª Drª Isis Didier Lins, for your invaluable guidance, patience, and support throughout my research journey. Your expertise, feedback, and encouragement were fundamental to this dissertation. I extend these greetings to Prof Dr Márcio Moura and all the CEERMA and PRH 38.1's team: I am truly grateful for the time you dedicated to reviewing my work, challenging my ideas, and helping me grow.

Finally, I acknowledge the broader academic community, including professors, peers, and institutions, whose work and resources have informed and enriched my research. This dissertation would not have been possible without the collective knowledge and inspiration drawn from scholars who have paved the way in this field. To everyone who played a role, big or small – thank you.

ABSTRACT

Ensuring the reliability of equipment and systems is of prime importance to all industries, in special the Oil and Gas (O&G) industry. This industry is known for its high complexity, operating in multiple uncertain and harsh conditions. Thus, the development of precise predictive models is essential. Pumps, turbines, compressors, among other rotating machinery, are widely used by the O&G industry and are considered critical parts of their systems. The bearing is a component common to all those machines and is responsible for up to 55% of the failures occurring in these machines. Therefore, this element is of major concern for optimization, in order to guarantee the system's Reliability, Availability, Maintainability and Safety (RAMS). Bearing elements have already been under focus of researches over time, but mostly over the bias of pure data-driven approaches. These methods, when dealing with complex systems or when available data does not cover all operating conditions, may result in false alarms that contradict the system's expected behavior. Those false alarms, then, may impact the prediction's confidence by operators. In order to solve the problems described, this research proposes a framework for bearing fault classification in a context applicable in the O&G industry. The methodology is composed of a Deep Learning (DL) model which is supplied with the system's known behavior via a threshold model, constrained within a customized loss function, an approach named Physics-Informed Deep Learning (PIDL). The approach is validated using two case studies: Paderborn University (PU) Bearing Data Center's Dataset and bearing vibration data from CEERMA's bearing vibration bench. Results show that the physical addition was capable of improving the performance of the pure-statistical approach, showing a 6% relative accuracy gain for case study 1. In case study 2, where accuracy gains were smaller, the PIDL model still demonstrated success when reducing the amount of misclassifications related to the extreme cases: healthy data classified as heavy damage, and vice-versa. On top of that, sensibility analyses were performed by varying the degree of guidance introduced by the threshold model's term. All these suggest the advantage of including available physical information into data-driven models.

Keywords: Physics-Informed Deep Learning. Bearing Vibration. Fault Classification. Vibration Analysis.

RESUMO

Garantir a confiabilidade de equipamentos e sistemas é de suma importância para todas as indústrias, especialmente para a indústria de Óleo e Gás (O&G). Essa indústria é conhecida por sua alta complexidade, operando em múltiplas condições adversas e com alto grau de incerteza. Portanto, o desenvolvimento de modelos preditivos precisos é essencial. Bombas, turbinas, compressores, entre outras máquinas rotativas, são amplamente utilizadas pela indústria de O&G e são consideradas partes críticas de seus sistemas. O rolamento é um componente comum a todas essas máquinas e é responsável por até 55% das falhas que ocorrem nessas máquinas. Portanto, esse elemento é de grande preocupação para otimização, a fim de garantir a Confiabilidade, Disponibilidade, Mantenabilidade e Segurança (RAMS) do sistema. Rolamentos já foram foco de diversas pesquisas ao longo do tempo, mas principalmente sob a ótica de abordagens puramente orientadas por dados, as chamadas *data-driven*. Esses métodos, ao lidarem com sistemas complexos ou quando os dados disponíveis não cobrem todas as condições operacionais, podem resultar em alarmes falsos que contradizem o comportamento esperado do sistema. Esses alarmes falsos, então, podem impactar a confiança na predição por parte dos operadores. Para resolver os problemas descritos, esta pesquisa propõe um *framework* para classificação de falhas em rolamentos em um contexto aplicável na indústria de O&G. A metodologia é composta por um modelo de Deep Learning (DL), o qual é alimentado com o comportamento conhecido do sistema por meio de um modelo de limiares, integrado em uma função perda personalizada, abordagem denominada Physics-Informed Deep Learning (PIDL). A abordagem é validada utilizando dois estudos de caso: o conjunto de dados da Universidade de Paderborn (PU) e dados de vibração de rolamentos gerados por uma bancada de vibração de rolamentos do Center for Studies and Trials in Risk and Environmental Modeling (CEERMA). Os resultados mostram que a adição física foi capaz de melhorar o desempenho da abordagem puramente estatística, apresentando um ganho de acurácia relativo de 6% no estudo de caso 1. No estudo de caso 2, onde os ganhos de acurácia foram menores, o modelo PIDL ainda demonstrou sucesso ao reduzir a quantidade de classificações incorretas relacionadas aos casos extremos: dados saudáveis classificados como contendo danos graves, e vice-versa. Além disso, foram realizadas análises de sensibilidade variando o grau de introdução do termo do modelo de limiar. Esses resultados sugerem a vantagem de incluir informações físicas em modelos orientados por dados, quando disponíveis.

Palavras-chave: Physics-Informed Deep Learning. Vibração de Rolamentos. Classificação de Falhas. Análise de Vibração.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

CEERMA	Center for Studies and Trials in Risk and Environmental Modeling
CCE	Categorical Cross-Entropy
CNN	Convolutional Neural Network
DL	Deep Learning
FFT	Fast Fourier Transform
ML	Machine Learning
NN	Neural Network
O&G	Oil and Gas
PHM	Prognostics and Health Management
PIDL	Physics-Informed Deep Learning
PINN	Physics-Informed Neural Networks
PU	Paderborn University
RAMS	Reliability, Availability, Maintainability and Safety
ReLU	Rectified Linear Unit activation function
RNN	Recurrent Neural Network
UFPE	Federal University of Pernambuco

CONTENTS

1 INTRODUCTION

Equipment and systems' reliability plays a central role in the success and competitiveness of the Oil and Gas (O&G) industry, and ensuring the reliability and safety of complex engineering systems is of prime importance (Yusuf *et al.*, 2014). This industry, specifically, commonly operate under harsh conditions, which often led to subpar performance of preventive and predictive approaches. As a result of this, decisions may include delayed maintenance activities, assumed critical risks or the adoption of bigger safety margins, which are not optimal for system management, possibly damaging the safety, reputation, cost or availability of the system (Arismendi *et al.*, 2021).

As machinery advances, it also grows in complexity while also often operating in harsh and uncertain conditions (Zhang; Wang, 2023). This complexity growth and variability can make systems more vulnerable, eventually leading to breakdowns (Orrù *et al.*, 2020). Thus, the capacity to create precise predictive models for system performance across various designs and operating conditions is essential and is often an important part of a Prognostics and Health Management (PHM) program (Barraza *et al.*, 2022) (Zhang *et al.*, 2023).

PHM is a methodology responsible for accessing the current state of an equipment or system, a process named diagnosis, in order to estimate its future health state, named prognosis (Pecht, 2009). The approach allows for early detection of faults, aiming to reduce downtime and maintenance operation costs, to assist proactive responses, and to enhance not only the productivity but also the Reliability, Availability, Maintainability and Safety (RAMS) of a system, which is an important step to achieve maintenance management in a system scope (Xu; Xu, 2011). Therefore, PHM is essential for ensuring optimal performance and longevity of equipment, especially in critical sectors like aviation, wind energy, and industrial robotics, where equipment failure can have severe economic and safety consequences (Wang *et al.*, 2023).

In regards to rotating machinery, such as pumps, turbines and compressors, those are usual to the O&G context and are considered critical parts of their systems (Shen *et al.*, 2021). These machines have an element in common, the bearing, which is responsible for their rotating capabilities and, thus, this rolling element is also widely present in the O&G industry (Petrovsky *et al.*, 2019) (Orrù *et al.*, 2020) (Aliyu *et al.*, 2022) (Barraza *et al.*, 2022). But, as pointed by Shen *et al.* (2021), the rolling element alone is the cause for up to 55% of failures in machines that contain those elements, consequently being a major concern when optimizing the proper availability of this industry. Due to this, bearing elements have been the focus of several studies over time, especially in terms of their vibration signals and in the field of predictive methodologies (Yuan *et al.*, 2020) (Ni *et al.*, 2023)

(Shutin *et al.*, 2024), showing an interest for such approaches.

According to Xu *et al.* (2023), predictive modeling methodologies fall into two main categories: physics-based approaches and Machine Learning (ML) based approaches. The first utilize the fundamental laws of nature to establish relationships between input parameters and system performance. These are well-suited for systems with well-understood processes but may face challenges in complex systems due to model simplification and computational complexity. On the other hand, ML approaches, such as Neural Network (NN) or Deep Learning (DL), leverage data collected from the system to make predictions, making them particularly useful for systems with poorly understood processes or where physics-based modeling is unfeasible. However, they may require substantial amounts of data and may struggle when extrapolating beyond the training set.

While Machine Learning, especially DL, models have shown promise in monitoring the health state of many types of equipment (Yan *et al.*, 2023) (Hou *et al.*, 2023) (Zhang *et al.*, 2024), the lack of integration of physical knowledge into these methods may result in false alarms, especially when available data is not enough to cover all operating conditions completely (Shen *et al.*, 2021). As a consequence, the prediction's confidence can be reduced if those false alarms contradict physical principles due to the model having insufficient characterization of the equipment under analysis, such as pumps (Aliyu *et al.*, 2022).

Hence, incorporating physical knowledge into the training process can enhance the interpretability and applicability of the models across different scenarios, as operators tend to better accept models that follow their expected knowledge of the system (Shen *et al.*, 2021). Also, as a result of the incorporation of such knowledge, models can learn effectively even without enough data of all operating conditions (Alzubaidi *et al.*, 2023). In order to solve those problems, there is a growth in the interest around Physics-Informed Machine Learning. Unlike purely data-driven, models with the incorporation of physical information prioritize features relevant to the expected behavior of the system, reducing misclassifications that contradict physical principles.

Therefore, this research focus on developing a framework to perform fault detection of bearings via Physics-Informed Deep Learning (PIDL) in a context applicable to the O&G industry. A regular Deep Learning model is constructed and constrained through its loss function by the physical laws of the system under analysis. The framework is then evaluated on bearings' vibration data from literature and generated using a vibration bench.

1.1 JUSTIFICATION AND RELEVANCE

The Oil and Gas industry has long been a core component of the global energy landscape, with this trend likely persisting for numerous years ahead (WANG *et al.*, 2024). Furthermore, this industry is crucial in terms of its substantial contribution to a country's economy and energy requirements, as stated by Mahmood *et al.* (2023); also, the authors emphasize the impact of an O&G infrastructure that is robust towards both predicted and unpredicted events: better environmental practices, continuous supply and more competitive prices. Therefore, improving reliability in this industry - whether by preventing failures or detecting them accurately - can significantly reduce its environmental, social, and economic impacts.

Bearing rolling elements is possibly considered the most critical part of a system that contains rotating machinery, including in the context of O&G, but are also primarily responsible for failures in rotating machines, with an estimated range of 45-55% of the failures being caused by them (Shen *et al.*, 2021). Among affected equipment, i.e., the ones that contain bearings, one may cite pumps, compressors, and turbines, all highly available in this industry and disposed of in processes such as irrigation, refining, transport, storage, manufacturing, and others.

Nonetheless, other approaches intrinsic to reliability assessment, such as parameter estimation, are also highly impactful in this context. Estimating Flow Rate many steps ahead (Khan *et al.*, 2020) (Franklin *et al.*, 2022), for example, allows the operation to react to any unwanted predicted variation in the flow. This is only an example of how impactful reliability assessment can be for the O&G industry as a whole.

As introduced, improving reliability for the context of Oil and Gas industry contributes directly in the social, environmental and economic aspects, supporting this research. In regards to the social impact, reliability assessment may reduce failures, unwanted variations in the production flow and product supply, which directly affects the society's routine. In terms of the environmental impact, it is crucial that every piece of the industry works as expected, in order to prevent hazards or leaks, which is strictly related to climate changes, possibility of extreme events and damage to the surrounding population and wildlife. Concerning the economic impact, as already discussed over this work, improving the RAMS of a system directly enhances its economic potential, thus having high impact in this aspect, which is a major concern not only for the company but also for the affected parts, such as the state, region and even the country. Therefore, this work's objectives are aligned with the strategic objectives of organizations.

Based on the aforementioned points, this research aims to develop a hybrid model, includ-

ing both statistical and known system information, to improve PHM in the O&G field. It is expected that, by correctly identifying the bearings' fault vibration patterns and including known behavior of the system into data-driven approaches, this work contributes to solving the aforementioned problems, while also providing valuable insights for the scientific community.

1.2 PROBLEM DESCRIPTION

Machine Learning, i.e., pure data-driven, models have proven successful in several tasks, such as: operation mode classification (Yuan *et al.*, 2020) (Orrù *et al.*, 2020), Remaining Useful Life prediction (Sikorska *et al.*, 2011) (Yan *et al.*, 2023), State of Health estimation (Hou *et al.*, 2023) (Zhang *et al.*, 2024), anomaly detection (Chandola *et al.*, 2009) (Chandola *et al.*, 2012) (Bakdi; Kouadri, 2017) (Aljameel *et al.*, 2022) (Bayazitova *et al.*, 2024), among others. However, in those models, available data is considered to be enough to completely characterize the event's behavior.

In the situations where data is scarce, statistical models alone may fail to model a certain operation condition, as it is generally difficult to extrapolate beyond training data. Data scarcity means not only lack of data as a whole, but also lack of data for specific cases. For example, if a large dataset fails to include abundant information related to a light damage operation condition, data-driven models will struggle to learn it, often misclassifying as either healthy or high damage. Thus, there is a potential for false alarms generated by Machine Learning, particularly Deep Learning, models used for monitoring the health state of various equipment types. Additionally, even if available data is enough, if a certain system is too complex, ML models' performance will be limited by its capability of extracting patterns (Shen *et al.*, 2021).

In industrial applications, like bearing fault detection, data may be limited or expensive to collect. A PIDL model leverages known physical principles in order to enhance robustness and performance, even with smaller datasets. Also, engineers and domain experts are more likely to trust a model that aligns with known physical principles. By embedding physical knowledge, the Physics-Informed Deep Learning model produces results that are not only accurate but also physically plausible, which is critical for decision-making in industrial applications (Shen *et al.*, 2021). Also, in real-world applications, vibration signals are often noisy and physical knowledge may improve fault detection and reduce false alarms. This is even more evident in the O&G industry, where bearings operate under diverse and challenging conditions.

In accordance with the above, the focus of this research is to address the problem of lack of integration of physical knowledge on pure data-driven Deep Learning models. By incorporating physics, it is expected that models better characterize the complete behavior of the system even in

situations where data is scarce (Alzubaidi *et al.*, 2023), while gaining advantages in terms of: 1) fewer data required; 2) faster training convergence; 3) reduced false alarms that are in disagreement with physics; 4) improved model confidence.

1.3 OBJECTIVE

1.3.1 General Objective

The main objective of this work is to develop a framework for bearing fault diagnosis in a context applicable to the O&G industry, formulated via Deep Learning models with the inclusion of additional physical information in order to improve its adequacy with physical expected behavior.

1.3.2 Specific Objectives

In order to fulfill this research's general objective, specific objectives are defined:

- To identify the best approaches for physical-statistical methods in the literature;
- To select, implement, and evaluate possible approaches – and combination of approaches – to perform bearing fault detection in the field of Physics-Informed Deep Learning;
- To find bearing vibration datasets in the context of the O&G industry;
- To project the experiment design and data extraction using a vibration experimental bench;
- To perform vibration analysis to the selected datasets;
- To implement, train, test, and analyze obtained results using the proposed method.

1.4 METHODOLOGY

The Problem Modeling method, which involves using mathematical techniques to describe some operation of a system (Gil, 2002), is the research classification concerning technical procedures. The Quantitative approach is the focus of this research, as it uses numbers to perform analysis and estimations (Gil, 2002). This study is of an applied nature, driven by practical interest, to solve real-world problems (Marconi; Lakatos, 2002). Regarding its objectives, the research can be classified as exploratory, implying procedures for investigating areas lacking knowledge and not extensively explored in the existing literature (Gil, 2002).

This study aims to identify bearing faults using Deep Learning methods supported by the bearings' expected vibration behavior. First, the datasets used are described and analyzed. Subsequently, each dataset is preprocessed using a sliding window, Hilbert transforms, and Fourier transforms, and their most important frequencies are selected based on the bearing faults characteristic

frequencies' equations, followed by a train-test split. Two identical DL models are constructed for each dataset, with their structures specific for each dataset's complexity, with the second copy using a customized loss function instead of the traditional one. This customized loss function is similar to the regular, but includes a penalty based on the prediction of a threshold model: this model predicts the expected class of a data point based solely on its frequency amplitudes around the bearing faults' characteristic frequencies. Then, the models are trained and the performance of the models are evaluated, compared and discussed.

1.5 WORK ORGANIZATION

The contents of the upcoming chapters of this dissertation are briefly described below:

- **Chapter 2:** the theoretical background of essential concepts, and the literature review and state-of-art of Physics-Informed Deep Learning approaches, especially in the context of bearings and/or O&G industry;
- **Chapter 3:** the detailed description of the datasets used and the proposed methodology for bearing fault classification using statistical and physical information;
- **Chapter 4:** the results obtained for each case study, the sensitivity analysis for the degree of introduction of known information and discussion;
- **Chapter 5:** the concluding remarks, contributions and future steps.

2 THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1 THEORETICAL BACKGROUND

2.1.1 Machine Learning

Machine Learning can be shortly described as learning through data. This process focuses on the development of algorithms and statistical models that enable computers to learn and improve their performance on a specific task through experience or data. The key idea behind is to enable computers to automatically learn and adapt from patterns or features in data, rather than relying solely on explicit instructions provided by programmers (Mitchell, 1997).

According to Mitchell (1997), ML is a computer program that relies on three key points: a task to be executed, which is the objective for the program to solve; a performance metric, which evaluates its capability at solving the task; and the experience, which is the set of data that the program extracts information from. Its applications, as stated by the author, include image and speech recognition, natural language processing, recommendation systems, finance, healthcare, and many others. The field continues to evolve rapidly, with ongoing research and advancements driving innovation in both algorithms and applications.

At its core, Machine Learning involves the process of training a model on a dataset to recognize patterns or relationships within the data, which involves feeding the model with labeled examples (input-output pairs) or unlabeled data, depending on the type of learning algorithm being used (Goodfellow *et al.*, 2016). During training, the model adjusts its internal parameters or structure to minimize the difference between its predictions and the actual outcomes. Once trained, the model can then be used to make predictions or decisions on new, unseen data, a process known as inference (Goodfellow *et al.*, 2016). According to the authors, ML models can be categorized, based on the training type, into several types, including:

- Supervised Learning: the model is trained on a labeled dataset, i.e., where each input is associated with a corresponding output. The goal is to learn a mapping from inputs to outputs, enabling the model to make predictions on new, unseen data;
- Unsupervised Learning: the model is trained on an unlabeled dataset, only providing it with the data input itself. The goal is to discover patterns or structures within the data without explicit guidance. Clustering and dimensionality reduction are common tasks in unsupervised learning;
- Semi-supervised Learning: this type of learning combines elements of both supervised and unsupervised learning, where the model is trained on a dataset that contains both labeled and

unlabeled data. This approach can be useful when labeled data is scarce or expensive to obtain;

- Reinforcement Learning: the model learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The goal is to learn to perform decision sequences, i.e., policies, in order to maximize or minimize certain cumulative reward.

The effectiveness of a model is typically evaluated based on its ability to generalize well to unseen data, also denominated generalization (Goodfellow *et al.*, 2016). It is a critical aspect of model evaluation as, normally, its primary goal is to make accurate predictions or decisions on data it hasn't encountered during training. The results obtained on the training data alone does not guarantee its effectiveness in real-world scenarios. A model that performs well on the training data but poorly on new data is said to be overfitted, meaning it has learned to capture noise or irrelevant patterns specific to the training set rather than generalizing to the underlying patterns present in the data (Goodfellow *et al.*, 2016).

2.1.2 Neural Networks

Neural Networks, proposed by McCulloch and Pitts (1943), represent a class of Machine Learning models that mimic the structure and function of the human brain, particularly in their ability to learn from data. At their core, NNs consist of interconnected computational units called neurons or nodes, organized in layers, which typically include an input layer, one or more hidden layers, and an output layer (Goodfellow *et al.*, 2016). Each neuron receives input signals, processes them using an activation function, and produces an output signal that is transmitted to neurons in the next layer. The strength of the connections between neurons, known as weights, determines how information flows through the network and influences the output produced by the model.

During training, Neural Networks learn to perform tasks by adjusting their weights based on the error between their predictions and the actual targets in the training data, i.e., a label, for supervised problems. This process, known as back-propagation (Goodfellow *et al.*, 2016), involves updating the weights using optimization algorithms such as gradient descent to minimize a predefined loss function. For example, in a multi-class classification problem, such as a fault detection based on three classes (e.g., healthy, light damage and heavy damage), the most suited loss function is the Categorical Cross-Entropy (CCE) (Murphy, 2022). For a single data point, it is calculated as in Equation 2.1.

$$CCE\ Loss_{single} = -\sum_{c=1}^{C} \left[Y_{true,c}\ \ln\left(Y_{pred,c}\right) \right] \tag{2.1}$$

Where: C is the number of classes in the problem; Y_{true} is the true label, such that $Y_{true,c}$ is 1 when the true label is c, and 0 otherwise; and $Y_{pred,c}$ is the predicted probability for the class c. For the total loss for a given dataset, the average loss is taken among all data points, as exemplified in Equation 2.2.

$$CCE\ Loss_{dataset} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\left[Y_{true,n,c}\ \ln\left(Y_{pred,n,c}\right)\right] \tag{2.2}$$

Where N is the number of data points in the dataset; C is the number of classes in the problem; Y_{true} is the true label, such that $Y_{true,n,c}$ is 1 when the true label of data point n is c, and 0 otherwise; and $Y_{pred,n,c}$ is the predicted probability for the class c of data point n.

One of the key points of Neural Networks is their ability to automatically extract meaningful features from raw data, making them well-suited for tasks such as image and speech recognition, natural language processing, and pattern detection. This is reinforced by Hornik (1991), which defines a Neural Network with a single hidden layer and a finite number of neurons as a universal approximator of any continuous function.

2.1.3 Deep Learning

Deep Learning is a sub-field of Machine Learning that focuses on the development and training of Neural Networks with multiple hidden layers, with the first apparitions of the term 'deep' being made by Hinton and Salakhutdinov (2006). In essence, these Deep Neural Networks emphasize the training and optimization of complex NNs, allowing them to effectively learn higher order representations of data, enabling them to solve complex tasks with high accuracy. The feature extraction potential is improved by the hierarchical organization of layers, where lower layers capture simple patterns and higher layers learn more abstract representations (Goodfellow *et al.*, 2016).

As a counterpoint to Hornik's definition of Neural Networks with one layer as universal approximators, which might need an infinitely large number of neurons (Hornik, 1991), Bengio (2009) proves that a Deeper Network can obtain similar results using significantly less neurons, requiring fewer parameters, computational cost and training data. Additionally, Deep Learning includes, relative to Neural Networks, different types of layers, suited for different cases (Goodfellow *et al.*, 2016). Some of them are:

- Fully Connected (or Dense) layer: dense layers are the original layers from Neural Networks, and also the simplest and most commonly used. In this layer, every neuron in the layer is connected to every neuron in the preceding layer. These layers are effective for learning complex

relationships between features in the data and also to model the network to a desired shape, such as the output shape.

- Convolutional Layer (Convolutional Neural Network (CNN)): convolutional layers are particularly well-suited for processing grid-like data such as images. They apply a series of learnable filters (kernels) across the input data to extract spatial hierarchies of features. Due to this, CNNs are translation-invariant layers, which help capture patterns that may be in different positions on different inputs, making them essential in tasks like image classification, object detection, and segmentation.
- Recurrent Layer (Recurrent Neural Network (RNN)): recurrent layers are designed to process sequential data by maintaining an internal state (hidden state) that captures information about previous elements in the sequence. RNNs can effectively model dependencies and temporal relationships in sequences of variable length, making them suitable for tasks such as natural language processing, time series prediction, and speech recognition.

The success of this family of models can be attributed to several factors, including the growth in the availability of large-scale datasets, advances in computational resources (e.g., GPUs and TPUs), and improvements in optimization algorithms (e.g., stochastic gradient descent and its variants) (Goodfellow *et al.*, 2016). Additionally, advanced techniques such as regularization, dropout, batch normalization, and transfer learning have contributed to enhancing its performance and generalization capabilities.

2.1.4 Physics-Based Machine Learning

Willard *et al.* (2022) named as Physics-Based Machine Learning the fusion between Machine Learning and science-based knowledge. This family of models are especially useful due to major problems underlying pure models. Pure data-driven approaches do not require in-depth knowledge about the physical system, but often struggle to perform on multiple operation conditions if the training data do not include enough information about all of them. On the other hand, pure physical approaches are essential when data is scarce, but are highly sensitive to models developed and require high level of knowledge about the system's behavior.

The Physics-Based Machine Learning models allow the construction of alternative modelings that can reproduce the physical behavior of the system with low computational cost and data requirement (Willard *et al.*, 2022). The authors suggest the categorization of these into the groups:

- Physics-informed loss functions: aim to integrate physical knowledge into the loss function used for Machine Learning training. This integration ensures that the model captures dynamics

consistent with physical laws. A common approach involves incorporating physical principles directly into the loss function. This group is also named Physics-Informed Machine Learning, also including the subgroups Physics-Informed Neural Networks and Physics-Informed Deep Learning.

- Physics-informed initialization strategies: involve initializing the model's state (weights and biases) using insights from physical knowledge. This approach helps prevent the model from getting stuck in local minima, as can occur with random initialization, and accelerates convergence during training.
- Physics-informed architectural design: focuses on constructing neural network architectures that maintain the characteristics of the problem being addressed. This can involve designing node connections to capture dependencies among variables or assigning physical significance to neurons within the neural network.
- Residual modeling approaches: involve training Machine Learning models to predict the errors (or residuals) generated by physics-based models. These models operate simultaneously using the same input signals.
- Hybrid physics-Machine Learning models: represent a broader approach that includes residual modeling. One method of developing such models involves treating the output of a pure physics-based model as an additional input for the physics-ML model. Alternatively, other strategies replace components of a physics-based model that are poorly modeled using physics with Machine Learning models.

This work opted to explore the PIDL approach. This decision was a result of internal collaboration with previous studied topics, and the addition of physical knowledge into existing Deep Learning models is a natural follow up to the research. On top of that, literature shows an upwards trend related to Physics-Informed Deep Learning research, showing interest by the academic and technical fields on this approach.

2.1.5 Physics-Informed Deep Learning

According to Thuerey *et al.* (2022), Physics-Informed Deep Learning is a category of models that introduce the dynamics of a physical system, or parts of it, into a Deep Learning model's loss function. The essence lies in formulating the loss function such that it incorporates the fundamental principles governing the system under consideration. This typically involves expressing physical laws or constraints within the loss function.

The learning process is similar to traditional DL models, which revolves around iteratively

evaluating the loss function to quantify the discrepancy between model predictions and desired outcomes. In the case of Physics-Informed training, these loss terms not only guide the optimization procedure towards minimizing prediction errors, like a regular Deep Learning model would, but also ensure that the learned model adheres to the underlying physical laws governing the system due to the incorporation of a physical term to it (Thuerey *et al.*, 2022).

What sets apart these loss terms is their ability to capture the intrinsic structure of the problem domain, constraining the solution space to physically meaningful regions. This is particularly relevant in scenarios where data may be scarce or noisy, and explicit incorporation of prior knowledge about the system's behavior is essential for robust learning. The term "Physics-Informed" encapsulates this approach, where the learning process is guided not only by empirical data, as in the original pure statistical approach, but also by the underlying physical laws or constraints (Shen *et al.*, 2021). This not only improves the final trained model, but also tends to speed up the training procedure, given the additional source of information to extract patterns from (Thuerey *et al.*, 2022).

As an example, one can consider a PIDL approach that consists of a DL classification model θ and a physics classification model ρ. Both models are independent on their predictions for a given input data X_i. In order for the Deep Learning model to be Physics-Informed, ρ model's predictions can be incorporated into θ's loss function. As a possible approach, one can calculate an alternative Categorical Cross-Entropy using the output of ρ, Y_ρ, as the true label vector and the output of θ, Y_θ, as the predicted probabilities for each class, as shown, for a single data point, in Equation 2.3.

$$CCE\ Loss_{custom} = -\sum_{c=1}^{C} \left[Y_{\rho,c}\ \ln\left(Y_{\theta,c}\right) \right] \tag{2.3}$$

Where C is the number of classes in the problem; ρ is the physics classification model; θ is the DL classification model; Y_ρ is the label predicted by ρ model for this data point: $Y_{\rho,c}$ is 1 if ρ predicted that this data point is of label c, and is 0 otherwise; and Y_θ is the probability vector predicted by θ model for this data point, $Y_{\theta,c}$ being the specific probability predicted for the class c. This Categorical Cross-Entropy Loss Function can be interpreted as the disagreement between the physical's expected behavior and the statistical prediction.

As even Physics-Informed models still require significant statistical information in their learning process (Ren *et al.*, 2023), the alternative CCE of Equation 2.3 cannot be used alone as the sole loss function during training. Thus, it can be combined to the original CCE of a Deep Learning model, shown in Equation 2.1, using a weight term α. The resulting Categorical Cross-Entropy is shown in Equation 2.4.

$$CCE\ Loss_{combined} = \left\{ -\sum_{c=1}^{C} \left[Y_{true,c}\ \ln\left(Y_{\theta,c}\right) \right] \right\} + \alpha \left\{ -\sum_{c=1}^{C} \left[Y_{\rho,c}\ \ln\left(Y_{\theta,c}\right) \right] \right\} \tag{2.4}$$

Where C is the number of classes in the problem; ρ is the physics classification model; θ is the DL classification model; Y_{true} is the true label, such that $Y_{true,c}$ is 1 when the true label is c, and 0 otherwise; $Y_{\rho,c}$ is 1 if ρ predicted that this data point is of label c, and is 0 otherwise; and $Y_{\theta,c}$ is the probability predicted by θ for the class c of this data point.

The interpretation of this combination is that the physical CCE is seen by the training process as a penalty. Originally, only the true labels Y_{true} were penalizing the loss function, guiding the training process to better match these labels. With the addition of the Physics-Informed term, the physical model's predictions Y_{ρ} also penalize it. This ensures that the model learns that the expected physical class given by ρ is also an important aspect to be mapped, but weighted by α, which controls the degree of importance of the physical's expected behavior, in relation to the true labels, for the training.

2.1.6 Bearing Vibration Monitoring

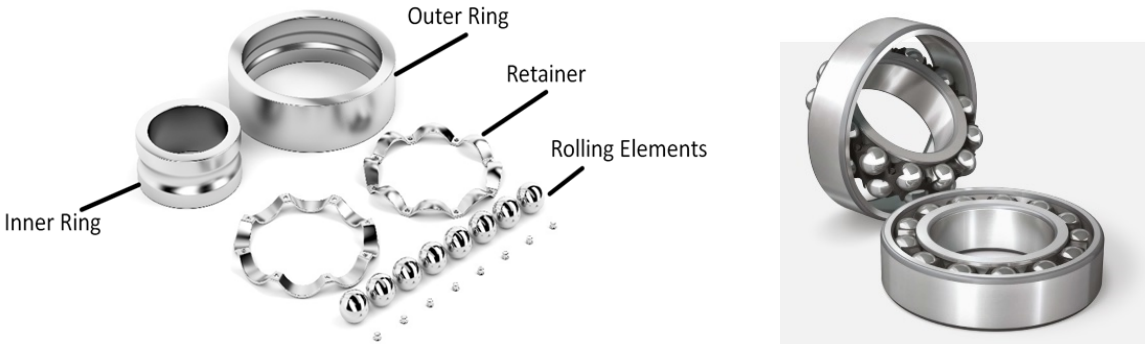
Machine vibration is a well-known and reliable way to monitor bearing condition (Jayaswal *et al.*, 2008). Bearings are largely responsible for the effective and dependable operation of mechanical transmission systems. Shafts and gears, among other rotating components, can move smoothly because they can support radial and axial loads. Self-aligning bearings are made to handle angular misalignments between the shaft and housing. Because of their special capacity to accept axial and radial misalignments, they are suited for applications where maintaining precise alignment is challenging or impossible. These bearings are often found in machines with flexible shafts or where alignment varies due to vibration, thermal expansion or component wear.

Extensive experiments are needed to gain a deeper understanding of the properties of vibrations in bearings, particularly in failure scenarios, as vibration monitoring has received more attention in recent years and has become more significant (Abbasion *et al.*, 2007). Vibration can be measured with vibration sensors, such as accelerometers and vibration speed transducers (Safizadeh; Latifi, 2014). Estimations should be taken on the direction or other structural components that strongly respond to the dynamic force and characterize the overall vibration of the machine. Figure 1a depicts a generic bearing consisting of an outside race, inner race, and moving component.

The reasons for bearing vibration include variations in external conditions over time between different parts. There are four types of inadequacies that can occur in a moving bearing,

Figure 1 – Rolling bearing's components.

(a) Generic Ball Bearing. (b) NSK 1205K C3 Bearing.



Source: a) TUOYUAN (Last access: 02/2025), b) NSK (Last access: 02/2025).

depending on where the fault occurs. The purported frequency of bearing defects is determined based on bearing parameters and rotational frequency. Each of these faults is associated with the following formulas used to calculate its specific frequency, as described in Table 1, where: N_B: number of rolling elements, B_D: ball diameter, P_D: bearing pitch diameter, ϕ: angle of contact, RPM: rotational speed.

Table 1 – Bearing characteristic frequencies: the frequencies in which failures occur.

Frequency Name	**Description**	**Equation**
Ball Pass Frequency Outer (BPFO)	Frequency at which failures occur in the outside lane.	$RPM\frac{N_B}{2}\left(1-\frac{B_D}{P_D}\cos\phi\right)$ (2.5)
Ball Pass Frequency Inner (BPFI)	Frequency at which failures occur in the inner lane.	$RPM\frac{N_B}{2}\left(1+\frac{B_D}{P_D}\cos\phi\right)$ (2.6)
Ball Spin Frequency (BSF)	Frequency at which the rolling elements themselves fail.	$RPM\frac{P_D}{B_D}\left(1-\left(\frac{B_D}{P_D}\cos\phi\right)^2\right)$ (2.7)
Fundamental Train Frequency (FTF)	Frequency at which a train's cage may fail.	$RPM\frac{1}{2}\left(1-\frac{B_D}{P_D}\cos\phi\right)$ (2.8)

Source: Randall and Antoni (2011).

According to Randall and Antoni (2011), envelope analysis stands as the primary method for bearing diagnosis, given that the raw signal typically lacks informative details about faults. This technique involves a filtering phase to remove frequency bands unrelated to the fault, followed by shifting the signal into the frequency spectrum, which emphasizes the repetitive nature of damage in rotating equipment. This shift is usually done via Fourier transform, using the Fast Fourier Transform (FFT) algorithm created by Cooley and Tukey (1965).

An enhancement to this analysis involves incorporating the Hilbert transform alongside

the time-to-frequency transform (Shen *et al.*, 2021) (Kanarachos *et al.*, 2017). This approach proves beneficial as it accentuates local features of the signal, producing an analytical representation of a real-valued signal. In the frequency domain, it introduces a 90º phase shift to all frequency components of a given function, aiding in the detection of instantaneous frequency changes by filtering out rapid oscillations from the signal.

2.2 LITERATURE REVIEW

2.2.1 Physics-Informed Deep Learning approaches for generic contexts

Bono *et al.* (2023) compared a simple autoencoder, a type of unsupervised DL model, with a Physics-Informed version of the same model in the context of structural health monitoring, showing that the Physics-Informed network exhibits greater precision in locating damage compared to the pure data-driven approach. Ni *et al.* (2023) also compared the performance of traditional statistic approaches to Physics-Informed ones using a novel approach called Physics-Informed Residual Network (PIResNet), which aims to learn the underlying physics embedded in train and test data, showing significant improvement in the results upon the usage of the latter. Boushaba *et al.* (2022) discuss two methods for detecting broken bars in induction motors, a Motor Current Signature Analysis with Convolutional Neural Networks, where measurements undergo frequency domain processing before training the CNN to ensure physical relevance, and a pure-statistic Principal Components Analysis in the time domain, applied to motor currents to perform anomaly detection via Q statistic, results showing that the CNN method offers more precise and reliable fault detection. The three works presented are valuable validation for Physics-Informed approaches as they all show that better results are yielded when combining physical knowledge into data-driven methods, incentivizing further research in this field.

Ren *et al.* (2023), similar to Bono *et al.* (2023), also introduce physical knowledge into an autoencoder model, but addressing sensor fault detection in Heating, Ventilation and Air Conditioning systems, showing higher fault detection rates and much lower false alarms. In this work, the authors highlight a key point when dealing with hybrid models: the trade-off between minimizing the statistical model's loss and the physical information's loss. Moderate reduction of the latter improved the generalization of the proposed method, whereas solely minimizing it compromised fault detection performance. This analysis suggests a need for other Physical-Informed approaches to carefully observe both losses' trade-offs while developing other models.

2.2.2 Physics-Informed Deep Learning approaches in the O&G industry context

Wang *et al.* (2023) also contribute to the Physics-Informed research field by demonstrating the effectiveness and robustness of a proposed Physics-Informed Neural Network, which involves establishing a discharge pressure model that incorporates physics information to describe instantaneous pressure changes in order to address the limitation of existing data-driven methods in fault severity identification for axial piston pumps, but offers no comparisons with regular NN models. Still in the context of pumps, Huang *et al.* (2020) use a hybrid Neural Network model for energy performance prediction of centrifugal pumps, which incorporates physical knowledge using a theoretical loss model, while also, differently from Wang *et al.* (2023), showing better results when compared to traditional NN models and even to a linear regression model. These works show great results upon using hybrid approaches for pumps, widely available in the O&G industry, which are a type of equipment that contains bearings – although their methodologies were not specifically designed for the bearings.

Aiming to replace or supplement costly and high-maintenance physical flow rate sensors, commonly used in O&G production systems, (Franklin *et al.*, 2022) introduce a novel application of Physics-Informed Neural Networks for the prediction of flow rates several steps ahead. Their method combines prior knowledge of system dynamics, represented by a phenomenological model, with the training of a Recurrent Neural Network, showing promising results. Their work is unique as no further works have been found that integrate physical knowledge into data-driven models for predicting flow rate, which shows a gap in the literature when dealing with real-world problems in the Oil and Gas industry's context via hybrid methodologies.

2.2.3 Physics-Informed Deep Learning approaches applied to bearings

Shen *et al.* (2021) propose a novel approach to improve bearing fault detection using a combination of physics-based knowledge and Deep Learning techniques. They highlight a limitation in existing Deep Learning methods, which often rely solely on data without incorporating the underlying physical principles of bearing faults. To overcome this limitation, the authors introduce a Physics-Informed Deep Learning approach that involves two main components: a threshold model and a Convolutional Neural Network, combined with a modified loss function design. The threshold model evaluates the health classes of bearings based on established physics principles related to bearing faults. The CNN then becomes more physically meaningful as it incorporates both data-driven features and domain-specific knowledge. This work demonstrates the effectiveness of a Physics-Informed Deep Learning approach in accurately detecting bearing faults, but the examples are applied to very simple

datasets, the Case Western Reserve University dataset (CWRU, 2019) and one generated by internal experiments. More complex datasets with operation modes that more accurately represent equipment of O&G industry, such as the PU Bearing Data Center's dataset (Lessmeier *et al.*, 2016), could be used to validate the proposed methodology further.

In the scope of developing synthetic data-driven dynamic models of fluid film bearings for rotor dynamics tasks, Shutin *et al.* (2024) test different models aiming to overcome computational inefficiencies of conventional numerical models. The research focuses on employing Machine Learning methods, particularly artificial Neural Networks, including single and multi-component models, along with Physics-Informed Neural Networks (PINN). A comparative analysis of these approaches in the context of rotor dynamics calculations shows that NNs prove flexible for managing model properties. In contrast, the multi-component approach enhances calculation speed and consistency. The PINN-based approach offers flexibility with unsupervised learning but may sacrifice some prediction accuracy. Overall, the study provides valuable insights into building models and comparisons between pure data-driven and PINN models. As the authors state, the study serves as a foundation for future research in this area, and further optimization of the models presented is possible.

Kim and Kim (2024b) present a Physics-Informed Deep Learning model designed to accurately diagnose bearing faults in noisy environments. The first innovation is a time-frequency multi-domain fusion block, designed to extract information from both time-domain vibration signals and their spectra in parallel. Then, the model parameters are informed by the physics of bearing faults, enhancing its performance by incorporating frequency-domain feature extraction. However, datasets were utilized in laboratory settings with the inclusion of Gaussian white noise, which may not fully replicate the conditions present in actual industrial environments. Therefore, future research should focus on collecting and analyzing fault-bearing vibration signals directly from various datasets to validate the practical applicability of the findings.

2.2.4 Overview

While many existing works focus on simple or single datasets, this work applies PIDL, similar to PINN but using Deep Learning as its underlying statistical model instead of Neural Network, to two datasets, under harder-to-identify conditions, which more accurately represents the high variability of operational conditions of bearings in the O&G industry. This addresses a critical gap in the literature, as industrial environments often involve more complex and noisy data.

Unlike some works, which did not compare their Physics-Informed Neural Networks to traditional NN models, this work provides a thorough comparison between PIDL and purely data-driven

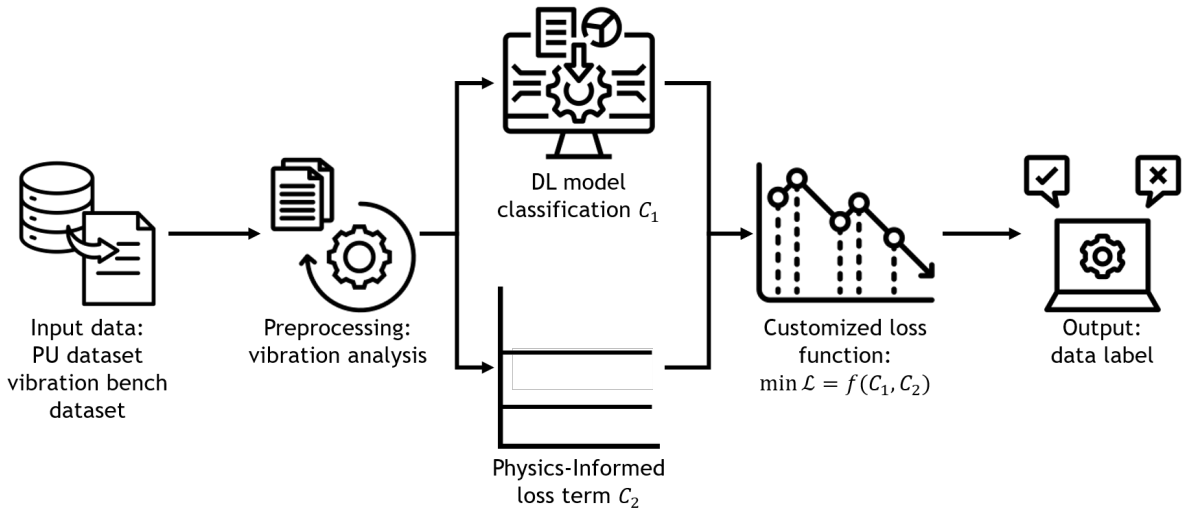
approaches. This comparison is crucial for demonstrating the added value of incorporating physical knowledge.

On top of that, much of the research focuses on other tasks, such as pumps and flow rate prediction. This work specifically targets bearing fault detection, which is critical for the reliability of rotating machinery in the O&G industry and is also highly applicable as bearings are present in multiple machines widely available in this context. This, alongside the previous points noted, enhances the practical relevance of the research. Thus, these points position this research as a valuable contribution to the ongoing development of hybrid models for fault detection in critical industrial equipment.

3 METHODOLOGY

Figure 2 illustrates the scope of the Physics-Informed model. The procedure follows the basic scheme of Artificial Intelligence modeling. This includes the steps of data loading and structuring, sequenced by applying preprocessing techniques, such as the Fourier Transform. Then, the Deep Learning architecture is built, with its loss function composed of the standard loss function for a classification problem, i.e., the CCE (Murphy, 2022), and a physics-informed loss term, resulting in a customized loss function. The model is trained on two datasets, and its prediction and results are analyzed and compared using pure statistical approaches. This methodology is also presented in Raupp *et al.* (2024a) and Raupp *et al.* (2024b) and is implemented in Python programming language version 3.10 using specialized libraries such as TensorFlow version 2.14 and Scikit-Learn version 1.3.1.

Figure 2 – Proposed framework.



Source: This work.

The adopted methodology focuses on incorporating a physical knowledge term into the loss function of a Deep Learning model. Here, this is done by constructing a threshold model based on the work of Shen *et al.* (2021), which guides the training step with the system's expected behavior. Thus, the principles of Physics-Informed Deep Learning approaches are attended (Thuerey *et al.*, 2022): combining system knowledge into model learning; reducing pure data and statistical dependence; and ensuring physically plausible results. Consequently, this work adopts the term "Physics-Informed".

The supporting model consists of linear separation thresholds that classify the input signal based solely on its highest amplitude in the frequency spectrum, i.e., if the amplitude is low, it is classified as healthy by this supporting model. The threshold for each class *c* is defined as the highest frequency amplitude among all data points of class *c* in the training set. For our case studies, the

classes evaluated were "Healthy", "Light Damage", and "Heavy Damage".

In terms of input data, several datasets were initially considered, such as Petrobras' 3W dataset (Vargas *et al.*, 2019) and the ESPset dataset (Pellegrini *et al.*, 2024), both related to real data obtained from the Oil and Gas industry. The former includes several failure modes, which could be adapted to a gradation condition, but does not include vibration data. The latter, on the other hand, is composed of vibration data, but with no gradation. Thus, both these datasets could not be included to evaluate the methodology proposed by this work. Thus, the choice of datasets had to keep these two points in mind.

Each step of the methodology presented in Figure 2 is further described in the following sections.

3.1 INPUT DATA

This section presents the datasets used to evaluate the proposed methodology.

3.1.1 Case Study 1: Paderborn University Dataset

The PU Bearing Data Center's dataset (Lessmeier *et al.*, 2016) is widely used as a benchmark dataset for Machine Learning models, especially in the field of predictive approaches (Zhao *et al.*, 2020) (Raupp *et al.*, 2024a). It includes vibration and current signals of bearings of type 6203 under several experiments designed to help research in the field. These bearings are not self-aligning.

The dataset is composed of 32 experiments in total, with varying criteria, based on Lessmeier *et al.* (2016):

- Type of damage: no damage, real damage, or artificially generated damage;
- Damaged bearing element: Outer Ring (OR), Inner Ring (IR), or both;
- Damage mode: fatigue or plastic deformation for real damaged bearings, electrical discharge machining or electric engraver for artificially generated damaged bearings;
- Damage combination: the type of damage occurring in the set, being either single damage, repetitive damage (a single damage type occurring multiple times), or multiple damage (different damage types occuring simultaneously);
- Damage extent: the intensity of the damage, ranging from levels 1 to 3, defined by the authors;
- Arrangement of repetitive/multiple damages: regular pattern, random distribution, or no repetition;
- Damage reach: whether or not a single bearing ball element is affected by all the combined damage.

From the total experiments, 6 were from healthy bearings, 12 from artificially introduced damage and 14 from accelerated lifetime tests. Each experiment was also conducted on four operational conditions, shown in Table 2. Some relevant specifications of the bearings in this dataset are shown in Table 3.

Table 2 – PU dataset's varying operational conditions. The parameters were defined by the authors of the experiment.

Number	Rotating speed (RPM)	Load torque (Nm)	Radial force (N)	Condition name
0	1500	0.7	1000	N15_M07_F10
1	900	0.7	1000	N09_M07_F10
2	1500	0.1	1000	N15_M01_F10
3	1500	0.7	400	N15_M07_F04

Source: Lessmeier *et al.* (2016).

Table 3 – PU dataset bearings' specifications.

Bearing model	6203
Inner diameter (d), mm	24.0
Outer diameter (D), mm	33.1
Pitch diameter (P_D), mm	28.55
Ball diameter (B_D), mm	6.75
No. of rolling elements (N_B)	8

Source: Lessmeier *et al.* (2016).

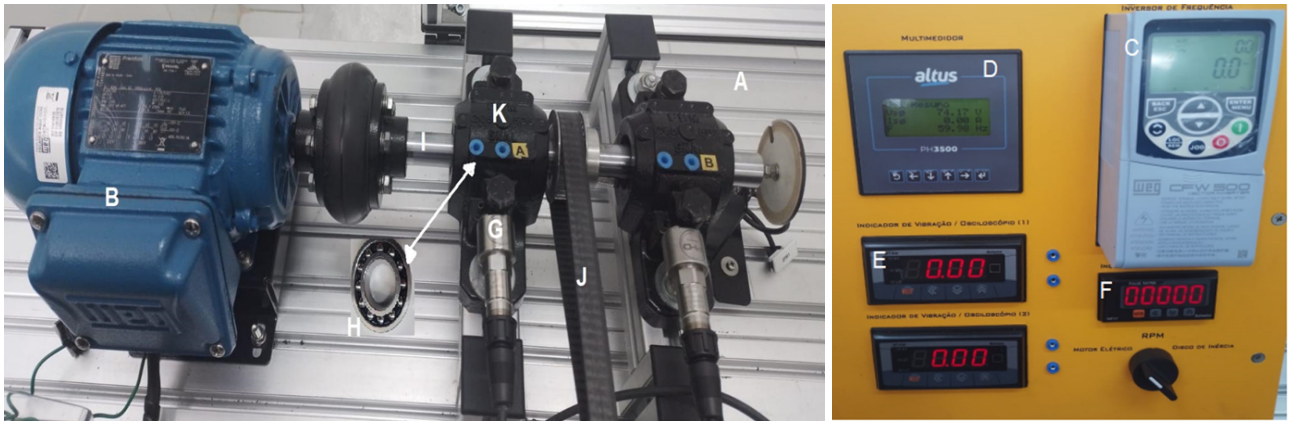
As briefly introduced in Chapter 3, our methodology comprises three classes: Healthy, Light Damage, and Heavy Damage. Among the multiple experiments, we chose to focus on inner ring damage, as it is harder to detect than the outer ring damage, according to the authors of the dataset (Lessmeier *et al.*, 2016). Also, in terms of the the damaged experiments, the dataset includes both simulated data and real data, with the latter obtained via accelerated lifetime tests. Thus, this work focused on the accelerated lifetime tests, as they represent damages caused by real bearing operation. Thus, we chose the experiments K004 (no damage) as our Healthy data, KI21 (extent of damage 1) as our Light Damage data, and KI16 (extent of damage 3) as our Heavy Damage data. This research selected the operation condition for all three data groups as 1500 RPM (25 Hz), 0.7 Nm load torque, and 1000 N radial force, or the condition number zero, as this include the most extreme condition for each variable considered. This selection is also presented in Raupp *et al.* (2024a).

3.1.2 Case Study 2: CEERMA Bearing Vibration Bench's Data

The data for the second experiment of this research was collected via a bearing vibration bench located in CEERMA – Federal University of Pernambuco (UFPE). The setup creates a transmission system via an induction motor through a frequency inverter, shown in Figure 3. The system comprises two bearings operating simultaneously in the housings shown by points A and K of Figure

3. Both bearings are interconnected through a synchronizer pulley with a guide, thus operating under the same frequency after being set in the inverter.

Figure 3 – Vibration bench's components.



Source: Raupp *et al.* (2024b).

This vibration bench was operated using the bearings of type NSK 1205K C3, shown in Figure 1b, which is defined as a double-row self-aligning deep-groove ball bearing. Both rows are contained within a spherical raceway located in the outer ring. This design enables the bearing to automatically adjust its alignment. Consequently, deviations in the alignment of the shaft with respect to the bearing housing do not harm performance (Raupp *et al.*, 2024b) (SNR, 1205K C3). The structure of the type of bearing used is shown in Figure 1 b), and its main geometric parameters can be found in Table 4.

Table 4 – Bearing (left) and accelerometer (right) specifications for vibration bench.

Bearing specifications		**Accelerometer specifications**	
Bearing model	NSK 1205K C3	Accelerometer model	603C01
Inner diameter (d), mm	25	Sensitivity	10.2 mV/(m/s²)
Outer diameter (D), mm	52	Measurement Range	± 490 m/s²
Pitch diameter (P_D), mm	38.5	Frequency Range	0.5 to 10000 Hz
Ball diameter (B_D), mm	7.14	Resonant Frequency	25 Hz
No. of rolling elements (N_B)	12	Broadband Resolution	3434 µm/sec²
		Non-Linearity	± 1%
		Transverse Sensitivity	≤ 7%

Source: Raupp *et al.* (2024b).

Using the magnetic method, the measuring process occurs via two accelerometers coupled to the base of each bearing's housing. The accelerometers used were piezoelectric, model 603C01, whose characteristics are shown in Table 4. The sensors measure 0.096mV per 1g of acceleration, translated to a Labjack U12 device. The sampling rate is 2048 samples per second, and a scan rate of 4096 Hz is used. All data collected, vibration data in acceleration units, in this experiment was acquired after setting the inverter frequency to 15 Hz or 900 RPM.

Similar to the data set presented in subsection 3.1.1, three damage levels are considered:

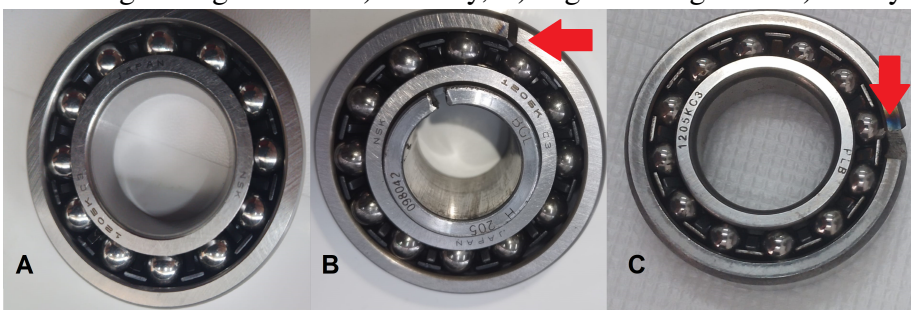
Healthy, Light Damage, and Heavy Damage (Raupp *et al.*, 2024b), whose damage levels are shown in Table 5 and Figure 4. For this case study, the bearing placed in housing A of Figure 3 is unchanged as always a Healthy bearing, and only the bearing in housing K is changed for the three experiments conducted. This experiment is also shown in Raupp *et al.* (2024b). Here, differently from Case Study 1, the Outer Race damage is used. This choice was derived from equipment limitations, as there was only one bearing with Inner Race damage available, thus not meeting the three damage levels' requirement of the methodology.

Table 5 – Bearing experiments' damage states.

Bearing state	**Damage extent**
Healthy (a)	No damage
Light damage (b)	Outer race damage 1mm
Heavy damage (c)	Outer race damage 3mm

Source: Raupp *et al.* (2024b).

Figure 4 – Bearing damage states: A) Healthy, B) Light Damage and C) Heavy Damage.



Source: Raupp *et al.* (2024b).

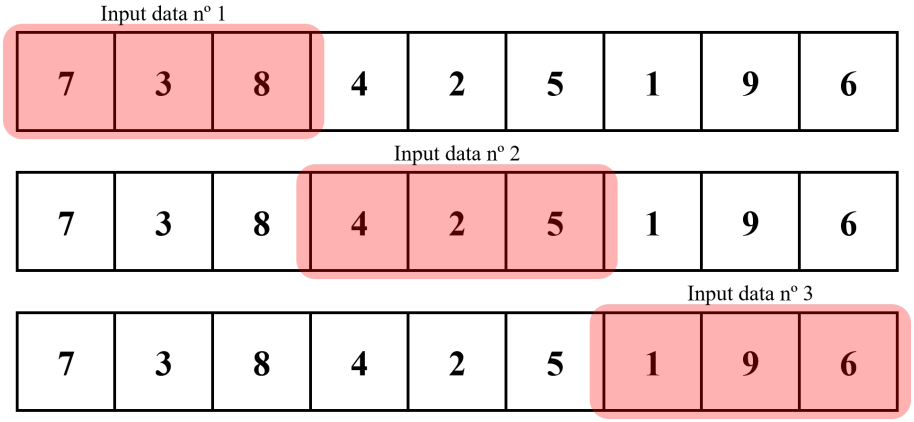
3.2 DATA PREPROCESSING

In order to better prepare our datasets for the Physics-Informed Deep Learning approaches, which incorporate the known behavior of the system into the models, the preprocessing is a crucial step (Song *et al.*, 2022) (He *et al.*, 2022) (He *et al.*, 2023) (Raupp *et al.*, 2024a) (Raupp *et al.*, 2024b). For both case studies, the general preprocessing approach is the same but fine-tuned according to each case, named study cases 1 and 2.

Both the PU dataset and the bearing vibration bench's data acquired are time series of a large length. In order to feed them to DL models, that large sequential amount of data must be split into segments of shorter length (Zhang *et al.*, 2019). This process is known as sliding window, segmentation, or windowing and is a crucial step of many approaches (Xu *et al.*, 2013) (Jaen-Vargas *et al.*, 2022) (Kim; Kim, 2024a) (Raupp *et al.*, 2024a) (Raupp *et al.*, 2024b). This technique is used in time series analysis to work with portions of the data at a time. An example of the procedure of the

sliding window is shown in Figure 5. Based on the example shown in the figure, given a time series of 9 data points and a sliding window of size 3 with no overlap, the first 3 time values of the time series will be considered 'input data 1', 4th to 6th values 'input data 2' and 7th to 9th values 'input data 3'; so, the processed time series consists of 3 data points, each consisting of 3 time values.

Figure 5 – Sliding window example.



Source: This work.

When dealing with case study 1, i.e., the PU dataset, the operation condition considered was 25 Hz and 64 kHz sampling rate, as opposed to 15 Hz operation speed and 2048 Hz sampling rate from the collected vibration bench's data. The sampling rate from case study 2 is the upper limit of the data acquisition device, hardware limitation, and the operation speed selected is the maximum operational speed in which the characteristic frequencies are still observable under the limited sampling rate.

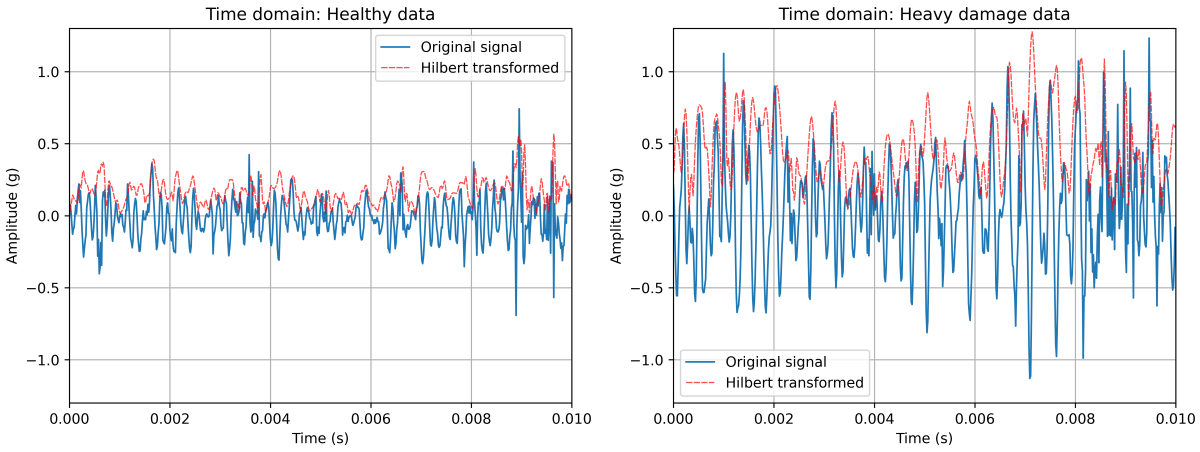
Thus, it is expected that higher sliding window sizes are required in the first case study as opposed to the latter. This is due to the nature of repeated events, such as rolling elements, occurring multiple times over the time series: each slice must include the fault information enough times for it to be accurately detected. Due to the much higher sampling rate, PU dataset's window will need many more points to include adequate fault information.

The procedure for choosing an appropriate value for the window size is to ensure that each window correctly contains multiple occurrences of the fault event of interest. In this work, the inner race and outer race faults are considered, for different case studies. Also, a very large window increases the complexity, as more harmonics of the signal will be considered for each window but much less data points will be available in the processed dataset. Based on the highest known frequency of these events, which is described later in this work, and in order to guarantee that multiple occurrences of each frequency appear in each window, we set the window sizes to 4096 points for case study 1 and 512 points for case study 2.

The next step is to perform envelope analysis of vibration data, responsible for transforming the signal from the time domain into the frequency domain, as shown by Lessmeier *et al.* (2016), Raupp *et al.* (2024a) and Raupp *et al.* (2024b), among others. This is a common procedure in bearing vibration analysis due to the repeating nature of the damage: the damage on a rolling element, due to the rolling process itself, has recurrent appearances in the signal, thus being better analyzed in the frequency spectrum.

The first step performed here is the application of a Hilbert transform, a common step of envelope analysis (Kanarachos *et al.*, 2017) (Zabin *et al.*, 2023). This transform constructs an analytical representation of a real-valued signal, which highlights its local features. The main impact of this, though, is seen after shifting to the frequency spectrum, as the Hilbert procedure provokes a phase shift of 90º to the signal, removing short-timed oscillations. This also allows the frequency information to be shifted to the left, allowing much more information to be obtained when analyzing a smaller frequency range. In the time-domain, a comparison example of the analytical signal constructed by the Hilbert transform and the original signal can be seen in Figure 6, for case study 1, and Figure 7, for case study 2. It can be noticed that the analytical signal tends to follow the upper limits of the original signal but filters sudden variations.

Figure 6 – Comparison of time-domain signal for PU dataset.
(a) Healthy data. (b) Heavy damage data.



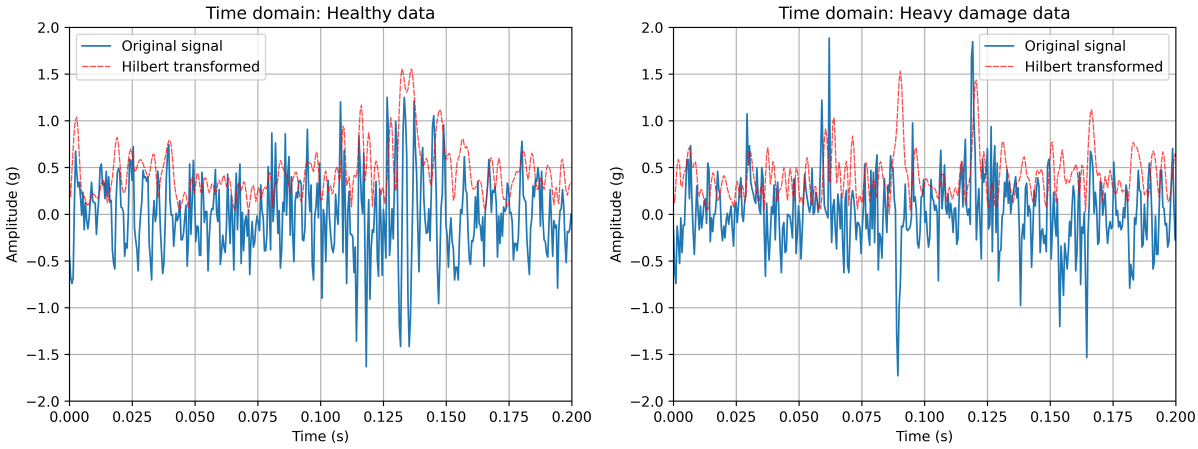
Source: This work.

Damages in bearings occur in known frequencies, as described in section 2.1.6, with the equations for each damage localization shown in Table 1. Through these equations and the bearing parameters for both case studies, Tables 3 and 4, we can calculate the characteristic frequencies for each, shown in Table 6.

Then, the frequency-domain spectrum is obtained by the Fourier Transform through the

Figure 7 – Comparison of time-domain signal for vibration bench's data.

(a) Healthy data. (b) Heavy damage data.



Source: This work.

Table 6 – Bearing characteristic frequencies for PU and vibration bench data.

Characteristic frequency	PU dataset	CEERMA's bench data
Ball Pass Frequency, Outer Race	76.3 Hz	73.3 Hz
Ball Pass Frequency, Inner Race	123.6 Hz	106.7 Hz
Ball Spin Frequency	99.8 Hz	78.1 Hz
Fundamental Train Frequency	9.5 Hz	6.1 Hz

Source: Raupp *et al.* (2024a) and Raupp *et al.* (2024b).

usage of the FFT algorithm (Cooley; Tukey, 1965) (Randall; Antoni, 2011) (Kim *et al.*, 2022). As introduced, this aims to better identify recurring patterns in the signal that indicate bearing damage. Each data point, after this transform, consists of a certain number, equal to the window size, of frequency values.

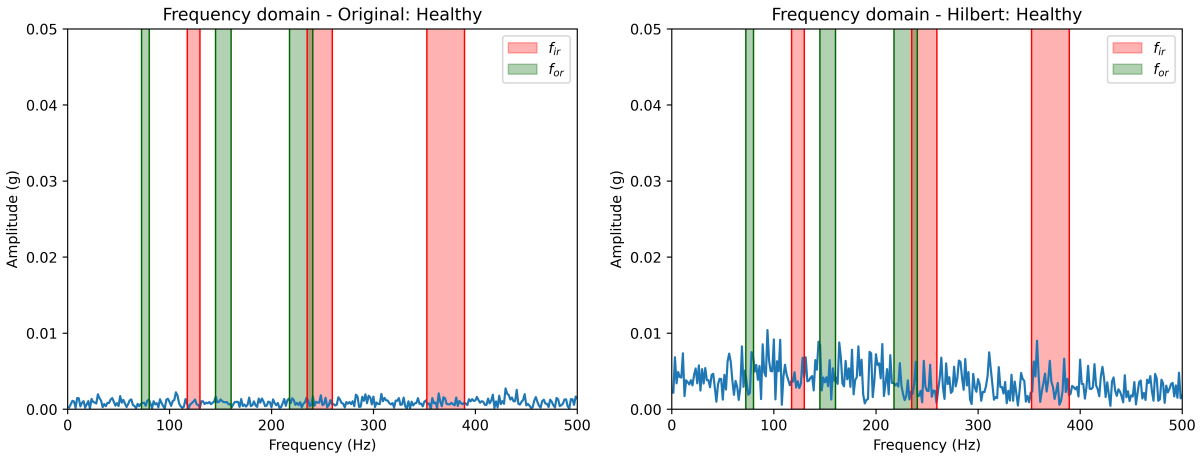
As a further preprocessing step, following the work of Shen *et al.* (2021), this research opts to enhance fault detection by using feature selection. From all the frequency values included in each data point, we select only specific frequencies: $\pm 5\%$ of the characteristic frequencies of the inner and outer races, also considering up to three harmonics (or multiples) for each. Including a bandwidth of $\pm 5\%$ instead of a single frequency value minimizes the effect of noise and small variations in the data acquisition (Shen *et al.*, 2021).

Figures 8 and 9 show the frequency-domain signal comparison between the original signal and its Hilbert-transformed variant for case studies 1 and 2, respectively; the blue line represent the frequency values, while the red and green sections represent the feature selection - only data values within the colored sections are maintained at the end of the preprocessing step. Here, the impact of the Hilbert transform is visible: the phase shift of 90º allows a large portion of the frequency information to be shifted to the left, allowing most of the frequency information to be obtained by analyzing fewer harmonics than the original signal would require: one can observe a crescent frequency trend beyond 500 Hz before the application of the Hilbert transform; this trend instead shifts to a decreasing trend

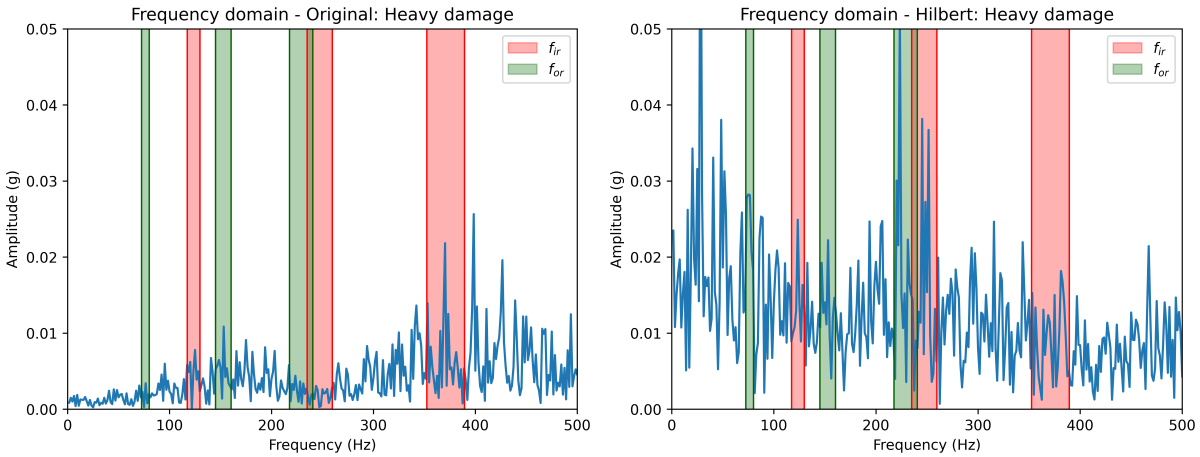
after the transform, showing that the frequency information shifted to the left. Thus, a smaller interval of frequency needs to be observed.

Figure 8 – Comparison of frequency-domain signal and selected sub-bands for PU dataset.

(a) Original signal: healthy. (b) Hilbert transformed: healthy.



(c) Original signal: heavy damage. (d) Hilbert transformed: heavy damage.



Source: This work.

The last preprocessing step is the division of the data into train and test sets. As common to Machine Learning approaches, this work adopted a 80%-20% split (Paulucio *et al.*, 2020) (Ghungrad *et al.*, 2023). Table 7 shows the final number of examples for each case study.

Table 7 – Train and test split for both case studies.

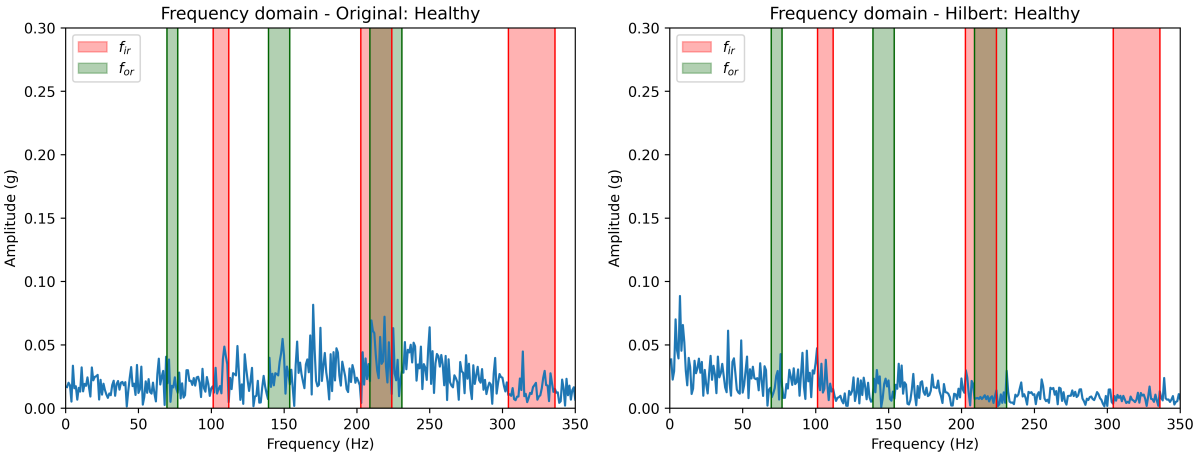
	PU dataset		**Vibration bench data**	
Subgroup	**# Examples**	**# Points per example**	**# Examples**	**# Points per example**
X_{train}	1647	27 (points within sub-bands)	5404	32 (points within sub-bands)
X_{test}	414	27 (points within sub-bands)	1352	32 (points within sub-bands)
Y_{train}	1647	3 (# classes)	5404	3 (# classes)
Y_{test}	414	3 (# classes)	1352	3 (# classes)

Source: Raupp *et al.* (2024a) and Raupp *et al.* (2024b).

Figure 9 – Comparison of frequency-domain signal and selected sub-bands for vibration bench's data.

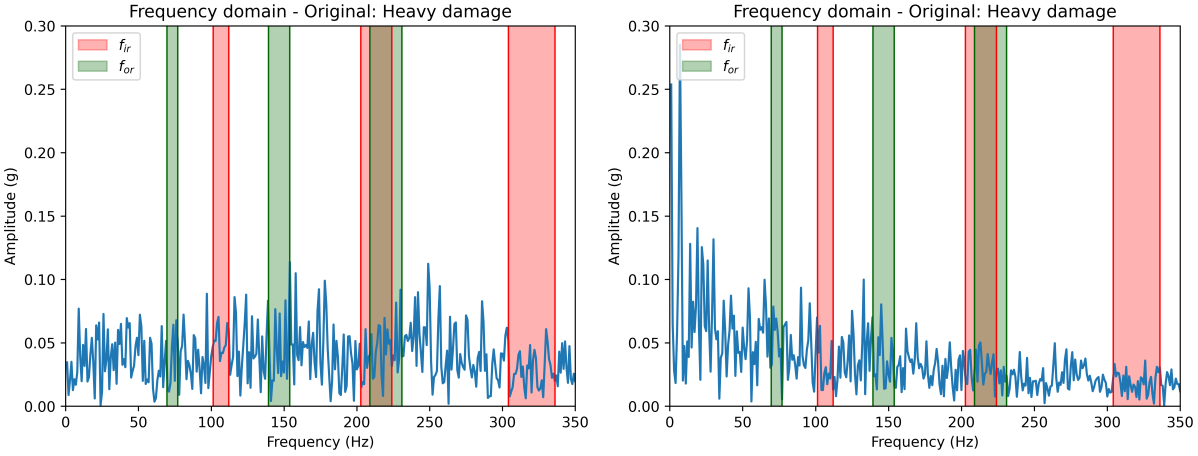
(a) Original signal: healthy.

(b) Hilbert transformed: healthy.



(c) Original signal: heavy damage.

(d) Hilbert transformed: heavy damage.



Source: This work.

3.3 NETWORK ARCHITECTURE

This research focuses on developing a framework for bearing fault detection in the context of the Oil and Gas industry by using a Physics-Informed Deep Learning model. Thus, our focus is not on creating a super-specialized network architecture but instead on validating the gains of the physics-informed term over the traditional pure statistical approach. Taking the PU dataset as an example, super-specialized models can reach over 99% accuracy (Zhao *et al.*, 2020), limiting the capability of observing the gains of the system's known behavior term.

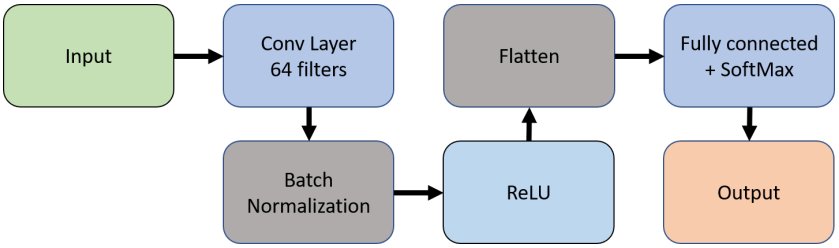
Therefore, we opt to construct base networks that can achieve good accuracy on their own but are not super optimized for each case. For each case study, the initial structure is a Convolutional block, composed of a CNN layer, a Batch Normalization and an activation function. For each network, it is started with one block, but increasing this number until a reasonable performance is reached, and ending with a fully connected layer - responsible for the classification. All the Convolutional Layers

used were one-dimensional, with stride 2 and padding 'same'.

The main argument for using a Convolutional Neural Network (CNN) is its property of invariability to translation (Goodfellow *et al.*, 2016). This means that, if we shift the input's position, the layer is still capable of detecting the same features as before the shift. This is a powerful property when we are not sure where the pattern will occur, thus, since we are considering several harmonics of the characteristic frequencies, we opted to use CNNs.

For the PU dataset, the network consists of a single CNN layer of 64 filters, followed by Batch Normalization and Rectified Linear Unit activation function (ReLU). Then, the output of the ReLU is flattened and fed to a fully connected layer with SoftMax activation function, which outputs the predicted class for the input data. Figure 10 shows the network architecture for this case study.

Figure 10 – Network architecture used for PU dataset.

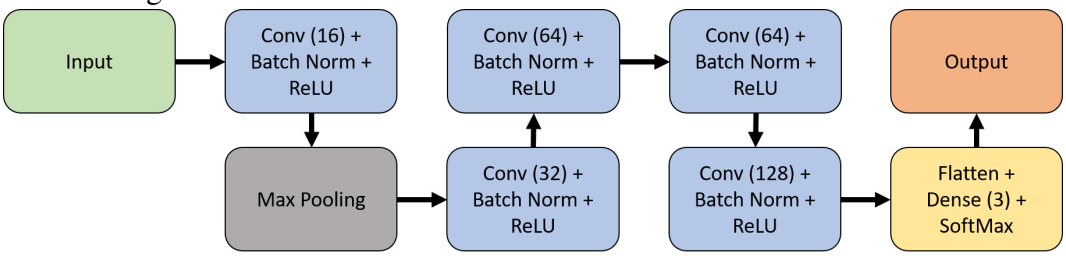


Source: Raupp *et al.* (2024a).

We identified that the vibration bench data was more intricate than the case study 1's data, possibly due to noise and residual misalignment after adjustments on the vibration bench. Thus, the network's architecture demanded more layers in order to deal with the added complexity. Then, we increased the number of convolution blocks – composed of a CNN, a Batch Normalization and a ReLU layers – to 5, with varying number of filters: 16, 32, 64, 64 and 128, respectively. The first block is different from the others, as it is followed by a Max Pooling layer, which halves the input dimension. This aims to force the model to learn only the most important features of the input, as most of its information will be lost by the pooling process. After the 5 Convolutional Neural Network blocks, the network follows the same procedure as case study 1: flatten, fully connected layer and SoftMax activation function, outputting the predicted class. This network architecture is shown in Figure 11.

For both case studies, two copies of the network architecture are built. The first one is kept as a purely statistical model, while the second is supplied with the system's known behavior through a customized loss function, described in the following section 3.4.

Figure 11 – Network architecture used for vibration bench's data.



Source: Raupp *et al.* (2024b).

3.4 CUSTOMIZED LOSS FUNCTION

The Physics-Informed term in our proposed approach refers to models that include a system's known behavior information by modifying its loss function, as described in section 2.1.4. Here, based on the work of Shen *et al.* (2021), the physical term is derived from a threshold model, which is included in the loss function of the DL model as an additional term (Raupp *et al.*, 2024a) (Raupp *et al.*, 2024b).

This threshold model comprises two thresholds: the maximum frequency amplitude for a data point to be considered healthy and the maximum frequency amplitude for a data point to be considered light damaged, based on the context where the number of classes C is equal to three: healthy, light damage, and heavy damage. The known behavior introduced here is that healthy data points are expected to show lower amplitudes within the filtered frequency sub-bands than light damage data ones, and, similarly, light damage data points are expected to be lower than high damage ones. The values for the thresholds are defined based on the maximum amplitude seen in the training set for the healthy – threshold 1 – and light damage – threshold 2 – classes. This procedure is applied to both case studies independently.

After constructing the threshold model, we implement the custom loss function. The main idea is that, for each input data, the Deep Learning classification model θ will make a class prediction based on its current weights. In parallel, the threshold model ρ will also predict the class for the same input data on the basis of its maximum frequency amplitude. The traditional loss function for Neural Networks in classification tasks is the Categorical Cross-Entropy between the true label of the input data and the label predicted by the model. Additionally, this work computes a different Categorical Cross-Entropy: using the threshold's output as the 'true label' and the Deep Learning's as the 'predicted label'. The interpretation of this procedure is that the model also considers the expected behavior of the system, i.e., the threshold model, as the true behavior when training the model. The exact procedure of the composed loss function is described below (Raupp *et al.*, 2024a) (Raupp *et al.*, 2024b):

1. CCE_θ is computed, for each data point, between the true label Y_{true} and the probabilities for each class predicted by the DL model Y_θ, according to Equation 2.1;
2. The threshold model outputs its prediction for the same input data Y_ρ based on the frequency amplitude of the signal;
3. CCE_ρ is computed for each data point, independently from CCE_θ, using the threshold's output label Y_ρ as true label and the probabilities for each class predicted by the DL model Y_θ as predicted label, according to Equation 3.1;

$$CCE_\rho = -\sum_{c=1}^{C} \left[Y_{\rho,c} \ln\left(Y_{\theta,c}\right) \right] \tag{3.1}$$

4. Filter CCE_ρ, based on Equation 3.2:
   4.1. Multiply by α all entries where Y_ρ is healthy ($argmax Y_\rho = 1$) but Y_θ is not ($argmax Y_\theta \neq 1$);
   4.2. Multiply by β all entries where Y_ρ is heavy damage ($argmax Y_\rho = 3$) but Y_θ is not ($argmax Y_\theta \neq 3$);
   4.3. Multiply by zero all other entries (no penalty added);

$$\begin{cases} CCE_\rho = -\alpha \sum_{c=1}^{C} \left[Y_{\rho,c} \ln\left(Y_{\theta,c}\right) \right], & \text{if } argmax Y_\rho = 1 \text{ and } argmax Y_\theta \neq 1 \\ CCE_\rho = -\beta \sum_{c=1}^{C} \left[Y_{\rho,c} \ln\left(Y_{\theta,c}\right) \right], & \text{if } argmax Y_\rho = 3 \text{ and } argmax Y_\theta \neq 3 \\ CCE_\rho = 0, & \text{otherwise} \end{cases} \tag{3.2}$$

5. Add both final CCEs, as shown in Equation 3.3.

$$\begin{cases} CCE_{combined} = \left\{ -\sum_{c=1}^{C} \left[Y_{true,c} \ln\left(Y_{\theta,c}\right) \right] \right\} + \\ \qquad \left\{ -\alpha \sum_{c=1}^{C} \left[Y_{\rho,c} \ln\left(Y_{\theta,c}\right) \right] \right\}, & \text{if } argmax Y_\rho = 1 \text{ and } argmax Y_\theta \neq 1 \\ CCE_{combined} = \left\{ -\sum_{c=1}^{C} \left[Y_{true,c} \ln\left(Y_{\theta,c}\right) \right] \right\} + \\ \qquad \left\{ -\beta \sum_{c=1}^{C} \left[Y_{\rho,c} \ln\left(Y_{\theta,c}\right) \right] \right\}, & \text{if } argmax Y_\rho = 3 \text{ and } argmax Y_\theta \neq 3 \\ CCE_{combined} = -\sum_{c=1}^{C} \left[Y_{true,c} \ln\left(Y_{\theta,c}\right) \right], & \text{otherwise} \end{cases} \tag{3.3}$$

Where C is the number of classes in the problem; ρ is the physics classification model; θ is the DL classification model; Y_{true} is the true label, such that $Y_{true,c}$ is 1 when the true label is c, and 0 otherwise; Y_ρ is the label predicted by ρ model for this data point: $Y_{\rho,c}$ is 1 if ρ predicted that this data point is of label c, and is 0 otherwise; and Y_θ is the probability vector predicted by θ model for this data point, $Y_{\theta,c}$ being the specific probability predicted for the class c.

As the main result of the usage of this procedure, it is expected that extreme classifications are minimized. These are defined as when the true class and the predicted class are completely opposites in regards to damage level, e.g., when healthy data is predicted as heavy damage or when heavy damage data are predicted as healthy. These cases are highly impactful in terms of model's confidence, as experts would fail in believing in the model if they see a nice-behaved data point – possibly healthy – being classified as heavy damage, and vice versa. As shown, the penalty is weighted – by α and β – when added to the traditional loss function. These weights range up to 10%. Thus, the model still relies significantly on statistical information: the expected behavior serves just as a training guide. Some works already state about the sensitivity of Physics-Informed Deep Learning models to the statistic-physics trade-off (Ren *et al.*, 2023), and this topic will be further discussed under the sensitivity analysis' subsections of the results chapter.

The term "Physics-Informed" is justified in this work due to the explicit incorporation of domain-specific physical knowledge, bearing characteristic frequencies and their amplitudes, into both the preprocessing procedure and the loss function construction. By preparing the input data to focus on the specific characteristic frequencies, and incorporating their amplitude into a custom loss function, penalizing predictions that deviate from expected physical behavior, the model is explicitly guided by expected physical behavior rather than relying solely on raw data. Thus, principles of PIDL methodologies are embedded (Thuerey *et al.*, 2022): including known information about the system into the learning process; reducing reliance on pure empirical training; and ensuring models are physically plausible.

4 RESULTS

This chapter presents the results and discussions related to the application of the methodology presented in chapter 3 on both case studies 1 and 2, shown in subsections 3.1.1 and 3.1.2, respectively. Here, we compare a purely statistical Deep Learning model with a proposed Physics-Informed Deep Learning one, which includes a system's known behavior into the DL's loss function via a threshold model. Both DL and PIDL models are identical except for their loss functions to better understand the gains of the physical information in the latter over the former. Each case study is presented separately in the following sections.

4.1 CASE STUDY 1: PADERBORN UNIVERSITY DATASET

The Paderborn University Bearing Data Center's Dataset, presented in subsection 3.1.1, comprises vibration signals of bearings under several experiments. For this research, we focused on three specific experiments under the following classes: healthy, light inner ring damage, and heavy inner ring damage.

After preprocessing the data, we identified the two threshold values – set as the maximum amplitude within the sub-bands in the training set among the healthy and light damage subsets, respectively – as 0.0494 and 0.0855, respectively. The models, described in section 3.3 and whose structure is shown in Figure 10, were trained for 20 epochs, using a batch size of 128 and ADAM optimizer with 0.001 initial learning rate. The learning rate is set based on values from literature, and the number of epochs is defined based on model convergence. The training time for all cases was below 5 seconds. The choice for a low number of epochs results from one of the expected advantages of PIDL approaches: the faster convergence due to the additional source of information.

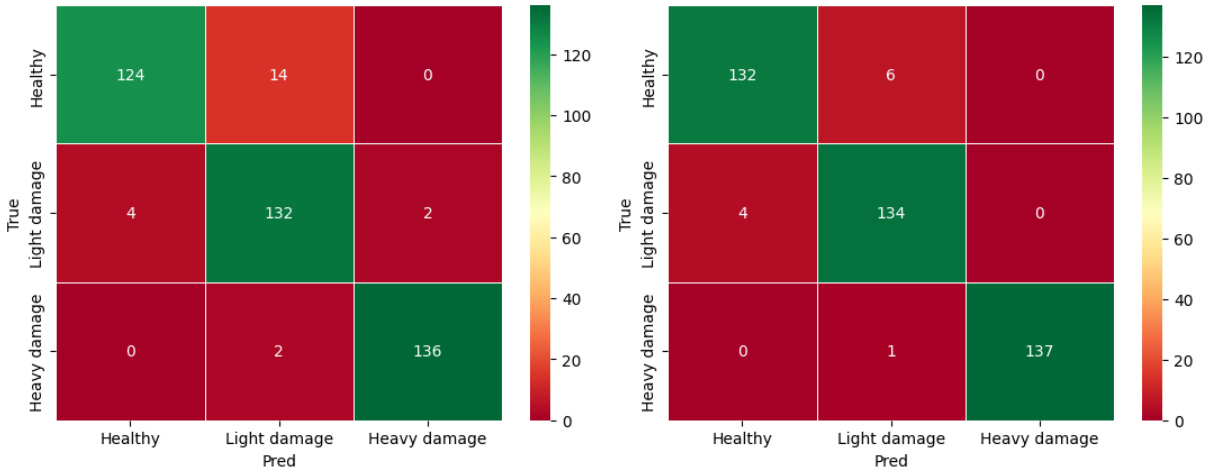
Figure 12 shows the confusion matrices for both the traditional Deep Learning and the proposed Physics-Informed Deep Learning models. For the low number of epochs used for training this dataset, the PIDL model was able to increase the accuracy of the predictions over the test set by a relative 2.8%, from 94.68% to 97.34%. When focusing on specific classes, one can observe that although all classes had performance improvements, the healthy one had the highest gain, ramping from 124 to 132 correct classifications, or a 6% relative growth. This indicates that the contribution based on the known behavior of the system is guiding the training process successfully. These results were obtained by empirically setting the penalty parameters as $\alpha = \beta = 0.02$, or 2% penalty.

Table 8 brings some classification metrics regarding the comparison between the models tested for Case Study 1: the traditional Deep Learning method and the Physics-Informed Deep

Figure 12 – Confusion matrices for PU dataset.

(a) Regular DL model: 94.68% accuracy.

(b) Physics-Informed Deep Learning model: 97.34% accuracy.



Source: Raupp *et al.* (2024a).

Learning one. For binary problems, true positive rate (sensitivity), true negative rate (specificity), positive predictive value (precision) and negative predictive value are commonly used when evaluating binary problems. Here, on a three-class problem, we use the generic terms 'true rate' (ratio of examples correctly classified as class c, from all examples of that class) and 'predictive value' (ratio of examples correctly classified as class c, from all examples classified as that class) for these calculations. The results show that, for every metric considered, the PIDL is superior to the purely statistical approach.

Table 8 – Comparison metrics for Case Study 1: Paderborn University Dataset.

			Healthy		**Light damage**		**Heavy damage**	
Model	**Acc.**	**Bal. Acc.**	**True Rate**	**Pred. Value**	**True Rate**	**Pred. Value**	**True Rate**	**Pred. Value**
DL	0.9468	0.9468	0.8985	0.9687	0.9565	0.8918	0.9855	0.9855
PIDL	**0.9734**	**0.9734**	**0.9565**	**0.9706**	**0.9710**	**0.9504**	**0.9927**	**1.0000**

Source: This work.

In order to proper validate the advantage of the PIDL over the DL approach, the same models were trained 10 times, each incurring a different train-test partition but using the same data to train both models. Since the data distribution is not assumed, a non-parametric test approach is chosen. Also, since the training procedure of both models uses the same dataset for each trial, a paired test approach is considered. Thus, the Wilcoxon Signed-Rank test was performed to compare the median of the accuracies obtained for each model when trained using the same dataset (Voorter *et al.*, 2023) (Ishida *et al.*, 2024). The null and alternative hypotheses for this test are presented in Equation 4.1

$$\begin{cases} H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 \\ H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2 \end{cases} \tag{4.1}$$

The null hypothesis represents that there is no statistical difference observed in the median accuracy of both groups. When this hypothesis is rejected, the alternative hypothesis is that statistical difference exists in the models' performances. The results are presented in Table 9. Based on the test statistic and the p-value presented, the null hypothesis is rejected, thus the Physics-Informed Deep Learning model is statistically different than its pure Deep Learning counterpart.

Table 9 – Wilcoxon Signed-Rank test for Case Study 1: Paderborn University Dataset.

Model	**Acc. 1**	**Acc. 2**	**Acc. 3**	**Acc. 4**	**Acc. 5**	**Acc. 6**	**Acc. 7**	**Acc. 8**	**Acc. 9**	**Acc. 10**
DL	0.9468	0.9589	0.9565	0.9614	0.9324	0.9396	0.9469	0.9348	0.9396	0.9662
PIDL	0.9734	0.9565	0.9614	0.9589	0.9638	0.9734	0.9686	0.9589	0.9734	0.9734
Test statistic	3.0									
p-value	0.009765625									

Source: This work.

4.1.1 Sensitivity Analysis

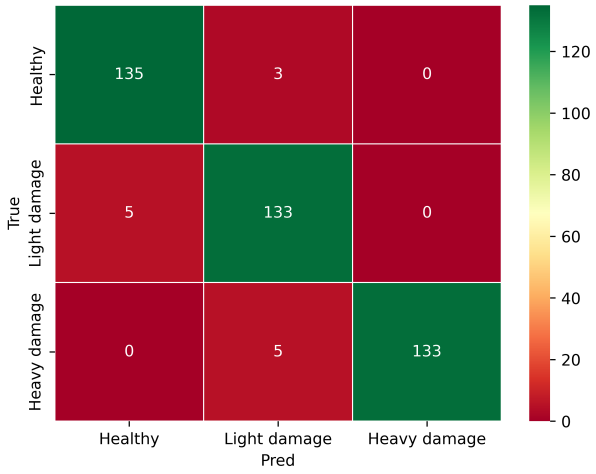
After the initial experiments, we varied the weights of physical contribution to better understand its impact on the prediction. This is an important step for Physics-Informed Deep Learning models because the trade-off between minimizing the statistical model's and physical information's losses may highly impact its generalization and fault detection performances, as pointed out by Ren *et al.* (2023). The core idea here is that since the loss is a minimizing function, it tends to update the parameters to obtain higher reductions in the value of the loss function. Thus, when adding an extra term to it, one must ensure that both terms have balanced, not necessarily equal, weights in terms of the minimizing function. If one term is comparatively much larger than the other, the optimization algorithm may only focus on updating the parameters to minimize that term exclusively, which, in our context, means focusing more on either statistical or physical information.

Based on the points noted above, this work performed two major sensitivity analyses experiments: #1 varying both weight parameters simultaneously and #2 fixing one parameter and varying the other. One may note that α penalizes DL's predictions that do not match the threshold's predictions when the threshold's prediction is healthy, regardless of data true label, as this comparison uses the threshold's prediction as an artificial true label for this term of the loss function, not the real true label. β operates similarly but it is related to the threshold's heavy damage predictions. Thus, the known behavior term guides the model so that the statistical prediction is similar to the prediction based on the frequency amplitude. Healthy data points still may have higher amplitudes than some heavy damage ones, so this known behavior guidance does not guarantee a perfect classification. This may occur in situations such as noisy data, imbalanced vibration bench, among others.

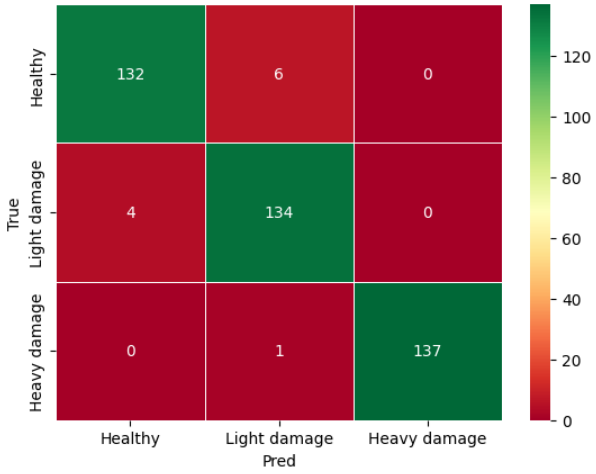
By varying α and β simultaneously, as seen in Figure 13, one may note that the performance of the prediction increases up to a certain point, $\alpha = \beta = 0.02$ in this case, and then starts to decrease. When outside the region of the best weight setup, the model tends to highly favor predicting data points as healthy. This is easily seen in the confusion matrices: in 13b, when the parameters are 0.02, a total of 136 points are predicted as healthy, while in 13d, when they are 0.05, 284 points are predicted this way.

Figure 13 – Confusion matrices – sensitivity analysis for PU dataset: varying α and β simultaneously.

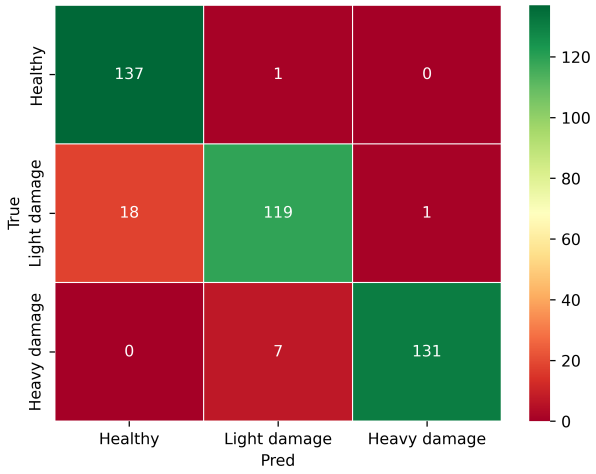
(a) $\alpha = \beta = 0.01$, 96.85% accuracy.



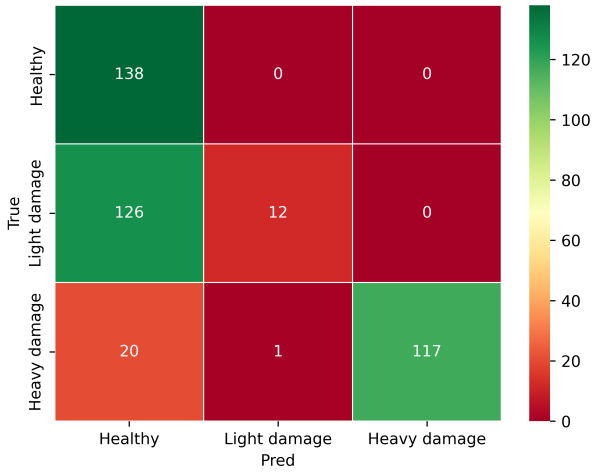
(b) $\alpha = \beta = 0.02$, 97.34% accuracy.



(c) $\alpha = \beta = 0.03$, 93.48% accuracy.



(d) $\alpha = \beta = 0.05$, 64.49% accuracy.



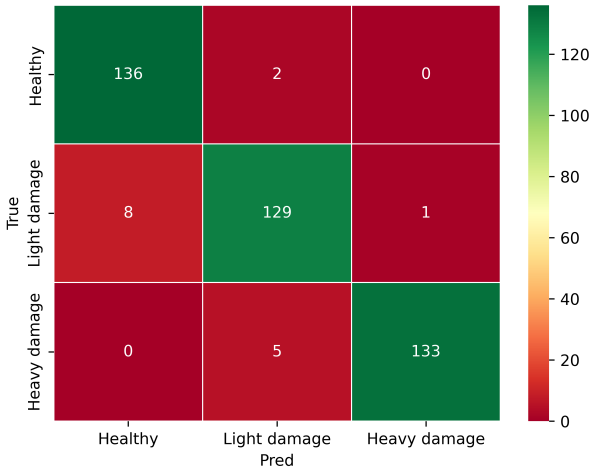
Source: This work.

Figure 14 represents the sensitivity analysis setup where one parameter is fixed and the other varies. For this setup, the fixed parameter is set as 0.02, which was the optimal point for experiment #1. In Figures 14a and 14b, α is fixed while β varies, while in 14c and 14d, α is the parameter under analysis.

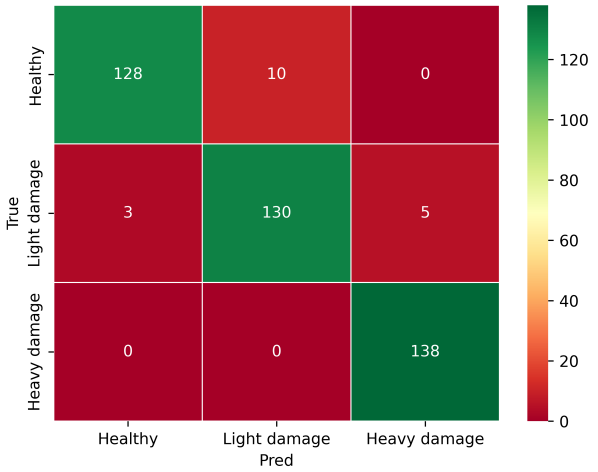
From Figures 14a and 14b, and comparing it with the baseline of Figure 13b, one must

Figure 14 – Confusion matrices – sensitivity analysis for PU dataset: varying α and β separately.

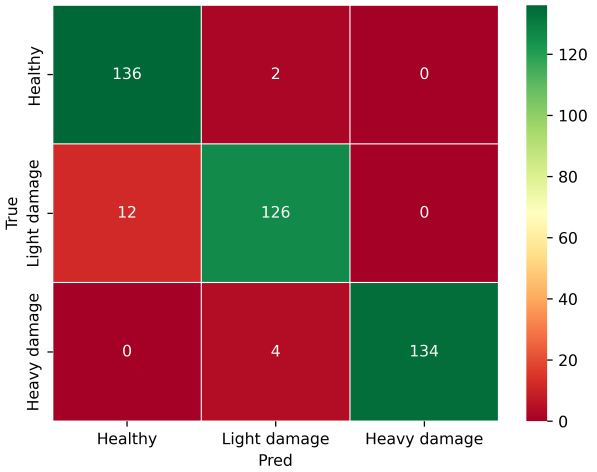
(a) $\alpha = 0.02$, $\beta = 0.01$, 96.14% accuracy.



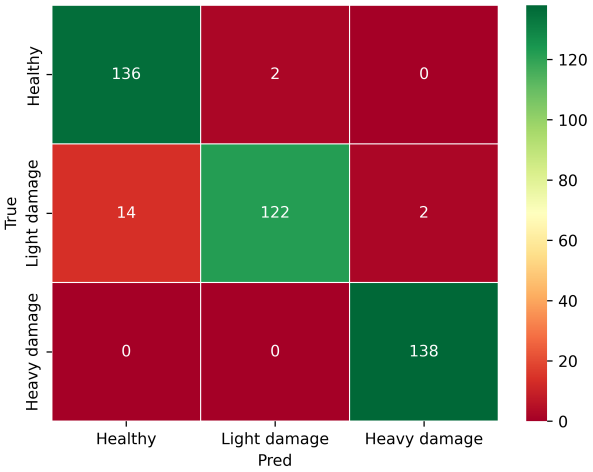
(b) $\alpha = 0.02$, $\beta = 0.03$, 95.65% accuracy.



(c) $\alpha = 0.01$, $\beta = 0.02$, 95.65% accuracy.



(d) $\alpha = 0.03$, $\beta = 0.02$, 95.65% accuracy.



Source: This work.

note that the variation of β directly impacts the shift of the prediction into healthy or heavy damage classes. By lowering β, true healthy data - the first horizontal line in the figures - are favored for the classification, i.e., there are much fewer misclassifications associated to this true label, while increasing this penalty parameter has the opposite effect - the third horizontal line in Figure 14b had no misclassifications.

Varying α, on the other hand, seems to directly impact the light damage class, as shown in Figures 14c and 14d. When the parameter is reduced, the growth in the misclassification is also present on the heavy class, while this does not occur when α is increased. By increasing α, both weights are set to moderate but impactful values, which ends up favoring the correct classification of the extreme true classes – healthy and heavy damage – but highly penalizing light damage predictions.

The behavior observed upon varying β follows the expectations: since it is responsible for the penalty related to the heavy damage class, lowering it should prioritize the healthy class

classification - which now has a higher penalty relatively, while increasing it should prioritize the heavy damage class, as shown by the analyses. On the other hand, varying α should have shown the opposite response, but instead the light damage class was penalized on both changes - decreasing or increasing.

4.2 CASE STUDY 2: CEERMA BEARING VIBRATION BENCH'S DATA

Similarly to case study 1, the CEERMA's Bearing Vibration Bench's Data, presented in subsection 3.1.2 is also used to evaluate the proposed framework. This dataset is collected via experiments using bearings of type NSK 1205K C3 under three damage levels: healthy, light outer ring damage, and heavy outer ring damage.

The models follow the same architecture described in section 3.3 and Figure 11. Due to the higher complexity of this dataset, possibly due to noise and residual misalignment, models were trained for a total of 50 epochs instead, using a batch size of 256 and ADAM optimizer with an initial learning rate of 0.001. The values for the first and second thresholds are set as 0.1048 and 0.1976, respectively. Please note that the thresholds are defined as the highest amplitude of each class in the training set - thus are results of data values. The training time was around 10-15 seconds.

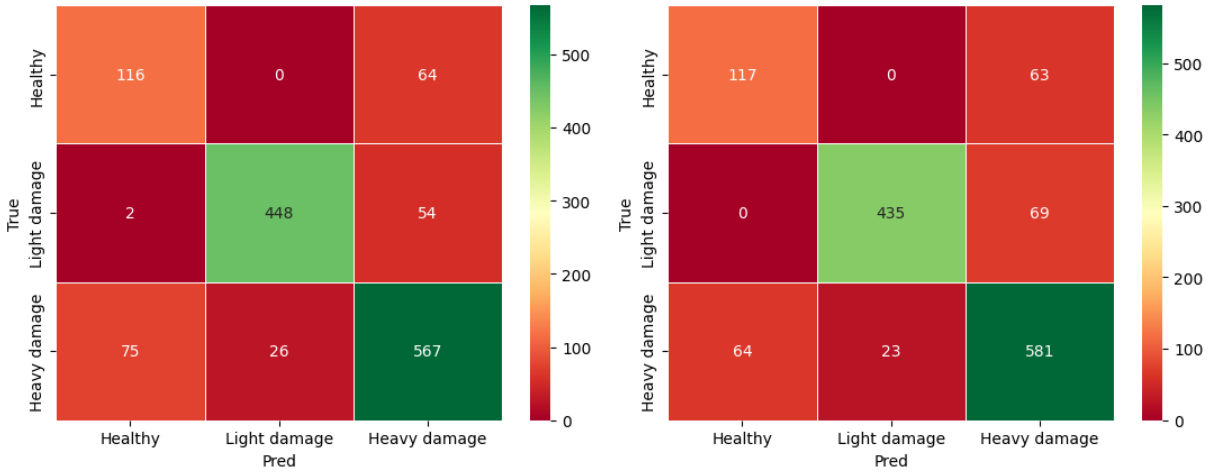
The advantages of the Physics-Informed Deep Learning approach for this case study are not in terms of general accuracy, as shown in Figure 15. Although the accuracy growth was minimal – from 83.6% to 83.8%, the main improvement is seen in the extreme-cases misclassifications: a 14.67% reduction in the number of heavy damage data points being classified as healthy and a 1.56% in the opposite direction. This directly impacts the model's credibility, as operators who come across a model that predicts a heavy damage input as healthy will fail to believe in further predictions of this same model, even if it generally has high accuracy. The results shown here were obtained by empirically setting the penalty parameters $\alpha = \beta = 0.05$, or 5% penalty.

Table 10 brings some classification metrics regarding the comparison between the models tested for Case Study 2: the traditional Deep Learning method and the Physics-Informed Deep Learning one. Again, since we are considering a three-class problem, we use here the generic terms 'true rate' (ratio of examples correctly classified as class c, from all examples of that class) and 'predictive value' (ratio of examples correctly classified as class c, from all examples classified as that class) for the calculations, instead of their binary counterparts (sensitivity, specificity, precision, negative predictive value). For most metrics considered, PIDL was superior to pure DL. The main downside for the proposed approach was related to its conservative behavior in regards to light damage: many data points from this class were predicted as heavy damage. This directly affected the metrics 'true light damage rate' – the ratio of true light damage examples that are correctly predicted as light damage –

Figure 15 – Confusion matrices for vibration bench data.

(a) Regular DL model: 83.6% accuracy.

(b) Physics-Informed Deep Learning model: 83.8% accuracy.



Source: Raupp *et al.* (2024b).

and 'heavy damage predictive value' – the ratio of examples predicted as heavy damage that are, in fact, heavy damage.

One of the possible explanations for the aforementioned behavior is the fact that our light damage class in the experiment is a 1mm fissure in the bearing. This, for some cases, might even be considered a large damage for a bearing, which might have been resulted into these misclassifications. In future experiments, more damage levels – especially at a lower degree – will be considered in order to better understand the methodology's behavior for those cases.

Table 10 – Comparison metrics for Case Study 2: CEERMA's Bearing Vibration Bench's Data.

			Healthy		Light damage		Heavy damage	
Model	**Acc.**	**Bal. Acc.**	**True Rate**	**Pred. Value**	**True Rate**	**Pred. Value**	**True Rate**	**Pred. Value**
DL	0.8365	0.7940	0.6444	0.6010	**0.8889**	0.9451	0.8488	**0.8277**
PIDL	**0.8380**	**0.7943**	**0.6500**	**0.6464**	0.8631	**0.9498**	**0.8698**	0.8149

Source: This work.

Similarly, as in Case Study 1, to properly validate the advantage of the PIDL over the DL approach, the same models were trained 10 times, each incurring a different train-test partition but using the same data to train both models. The Wilcoxon Signed-Rank test for paired samples, whose hypotheses are presented in 4.1, was performed to compare the median of the accuracies obtained for each model when trained using the same dataset (Voorter *et al.*, 2023) (Ishida *et al.*, 2024). The results are presented in Table 11.

Similarly to Case Study 1, the null hypothesis was rejected, thus showing that the PIDL model was statistically different than the DL model, in terms of their accuracy medians.

Table 11 – Wilcoxon Signed-Rank test for Case Study 2: CEERMA's Bearing Vibration Bench's Data.

Model	Acc. 1	Acc. 2	Acc. 3	Acc. 4	Acc. 5	Acc. 6	Acc. 7	Acc. 8	Acc. 9	Acc. 10
DL	0.8365	0.8454	0.8409	0.8360	0.8380	0.8091	0.8106	0.8158	0.7825	0.7929
PIDL	0.8380	0.8521	0.8654	0.8447	0.8536	0.8321	0.8580	0.8299	0.8202	0.8129
Test statistic	0.0									
P-value	0.001953125									

Source: This work.

4.2.1 Sensitivity Analysis

For this case study, the parameters that control the known behavior's contribution are also analyzed. Again, two setups are considered: #1 varying α and β simultaneously; and #2 fixing one parameter and varying the other.

By varying α and β simultaneously, as seen in Figure 16, the performance decreases as the parameters increase. When the weights are 0.02, in Figure 16a, the accuracy is of 84.76%, higher than the 83.8% shown by 0.05 penalty (Figure 16b). This, although, comes at the cost of wrong extreme predictions: a 64% increase in the number of heavy damage inputs classified as healthy, which is the most critical misclassification possible, as it might result in a machine failure. Even a 54% reduction in healthy to heavy misclassifications is not enough to overcome this penalty. Over-increasing the parameters, as seen in Figures 16c and 16d, over-specializes in the healthy true class and fails to learn the heavy damage patterns correctly.

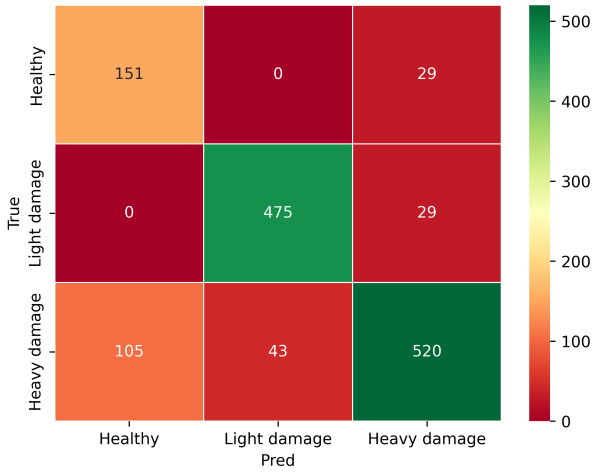
Next, the sensitivity analysis is also performed by varying one parameter at a time while the other is set at 0.05. Figure 17 shows the confusion matrices for these experiments. In Figures 17a and 17b, α is fixed while β varies, while in 17c and 17d, α is the parameter under analysis.

Figures 17a and 17b, when compared to the baseline shown in Figure 16b, show that a trend similar to case study 1 is seen here. When β is lowered, the true healthy class is favored in the classifications, as opposed to when this weight is increased, where the opposite trend occurs.

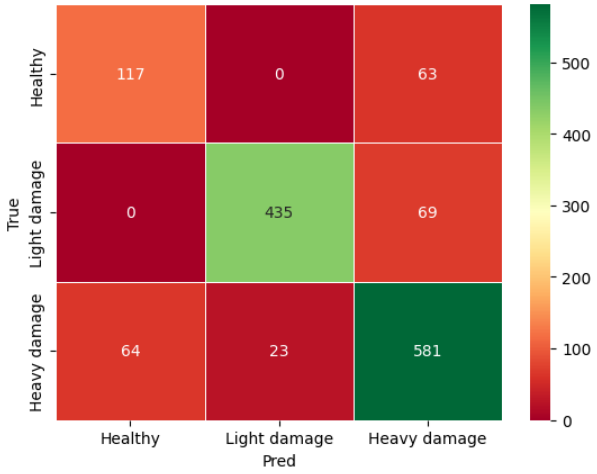
When α varies, however, a trend different than case study 1 is shown. By lowering the parameter, the heavy damage true class is favored, which was the expected behavior since this parameter tends to control the penalty associated to the healthy class - since the penalty is lowered, there is less effort by the network to correctly classify this class. Additionally, when this weight is increased, both the true healthy and true light damage are improved, also following the expected behavior. Thus, this dataset's results show a trend that better follows the expectations, differently than in case study 1, where varying α had a drop in the accurate classifications for true light damage class, not seen for this dataset. This suggests that the parameter's impact is much more related to the dataset signals' amplitudes, which are classified by the threshold model and may incur the penalty if the DL model's prediction deviates from it.

Figure 16 – Confusion matrices – sensitivity analysis for vibration bench data: varying α and β simultaneously.

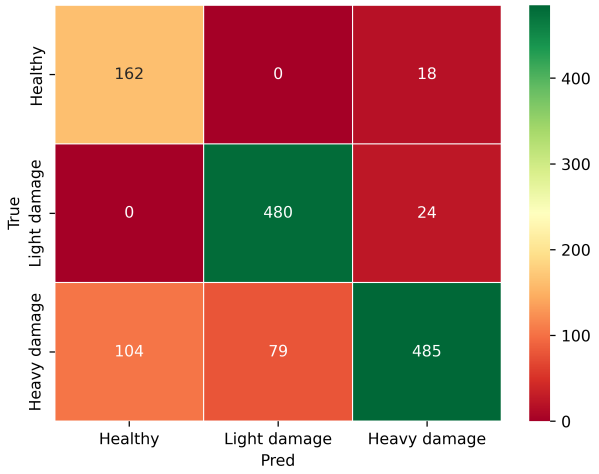
(a) $\alpha = \beta = 0.02$, 84.76% accuracy.



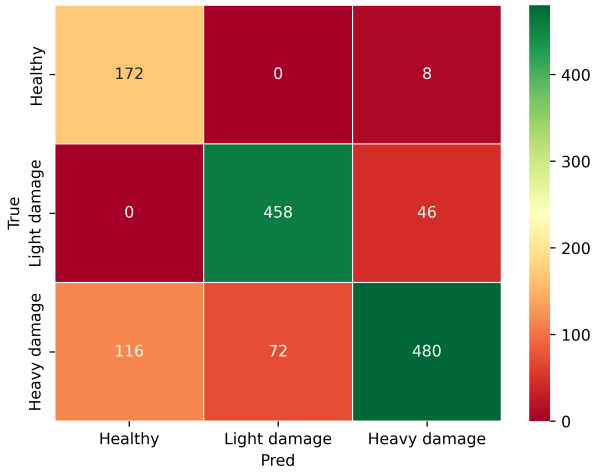
(b) $\alpha = \beta = 0.05$, 83.8% accuracy.



(c) $\alpha = \beta = 0.07$, 83.36% accuracy.



(d) $\alpha = \beta = 0.10$, 82.1% accuracy.



Source: This work.

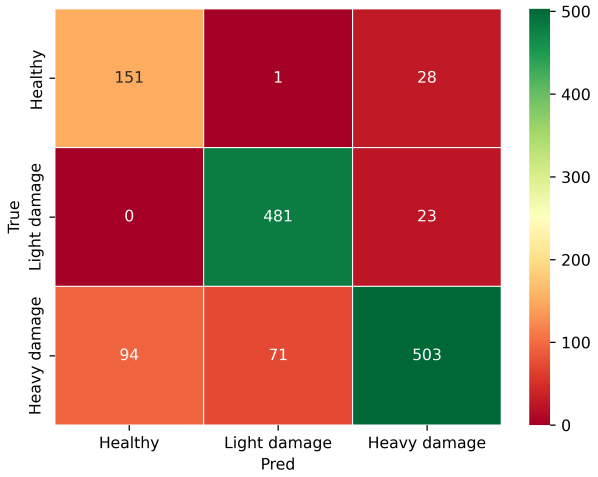
4.3 DISCUSSION

This chapter presented the results obtained by a traditional Deep Learning model and a proposed Physics-Informed Deep Learning approach for two case studies: Paderborn University Dataset and CEERMA's bearing vibration bench data. For both cases, sensitivity analyses were performed in order to evaluate the behavior of the degree of known information being introduced during the training.

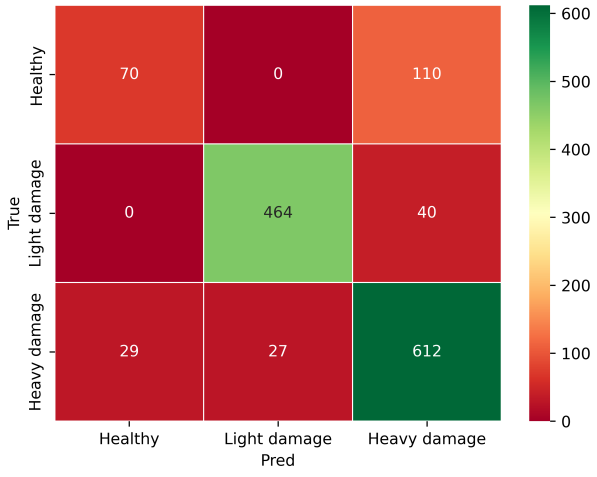
For both case studies, the PIDL proposed approach was able to improve the results obtained by the pure data-driven DL model. This suggests that, when correctly optimized for the specific case under analysis, the Physics-Informed have potential to enhance the performance of statistical models. On the other hand, some considerations are defined regarding the usage of hybrid models:

Figure 17 – Confusion matrices – sensitivity analysis for vibration bench data: varying α and β separately.

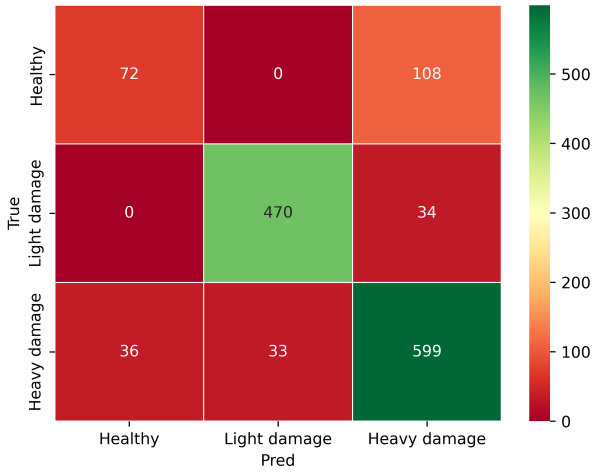
(a) $\alpha = 0.05$, $\beta = 0.03$, 83.95% accuracy.



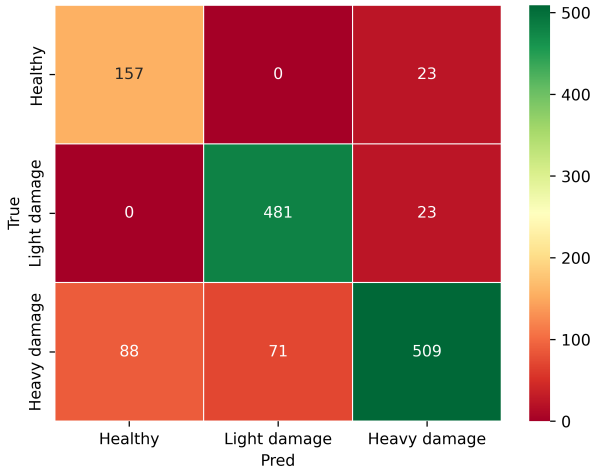
(b) $\alpha = 0.05$, $\beta = 0.07$, 84.76% accuracy.



(c) $\alpha = 0.03$, $\beta = 0.05$, 84.39% accuracy.



(d) $\alpha = 0.07$, $\beta = 0.05$, 84.83% accuracy.



Source: This work.

- Those models are constructed for a specific application. The threshold model proposed here is applicable for bearing vibration data that follows a damage-growth scheme, where the classes represent the degree of evolution of a given failure (healthy $\rightarrow$ light damage $\rightarrow$ heavy damage). Changing the application requires adaptation of the known behavior term and its preprocessing.
- The proposed framework here introduces two hyperparameters, α and β. Our sensitivity analysis provided some guidance about the relationship of those parameters regarding the prediction's performance, but still did not reach an optimal value for all cases. Thus, they still must be optimized for each specific application.

5 CONCLUSION

Rotating machinery – pumps, turbines, compressors, etc. – are widely used in the Oil and Gas industry, being considered critical components of their systems. Common to all of them, the bearing component is responsible for up to 55% of the failures occurring in these machines, as pointed by Shen *et al.* (2021). Thus, this element is of major concern for optimization, in order to guarantee the Reliability, Availability, Maintainability and Safety of this industry.

In this context, bearing elements have already been under focus of researches over time, but mostly over the bias of pure data-driven approaches (Yuan *et al.*, 2020) (Ni *et al.*, 2023) (Shutin *et al.*, 2024). These methods may result in false alarms that contradict the system's expected behavior (Shen *et al.*, 2021) (Raupp *et al.*, 2024a) (Raupp *et al.*, 2024b), which could impact the prediction's confidence (Aliyu *et al.*, 2022).

In order to solve the problems described, this research proposed a framework for bearing fault classification in the context of the O&G industry. The methodology is composed of a DL model which is supplied with the system's known behavior via a threshold model, constrained within a customized loss function. The approach was validated using two case studies: Paderborn University Bearing Data Center's Dataset and bearing vibration data from CEERMA's bearing vibration bench.

Results show that the physical addition was capable of improving the results of the purely statistical approach, showing a 2.8% relative accuracy gain for case study 1, but increasing up to 6.4% for the healthy class, which is one of the two most important classes to analyze in order to prevent false alarms. In case study 2, where accuracy gains were smaller, the PIDL model still demonstrated success when reducing the amount of misclassifications related to the extreme cases: healthy data classified as heavy damage, and vice-versa. On top of that, sensitivity analyses were performed by varying the degree of guidance introduced by the threshold model's term.

The main contributions of this research can be listed:

- Contribution to the scientific community in regards to studies and applications of novel Physics-Informed Deep Learning approaches, and vibration analyses in the bearing context, widely available in the Oil and Gas context;
- Development of a model framework for anomaly detection with good performance and physical adequacy on bearings;
- Publication of scientific articles in national and international conferences;
- Impact in the social aspect by providing a methodology that can possibly reduce failures, thus minimizing unwanted variations in the product supply for the society;

- Impact in the environmental aspect by increasing the likelihood of damage being identified in machines, preventing leaks and hazards that may impact the surrounding population and wildlife;
- Impact in the economic aspect for both the companies and the society, as improving the RAMS of a system directly impacts its economic potential, positively affecting its region's economy.

Lastly, the following future steps can be listed:

- To further study PIDL approaches for different contexts, on account of hybrid models being tied to their specific cases, in special the anomaly detection task;
- To identify different preprocessing approaches for the bearing damage-level scheme;
- To pursue bearing datasets in the Oil and Gas industry to further validate this work's proposed methodology;
- To conduct further bearing experiments using CEERMA's bearing vibration bench;
- To create an application comprising the developed work. The application will allow the automated procedure of data preprocessing, training and testing given data provided and bearing parameters.

REFERENCES

ABBASION, S.; RAFSANJANI, A.; FARSHIDIANFAR, A.; IRANI, N. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. **Mechanical Systems and Signal Processing**, v. 21, n. 7, p. 2933–2945, 2007. ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2007.02.003.

ALIYU, R.; MOKHTAR, A. A.; HUSSIN, H. Prognostic health management of pumps using artificial intelligence in the oil and gas sector: A review. **Applied Sciences**, v. 12, n. 22, 2022. ISSN 2076-3417. https://doi.org/10.3390/app122211691.

ALJAMEEL, S. S.; ALOMARI, D. M.; ALISMAIL, S.; KHAWAHER, F.; ALKHUDHAIR, A. A.; ALJUBRAN, F.; ALZANNAN, R. M. An anomaly detection model for oil and gas pipelines using machine learning. **Computation**, v. 10, n. 8, 2022. ISSN 2079-3197. https://doi.org/10.3390/computation10080138.

ALZUBAIDI, L.; BAI, J.; AL-SABAAWI, A.; SANTAMARÍA, J.; ALBAHRI, A. S.; AL-DABBAGH, B. S. N.; FADHEL, M. A.; MANOUFALI, M.; ZHANG, J.; AL-TIMEMY, A. H.; DUAN, Y.; ABDULLAH, A.; FARHAN, L.; LU, Y.; GUPTA, A.; ALBU, F.; ABBOSH, A.; GU, Y. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. **Journal of Big Data**, v. 10, n. 1, p. 46, Apr 2023. ISSN 2196-1115. https://doi.org/10.1186/s40537-023-00727-2.

ARISMENDI, R.; BARROS, A.; GRALL, A. Piecewise deterministic markov process for condition-based maintenance models — application to critical infrastructures with discrete-state deterioration. **Reliability Engineering & System Safety**, v. 212, p. 107540, 2021. ISSN 0951-8320. https://doi.org/10.1016/j.ress.2021.107540.

BAKDI, A.; KOUADRI, A. A new adaptive pca based thresholding scheme for fault detection in complex systems. **Chemometrics and Intelligent Laboratory Systems**, v. 162, p. 83–93, 2017. ISSN 0169-7439. https://doi.org/10.1016/j.chemolab.2017.01.013.

BARRAZA, J. F.; BRäUNING, L. G.; PEREZ, R. B.; MORAIS, C. B.; MARTINS, M. R.; DROGUETT, E. L. Deep learning health state prognostics of physical assets in the oil and gas industry. **Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability**, v. 236, n. 4, p. 598–616, 2022. https://doi.org/10.1177/1748006X20976817.

BAYAZITOVA, G.; ANASTASIADOU, M.; dos Santos, V. D. Oil and gas flow anomaly detection on offshore naturally flowing wells using deep neural networks. **Geoenergy Science and Engineering**, v. 242, p. 213240, 2024. ISSN 2949-8910. https://doi.org/10.1016/j.geoen.2024.213240.

BENGIO, Y. Learning deep architectures for ai. **Foundations and Trends® in Machine Learning**, v. 2, n. 1, p. 1–127, 2009. ISSN 1935-8237. http://doi.org/10.1561/2200000006.

BONO, F. M.; RADICIONI, L.; CINQUEMANI, S.; BOMBACI, G. A comparison of deep learning algorithms for anomaly detection in discrete mechanical systems. **Applied Sciences**, v. 13, n. 9, 2023. ISSN 2076-3417. https://doi.org/10.3390/app13095683.

BOUSHABA, A.; CAUET, S.; CHAMROO, A.; ETIEN, E.; RAMBAULT, L. Comparative study between physics-informed CNN and PCA in induction motor broken bars MCSA detection. **Sensors (Basel)**, MDPI AG, v. 22, n. 23, p. 9494, dec. 2022.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 41, n. 3, jul. 2009. ISSN 0360-0300. https://doi.org/10.1145/1541880.1541882.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection for discrete sequences: A survey. **IEEE Transactions on Knowledge and Data Engineering**, v. 24, n. 5, p. 823–839, 2012. https://doi.org/10.1109/TKDE.2010.235.

COOLEY, J.; TUKEY, J. An algorithm for the machine calculation of complex fourier series. **Mathematics of Computation**, v. 19, n. 90, p. 297–301, 1965. https://doi.org/10.2307/2003354.

CWRU. **Case Western Reserve University Bearing Data Center**. 2019. https://engineering.case.edu/bearingdatacenter/download-data-file.

FRANKLIN, T. S.; SOUZA, L. S.; FONTES, R. M.; MARTINS, M. A. A physics-informed neural networks (pinn) oriented approach to flow metering in oil wells: an esp lifted oil well system as a case study. **Digital Chemical Engineering**, v. 5, p. 100056, 2022. ISSN 2772-5081. https://doi.org/10.1016/j.dche.2022.100056.

GHUNGRAD, S.; FAEGH, M.; GOULD, B.; WOLFF, S. J.; HAGHIGHI, A. Architecture-driven physics-informed deep learning for temperature prediction in laser powder bed fusion additive manufacturing with limited data. **Journal of Manufacturing Science and Engineering**, v. 145, n. 8, p. 081007, 04 2023. ISSN 1087-1357. https://doi.org/10.1115/1.4062237.

GIL, A. C. **Como elaborar projetos de pesquisa**. São PauloAtlas, 2002.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. http://www.deeplearningbook.org.

HE, Y.; SU, H.; ZIO, E.; FAN, L.; ZHANG, J. A physics-informed training approach for data-driven method in remaining useful life estimation. **2022 6th International Conference on System Reliability and Safety (ICSRS)**, p. 500–504, 2022. https://doi.org/10.1109/ICSRS56243.2022.10067722.

HE, Y.; SU, H.; ZIO, E.; PENG, S.; FAN, L.; YANG, Z.; YANG, Z.; ZHANG, J. A systematic method of remaining useful life estimation based on physics-informed graph neural networks with multisensor data. **Reliability Engineering & System Safety**, v. 237, p. 109333, 2023. ISSN 0951-8320. https://doi.org/10.1016/j.ress.2023.109333.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. **Science**, v. 313, n. 5786, p. 504–507, 2006. https://doi.org/10.1126/science.1127647.

HORNIK, K. Approximation capabilities of multilayer feedforward networks. **Neural Networks**, v. 4, n. 2, p. 251–257, 1991. ISSN 0893-6080. https://doi.org/10.1016/0893-6080(91)90009-T.

HOU, X.; GUO, X.; YUAN, Y.; ZHAO, K.; TONG, L.; YUAN, C.; TENG, L. The state of health prediction of li-ion batteries based on an improved extreme learning machine. **Journal of Energy Storage**, v. 70, p. 108044, 2023. ISSN 2352-152X. https://doi.org/10.1016/j.est.2023.108044.

HUANG, R.; ZHANG, Z.; ZHANG, W.; MOU, J.; ZHOU, P.; WANG, Y. Energy performance prediction of the centrifugal pumps by using a hybrid neural network. **Energy**, v. 213, p. 119005, 2020. ISSN 0360-5442. https://doi.org/10.1016/j.energy.2020.119005.

ISHIDA, S.; FUJIWARA, Y.; MATTA, Y.; TAKEI, N.; KANAMOTO, M.; KIMURA, H.; TSUJIKAWA, T. Enhanced parameter estimation in multiparametric arterial spin labeling using artificial neural networks. **MAGNETIC RESONANCE IN MEDICINE**, v. 92, n. 5, p. 2163–2180, NOV 2024. ISSN 0740-3194. https://doi.org/10.1002/mrm.30184.

JAEN-VARGAS, M.; LEIVA, K. M. R.; FERNANDES, F.; GONCALVES, S. B.; SILVA, M. T.; LOPES, D. S.; OLMEDO, J. J. S. Effects of sliding window variation in the performance of acceleration-based human activity recognition using deep learning models. **PEERJ COMPUTER SCIENCE**, v. 8, 2022. https://doi.org/10.7717/peerj-cs.1052.

JAYASWAL, P.; WADHWANI, A. K.; MULCHANDANI, K. B. Machine fault signature analysis. **International Journal of Rotating Machinery**, v. 2008, n. 1, p. 583982, 2008. https://doi.org/10.1155/2008/583982.

KANARACHOS, S.; CHRISTOPOULOS, S.-R. G.; CHRONEOS, A.; FITZPATRICK, M. E. Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and hilbert transform. **Expert Systems with Applications**, v. 85, p. 292–304, 2017. ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2017.04.028.

KHAN, M. R.; TARIQ, Z.; ABDULRAHEEM, A. Application of artificial intelligence to estimate oil flow rate in gas-lift wells. **Natural Resources Research**, v. 29, n. 6, p. 4017–4029, Dec 2020. ISSN 1573-8981. https://doi.org/10.1007/s11053-020-09675-7.

KIM, D. W.; LEE, E. S.; JANG, W. K.; KIM, B. H.; SEO, Y. H. Effect of data preprocessing methods and hyperparameters on accuracy of ball bearing fault detection based on deep learning. **ADVANCES IN MECHANICAL ENGINEERING**, v. 14, n. 2, 2022. ISSN 1687-8132. https://doi.org/10.1177/16878132221078494.

KIM, M.; KIM, H. A dynamic analysis data preprocessing technique for malicious code detection with tf-idf and sliding windows. **ELECTRONICS**, v. 13, n. 5, 2024.

KIM, Y.; KIM, Y.-K. Physics-informed time-frequency fusion network with attention for noise-robust bearing fault diagnosis. **IEEE Access**, v. 12, p. 12517–12532, 2024. https://doi.org/10.1016/0893-6080(91)90009-T.

LESSMEIER, C.; KIMOTHO, J. K.; ZIMMER, D.; SEXTRO, W. 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. **PHM Society European Conference**, v. 85, n. 1, 2016. https://doi.org/10.36001/phme.2016.v3i1.1577.

MAHMOOD, Y.; AFRIN, T.; HUANG, Y.; YODO, N. Sustainable development for oil and gas infrastructure from risk, reliability, and resilience perspectives. **Sustainability**, v. 15, n. 6, 2023. ISSN 2071-1050. https://doi.org/10.3390/su15064953.

MARCONI, M. d. A.; LAKATOS, E. M. **Técnicas de pesquisa : planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados**. São PauloAtlas, 2002. ISBN 978-85-224-5152-4.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115–133, Dec 1943. ISSN 1522-9602. https://doi.org/10.1007/BF02478259.

MITCHELL, T. M. **Machine Learning**. New YorkMcGraw Hill, 1997.

MURPHY, K. P. **Probabilistic Machine Learning: An introduction**. MIT Press, 2022. probml.ai.

NI, Q.; JI, J.; HALKON, B.; FENG, K.; NANDI, A. K. Physics-informed residual network (piresnet) for rolling element bearing fault diagnostics. **Mechanical Systems and Signal Processing**, v. 200, p. 110544, 2023. ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2023.110544.

NSK. **SNN SERIES PILLOW BLOCKS: HIGH-PERFORMANCE MODULAR BEARING HOUSINGS**. Last access: 02/2025. https://www.nsk.com/content/dam/nsk/am/en_us/documents/bearings-americas/SNN-Series-Pillow-Blocks.pdf.

ORRù, P. F.; ZOCCHEDDU, A.; SASSU, L.; MATTIA, C.; COZZA, R.; ARENA, S. Machine learning approach using mlp and svm algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. **Sustainability**, v. 12, n. 11, 2020. ISSN 2071-1050. https://doi.org/10.3390/su12114776.

PAULUCIO, L. S.; PAIXãO, T. M.; BERRIEL, R. F.; SOUZA, A. F. D.; BADUE, C.; OLIVEIRA-SANTOS, T. Product categorization by title using deep neural networks as feature extractor. **2020 International Joint Conference on Neural Networks (IJCNN)**, p. 1–7, 2020. https://doi.org/10.1109/IJCNN48605.2020.9207093.

PECHT, M. Prognostics and health management of electronics. In: ____. **Encyclopedia of Structural Health Monitoring**. John Wiley & Sons, Ltd, 2009. ISBN 9780470061626. https://doi.org/10.1002/9780470061626.shm118.

PELLEGRINI, M.; VAREJAO, F.; RODRIGUES, A.; MELLO, L. H. S.; SILVA, L. **ESPset**. 2024. https://doi.org/10.17632/m268jsw339.1.

PETROVSKY, E. A.; BASHMUR, K. A.; SHADCHINA, Y. N.; BUKHTOYAROV, V. V.; TYNCHENKO, V. S. Study of microrelief forming technology on sliding bearings for oil and gas centrifugal units. **Journal of Physics: Conference Series**, IOP Publishing, v. 1399, n. 5, p. 055032, dec 2019. https://doi.org/10.1088/1742-6596/1399/5/055032.

RANDALL, R. B.; ANTONI, J. Rolling element bearing diagnostics—a tutorial. **Mechanical Systems and Signal Processing**, v. 25, n. 2, p. 485–520, 2011. ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2010.07.017.

RAUPP, L. S.; LINS, I. D.; MAIOR, C. B. S. Physics-informed deep learning approach for fault detection in the oil and gas industry. **Rio Oil and Gas - ROG.e 2024**, Rio de Janeiro, Brazil, n. 3714, 2024. https://doi.org/10.48072/2525-7579.roge.2024.3714.

RAUPP, L. S.; LINS, I. D.; MAIOR, C. B. S.; CAVALCANTI, T.; MOURA, M. J. das C.; LEITE, G. de N. P. Bearings' vibration fault detection via physics-informed deep learning. **Probabilistic Safety Assessment and Management & Asian Symposium on Risk Assessment and Management - PSAM17 & ASRAM2024**, Sendai, Japan, n. 1247, 2024. https://www.iapsam.org/PSAM17/program/Program.html#CH0293.

REN, H.; XU, C.; LYU, Y.; MA, Z.; SUN, Y. A thermodynamic-law-integrated deep learning method for high-dimensional sensor fault detection in diverse complex hvac systems. **Applied Energy**, v. 351, p. 121830, 2023. ISSN 0306-2619. https://doi.org/10.1016/j.apenergy.2023.121830.

SAFIZADEH, M.; LATIFI, S. Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. **Information Fusion**, v. 18, p. 1–8, 2014. ISSN 1566-2535. https://doi.org/10.1016/j.inffus.2013.10.002.

SHEN, S.; LU, H.; SADOUGHI, M.; HU, C.; NEMANI, V.; THELEN, A.; WEBSTER, K.; DARR, M.; SIDON, J.; KENNY, S. A physics-informed deep learning approach for bearing fault detection. **Engineering Applications of Artificial Intelligence**, v. 103, p. 104295, 2021. ISSN 0952-1976. https://doi.org/10.1016/j.engappai.2021.104295.

SHUTIN, D.; KAZAKOV, Y.; STEBAKOV, I.; SAVIN, L. Data-driven and physics-informed approaches for improving the performance of dynamic models of fluid film bearings. **Tribology International**, v. 191, p. 109136, 2024. ISSN 0301-679X. https://doi.org/10.1016/j.triboint.2023.109136.

SIKORSKA, J.; HODKIEWICZ, M.; MA, L. Prognostic modelling options for remaining useful life estimation by industry. **Mechanical Systems and Signal Processing**, v. 25, n. 5, p. 1803–1836, 2011. ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2010.11.018.

SNR. **Características técnicas**. Rolamentos autocompensadores de esferas, 1205K C3. https://eshop.ntn-snr.com/pt/product/1205KC3-SNR/1205.KC3.

SONG, H.; ZHU, J.; WEI, C.; WANG, J.; DU, S.; XIE, C. Data-driven physics-informed interpolation evolution combining historical-predicted knowledge for remaining oil distribution prediction. **Journal of Petroleum Science and Engineering**, v. 217, p. 110795, 2022. ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2022.110795.

THUEREY, N.; HOLL, P.; MUELLER, M.; SCHNELL, P.; TROST, F.; UM, K. **Physics-based Deep Learning**. 2022. https://doi.org/10.48550/arXiv.2109.05237.

TUOYUAN. What is the deep groove ball bearing? **Xtybearing**, Last access: 02/2025. https://xtybearing.com/what-is-the-deep-groove-ball-bearing/.

VARGAS, R. E. V.; MUNARO, C. J.; CIARELLI, P. M.; MEDEIROS, A. G.; AMARAL, B. G. do; BARRIONUEVO, D. C.; ARAúJO, J. C. D. de; RIBEIRO, J. L.; MAGALHãES, L. P. A realistic and public dataset with rare undesirable real events in oil wells. **Journal of Petroleum Science and Engineering**, v. 181, p. 106223, 2019. ISSN 0920-4105. http://www.sciencedirect.com/science/article/pii/S0920410519306357.

VOORTER, P. H. M.; BACKES, W. H.; GURNEY-CHAMPION, O. J.; WONG, S.-M.; STAALS, J.; OOSTENBRUGGE, R. J. van; THIEL, M. M. van der; JANSEN, J. F. A.; DRENTHEN, G. S. Improving microstructural integrity, interstitial fluid, and blood microcirculation images from multi-b-value diffusion mri using physics-informed neural networks in cerebrovascular disease. **MAGNETIC RESONANCE IN MEDICINE**, v. 90, n. 4, p. 1657–1671, OCT 2023. ISSN 0740-3194. https://doi.org/10.1002/mrm.29753.

WANG, Z.; FAN, Z.; CHEN, X.; LI, Y.; FAN, Z.; WEI, Q.; PENG, Y.; LIU, B.; YUE, W.; WANG, X.; XIONG, L. Global oil and gas development situation, trends and enlightenment in 2023. **Petroleum Exploration and Development**, v. 51, n. 6, p. 1536–1555, 2024. ISSN 1876-3804. https://doi.org/10.1016/S1876-3804(25)60558-1.

WANG, Z.; ZHOU, Z.; XU, W.; SUN, C.; YAN, R. Physics informed neural networks for fault severity identification of axial piston pumps. **Journal of Manufacturing Systems**, v. 71, p. 421–437, 2023. ISSN 0278-6125. https://doi.org/10.1016/j.jmsy.2023.10.002.

WILLARD, J.; JIA, X.; XU, S.; STEINBACH, M.; KUMAR, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 55, n. 4, nov. 2022. ISSN 0360-0300. https://doi.org/10.1145/3514228.

XU, J.; XU, L. Health management based on fusion prognostics for avionics systems. **Journal of Systems Engineering and Electronics**, Journal of Systems Engineering and Electronics, v. 22, n. 3, p. 428, 2011. https://doi.org/10.3969/j.issn.1004-4132.2011.03.010.

XU, K.; JIANG, Y.; TANG, M.; YUAN, C.; TANG, C. Presee: An mdl/mml algorithm to time-series stream segmenting. **SCIENTIFIC WORLD JOURNAL**, 2013. ISSN 1537-744X. https://doi.org/10.1155/2013/386180.

XU, Y.; KOHTZ, S.; BOAKYE, J.; GARDONI, P.; WANG, P. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. **Reliability Engineering & System Safety**, v. 230, p. 108900, 2023. ISSN 0951-8320. https://doi.org/10.1016/j.ress.2022.108900.

YAN, J.; HE, Z.; HE, S. Multitask learning of health state assessment and remaining useful life prediction for sensor-equipped machines. **Reliability Engineering & System Safety**, v. 234, p. 109141, 2023. ISSN 0951-8320. https://doi.org/10.1016/j.ress.2023.109141.

YUAN, L.; LIAN, D.; KANG, X.; CHEN, Y.; ZHAI, K. Rolling bearing fault diagnosis based on convolutional neural network and support vector machine. **IEEE Access**, v. 8, p. 137395–137406, 2020. https://doi.org/10.1109/ACCESS.2020.3012053.

YUSUF, Y. Y.; MUSA, A.; DAUDA, M.; EL-BERISHY, N.; KOVVURI, D.; ABUBAKAR, T. A study of the diffusion of agility and cluster competitiveness in the oil and gas supply chains. **International Journal of Production Economics**, v. 147, p. 498–513, 2014. ISSN 0925-5273. Building Supply Chain System Capabilities in the Age of Global Complexity: Emerging Theories and Practices. https://doi.org/10.1016/j.ijpe.2013.04.010.

ZABIN, M.; CHOI, H.-J.; UDDIN, J. Hybrid deep transfer learning architecture for industrial fault diagnosis using hilbert transform and dcnn-lstm. **JOURNAL OF SUPERCOMPUTING**, v. 79, n. 5, p. 5181–5200, 2023. ISSN 0920-8542. https://doi.org/10.1007/s11227-022-04830-8.

ZHANG, B.; LIU, W.; CAI, Y.; ZHOU, Z.; WANG, L.; LIAO, Q.; FU, Z.; CHENG, Z. State of health prediction of lithium-ion batteries using particle swarm optimization with levy flight and generalized opposition-based learning. **Journal of Energy Storage**, v. 84, p. 110816, 2024. ISSN 2352-152X. https://doi.org/10.1016/j.est.2024.110816.

ZHANG, L.; WANG, J. Intelligent safe operation and maintenance of oil and gas production systems: Connotations and key technologies. **Natural Gas Industry B**, v. 10, n. 3, p. 293–303, 2023. ISSN 2352-8540. https://doi.org/10.1016/j.ngib.2023.05.006.

ZHANG, S.; LUO, M.; QIAN, H.; LIU, L.; YANG, H.; ZHANG, Y.; LIU, X.; XIE, Z.; YANG, L.; ZHANG, W. A review of valve health diagnosis and assessment: Insights for intelligence maintenance of natural gas pipeline valves in china. **Engineering Failure Analysis**, v. 153, p. 107581, 2023. ISSN 1350-6307. https://doi.org/10.1016/j.engfailanal.2023.107581.

ZHANG, S.; YE, F.; WANG, B.; HABETLER, T. G. **Semi-Supervised Learning of Bearing Anomaly Detection via Deep Variational Autoencoders**. 2019. https://doi.org/10.48550/arXiv.1912.01096.

ZHAO, Z.; LI, T.; WU, J.; SUN, C.; WANG, S.; YAN, R.; CHEN, X. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. **ISA Transactions**, v. 107, p. 224–255, 2020. ISSN 0019-0578. https://doi.org/10.1016/j.isatra.2020.08.010.