Explainable structural health monitoring(SHM) for damage classification based on vibration.

by

Sai Ganesh Reddy Kongala

A Master's thesis submitted to the
Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
in partial fulfillment of the requirements for the degree of

MSc in Business Information Technology- Data Science & Business

Faculty of Electrical Engineering, Mathematics
& Computer Science (EEMCS)

University of Twente

Enschede, Overijssel, The Netherlands

22nd May, 2024

ABSTRACT

Traditional SHM methods often face challenges in scaling and accuracy due to their dependency on manual inspection and non-explainable ML models. Recent advancements in big data analytics, machine learning, cloud data storage systems, and wireless sensor network technologies have significantly expanded the potential for data-driven Structural health Monitoring (SHM) systems for large-scale engineering projects. The application of machine learning methods in SHM has seen considerable growth in recent years. However, given the cross-domain nature of this technology, there emerges a critical need for Explainable Artificial Intelligence (XAI) within the realm of SHM. This study investigates the application of XAI frameworks to enhance the interpretability and reliability of SHM systems, specifically targeting damage classification using vibration data. This research employs accelerometer sensor networks to collect vibration data from composite structures, utilizing various ML models such as Random Forest, Support Vector Machine, XGBoost, Convolutional Neural Networks (CNN), and Transformer networks. The study underscores the superior performance of advanced models, particularly CNNs and Transformer networks, in accurately identifying structural defects. The thesis also addresses the challenges of data corruption in SHM systems. Data corruption, caused by sensor faults, transmission errors, environmental interference, and software bugs, can lead to false positives and false negatives, severely impacting the reliability of SHM systems. To mitigate these effects, robust data validation and cleaning mechanisms are implemented, including anomaly detection algorithms. Future research directions include optimizing sensor coverage, integrating advanced sensing technologies, and fostering interdisciplinary collaborations to develop economically viable and technically proficient SHM solutions.

AUTHOR'S DECLARATION

I, Sai Ganesh Reddy Kongala, declare that this thesis titled "Explainable Structural Health Monitoring (SHM) for Damage Classification Based on Vibration Data" and the work presented in it are my own and have been generated by me as the result of my own original research, conducted during my thesis internship at Tarucca as part of my Master study program at the University of Twente.

This work was done wholly or mainly while in candidature for a research degree at this University. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated. Where I have consulted the published work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work. I have acknowledged all main sources of help. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself. Either none of this work has been published before submission, or parts of this work have been published as listed below.

I authorize the University of Twente to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Twente to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

Sai Ganesh Reddy Kongala

Acknowledgements

Completing this thesis would not have been possible without the support, guidance, and encouragement of many individuals. I would like to take this opportunity to express my heartfelt gratitude to all those who have contributed to this journey.

First and foremost, I extend my deepest appreciation to my mentors Faizan and Abhista. Your expertise, patience, and unwavering support have been invaluable throughout this process. I am particularly grateful to Faizan, whose insights and constructive feedback have shaped the direction of my research. Your commitment to my academic growth has been both inspiring and motivating.

I am also immensely grateful to my mentor and colleagues at Tarucca. Your practical guidance and professional insights have significantly enriched my understanding and application of theoretical concepts. Special thanks to Hernani Costa, whose mentorship and support during my internship were instrumental in enhancing my skills and confidence in the field.

To my friends, thank you for your constant encouragement and understanding during the challenging times. Your companionship and the countless discussions we had, both academic and otherwise, have been a source of great comfort and inspiration. I am especially grateful to the friends who made me feel at home despite being away from my home country. Your support helped me stay connected and positive throughout this journey. Special thanks to Sonakshi, Rakshita, and Saran for always being there to lend a listening ear and for the late-night study sessions that kept me going. I apologize if I was ever annoying, and a special sorry to Sonakshi for any moments of frustration I may have caused.

I owe a profound debt of gratitude to my family, whose love and support have been my foundation. To my parents, Jogi Reddy and Vijaya laxmi, your belief in me has been unwavering, and your sacrifices have not gone unnoticed. Thank you for instilling in me the values of hard work and perseverance. To my uncle, Vishnuvardhan Reddy, for his unwavering support and encouragement. Your guidance and wisdom have been invaluable, and your belief in me has been a constant source of motivation. To my brother, Sai Shashank, thank you for stepping in and being there for our parents, especially during the difficult time of our grandfather's passing. Your ability to support and comfort them in my absence has meant the world to me. Your strength and dedication have been truly admirable.

Finally, I would like to thank everyone who has wished me well and supported me throughout this journey. Your good wishes and encouragement have been deeply appreciated and have played a significant role in helping me reach this milestone.

Sai Ganesh Reddy Kongala

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

CNN Convolutional Neural Networks. 13

CRISP-DM CRoss-Industry Standard Process for Data Mining. vi, viii, 23–27

DSF Damage sensitive features. 10, 55

LRP Layer-wise relevance propagation. 13, 55, 56

SHM Structural health Monitoring. i, 1, 3, 6, 8, 9, 14, 23

SVM Support Vector Machine. 12, 13

XAI Explainable Artificial Intelligence. i, 3, 8

1

INTRODUCTION

1.1. BACKGROUND

Traditionally, the SHM for large-scale civil structures, such as bridges and dams, as well as mega-projects like power plants and manufacturing units, has been conducted through human intervention or manual inspection. However, as the scale of these projects increases, the reliance on human presence for monitoring the health of these structures becomes challenging, time-consuming, and expensive process. Recent advancements in computational capabilities, along with improvements in sensors, communication, and storage technologies, have facilitated the implementation of comprehensive, data-driven SHM systems [6] to monitor the maintenance of infrastructure with composite structures. The data driven approach depends in various sensors to collect data, Mukhopadhyay and Ihara [7] highlights the most commonly utilized sensors in SHM system development that include strain gauges, piezoelectric sensors, accelerometers, and acoustic sensors. These devices are instrumental in extracting features such as strain displacements, accelerations, and acoustic vibration of structures. In our case, we have installed an accelerometer sensor network to detect vibration features of structure at multiple positions.

Today, data-driven SHM is pivotal in ensuring the maintenance and safety of various composite structures across engineering disciplines. The evolution stages of an SHM system, as theorized by Tibaduiza Burgos et al. [1], categorize SHM systems into hierarchical stages based on the extent to which they are data-driven. This transformation is represented as a hierarchical pyramid 1.1, where each ascending level signifies a more advanced stage of data collection and analysis. The stages are briefly described below.

- Damage Detection: At the base level of the pyramid, the primary goal is the detection of damage within the structure. This is commonly achieved by strategically positioning

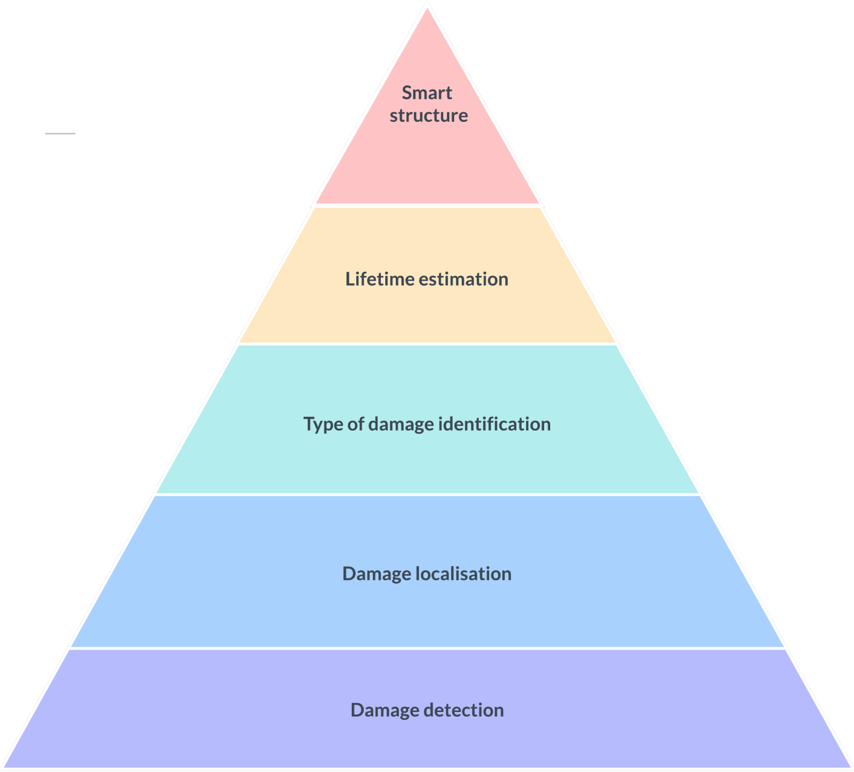


Figure 1.1: SHM evolution pyramid [1]

sensors at locations sensitive to damage detection. The data collected by these sensors, which includes various features(acceleration, displacement), is used to determine the damage status of the structure. However, at this stage, the exact location of the damage is not yet identified.

- Damage Localization: Once damage is detected, the next stage is to pinpoint its location. Damage localization is more complex than detection and often requires a denser network of sensors. These sensors are placed in various sub-parts of the composite structure to gather more detailed data, enabling the precise identification of where the damage has occurred.
- Type of Damage Identification: Different structures can suffer from various types of damages like cracks, corrosion, delamination, or breakage. At this stage, the specific type of damage is identified. This requires a more sophisticated analysis of the data collected. For instance, different types of damage can affect the structure's response in unique ways, and understanding these patterns is key to identify the type of damage. This stage may involve the use of advanced algorithms and machine learning techniques to interpret the sensor data.
- Lifetime Estimation: This stage involves predicting the remaining useful life of components in structure. By comprehensively understanding the extent and type of damage,

the material properties of the structure and prevailing environmental conditions, it is possible to estimate the remaining functional lifespan of the structure accurately. This stage is crucial for planning maintenance, repairs, or replacements before catastrophic failures occur.

- Smart Structures: At the top of the pyramid are smart structures. Smart structures are designed from the outset with a SHM perspective. This could include structural components that adapt their properties in response to damage, thereby maintaining structural integrity. This is achieved with the integration of various disciplines such as mechanics , structural properties, sensors, and control systems.

Traditional Machine learning models, especially deep learning models, are often seen as " black boxes " due to their complex and opaque decision-making processes. XAI seeks to enhance the acceptance and clarity of Machine Learning (ML) algorithms by offering explanations and interpretations. In SHM, this means engineers and decision-makers of diverse backgrounds can understand why an AI system has identified a potential issue or recommended a particular course of action. Integrating explainable ML algorithms into contemporary smart SHM systems is helpful to improve the acceptance of AI solutions in SHM problems. There is a notable shortage of literature for implementing XAI methods in data-driven SHM applications.

1.2. RESEARCH QUESTIONS:

To effectively achieve the research objective, a series of research questions have been developed. These questions provide a structure for examining the relevant existing literature and formulating solutions to to enhance the accuracy and interpretability of SHM models in detecting and localizing structural defects through vibration analysis.

MAIN RESEARCH QUESTION:

How to develop an interpretable data driven SHM model for damage classification that uses vibration data?

SUB RESEARCH QUESTION:

- What are the current ML models and techniques used in the SHM methods to identify defects based on accelerometer data?
- What ML methods are suitable for damage localisation ?
- How can we integrate XAI frameworks to explain SHM models?

1.3. THESIS STRUCTURE

This thesis is divided into six chapters, each unfolding systematically to cover various aspects of the research. Chapter 1 sets the stage by offering background context and posing the research questions. Chapter 2 conducts a comprehensive literature review to tackle these questions, while pinpointing existing research gaps. Also, introduces theoratical background of analytical models utilized in the research, including model-agnostic explanatory methods such as SHAP and LIME.

Chapter 3 outlines the core research methodology employed, specifically the CRISP DM framework tailored for engineering data mining. Chapter 4 explores the process of data collection and the initial steps involved in Exploratory Data Analysis (EDA), which include data preprocessing, feature engineering, and the examination of various models for data modeling and individual classification, along with their respective evaluations.

Building on this, Chapter 5 elaborates on the findings from applying specific machine learning techniques to predict customer investment behaviors, highlighting the insights gained through SHAP analysis. Conclusively, Chapter 6 synthesizes the thesis's findings, addressing the research questions and emphasizing both the scholarly and practical contributions of the study. Additionally, it critically evaluates the limitations encountered and suggests directions for future research endeavors.

2

BACKGROUND

This chapter begins with the systematic literature review we conducted to address the questions posed in Chapter 1. Later, the chapter provides an overview of various machine learning and time series analysis methodologies that have been applied in subsequent chapters. By examining these methodologies, we aim to offer insights into how they have been adapted and optimized for Structural Health Monitoring (SHM) applications. Furthermore, this chapter will analyze the evaluation methods used to assess the performance of machine learning models in time series analysis. Through this analysis, we intend to elucidate the standard metrics and validation strategies, such as cross-validation and techniques specific to time series, that underscore the robustness of machine learning predictions in SHM scenarios. This chapter not only serves to inform future research directions but also aims to bridge gaps in the literature, paving the way for advancements in the use of interpretable machine learning methods for SHM.

2.1. REVIEW METHODOLOGY

In this section, we discuss the selection of databases, the queries used for literature search, and the criteria for determining eligibility. In the literature review multiple databases were considered instead of one to minimize bias. These include Scopus, Web of Science, and IEEE Xplore. To import relevant articles, specific modifiers were applied during the database search. These modifiers included language, restricting the search to English-only articles. Accessibility was another criterion, ensuring the articles were obtainable through University of Twente subscriptions. Finally, the publication date was a crucial modifier; we limited our search to papers published after 2010. This was done to exclude outdated methods from our review.

2.1.1. SEARCH QUERIES:

The train of thought underlying the formulation of search queries is to comprehensively cover academic literature related to data-driven SHM systems. Recognizing the broad scope of this topic, it is essential to narrow our focus on at least two fronts:

1. The types of sensors used in experiments, particularly those measuring acceleration of structural vibration,
2. The specific structures under study. Consequently, our research primarily considers studies related to wind turbine blades and other related composite structures with rotating components.

Key words such as ‘Structural Health Monitoring ’ and ‘accelerometer’, as detailed in the list below, were utilized in the search process. These keywords were combined using the logical operators ‘AND’ and ‘OR’ to optimize the search results. The search process was repeated for Scopus; Web of Science; IEEE explore; dblp. Query results are exported in RIS(Research Information Systems) format.

1. Structural AND Health AND Maintenance OR Management AND accelerometer
2. Structural AND Health AND Maintenance OR Management AND accelerometer AND Rotate
3. Structural AND Health AND Maintenance OR Management AND XAI OR explainable OR interpretable
4. Structural AND Health AND Maintenance OR Management AND accelerometer AND Data-driven

The collected data from various databases were amalgamated using Covidence, a software tool designed to streamline and manage literature from multiple sources. Covidence begins by identifying duplicates in the initially imported set of references and maintains this check throughout the importation process. The first unique instance of a reference is treated as the primary reference, with any subsequent instances marked as duplicates. Covidence then continuously checks for duplicates in each new import, cross-referencing both the current import batch and all previous ones. The software performs these checks by comparing titles, publication years, volumes, and author names.

Once uploaded RIS format results, Covidence removes the duplicate papers repeated in these query results. We had combined count of 257 entries uploaded to Covidence from all the selected databases, Covidence removed duplicates and reduced the number to 186. In the next phase, title and abstract screening, here 32 papers are selected for full text review. Selection criteria for these selection and exclusion is documented in Section 2.1.2. From the 32 selected for full text review, they are further reduced to 25 after full text review of these papers. These steps are visualised a PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart Figure 2.1.

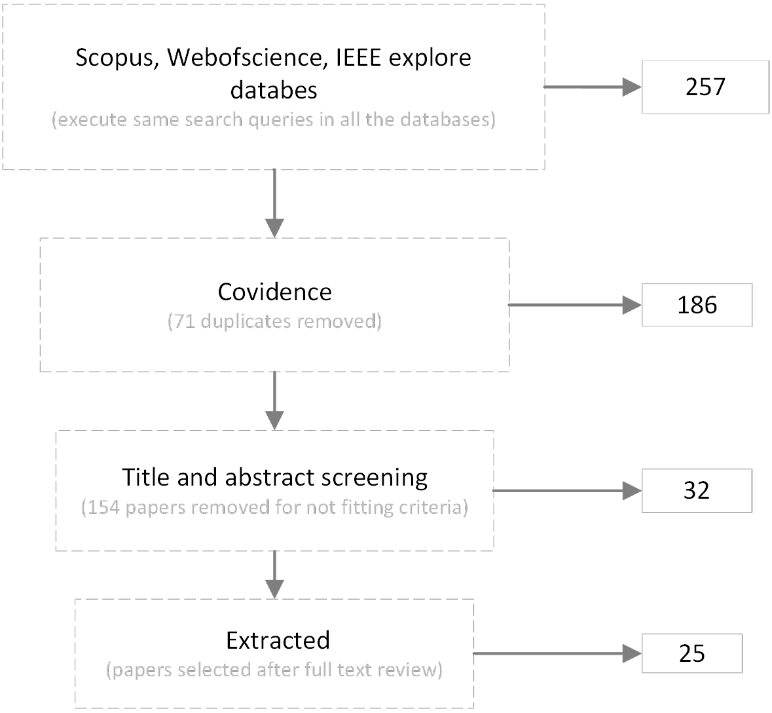


Figure 2.1: Systematic Literature Review flow diagram

2.1.2. SELECTION CRITERIA:

This, section we establish eligibility criterion considered to exclude or include a paper in our systematic literature review. The objective of this sub-section is to have systematic knowledge of why we are including certain papers in the review, and while excluding other. We are including most papers regarding XAI in SHM. Because as shown in Section 2.2, there is limited literature available for implementing XAI techniques in SHM. On the other hand, literature for data-driven techniques originates from diverse backgrounds. To narrow the scope of our study, we have selectively included papers that conduct experiments on Dynamic Load-Generating Components of composite structures. These are elements within a mechanical or structural system that produce forces or loads as a result of their motion or operation. In this process we are excluding stationary structures, like bridges. Additionally, our scope explicitly excludes papers with engineering informed modeling of SHM and papers without any data-driven methods.

Table 2.1: Table containing a summary of the criteria used to select the articles

Criteria	**Decision**
The pre-defined keywords are included in the title, abstract or in the keyword list of the paper	Included
The paper was written in English	Included
If the paper is on XAI in SHM	Included
If experiments focus on Dynamic Load-Generating structures	Included
If experiments focus on stationary structures	excluded

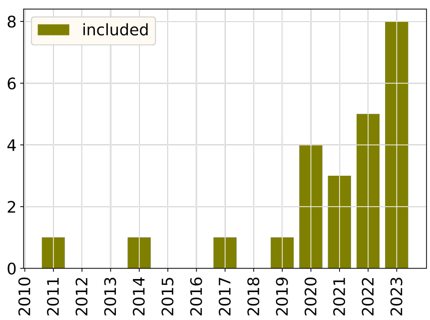
The paper was published before 2010	Excluded
Duplicates of an original paper	Excluded
The paper was not available through UT subscription	Excluded

2.2. RELEVANT TRENDS IN LITERATURE

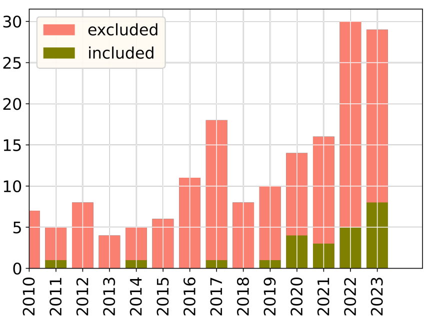
2.2.1. YEAR WISE DISTRIBUTION:

The year-wise distribution of selected papers depicted in Figure 2.2a. We can see the sudden surge of research in data-driven SHM in recent times. Extracted papers, literature which we are specifically interested can be seen in comparison with excluded papers Figure 2.2b. This show, there's nouveau trend in usage of selected keywords**??** in academic literature. Literature regarding usage of XAI in SHM context is seen only in last couple of years, which may suggest a growing interest in explainability among institutions. Among the selected papers published in time span of 2010-2023, there are few years, in early half of the range there are no papers regarding XAI implementation in context of SHM. We can conclude from this there are only efforts in recent times to publish interpretable SHM systems.

figure



(a) selected papers



(b) selected vs excluded papers

Figure 2.2: Temporal trend

2.3. DOMINANT THEMES

In this section, we outline the dominant themes observed in the literature pertaining to data-driven SHM and XAI in SHM. Our analysis will be conducted from the following perspectives: the setting in which the studies were conducted, the data-driven techniques adopted and the observed themes in the implementation of XAI. The settings of experiments can be categorised based on how the experiment in generating its data from, this can be developed multiple settings, most common models are following:

2.3.1. CONTEXTUAL ANALYSIS OF EXPERIMENTAL SETTING:

In this subsection, we examine various experimental setups encountered in the literature on data-driven SHM . A key feature distinguishing these setups is the methodology employed for data collection in the experiments. The three primary types of setups identified are: Simulation based, Physical prototypes and real-time monitoring. These data collection methods are compared in Table 2.2.

Table 2.2: Differences between experimental setups

Aspect	Simulation prototype	Physical prototypes	Real-time monitoring
Methodology	Data generated using digital prototypes, often through finite element models.	Utilizes scaled physical models of structures in controlled lab environments.	Involves monitoring structures in operational environments without induced vibrations
Advantages	Ability to generate data under controlled factors, effective for simulating damaged behavior in structures.	Facilitates study of material responses under stress and damage with controlled vibrations, allows for various damage inductions.	Provides data from natural forces, leading to variability and complexity in data, offering a realistic scenario for SHM systems.
Limitations	Difficulty in replicating real-world scenarios perfectly, leading to discrepancies between simulated and actual field data.	Size constraints of lab-tested structures affect data accuracy, challenges in extrapolating to full-scale structures.	Unpredictable and uncontrollable natural forces, inability to deliberately inflict damage for monitoring purposes.

- **Simulation:** In these experiments, data is generated using digital prototypes. [8, 9] and [10] utilise finite element models to simulate data. The primary advantage of simulated experimental setups is their ability to generate data under various controlled factors. This approach is particularly effective in simulating the damaged behavior of structure, a scenario that is challenging to replicate in physical setups. Based on this simulated data, multi-class SHM models are developed.
  However, simulation-based experiments also present certain limitations. A major challenge is the difficulty in replicating real-world scenarios perfectly. In actual conditions, structures face a multitude of factors, including varying load conditions, environmental influences, and unpredictable events, which are often challenging to fully simulate. This gap in replicating real-world conditions leads to inevitable discrepancies between the simulated data and actual field data. Such discrepancies can impact the effectiveness of SHM systems when deployed in real-world scenarios.

- **Use of Physical prototypes:** This is most commonly observed setting in literature[11–22]. In a controlled lab environment, scaled physical models of structures are used. These models are subjected to controlled vibrations using actuators, allowing for precise manipulation of frequency, amplitude of vibration. This approach facilitates the study of material responses under various stress and damage. Similar to simulation we can induce various types of damages. Sensors networks on these models collect data on vibration responses. Based on this data SHM systems area developed and validated.
  However, this approach is not without limitations. A significant challenge is the constraint imposed by the size of structures that can be realistically tested in a lab. The need to scale down models can affect the accuracy of the data, as size is a critical factor in-

fluencing stress distribution and vibration characteristics. This issue of scaling presents challenges in extrapolating findings from these models to full-scale structures. Unlike in lab setup , it is also not feasible to induce controlled vibrations on structures in real-time field conditions.

- **Real-time monitoring:** This scenario is less commonly covered in the literature, as noted by [23]. Unlike lab-based experiments, real-time monitoring does not permit the use of actuators to induce controlled vibrations. In operational environments, structures are subject to natural forces like wind, seismic activity, which are inherently unpredictable and uncontrollable. Consequently, the resulting data tends to be more variable and complex to analyze. Unlike laboratory settings where damage can be intentionally inflicted to study its impact, real-world structures cannot be damaged deliberately for monitoring purposes. This limits SHM systems development with real-time monitoring data.

2.3.2. DAMAGE SENSITIVE FEATURES:

Damage sensitive features (DSF) are features extracted from raw data which are helpful for models to predict or quantify damage. These damage discriminative features can be classified into two types, as shown in Figure 2.3 domain knowledge based and automatic discriminative feature extraction.

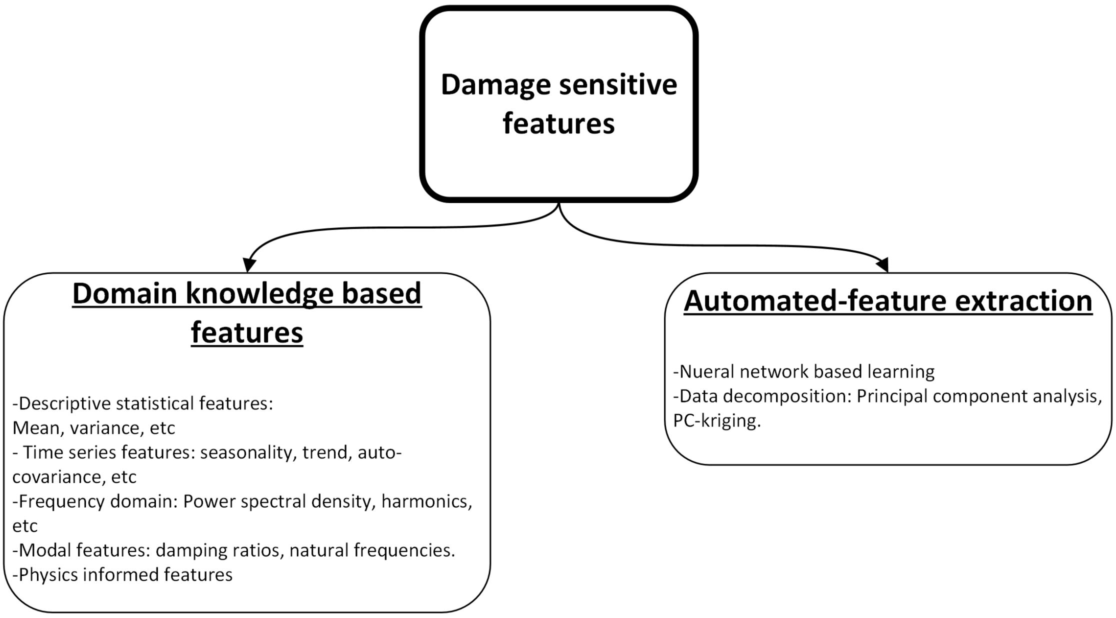


Figure 2.3: Damage sensitive features

DOMAIN KNOWLEDGE BASED:

These are manually extracted features, features are manually identified, selected, and extracted from the data by experts or analysts, rather than being automatically derived by an algorithm. These are features or properties extracted from raw data, which strongly help models to discriminate damage classification. These features observed in literature review, can be classified into three categories based on domain knowledge statistical features, Modal features, physics informed features. Formula-based features are often easy to understand and helpful building interpretable models.

- **Statistical features:** The statistical features of data in time or frequency domain tend to give some information regarding health of structure. Raw data samples can be represented with an array of descriptive statistical features of data in both time and frequency domain. correlation and auto-correlation features also can help with damage discriminating model. In frequency domain features like power spectral density or frequency domain representations can be crucial. The study by Xu et al. [24] utilizes harmonic frequencies of rotational frequency as a damage-sensitive feature. Peng et al. [25] calculate the similarity between data distributions, using this similarity as an indicator of the health state of wind turbines. Movsessian et al. [26] develop covariance based damage indicators.

- **Modal features:** This technique involves extracting changes in modal properties [11, 13] of a structure, such as natural frequencies, damping ratios, and mode shapes. Modal analysis is used to determine these characteristics, which are sensitive to changes in the physical properties of a structure, such as mass, stiffness, and damping. Any significant changes in these modal parameters can be indicative of damage.

- **Physics-Informed Damage Detection:** This approach integrates the principles of physics with data-driven methods. Cross et al. [27] combines the understanding of physical behavior of structures with data analytics and machine learning techniques to detect, locate, and quantify damage. This method often involves the use of computational models that are informed by physical laws (like mechanics of materials) and are calibrated or updated based on actual data from the structure.

AUTOMATIC DISCRIMINATIVE FEATURE EXTRACTION:

Automatic discriminative feature extraction process, automation in feature extraction involves using algorithms to automatically identify and extract these features, without requiring manual intervention. Other than neural networks learning discriminative features, [10] and [22] employ (Principal component analysis(PCA) and Polynomial Chaos-Kriging respectively) to extract descriminative features.

- **Principal compenent analysis(PCA):** The data is linearly transformed onto a new coordinate system such that the directions capturing the largest variation in the data can be easily identified.

- **PC-Kriging:** Polynomial Chaos Expansion(PCE) is a method used in uncertainty quantification. It represents a model's output as a series of polynomial functions of random variables. Raw data is reduced in dimensonality by polynomial chaos expansions. Kriging is then applied to the coefficients of the polynomial chaos expansion, rather than to the original data points.

2.3.3. ANALYSIS OF DATA-DRIVEN TECHNIQUES IN SHM:

Data-driven techniques in SHM literature are primarily divided into two categories: Statistical Hypothesis Testing and Machine Learning-Based Damage Detection.

HYPOTHESIS TESTING:

IIn this category, approaches for statistical analysis are highlighted, as exemplified by the methodologies developed by [28] and [19]. [19] focuses on the quantitative evaluation of damage using statistical performance metrics. These metrics are obtained from the analysis of dynamic responses in structures. The method involves scrutinizing changes in statistical properties, which can reveal anomalies potentially indicative of damage. Features derived from sensor data, aiding in the classification of a structure's health status, are thus regarded as health or damage indicators. Predominantly relying on domain-specific knowledge, these methods are effective for feature extraction. However, when dealing with real-time data subject to environmental influences, the task of accurately classifying the health condition using these descriptive features becomes more challenging.

Based on [28], learning and inspection phases are null and alternative hypothesis. In learning phase the selected characteristic quantities(Q) of healthy(Q_h), damaged(Q_d), various damaged levels($Q_{d1}, Q_{d2}, ..$) and unknown state(Q_u). The problem can now be framed as hypothesis testing statemnts as:

$$H_0 : Q_u \approx Q_o \textit{ Null hypothesis, unknown state of structure is healthy}$$
$$H_1 : Q_u{\approx}Q_o \textit{ Alternative hypothesis, unknown state of structure is damaged}$$
$$H_{d1} : Q_u \approx Q_{d1} \textit{ unknown state of structure is damage type 1}$$
$$H_{d1} : Q_u \approx Q_{d2} \textit{ unknown state of structure is damage type 2}$$

MACHINE LEARNING METHODS:

This approach utilizes machine learning algorithms to analyze data collected from structures. It takes both domain knowledge features and learns automatically discriminating features. These models are capable of learning complex patterns and relationships within the data, facilitating applications such as predictive maintenance, anomaly detection, and damage identification. A variety of data sources, including sensor data, environmental factors, and operational conditions, can be utilized for training predictive machine learning models.
Most commonly used models in literature can be observed in Table 1. Principal Component Analysis (PCA) is frequently employed for dimensionality reduction of raw data, as seen in studies by [10, 17, 18]. However, in [19] uses Singular spectrum analysis to decompose the raw vibration data.
In literature after extracting required features from senor data, following methods are used to classify the structure as shown in Table 2.3. KNN, Support Vector Machine (SVM), XGBoost

and neural networks classifiers are used on tabular statistical features, Convolutional Neural Networks (CNN) are specifically used in cases where tabular data or statistical feature matrix is transformed into image data. Features like short term fourier transform spectrum images are used as DSF. Among other neural networks [9]implemented bi-directional LSTM(Long short-term memory) and [21] adapted transformer to train on sequential time series data.

Table 2.3: ML techniques in literature

Features	Machine learning techniques in literature	Papers
Tabular features	KNN, SVM, XGBoost and neural network classifiers	[8, 10, 14, 17, 18, 20, 23, 26]
Feature matrix	CNN	[15, 29]
Sequential data	LSTM and Transformmer	[9, 21]

2.3.4. XAI TECHNIQUES FOR DATA-DRIVEN SHM SYSTEMS:

Though there is limited literature for implementation of XAI in SHM, common methods used are Layer-wise relevance propagation (LRP), SHAP, LIME.

- **LRP:** This method is employed to explain the decision-making process in deep neural networks. Studies by [15, 29] transform the feature matrix into a grayscale image for each sample. These image samples are classified using CNN. LRP works by reverse propagating a network's prediction to interpret it, adhering to a predefined set of rules as outlined in [30]. LRP explanations highlight the most relevant features of an image that influenced the prediction results.

- **SHAP(SHapley Additive exPlanations):** This isa model agnostic method in XAI doamin. It is grounded in the Shapley values, derived from game theory, to interpret predictions. SHAP assigns a Shapley value to each feature value of an instance, representing the feature's contribution to the prediction's deviation from the average outcome. Studies by [8, 31] and [26] have employed SHAP tools for interpreting their results.

- **LIME(Local Interpretable Model-agnostic Explanations):** LIME helps in understanding complex, black-box machine learning models by approximating them locally with interpretable models. It is used to explain individual predictions, making it easier to understand why a model made a particular decision. explanations are local, meaning they explain individual predictions and don't necessarily provide a global view of the model's behavior.

2.4. GAP ANALYSIS

Significant findings from the review highlighted three primary areas: damage detection, which has seen an uptick in research, particularly utilizing sensors and mechanical signals; damage classification, enhanced by multi-sensor networks aiding in more precise localization; and the emergent role of Explainable Artificial Intelligence (XAI) in SHM, with tools like LIME and SHAP beginning to illuminate the contributions of specific frequencies and features in classification tasks.

However, the literature review also underscored notable gaps, particularly the nascent stage of XAI within SHM and a conspicuous scarcity of real-time data applications in existing studies. This gap not only underscores the field's nascent recognition of XAI's importance but also highlights a critical avenue for future research— the validation of SHM models using real-time data. Such efforts are vital for advancing SHM's capabilities and ensuring the practical applicability and understanding of these complex models, thus providing a fertile ground for future investigations to enrich the domain's literature and practical implementations.

2.5. MACHINE LEARNING METHODS

This section explores the theoretical underpinnings of the machine learning algorithms applied in this study, which were selected to answer the posed research questions. The choice of these algorithms stems from a detailed evaluation of machine learning strategies as revealed through the Systematic Literature Review (SLR). Specifically, these algorithms were chosen for their relevance to the domain of SHM. Each algorithm offers a unique method of modeling and classification, with specific strengths that correspond to the aims of this research. Through an examination of their principles and capabilities, this section endeavors to thoroughly elucidate the approaches used to construct the machine learning models for this research framework.

2.5.1. RANDOM FOREST CLASSIFIER

Applying a Random Forest Classifier to time series data involves unique considerations compared to its application on static datasets. Random Forest is an ensemble learning method, that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes.

2.5.2. SUPPORT VECTOR MACHINE CLASSIFIER

Support Vector Machine (SVM) is a powerful and versatile machine learning model used for classification, regression, and outlier detection tasks. The goal of SVM is to find the best separating hyperplane that divides the data points of one class from another with the maximum margin. Hyperplane is like a line (in 2D) or a plane (in 3D) that separates data points. The equation 2.1 for this hyperplane in case of binary classfication. Where, w is direction and b is offset from origin.

$$w \cdot x - b = 0 \tag{2.1}$$

Support vectors are the data points that are closest to the hyperplane and directly influence its position and orientation. These points essentially support the hyperplane.

2.5.3. XGBOSST CLASSIFIER

XGBoost, or Extreme Gradient Boosting, is an advanced implementation of gradient boosting that is widely used in machine learning for regression, classification, and ranking tasks. It builds upon the principle of boosting, where decision trees are added sequentially to correct the errors made by previous models until a highly accurate ensemble model is constructed.

Components of XGboost:

- The **objective function** that XGBoost optimizes consists of a loss function and a regularization term. The loss function evaluates how well the model predicts the target variable, and the regularization term controls the model's complexity to prevent overfitting.

$$Obj(\Theta) = L(\Theta) + \Omega(\Theta) \tag{2.2}$$

  where:

  L is loos function,

  Ω is regularization term

  Θ represents parameters of model

- XGBoost performs **optimization** by calculating gradient and Hessian of loss function. This allows for more efficient and accurate updates. The update for the weight of each leaf in a tree.

$$\omega_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda}$$

  I_j is the set of indices of instances in j^{th} leaf.

  g_i, h_i are the first and second order gradients of .

- XGBoost uses a depth-first approach and prunes trees backward. It starts by growing the tree to a maximum depth and then prunes it using the gain from each split. The gain is calculated using the gradient and Hessian, and splits with negative gain are pruned.

2.5.4. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNNs) are a class of deep neural networks, highly effective for analyzing visual imagery. To apply CNNs to time series data, we typically use 1D convolutional layers instead of the 2D layers used for image data. In a 1D convolutional layer, the filters slide across the time series data, processing one-dimensional segments of the data to capture temporal patterns and dependencies. This process can be particularly effective for identifying local patterns or trends within the time series data that are indicative of broader behaviors or future events.

KEY COMPONENTS OF A 1D CNN FOR TIME SERIES

- **1D Convolutional Layers:** These layers apply a series of learnable filters to the input time series data. Each filter captures specific features at different scales or time frames (e.g., short-term trends vs. long-term patterns).

- **Activation Functions:** Similar to CNNs for image data, activation functions introduce non-linearity to the model, enabling it to learn more complex temporal patterns. ReLU and its variants are commonly used. In the hidden layers, activation functions help the network learn complex patterns by allowing each neuron to activate (i.e., output a non-zero value) in response to a specific feature or pattern in the input data. The choice of activation function in the output layer depends on the type of classification task.

  - **Softmax:** For multi-class classification tasks, where each instance can belong to exactly one class out of many, the softmax function is used. It converts the output scores from the final layer into probabilities by taking the exponential of each output and then normalizing these values by dividing by the sum of all the exponentials. This ensures that the output values are in the range (0, 1) and sum up to 1.

  - **Sigmoid:** For binary classification tasks, the sigmoid function is used in the output layer. It maps the output value to a probability score between 0 and 1, indicating the likelihood of the instance belonging to the positive class. For multi-label classification, where each instance can belong to multiple classes, the sigmoid function can be applied independently to each output neuron.

- **Pooling Layers:** 1D pooling layers reduce the dimensionality of the data, summarizing the features extracted by the convolutional layers. This step helps in reducing computation and in extracting more global features from the local ones identified by the convolutional layers.

- **Normalization and Regularization:** Techniques like Batch Normalization and Dropout are often used to improve training stability and mitigate overfitting.

2.5.5. TRANSFORMER

The Transformer model, introduced in the paper "Attention is All You Need" by Vaswani et al. in 2017, leverages self-attention mechanisms to process sequential data, offering advantages in capturing long-range dependencies and parallelising computations, which are crucial for efficiently handling time series data.

Transformer uses encoder-decoder architecture, both the encoder and decoder are composed of a series of identical layers that include self-attention mechanisms and feed-forward neural networks.

TRANSFORMER ARCHITECTURE FOR TIME SERIES CLASSIFICATION:

- **Input layer:** The raw time series data or extracted features are input to the model. If the raw data is used, it may be beneficial to apply normalization or standardization.
- **Positional Encoding:** To maintain the sequence's temporal order, positional encodings are added to the input embeddings.
- **Encoder Stack:** Consists of multiple layers, each containing two sub-layers: a multi-head self-attention mechanism and a position-wise fully connected feed-forward network. Normalization is applied before each sub-layer, and residual connections are used around each of the two sub-layers.
- **Decoder Stack:** While the original Transformer includes a decoder stack for sequence-to-sequence tasks (like translation), for classification tasks, the decoder stack might be simplified or omitted. Instead, the output of the encoder stack can be directly connected to a classification head.
- **Classification Head:** The output from the Transformer encoder is fed into a classification head, which usually consists of one or more dense layers followed by a softmax layer for class probability prediction.
- **Output Layer:** : Outputs the class predictions for the input time series.

2.6. TIME SERIES ANALYSIS METHODS:

2.6.1. DTW DISTANCE:

Dynamic time warping (DTW) is an algorithm for measuring similarity between two time-series or any sequential data. DTW finds optimal alignment between two sequences, minimizing distance between two sequences. **DTW distance** is a relative measure, calculated between two timeseries. To calculate DTW distance, it follows steps shown in Algorithm 1.

It starts by creating a matrix where each cell represents the distance between points in the two sequences, initializing the borders appropriately. The algorithm then fills this matrix by accumulating the minimum distance from either the left, above, or diagonal cell for each point, effectively mapping points from one sequence to the other in a way that minimizes the overall distance. Finally, by backtracking from the bottom right to the top left of the matrix, DTW finds the path that represents the best alignment between the sequences is visualised by calculating DTW distance between two undersampled signals from a sample in Figure 2.4a and Figure 2.4b.

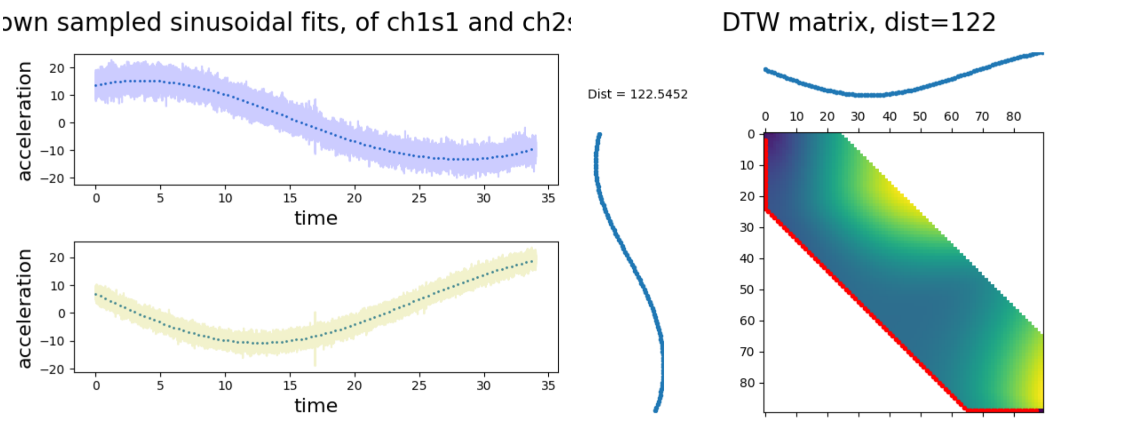
2.6.2. STATIONARISING SIGNAL USING CURVE FIT:

Stationarising a time series involves removing the effects of seasonal variations to better analyze the underlying trends. One way to do this is by using curve fitting, where you model the seasonal pattern and then adjust the data to remove this pattern.

Algorithm 1 Dynamic Time Warping Distance Calculation

Result: The DTW distance between sequences s and t
Input: Two sequences $s[1..n]$ and $t[1..m]$
Function DTWDistance(s, t)**:**

```
    /* Initialization of the DTW matrix with infinity                         */
    DTW ← array[0..n,0..m]  for i ← 0 to n do
        for j ← 0 to m do
            DTW[i, j] ← ∞
        end
    end
    DTW[0,0] ← 0 /* Populating the DTW matrix                                */
    for i ← 1 to n do
        for j ← 1 to m do
            cost ← d(s[i], t[j])  DTW[i, j] ← cost + min(DTW[i−1, j], DTW[i, j−1], DTW[i−
             1, j−1])
        end
    end
    return DTW[n, m]
```



(a) (a) Down sampled Sinusoidal fits

(b) (b) DTW matrix to calculate distance

- **Identify the Seasonal Pattern:** First, you need to identify the seasonal pattern. This could be done through plotting and visually inspecting the data or by using statistical tests.
- **Model the Seasonal Pattern:** Use curve fitting to model the seasonal pattern. Common models include sinusoidal functions for yearly patterns or polynomials for more complex patterns.
- **Fit the Model to the Data:** Use the curve fitting function from a library like *scipy.optimize* to fit the model to your data.
- **Deseasonalize the Data:** Subtract the fitted model from the original data to remove the seasonal effect, leaving you with the deseasonalized time series.

```
from scipy.optimize import curve_fit
```

```
func = ... #Fuction which represents seasonal pattern

popt, pcov = curve_fit(func, data)
#popt is array of optimal values which fit the function for given data
#pcov estimated covariance of popt array.

ydata=func(**popt, xdata)
```

WELCH POWER SPECTRAL DENSITY:

Welch's method for estimating the power spectral density (PSD) represents a significant advancement in signal processing, particularly in analyzing the power of a signal across various frequencies. This method, an adaptation of the periodogram approach, enhances the estimate's reliability by reducing its variance through averaging modified periodograms of overlapped segments of the signal. Welch's method can be described as follows:

- **Segmentation:** The time series $x(n)$, where n is the discrete time index, is divided into L overlapping segments. Each segment contains M points, with adjacent segments overlapping by D points.
- **Windowing:** Each segment i is multiplied by a window function $w(n)$ to reduce spectral leakage. This process yields windowed segments $x_i(n) = x(n)w(n)$, where n ranges over the segment length M.
- **Averaging:** The PSD estimate is obtained by averaging the modified periodograms of all segments:
$$P(f) = \frac{1}{L}\sum_{i=1}^{L} P_i(f)$$
- **Normalisation:** The PSD is normalised by sampling frequency to maintain consistency.

2.6.3. DICKEY–FULLER TEST:

The Dickey-Fuller test is a statistical test used to determine if a time series is stationary. A time series is considered stationary if its statistical properties, such as mean, variance, and autocorrelation, are constant over time. Stationarity is an important assumption in time series analysis because many statistical models require the time series to be stationary to make valid inferences.

AUGMENTED DICKEY-FULLER TEST:

ADF test is a formal statistical test for stationarity, specifically designed to help identify whether a time series has a unit root, indicating it is non-stationary. The presence of a unit root means the time series can be unpredictable and may hinder the effectiveness of many time series forecasting methods.

ADF test accounts for higher-order correlation by adding lagged differences of the time series. ADF test expands on the Dickey-Fuller test by including lagged differences of the time series in the regression equation to account for autocorrelation. The regression equation used in the ADF test is:

$$\Delta = \alpha + \beta t + \gamma y_{t-1} + \gamma_1 \Delta_{y_{t-1}} + \cdots + \gamma_{p-1} \Delta_{y_{t-p+1}} + \epsilon_t$$

where, y_t is time series, Δ is difference operator, α is constant, β is coefficent on the time trend, ϵ_t is eroor or noise.

$$H_o: \textit{ There is a unit root}$$

$$H_1: \textit{ Alternative hypothesis, there is no unit root and the time series is stationary.}$$

ADF test interpretation:

- **ADF Statistic:** A more negative value suggests stronger evidence against the null hypothesis, indicating stationarity.
- **p-value:** If the p-value is less than a chosen threshold (e.g., 0.05), we reject the null hypothesis, suggesting the time series does not have a unit root and is stationary.
- **Critical Values:** The ADF statistic should be compared to these values. If the ADF statistic is less than the critical value, we reject the null hypothesis.

2.7. EVALUATION METHODS:

- **Confusion Matrix:** This is a table(Figure 2.5) that shows the number of correct and incorrect predictions categorized by their actual classes. It provides insight into the types of errors made by the model. Matrix, which consists of four components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).
  - **True positive:** Predicted value is positive and true value is also positive.
  - **False positive:** False positive or Type 1 error, when predicted value is positive but actual value is negative.
  - **False Negative:** False positive or Type 2 error, the predicted value is negative but it is actually positive.
  - **True Negative:** The predicted value is negative and actual value is also negative.
- **Accuracy:** This is the simplest metric, measuring the proportion of correctly predicted observations to the total observations.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

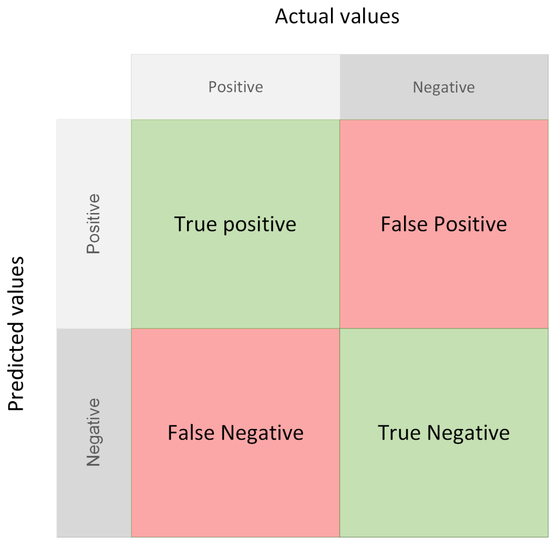


Figure 2.5: Confusion matrix for binary classification

- **Precision:** Precision measures the proportion of correctly predicted positive observations to the total predicted positives. It's crucial when the cost of false positives is high.

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

- **Recall:** Recall measures the proportion of correctly predicted positive observations to all the actual positives. It's important when the cost of false negatives is high.

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score:** The F1 Score is the harmonic mean of precision and recall, providing a balance between the two. It's useful when you need a balance between precision and recall and there's an uneven class distribution.

$$\textbf{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- **ROC (Receiver Operating Characteristic) Curve:** This is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. The area under the ROC curve (AUC-ROC) is a measure of the test's ability to distinguish between the classes. A higher AUC indicates a better model performance.

2.8. EXPLAINABLE AI METHODS:

XAI techniques in SHM aim to make the decision-making process of AI models understandable to human experts. This is crucial for ensuring safety, optimizing maintenance operations, and justifying interventions. By providing insights into how AI models derive their conclusions, XAI helps in building trust among engineers and stakeholders, facilitating the adoption of AI in critical infrastructure monitoring, and enabling timely and informed decision-making to prevent failures and extend the lifespan of structures.

2.8.1. SHAP

SHAP (SHapley Additive exPlanations) is a method used in the field of explainable AI (XAI) to interpret machine learning models. It assigns each feature an importance value for a particular prediction, making the model's output interpretable in terms of its input features. SHAP values are grounded in cooperative game theory and offer a unified measure of feature importance.
Shapley value, a concept from cooperative game theory. It represents the average marginal contribution of a feature across all possible combinations of features. For a prediction model, the Shapley value of a feature value is the average change in the prediction that the feature value causes by being included in the model.

$$\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S \cup i) - f(S)]$$

Where, F is the set of all features, S is subset of features excluding i, $|S|$ is the size of S and $|F|$ is total number of features. $[f(S \cup i) - f(S)]$, represents contribution of features when added to a subset of features S.

PLOTTING SHAP VALUES

- **SHAP Summary Plot:** Shows the importance of each feature across all the data points. It displays each point's SHAP value on a color scale, indicating the impact of that feature on the model's output.

3

METHODOLOGY

This chapter outlines methodologies undertaken to answer the research questions. It begins with adopting the CRISP-DM framework to SHM problems. This involves a systematic approach to using data mining techniques for monitoring and assessing the health of structures. SHM aims to detect, localize, classify, and assess. damage and degradation over time to ensure safety and efficiency while reducing maintenance costs.

3.1. CRISP-DM

CRISP-DM was introduced as a cross industry template for data mining in [3], Figure 3.1 illustrates basic framework of CRISP-DM. Huber et al. [2] expanded the basic framework to adapt engineering applications, Figure 3.2 depicts expanded CRISP-DM framework. Applying the CRISP-DM framework to SHM entails a systematic methodology for leveraging data mining techniques as theorised by [4] to monitor and evaluate the health of structures, with the goal of detecting, localizing, classifying, and assessing damage and degradation over time. This approach not only aims to ensure safety and efficiency but also seeks to minimize maintenance costs by tailoring CRISP-DM specifically for the unique challenges and objectives of SHM.

3.1.1. BUSINESS UNDERSTANDING

In CRISP-DM business understanding phase sets the direction for rest of the steps by understanding requirements and then translating this knowledge in terms of data mining. In the context of SHM, Gordan et al. [4] translates this step to target identification. Defining specific objectives for SHM, such as early damage detection, life expectancy prediction, maintenance need identification, or structural performance under various loads and conditions. Understanding requirements from a structural engineering perspective, including safety criteria, and operational demands. In our work this literature review and process of framing the research questions Chapter 1.

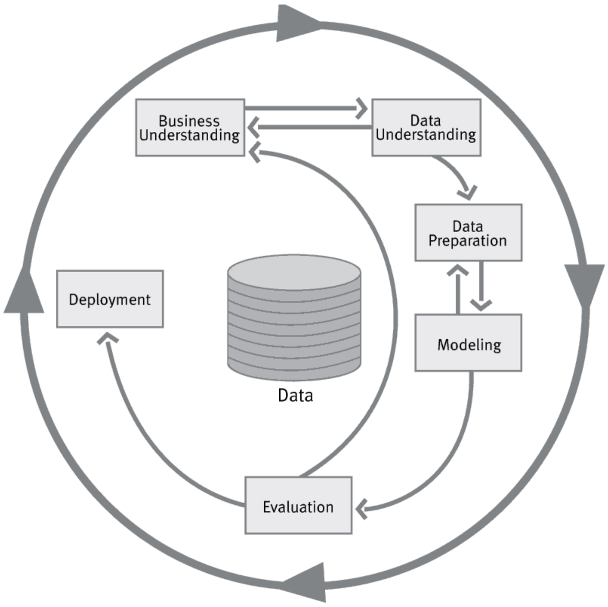


Figure 3.1: CRISP-DM framework [2]

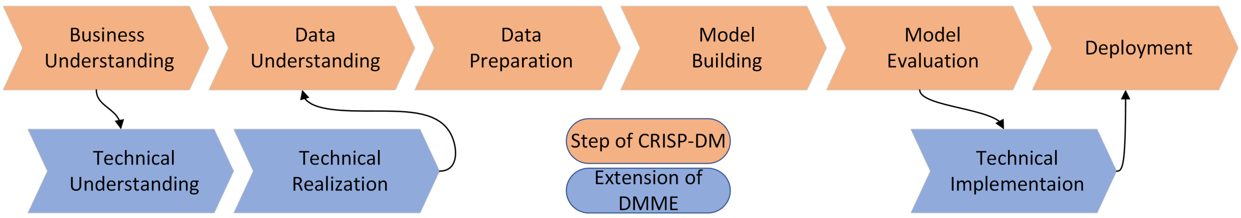


Figure 3.2: Data Mining Methodology for engineering applications [3]

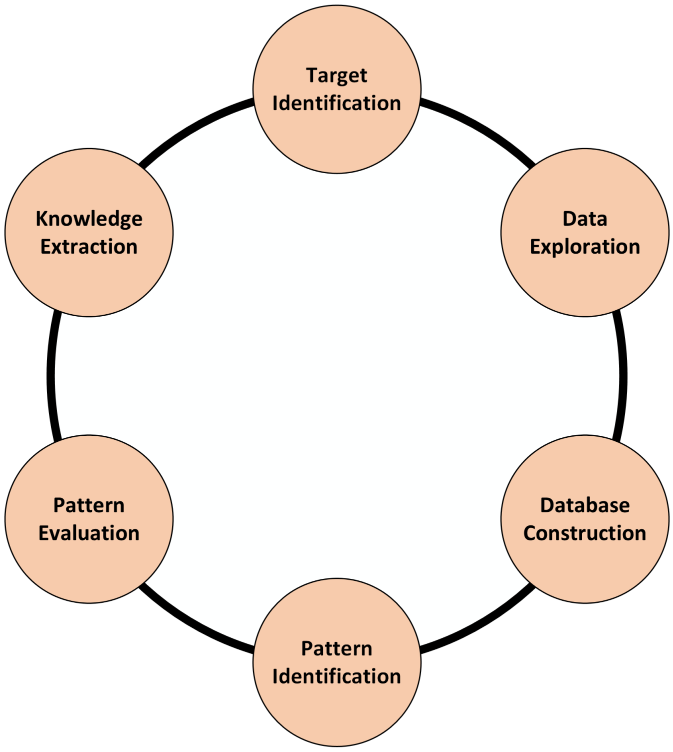


Figure 3.3: Generalised CRISP-DM for SHM [4]

3.1.2. TECHNICAL UNDERSTANDING:

This task aims to convert business objectives into quantifiable technical objectives, compile existing knowledge on relevant physical phenomena and process effects, and outline a plan for experimentation. The process encompasses several critical steps, starting with an examination of the system's structure, processes, and pertinent parameters. It then involves setting clear technical objectives and identifying target variables, from which specific technical analysis tasks are derived. An important part of this task is the collection and documentation of existing expertise concerning the relevant physical effects and foundational conditions. It also requires specifying the physical parameters and effects that will be crucial for future measurements, along with devising measurement strategies for each significant physical parameter. The culmination of this process is the creation of a detailed experimental plan to guide the execution of measurements.

3.1.3. TECHNICAL REALIZATION:

The purpose of this phase is to evaluate and choose the appropriate measurement methodologies and to implement the planned experiments. This phase is designed to produce data that encompasses all necessary details and characteristics for the following data analysis phase, aimed at achieving the specified business objectives. The steps involved include constructing a technical testing framework by identifying the most suitable data collection methods from the proposed measurement concepts. Following this, the experimental plan is executed. Documentation of every aspect of the data collection process is crucial, including any technical constraints, potential error sources, and the overall quality of the data gathered.

3.1.4. Data understanding

Once we understand business problems, the next step is understanding available data. This phase is crucial as it lays the foundation for how well you can prepare and model your data. It involves collecting, exploring, and familiarizing yourself with the data to identify quality issues, discover first insights into the data, or detect interesting subsets to form hypotheses for hidden information. The collection of data and integrating various datasets and exploring this data in different dimensions like in time and frequency domains. Another important aspect of this step is to identify quality issues in data. This part of methodology is addressed in Chapter 4.

3.1.5. Data preparation

Combine data from disparate sources into a unified database, ensuring consistency and accessibility for analysis. Address data quality issues, such as missing values, outliers, and noise. Standardize formats and scales to make the data suitable for further analysis. Develop features that are indicative of structural health, such as derived statistics from sensor readings, frequency domain features, or time series features capturing temporal patterns. Important process at this step is labelling the data and train test splitting data. This helps in unbiased evaluation of the model.

3.1.6. Modeling

Once the data is prepared for training the next step is to apply machine leaning or statistical models on the data. Choose appropriate data mining and machine learning models suited for SHM tasks. This may include regression models for predicting life expectancy, classification models for damage detection, or clustering for anomaly detection. Train models using historical data, ensuring that they are validated using techniques suitable for time series data, such as time-based cross-validation.

The next step after model evaluation is interpreting the results. To make the model interpretabile, we use SHAP method. This provides insights into the output of machine learning models. SHAP is grounded in the Shapley value—a concept from cooperative game theory. SHAP library comes with visualisation which help us understand feature importance in multiple dimensions.

3.1.7. Evaluation

This phase plays a crucial role in assessing the performance of developed models, using metrics like accuracy, precision, and recall for classification, or MSE and R-squared for regression. This phase also involves reviewing the entire modeling process to ensure proper methods were used and deciding if the model needs further refinement or is ready for deployment.

3.1.8. Technical implementation:

This task focuses on making the validated model operational during production by feeding it real-time data. It involves converting the data acquisition methods outlined in the "Technical Realization" phase of DMME into a system capable of functioning in real time. Key activities include verifying the long-term data streaming capabilities of all employed sensor technologies. This verification process covers ensuring adequate power supply, managing potential connection losses, and incorporating self-monitoring features, possibly through the use of redundant sensors. Additionally, it requires the development or selection of a software framework that can efficiently collect, pre-process, and analyze data streams from machine control systems and sensors over extended periods.

3.1.9. Deployement

Integrate the models into the SHM system, ensuring that they can process data in real-time or near-real-time as required. Set up systems to monitor the performance of the SHM system, including automated alerts for detected issues and regular model updates based on new data or structural changes.

4

EXPERIMENT SETUP

This chapter outlines the experimental framework used in the study, offering an in-depth look at data comprehension and the Exploratory Data Analysis performed to uncover insights within the dataset. Additionally, it clarifies the basic needs for the suggested solution, and describes the critical features and attributes that the solution must possess.

4.1. TECHNICAL UNDERSTANDING:

Wind turbine are complex structures designed to convert wind energy into mechanical power efficiently. Wind turbines can vary in size from small applications powering individual homes to large utility-scale turbines contributing to the electrical grid. Wind turbine with its main components depicted in Figure 4.1, and below we describe these components:

- **Rotor Blades:** These capture the wind's energy and transfer it to the rotor hub. Typically, turbines have two or three blades made of lightweight, durable materials like fiberglass or carbon fiber.
- **Rotor Hub:** The hub connects the blades and transfers the rotational motion to the generator through the main shaft.
- **Nacelle:** Mounted on top of the tower, the nacelle houses the generator, gearbox (in some turbines), drive train, and other mechanical components. It also contains control systems that adjust the blade pitch angle and yaw orientation system changes the Nacelle direction to maximize efficiency.
- **Tower:** The structure that supports the nacelle and rotor assembly. Towers are made from steel or concrete and are designed to elevate the turbine components to capture higher wind speeds at higher altitudes.

The wind turbine blades are engineered with several sections, each serving a specific function

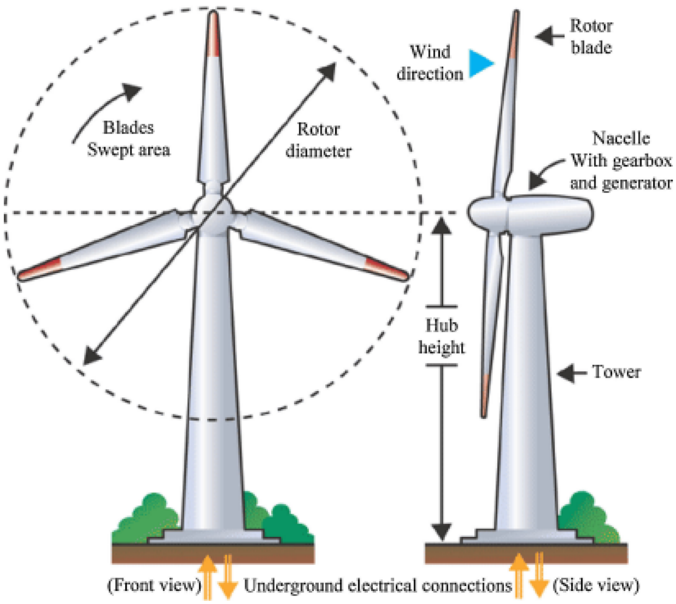


Figure 4.1: Wind turbine front and side view [5]

in aerodynamics and structural integrity. The sections of wind turbine blade are depicted in Figure 4.2 and Figure 4.3. Here's an overview of the different sections and features commonly found in wind turbine blade literature [32] and [33]:

- **Root Section:** This is the innermost part of the blade that attaches to the rotor hub. The root section is characterized by its large diameter and is engineered to withstand substantial forces and moments transmitted between the blade and the hub. It usually has a circular or polygonal cross-section for a secure connection.
- **Maximum Chord Section:** The chord of an airfoil is the straight line connecting its leading and trailing edges. The maximum chord section is where the blade's cross-section is widest. This section is crucial for generating lift and is often located near the root to maximize the blade's aerodynamic efficiency while minimizing structural loads.
- **Suction Side:** The suction side is the part of the airfoil (blade cross-section) that faces away from the direction of the wind. It's typically smoother and more curved than the pressure side. Air moving over the suction side speeds up and decreases in pressure, contributing to the lift generated by the blade.
- **Pressure Side:** The pressure side faces the oncoming wind and is the part of the airfoil that experiences increased air pressure. It is usually less curved than the suction side. The difference in pressure between the suction side and the pressure side creates lift, propelling the blade and turning the rotor.
- **Leading Edge:** The leading edge is the front part of the blade that first contacts the wind. It is designed to be aerodynamic to reduce drag and is often reinforced to withstand impact from airborne particles (e.g., rain, hail, debris).

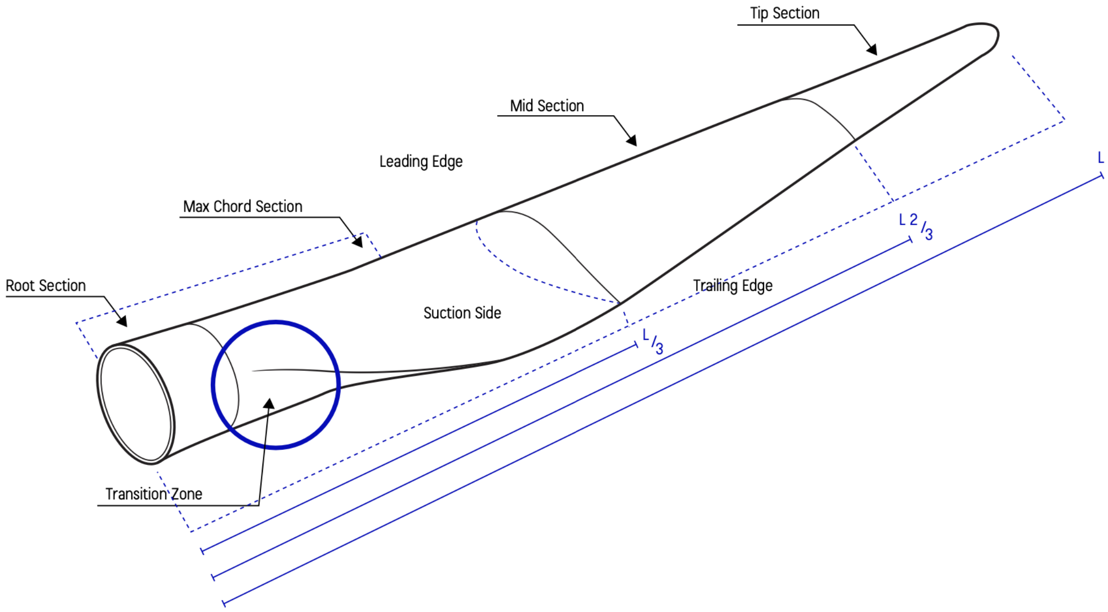


Figure 4.2: Wind turbine blade sections

- **Trailing Edge:** The trailing edge is the rear part of the blade where the air flowing over the suction and pressure sides meets. It is thinner than the leading edge and optimized for a smooth airflow exit, minimizing turbulence and drag.
- **Tip Section:** The outermost part of the blade, farthest from the rotor hub. The tip section moves the fastest and generates a significant portion of the blade's lift. It is often tapered and twisted to ensure optimal performance across various wind speeds and to reduce noise and vibration.
- **Spar Cap:** A structural component running along the length of the blade, usually on both the suction and pressure sides, providing strength and stiffness to resist bending moments.
- **Shear Web:** Internal structures that connect the pressure side and suction side of the blade, providing shear stiffness and contributing to the overall structural integrity of the blade.

4.2. TECHNICAL IMPLEMENTATION:

Data is gathered via a series of accelerometers embedded within a wind turbine mechanism identified to have a defective blade. This turbine features blades each extending 45 meters, with every blade hosting a collection of 6 sensors, totaling 18 sensors across the system for data accumulation. These sensors are organized as s1, s2, s3, s4, s5, s6 within each blade, with each blade acting as a distinct channel (e.g., the first sensor on the first blade is denoted as ch1s1). Sensors sharing the same number are positioned identically across all blades. The sensors and

their respective placements and direction of measurement is summarised in Table 4.1.
Each of these sensors is a uni-directional accelerometer, encased within cubes for placement on the blades. Specifically, a cube containing only the second sensor is located 5 meters from the blade's root on the pressure side. Another cube, positioned 15 meters from the root (one-third down the blade), houses accelerometers (sensors 1, 3, 4) that monitor in three different directions. At the same 15-meter mark, but on the suction side, another cube is positioned to measure acceleration in two directions (sensors 5, 6). Sensor 1 is aligned with the blade's length (z-axis), while sensors 2, 4, and 6 track acceleration across the leading and trailing edges (x-axis). Meanwhile, sensors 3 and 5 capture movements between the suction and pressure sides (y-axis).

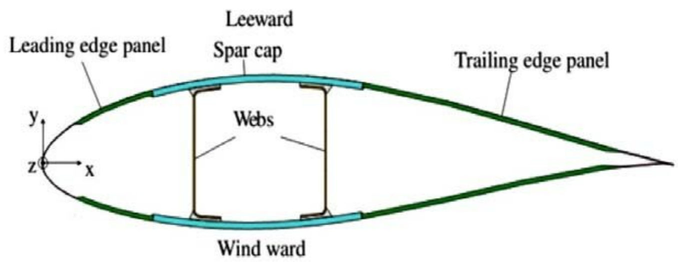


Figure 4.3: Wind turbine blade cross-sectional view

Table 4.1: Sensor placements

Sensor	Placement	Direction
Sensor 1	15m from Root section, Wind ward side	Z-axis
Sensor 2	5m from Root section, Wind ward side	X-axis
Sensor 3	15m from Root section, Wind ward side	Y-axis
Sensor 4	15m from Root section, Wind ward side	X-axis
Sensor 5	15m from Root section, Leeward side	Y-axis
Sensor 6	15m from Root section, Leeward side	X-axis

4.3. DATA UNDERSTANDING

4.3.1. WIND TURBINE BLADE MOVEMENT:

During the motion of wind turbine blade as seen in front view in Figure 4.1 the motion is in zx-plane of wind turbine blade. Senors 1, measuring in z-axis and Sensors 2, 4, 6 measuring in x-axis capture this dominant motion of wind turbine blade. Once we neglect this dominant motion from wind turbine blade movement the blade is subjected to loads based on their direction and the type of stress they cause. Three primary types of loads are edgewise, flapwise, and torsional. These loads play a significant role in the structural design and operational integrity of wind turbines. Let's explore each of these load types in more detail:

- **Edgewise vibration:** Edgewise loads act parallel to the plane of rotation and are mainly due to wind pressure hitting the side of the blade, gravity (when the blade is in the horizontal position), and centrifugal forces. While generally smaller than flapwise loads, edgewise loads are significant, especially in large blades where the effects of gravity and centrifugal forces become more pronounced. Sensors 2, 4, 6 capture edgewise vibrations of the blade.

- **Flapwise vibration:** Flapwise loads act perpendicular to the plane of rotation, that is in y-axis of wind turbine blade as shown in Figure 4.3. These are primarily caused by the aerodynamic lift forces that enable the rotor to turn. Flapwise loads can lead to bending of the blades in the direction of the wind flow. They are significant because they directly relate to the primary function of the blades—capturing wind energy. Sensors 3, 5 measure flapwise vibration.

- **Torsional vibrations:** Torsional loads refer to the twisting forces that act along the length of the blade. These can be caused by variations in wind speed across the blade length (wind shear), changes in blade pitch, and aerodynamic imbalances. Sensor 1 measuring along z-axis of blade measures this vibration.

Table 4.2: Vibrations

Vibration	**Sensors**	**Direction**
Edgewise vibration	Sensors 2, 4 and 6	X-axis
Flapwise vibration	Sensors 3 and 5	Y-axis
Torsional vibrations	Sensor 1	Z-axis

WIND TURBINE BLADE ROTATION:

The motion of a wind turbine blade is a complex combination of the blade's rotation around the turbine hub and its vibrations. As the blades rotate to capture wind energy, they also experience various forces that cause them to bend and vibrate. The rotation which is also the dominant motion of the blade can be observed in sensors 1, 2, 4 and 6. The acceleration responses of these sensors can be seen in Figure 4.4. So, one complete rotation of blade can be represented as one sinusoidal wave in accelerometer response signal. We can represent this dominant motion as sinusoidal-fit of the response signal. Sensor 1 and 2, 4, 6 in a same blade have phase difference of 90°, since they are measuring in perpendicular axes. In case of comparing blades, in a tri-blade wind turbine system angle which has between adjacent blades is 120°. This implies, that when comparing the readings from the same sensors on different blades, there is a 120° phase difference as shown in Figure 4.6.

ESTIMATING ROTATIONAL FREQUENCY:

In movement of a wind turbine blade primary load is the rotational motion to generate energy, this is also reflected in frequency response of the acceleration signal. When we represent acceleration data in frequency domain rotational frequency is dominant frequency with maximum

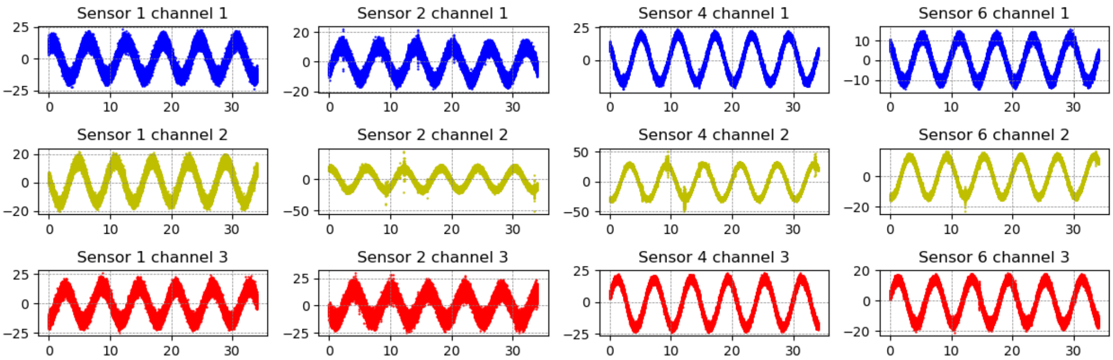


Figure 4.4: Accelerometer measuring in yz-axes of a sample

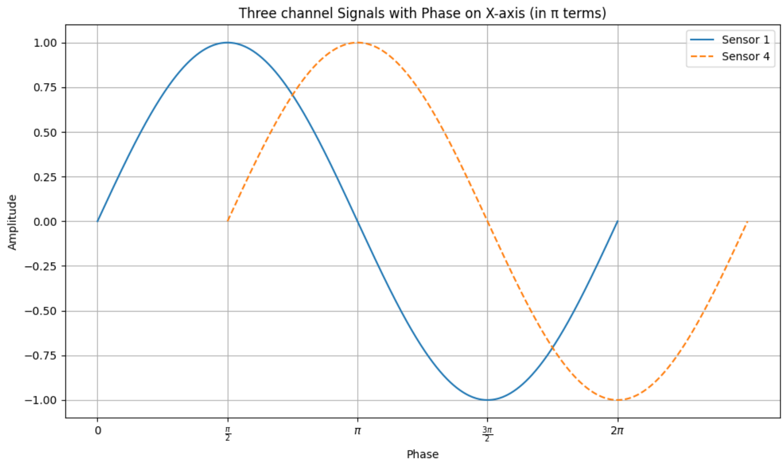


Figure 4.5: Comparing phase difference of sinusoidal fits of sensor 1 and 4 of channel 1

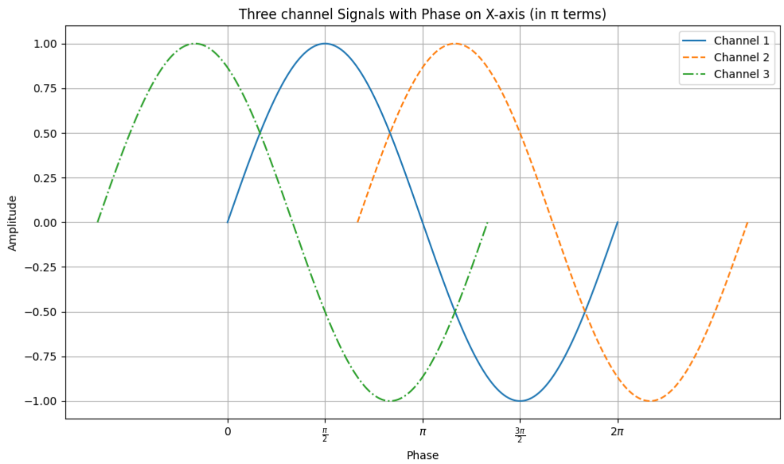


Figure 4.6: Comparing phase difference of sensor 1 sinusoidal fits, across channels

amplitude. This can be observed observed across the sensors in a sample. In Figure 4.7 frequency bin with maximum amplitude is $0.18Hz$, so estimated RPM of this sample is 60 times 0.18.

Algorithm 2 Rotational frequency estimation

Data: Signal time series of a sensor with data points yy and time order tt
Result: Rotations per minute rpm
RPMestimate yy, tt**:**
$fs \leftarrow 1/(tt[2]-tt[1])$ $ff \leftarrow$ np.fft.fftfreq(len$(tt), (tt[2]-tt[1]))$ $Fyy \leftarrow$ abs(np.fft.fft(yy)) $max\_index \leftarrow$ np.argmax$(Fyy[1:]) + 1$ $max\_freq \leftarrow$ abs$(ff[max_index])$ **return** $max\_freq * 60$

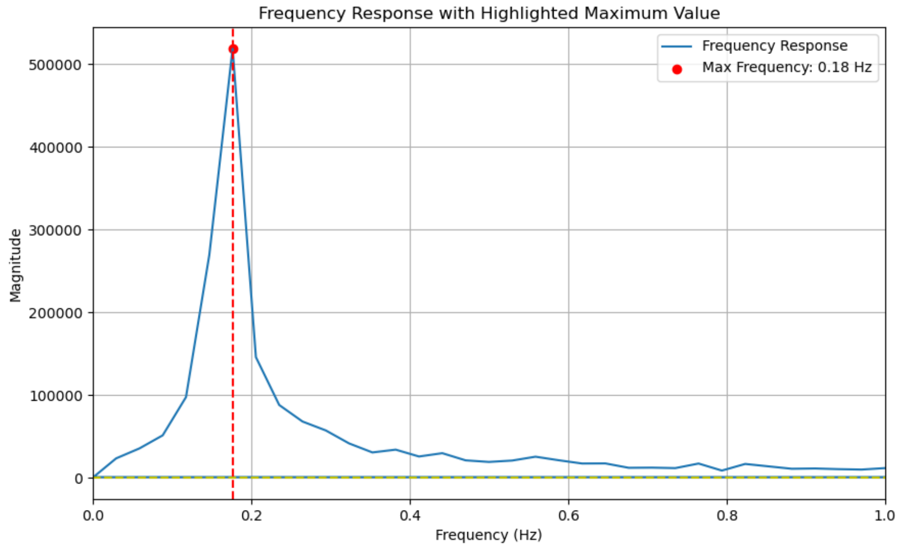
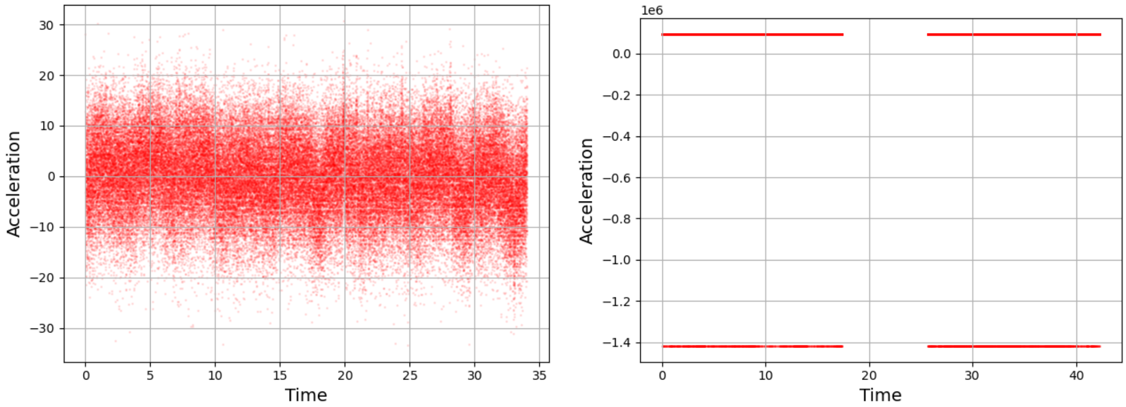


Figure 4.7: Estimating dominant frequency

4.3.2. DATA INCONSISTENCY:

Data collection via sensors commenced on June 17th, 2023. However, the dataset exhibits some discrepancies in the samples collected from some of sensors during the data acquisition process. The observed inconsistencies shown in Figure 4.8a and Figure 4.8b, corruptions noticed can be categorized into several types of corruption:

- **Extreme Values:**A few sensors recorded extreme values of the orders of 10^6, which stand out as anomalies. These outliers can easily be identified by examining basic statistical parameters such as the maximum, minimum, and range of the sensor data.
- **Asynchronous Vibration:**This category differs from the extreme values in that the recorded data falls within the typical range of accelerometer readings but lacks synchronization with other sensor channels. While the data from two channels might resemble sinusoidal waves, the affected sensor displays stochastic vibrations that do not align with the readings from other channels. Identifying this type of discrepancy involves comparing statistical features such as range, root mean square (RMS), standard deviation, and Dynamic Time Warping (DTW) distance between the blades.
- **Time Gap:** This issue is characterized not by data corruption per-se but by interruptions



(a) (a) Asynchronous vibration

(b) (b) Extreme values with time gap

in the data collection process. Instead of a continuous 30-second sample, there are instances where the data is split into two 15-second segments separated by a 20-second gap. Such occurrences should be treated as distinct samples for analysis purposes.

The initial phase in the data processing flow involves excluding corrupted samples from subsequent modeling efforts. This entails the elimination of asynchronous samples and those exhibiting extreme values. The process commences with the manual labeling of data, which involves examining the time series plots of the samples. Following this, classifiers that have been trained on manually labeled data are utilized to distinguish between healthy and unhealthy samples. As depicted in the data flowchart Figure 4.9, we construct and evaluate models based on statistical features alongside neural network models trained on time series data to identify corrupted samples.

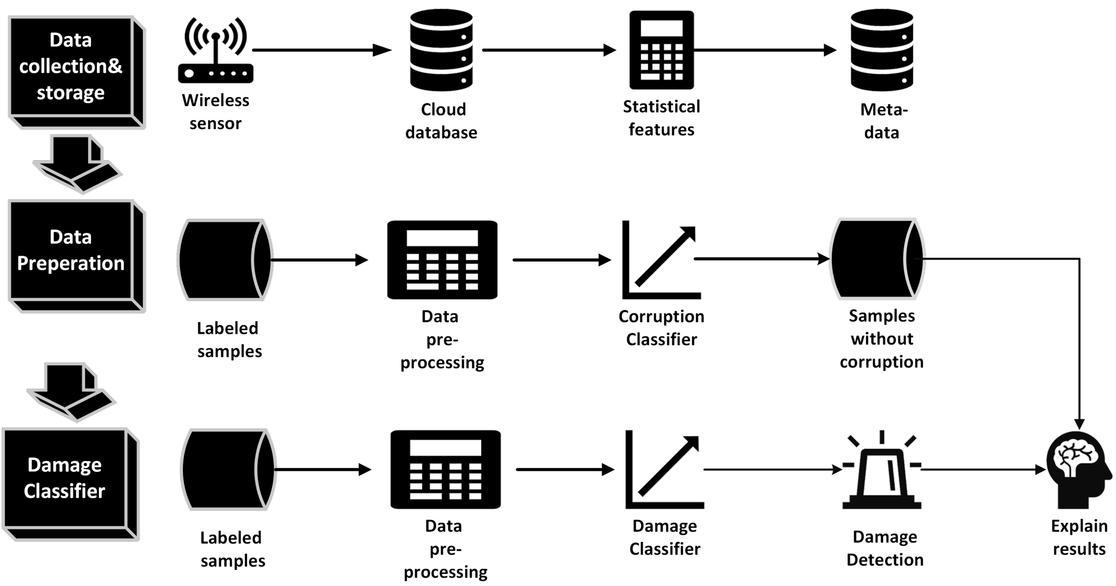


Figure 4.9: Overview of methodology

4.4. DATA PREPARATION

In our research, we utilized three distinct types of models, each based on the specific nature of the data on which they were trained. The process of preparing the data was segmented into three sub-categories, tailored to accommodate the different models we employed.

- Statistical features
- Sequential features
  - Time series sequences
  - Frequency response

4.4.1. STATISTICAL FEATURES BASED CLASSIFIER:

In the data analysis process, **feature engineering** is essential for preparing the dataset for the predictive model. This involves extracting important statistical features from the time series data, capturing the dataset's distribution. During feature engineering, we extract relevant statistical features from the time series data. The computed features are described below:

- **DTW distance:** We consider DTW distance measure of a sensor response from similar senors in remaining blades to measure the similarity of the sensor response.

$$dtw_{channel1} = \frac{dtw_{channel\ 1\ and\ 2} + dtw_{channel\ 1\ and\ 3}}{2}$$

- **Standard deviation:** To measure variation of data from it's mean. The standard deviation of the given time series data is the value that quantifies the dispersion of the data points around their mean, indicating how spread out the values are in the series.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)}{N}}$$

σ = standard deviation;
N = length of the time series;
x_i = data points in timeseries;
μ = mean of time series data points;

$$\mu = \frac{\sum x_i}{N}$$

- **Root mean square value:** The Root Mean Square (RMS) value is a statistical measure of the magnitude of a varying quantity. It is particularly useful in contexts where variations are both above and below a mean value, especially when dealing with waveforms or alternating amplitude.

$$RMS = \sqrt{\frac{1}{n}\sum x_i^2}$$

- **Inter-quartile range:**The Interquartile Range (IQR) is a measure of statistical dispersion, which is the spread of the data points in a dataset. The IQR is calculated as the difference between the 75th percentile (the upper quartile, Q_3) and 25th percentile (the lower quartile, Q_1)of the data. This range covers the middle 50% of the data points and is used to evaluate the variability while ignoring outliers and extreme values. The IQR is particularly useful in identifying outliers and understanding the overall spread of the central portion of a dataset.

- **Skewness:** Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, negative, or undefined. In a perfectly symmetrical distribution, the skewness would be zero. The signals with high skewness are with varying RPM.

$$\mu_3 = \frac{\sum_i^N (X_i - \mu)}{(N-1) * \sigma^3}$$

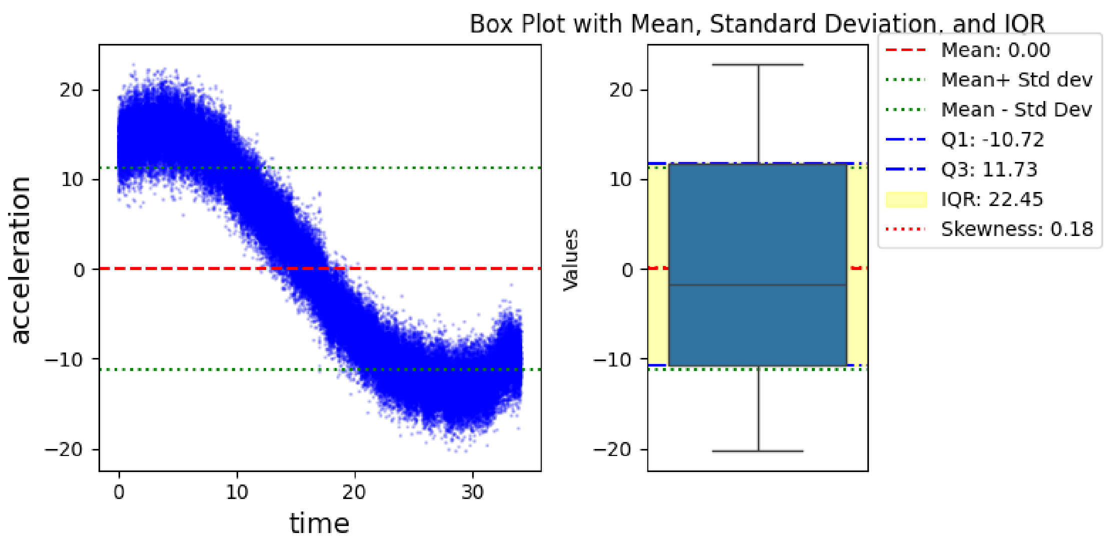


Figure 4.10: Statistical features extracted from sample

EXPLORATORY DATA ANALYSIS OF STATISTICAL FEATURES

The correlation matrix of statistical features along with labels, as illustrated in Figures Figure 4.12a, reveals none of the metrics have strong correlation with the labels. This matrix aids in understanding the relationships among variables, yet it does not necessarily imply causation. Figure Section 4.4.1 presents distribution plots for selected statistical features, including standard deviation, skewness, inter-quartile range, RMS, and range, which are log-transformed to correct for skewness and facilitate the understanding of their distribution. Additionally, Figure Figure 4.12b depicts the distribution of manually labeled samples for corruption classifier, comprising 250 healthy and 328 corrupted datapoints.

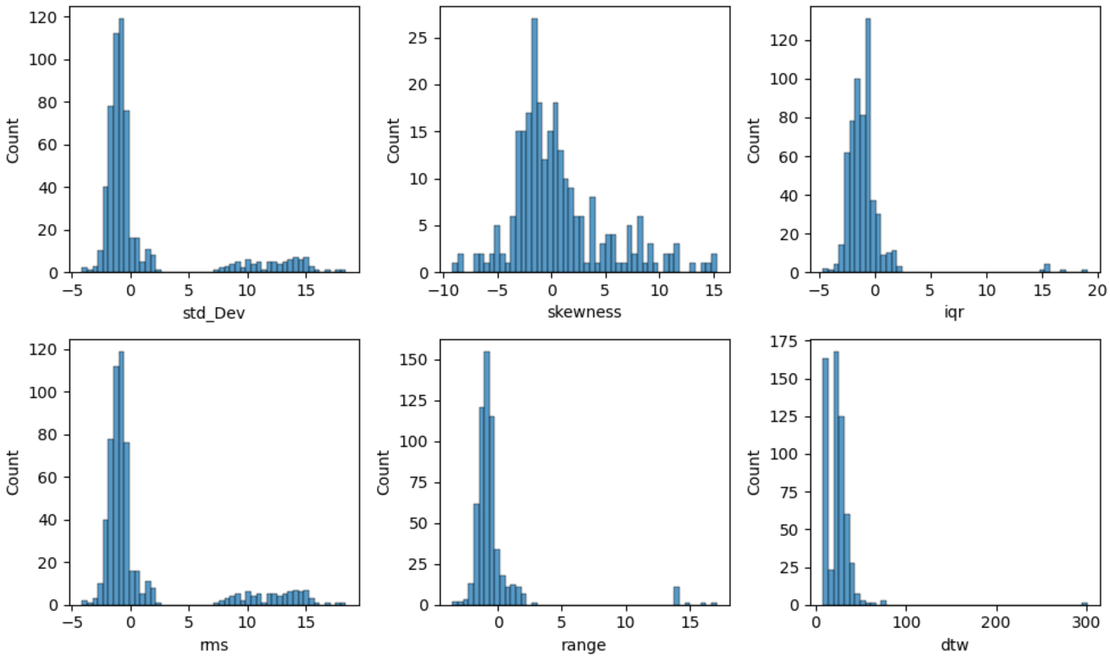
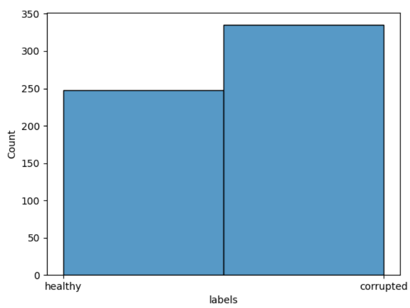


Figure 4.11: Distribution of features

(a) a. Correlation between features



(b) b. Labels

4.4.2. SEQUENTIAL FEATURES:

We use two kinds of sequential data in our experiments. One is time series data, responses from senors and other being frequency domain representation of these. In case of training with time series data we decimated raw data in time domain from 90,000 data points to 900 data points by selecting every 100^{th} point to avoid time complexity.

The frequency response of data is used to train Transformer neural network. In the context of training Transformer neural networks, leveraging the frequency response of datasets as input has been identified as a strategic approach to enhance learning efficiency. Transformers, when trained on decimated raw data, often struggle to identify underlying patterns, a challenge primarily attributed to the inherent noise present within raw datasets. Contrarily, frequency response sequences, characterized by their reduced noise levels compared to raw data, offer a more conducive environment for Transformer models to learn. This distinction underscores the potential of frequency domain representations to facilitate more effective pattern recognition by Transformer neural networks, suggesting that preprocessing data to emphasize frequency response may significantly improve learning outcomes. This approach aligns with the broader understanding that data quality and representation critically influence the performance of deep learning models, particularly in complex tasks where pattern discernment is fundamental. Welch power spectral density response of the signal is estimated with parameters shown in Table 4.3, to obtain frequency domain representation of the data.

Table 4.3: Welch estimation parameters

Parameter	**Value**
Signal	Raw data (90,000 data points)
Length of segment (1)	1000
Sampling frequency	20
Averaging method	Median

Features extracted from raw data are not very helpful to learn about damage classification. Shown in Figure 4.9, damage classification can be done with features extracted from stationary signal. Detrending helps in making the time series stationary:

- Removing trend and seasonal properties.
- Standardise the data
- Extract statistical and frequency response features from the raw signal

Data may exhibit sinusoidal trends or seasonal patterns. To address this, we can eliminate these trends by subtracting a sinusoidal fit from the signal, thereby rendering the data stationary. This process also effectively removes rotational motion from a blade in a structural context, allowing for the identification of patterns in the stationary vibration signal. The curve fit equation to be removed from signal can be assumed in the form as shown below where, A is amplitude, w is frequency of signal, t is time index, ϕ being phase difference, and C is a constant.

$$A * sin(wt + \phi) + C$$

Signal can be decomposed into a sinusoidal component and stationary component, stationarity of vibration is tested by Dickey Fuller test mentioned in Chapter 3.

These stationary signals frequency responses are standardised using standard scaler test mentioned in Chapter 3. The responses are labeled as binary labels if the data collected is from damaged blade or healthy blade. This response is trained on transformer with similar architecture as we used for corrupted sample detection. Similarly, statistical features extracted from stationary signal along with Standard deviation, Root mean square, Inter-quartile range, skewness we also consider normalised responses of rotational frequency and it's harmonics.

4.5. MODELING:

Based on kinds of data, we have to model for statistical data and sequential data. Simultaneously, we need to model for classifying corrupted samples and later for damage detection as show in Figure 4.13.

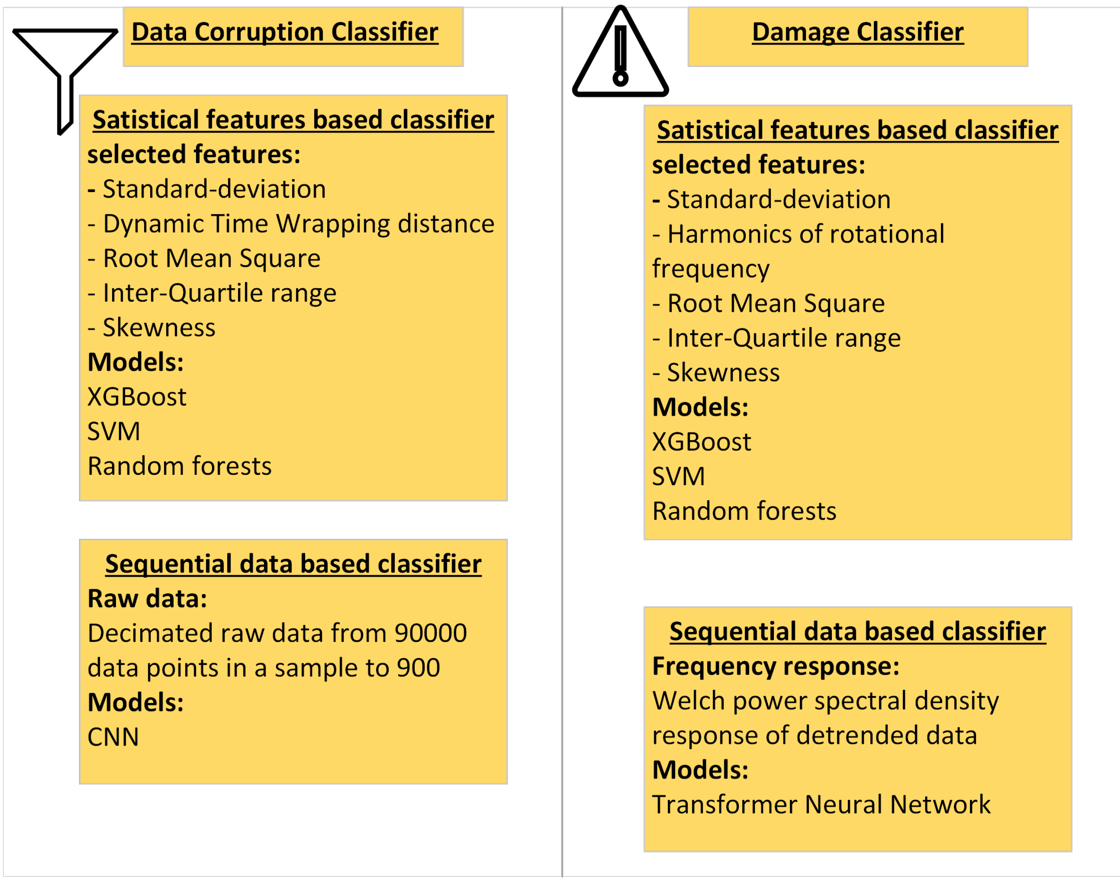


Figure 4.13: Models

4.5.1. STATISTICAL FEATURE BASED MODELS:

We used tabular classifier models to train both corruption and damage detection. For corruption classifier we chose, standard deviation, average of DTW distance, RMS, IQR and skewness but for damage classifier DTW is redundant as data is detrended. We included frequency response amplitudes of harmonic frequencies.

We applied Random Forest, SVM, and XGBoost classifier models to a normalized array of extracted statistical features. These baseline models are compared and later the XGBoost model underwent tuning via RandomSearchCV on grid of values Table 4.4. We then compared these models based on various parameters to identify the most effective method for classification.

Table 4.4: XGB optimisation parameters

Parameter	**Range**
max depth	3 to 18 at steps of 3
gamma	1 to 9 at steps of 1
reg alpha	40 to 180 at steps of 20
reg lambda	0 to 1 at steps of 0.2
colsample bytree	0.5 and 1
min child weight	0 to 10 at steps of 2

4.5.2. SEQUENTIAL DATA BASED MODELS:

As shown in Figure 4.13, sequential data is used to train CNN and transformer models. Data is under sampled ten times to reduce computational and time complexity for convolutional neural network in classifying corrupted data.

Architecture of a convolutional neural network (CNN) designed for processing time series data with an input shape of 900 time steps and a single feature per time step. The model's architecture comprises a series of convolutional layers, each followed by batch normalization and ReLU activation functions to introduce non-linearity and normalize the outputs for improved training stability. This pattern is repeated two more times, followed by a global average pooling layer condenses the feature maps from each of the 900 time steps into a single 64 feature vector by averaging, effectively reducing the data's dimensionality and preparing it for classification. The model concludes with a dense layer consisting of two neurons, intended for binary classification. Model optimized with Adam optimizer and loss function used is sparse categorical crossentropy. Model is evaluated by sparse categorical accuracy.

Architecture of a convolutional neural network (CNN) designed for processing time series data with an input shape of 900 time steps and a single feature per time step. The model's architecture comprises a series of convolutional layers, each followed by batch normalization and ReLU activation functions to introduce non-linearity and normalize the outputs for improved training stability. This pattern is repeated two more times, followed by a global average pooling layer condenses the feature maps from each of the 900 time steps into a single 64 feature vector by averaging, effectively reducing the data's dimensionality and preparing it for classification. The model concludes with a softmax dense layer consisting of two neurons, intended for binary classification.

We developed a Transformer-based neural network tailored for sequential data analysis, capable of processing frequency response sequences of 257 elements with a single feature dimension. Our model initiates with a dense layer to enhance the representational capacity of the input data, followed by the strategic incorporation of positional encoding to capture sequence order—crucial for the Transformer's self-attention mechanism. The architecture comprises three Transformer blocks, each featuring a multi-head attention module with four heads and a subsequent feed-forward network, integrating residual connections and layer normalization to facilitate efficient learning. The sequence information is condensed via global average pooling and then classified using a dense softmax layer for binary classification. CNN and transformer models optimised with the Adam optimizer and evaluated on "sparse categorical accuracy" through "sparse categorical crossentropy" loss.

4.6. EVALUATION:

4.6.1. PERFOMANCE:

Key metrics to evaluate classifier models, Confusion matrices, Accuracy, Precision, recall, F1 score and ROC. For, tabular data based models first the base-line model are compared using cross validated metrics, best among them is further fine-tuned using Random search algorithm. In case of Neural networks we used sparse categorical values to calculate and cross validate accuracy.

4.6.2. EXPLAINABILITY:

To demystify the rationale behind blackbox model predictions, particularly in complex scenarios, it is crucial to render the decision-making process of models transparent. In this context, the utilization of post-hoc explanation methodologies becomes indispensable. In our study, SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are employed to explain these results.

We used SHAP to explain tabular data models, by visulaising shap values using visulisation techniques provided in SHAP library. Bar plot to explain importance of a particular feature. Beeswarm plots, where feature values are compared to SHAP values which give us insights into how model prediction and feature values are related. The Neural network models are explained using LIME, by visualising top 10 features and how they are influencing the model predictions.

5

RESULTS

5.1. CORRUPTION CLASSIFIER:

5.1.1. PERFORMANCE:

This segment delves into the efficacy of models designed for tabular and sequential data in identifying corrupt samples within a dataset. We trained and assessed these models on manually tagged datasets, comprising 500 samples for tabular data and 1,000 samples for models based on neural networks. Subsequently, these models were applied to label the entire dataset, and the outcomes of this process are examined in the following discussion.

Table 5.1: Evaluation of models to detect corrupted data

Model	Accuracy	Precision	Recall	F1	ROC AUC
Random Forest classifier	0.90	0.91	0.93	0.90	0.90
SVM classifier	0.85	0.89	0.87	0.87	0.89
XGBoost classifier	0.92	0.90	0.96	0.93	0.93
XGBoost classifier (fine tuned)	0.94	0.91	0.97	0.94	0.93
CNN	0.96	0.97	0.92	0.94	0.94
Transformer	0.95	0.97	0.95	0.94	0.94

Table Table 5.2 presents evaluation of models designed for detecting corrupted data within a dataset, various models were assessed based on their performance metrics including Accuracy, Precision, Recall, F1 Score, and ROC AUC. The Random Forest classifier showed commendable results with an accuracy of 90%, precision of 91%, recall of 93%, an F1 score of 90%, and an ROC AUC of 90%. Following this, the SVM (Support Vector Machine) classifier demonstrated an accuracy of 85%, precision of 89%, recall of 87%, an F1 score of 87%, and an ROC AUC of 89%.

The XGBoost classifier initially reported an accuracy of 92%, precision of 90%, recall of 96%, an F1 score of 93%, and an ROC AUC of 93%. Further improvement was noted in the XGBoost classifier upon fine-tuning, achieving an accuracy of 94%, precision of 91%, recall of 97%, an F1 score of 94%, and maintaining an ROC AUC of 93%.
The Neural Network models outperformed others with better metrics. These results highlight the efficacy of advanced models, particularly CNN and Transformer models, in accurately identifying corrupted samples in dataset.

5.1.2. EXPLAINABILITY:

To interpret the prediction of data corruption by machine learning models, we utilized the XAI frameworks SHAP and LIME to determine the influential features in classifying samples. Figure 5.1 highlights the primary contributors—IQR, DTW, and the rotational frequency of wind turbines—to the determination of sample data corruption. IQR identifies samples with extreme values, DTW distance measures the similarity across three channels, and rotational frequency is a significant factor as the data encompasses various IQR and RMS values per rotational frequency.

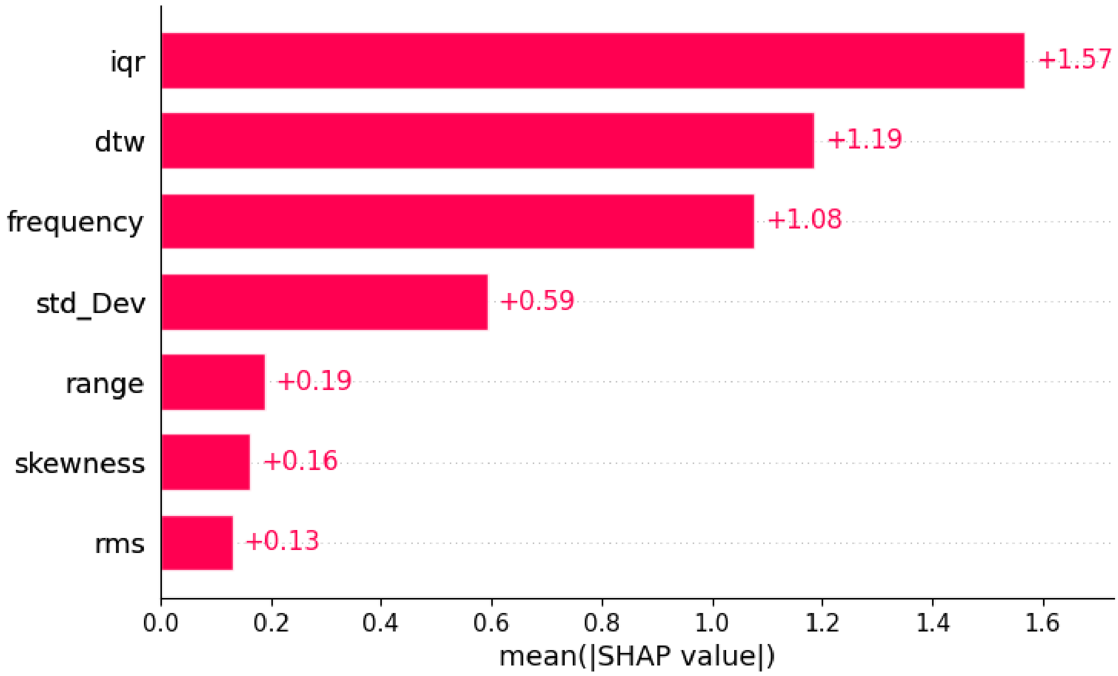


Figure 5.1: SHAP values bar plot of XGBoost model

In the beeswarm plot displayed in Figure 5.2, which illustrates how feature values are influencing SHAP values. Samples with low IQR, though spread across the x-axis, are denser and more likely to predict uncorrupted data, whereas samples with high IQR are densely clustered on the corruption side. For DTW, high values lean towards predicting uncorrupted data, while low DTW values are densely associated with corruption. Regarding frequency, instances are distributed throughout the x-axis, a pattern that holds true for the other features as well.

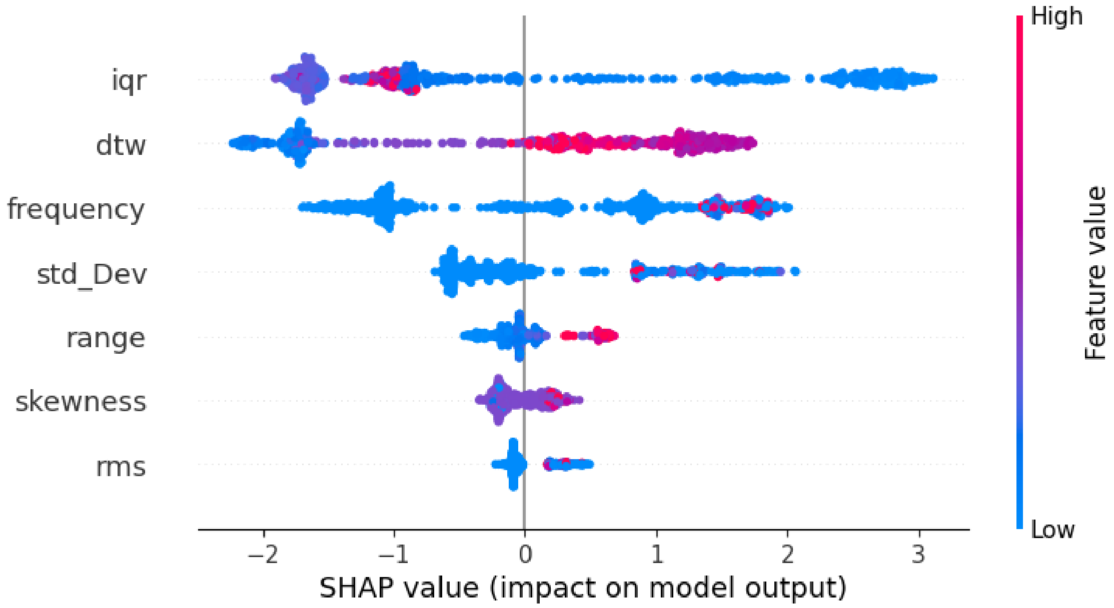


Figure 5.2: SHAP values, beeswarm plot of XGBoost model

The LIME framework is used to analyze decimated time series data trained on neural networks, as illustrated in the plots for uncorrupted data Figure 5.3 and corrupted samples Figure 5.4 shown below. These plots reveal the top 10 time-related features that influenced the classification decisions for the respective instances.

5.2. DAMAGE CLASSIFIER:

5.2.1. PERFORMANCE

Unlike the corruption classification for damage detection, the tabular data doesn't help us much with classification of damage detection. Similarly, time domain data don't seem to learn any patterns to detect damage from data even for lower decimation factors. But the frequency domain representations of the data show good results in damage detection.

Table 5.2: Evaluation of models to detect damage data

Model	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC AUC**
Random Forest classifier	0.70	0.72	0.70	0.75	0.68
SVM classifier	0.70	0.71	0.68	0.74	0.67
XGBoost classifier	0.74	0.72	0.68	0.73	0.69
XGBoost classifier (fine tuned)	0.75	0.72	0.72	0.73	0.71
Transformer	0.98	0.97	0.98	0.99	0.96

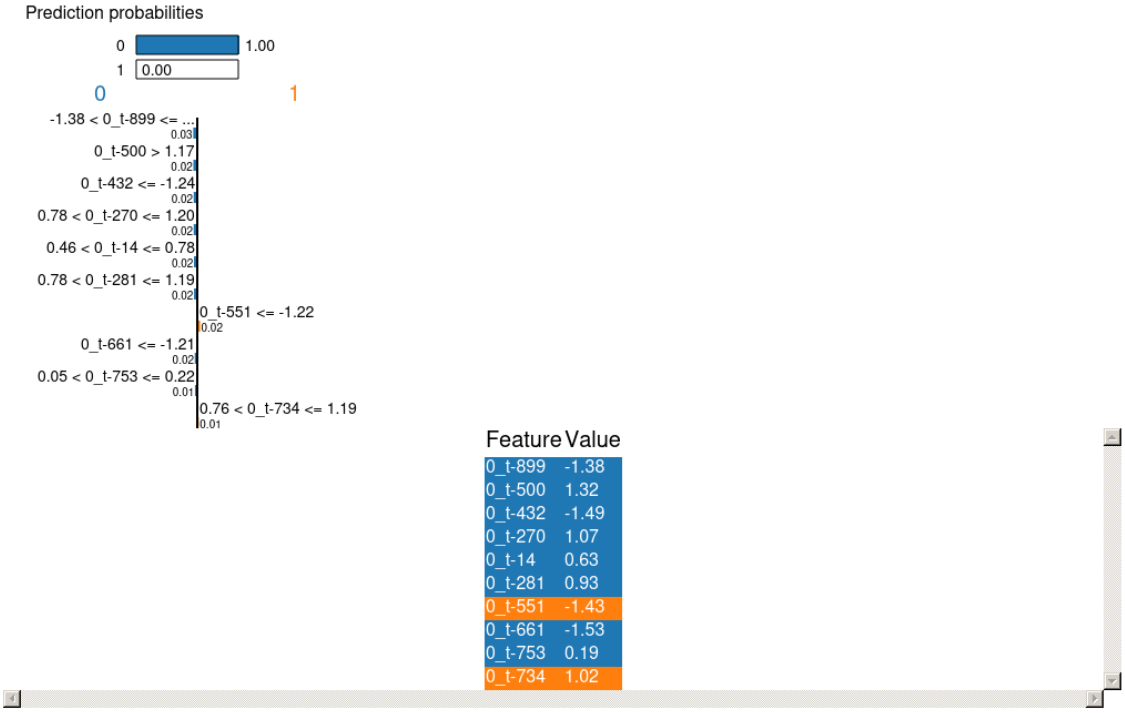


Figure 5.3: LIME feature values of top 10 features of corrupted sample

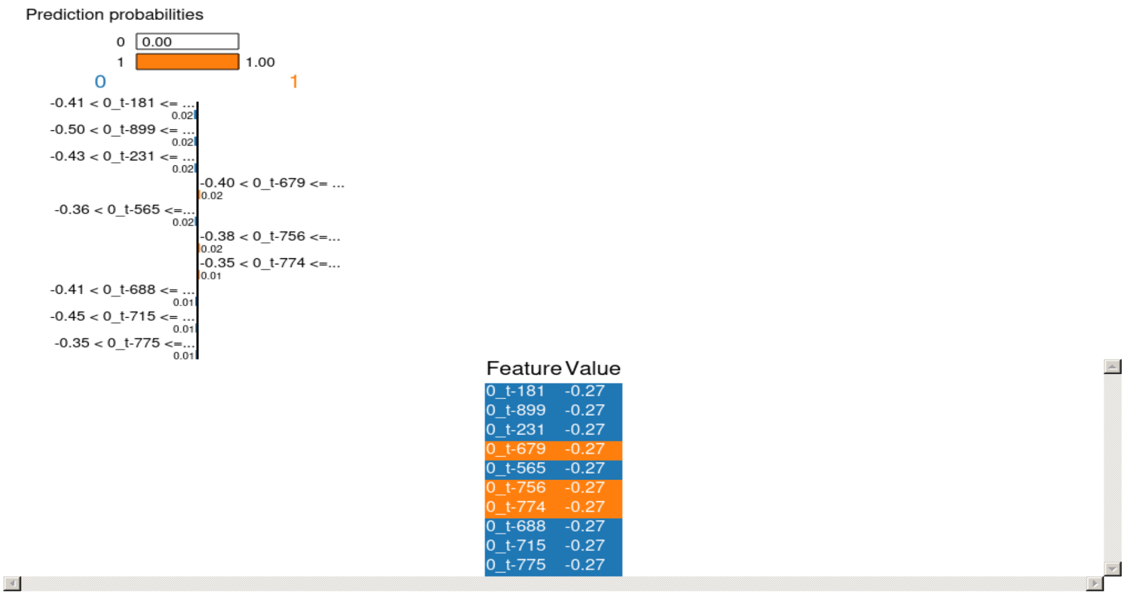


Figure 5.4: LIME feature values of top 10 features of corrupted sample

5.2.2. TESTING MODELS ON UNTRAINED DATA:

Comparing the models Figure 5.5 by extending the anomaly detection on 10,000 samples, which were not part of training process. Channel 2 sensors show high corrupted samples. Tabular data based XGBoost model detects bit more corrupted samples than models trained on time and frequency domain raw data.

Deploying damage detection model shown in Figure 5.6. Frequency response data of 2,000 samples, which were not part of training process are considered. As expected channel 3(damaged blade) sensors are classified as damage samples in significantly high numbers.

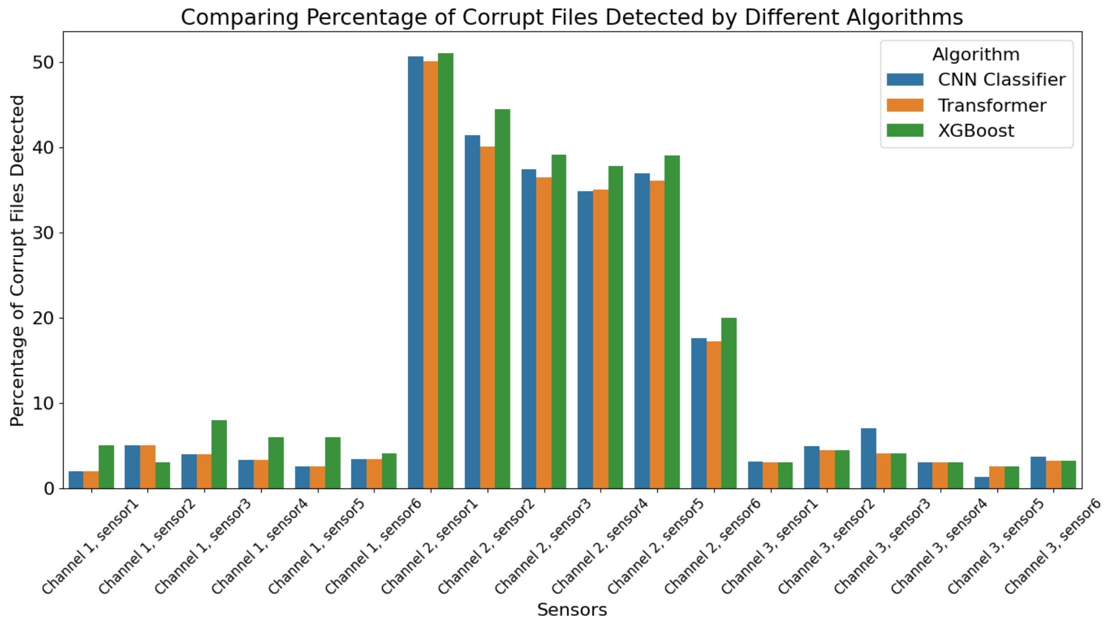


Figure 5.5: Comparing percentage of corrupt files detected through different models

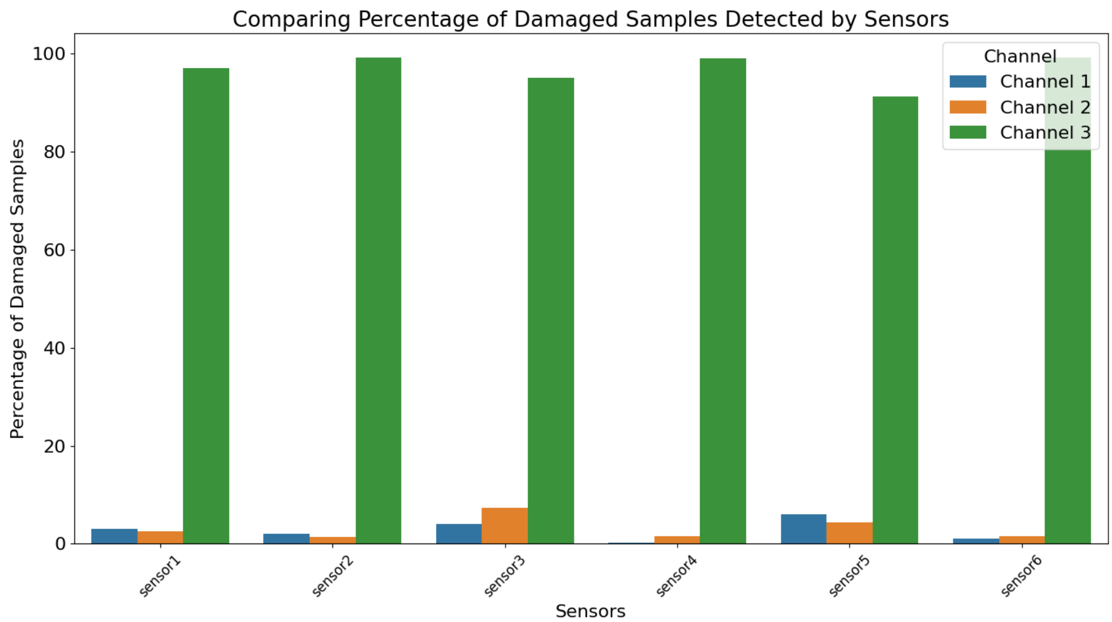


Figure 5.6: Comparing percentage of damaged samples detected from frequency domain responses

6
CONCLUSION

This thesis has explored the application of explainable artificial intelligence (XAI) frameworks, particularly SHAP and LIME, in structural health monitoring (SHM) systems using vibration data to predict damage and ensure the integrity of engineering structures. The comprehensive literature review, experimental setups, and methodological applications discussed in the preceding chapters underscore the emerging capabilities of machine learning models in advancing the field of SHM.

Throughout this study, several key findings have highlighted the transformative impact of XAI in SHM systems. The integration of XAI has significantly enhanced the transparency and interpretability of predictive models, which is crucial for building trust among engineers and stakeholders who rely on these systems for critical decision-making. The use of sophisticated machine learning algorithms like XGBoost and convolutional neural networks has also demonstrated superior performance in detecting and classifying damage, benefiting from the robust feature extraction and selection capabilities that XAI methods facilitate. Moreover, the practical application of these systems in monitoring large-scale structures, such as wind turbines, offers a promising framework for future deployments. The insights provided by XAI into model predictions are invaluable, helping to fine-tune systems for better accuracy and reliability.

6.1. PRACTICAL IMPLICATIONS OF DATA CORRUPTION IN SHM SYSTEMS:

Data corruption represents a significant challenge in the effective implementation of SHM systems, particularly when these systems rely on accurate and reliable data to make critical decisions regarding the health and safety of infrastructure. Corruption in data can occur due to a variety of factors, including sensor faults, transmission errors, environmental interference, and software bugs, among others. The implications of such corruption are profound, particularly in fields requiring high reliability and precision.

Corruption in data can lead to incorrect assessments of structure health, potentially resulting in two primary adverse outcomes: false positives and false negatives. False positives, where the system incorrectly identifies damage that does not exist, can lead to unnecessary inspections and maintenance, thus wasting resources and increasing operational costs. More critically, false negatives, where actual damage goes undetected, can pose serious risks to the safety and integrity of structures, potentially leading to catastrophic failures.
To mitigate the effects of data corruption, SHM systems must incorporate robust data validation and cleaning mechanisms. Implementing anomaly detection algorithms that identify and isolate corrupted data before it impacts the system is crucial. Moreover, the development of adaptive models that can handle some level of data imperfection without significant performance degradation is also vital. Techniques such as data imputation, robust outlier detection, and advanced noise filtering can enhance the resilience of SHM systems against corrupted data.

6.1.1. ROLE OF XAI IN ADDRESSING DATA CORRUPTION:

Explainable AI (XAI) can play a crucial role in addressing data corruption by making the decision-making processes of SHM systems transparent. By using XAI frameworks like SHAP and LIME, engineers can understand how the model processes data, including how it deals with anomalies and potential data corruption. This transparency allows for more informed adjustments to the model and the data processing pipeline, enhancing the system's ability to accurately assess structural health despite the presence of corrupted data.

In conclusion, addressing data corruption is fundamental to the deployment of reliable SHM systems. By implementing robust mechanisms for data validation, maintaining sensor integrity, and employing XAI, the practical implications of data corruption can be significantly mitigated, thereby ensuring the safety, efficiency, and reliability of critical infrastructure monitoring.

6.2. ANSWERS TO RESEARCH QUESTIONS:

How can we implement explainable structural health monitoring (SHM) for damage classification based on vibration data?

The research demonstrated that implementing explainable SHM for damage classification using vibration data is feasible and highly effective with the integration of XAI methods like SHAP and LIME. These tools enhanced the interpretability of complex machine learning models, such as XGBoost, CNNs and transformers, allowing for greater transparency in damage detection and classification processes. The application of these XAI frameworks provided valuable insights into the feature contributions and decision-making processes of the models, ensuring that the SHM systems are not only accurate but also trustworthy.

What are the current ML models and techniques used in the SHM methods to identify defects *based on accelerometer data?*

The study identified several machine learning models and techniques prominently used in SHM, including Random Forest, Support Vector Machines, XGBoost, Convolutional Neural Networks and transformers. These models proved capable of effectively identifying defects in structures from accelerometer data, especially when enhanced by feature extraction techniques that improve data interpretability and model accuracy.

What ML methods are suitable for damage localization?

The study faced challenges in effectively localizing structural damage due to an insufficient number of sensors. This limitation significantly hampered the precision and accuracy of the damage localization process, a critical component in structural health monitoring (SHM) systems. The lack of extensive sensor coverage meant that data collected was not comprehensive enough to accurately pinpoint the specific areas of damage, leading to potential gaps in the monitoring process. This shortfall underscores the importance of deploying a well-distributed sensor network to ensure that SHM systems can perform their intended functions effectively, particularly in large-scale infrastructure projects where the identification of exact damage locations is crucial for timely and cost-effective maintenance and repairs.

How can we integrate XAI frameworks to explain SHM models?

successfully achieved, providing clear explanations of model predictions. These explanations were instrumental in validating the models' reliability and in facilitating their acceptance among engineers and decision-makers. The use of XAI allowed for a deeper understanding of the underlying dynamics of the models, which is crucial for critical applications where safety and precision are paramount.

These conclusions align with the specific research questions posed at the beginning of the study and provide targeted insights based on the comprehensive analysis conducted throughout the thesis. This structured approach to concluding ensures that each research question is addressed with clear evidence and outcomes from the research.

6.3. FUTURE IMPLICATIONS:

The study has provided valuable insights into the application of XAI in SHM systems and demonstrated the potential of advanced machine learning models. However, the challenge of accurately localizing damage due to insufficient sensor coverage has highlighted a critical area for improvement in SHM systems. Addressing this gap is essential for enhancing the functionality and reliability of future deployments, particularly in monitoring large-scale structures like wind turbines.

6.3.1. ENHANCING SENSOR COVERAGE

Future research should focus on optimizing sensor deployment strategies to ensure comprehensive coverage and data collection capabilities. Developing guidelines for sensor placement that maximize the area covered while minimizing costs and logistical complexities could significantly improve damage localization accuracy.

Incorporating advanced sensing technologies, such as distributed fiber optic sensors or wireless sensor networks, could provide a higher resolution of data across structures. These technologies offer the potential for real-time monitoring and a greater density of data points, which could enhance the SHM system's ability to detect and localize structural damages more precisely.

6.3.2. DEVELOPING ADAPTIVE SHM SYSTEMS

There is a need to develop adaptive SHM systems that can adjust to varying data availability and quality. Such systems could use machine learning algorithms to optimize their performance based on the data input they receive, which would be particularly useful in scenarios with limited sensor data.

Further exploration into the integration of XAI methods to enhance the interpretability of SHM systems remains a priority. Future developments should aim to make these systems more transparent and understandable to a broader range of stakeholders, including those without deep technical knowledge of machine learning, facilitating better decision-making.

6.3.3. COLLABORATIVE AND INTERDISCIPLINARY APPROACHES

Adopting a collaborative and interdisciplinary approach that includes engineers, data scientists, and industry professionals could lead to more innovative solutions to the challenges faced in SHM. Collaborations can also help in designing SHM systems that are not only technically proficient but also economically viable and easy to implement on a larger scale.

In conclusion, while the study has made significant contributions to the field of SHM through the use of XAI and machine learning, the challenges encountered with sensor coverage and damage localization highlight the need for continued research and development. By addressing these challenges, future SHM systems can achieve higher accuracy, reliability, and practicality in real-world applications.

References

[1] D. A. Tibaduiza Burgos, R. C. Gomez Vargas, C. Pedraza, D. Agis, F. Pozo. Damage identification in structural health monitoring: A brief review from its implementation to the use of data-driven applications. Sensors 20 (2020). doi:10.3390/s20030733.

[2] S. Huber, H. Wiemer, D. Schneider, S. Ihlenfeldt. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP 79 (2019) 403–408. doi:10.1016/j.procir.2019.02.106.

[3] C. Shearer. The crisp-dm model: the new blueprint for data mining. Journal of data warehousing 5 (2000) 13–22.

[4] M. Gordan, H. A. Razak, Z. Ismail, K. Ghaedi, Z. X. Tan, H. H. Ghayeb. A hybrid ann-based imperial competitive algorithm methodology for structural damage identification of slab-on-girder bridge using data mining. Applied Soft Computing 88 (2020) 106013. doi:https://doi.org/10.1016/j.asoc.2019.106013.

[5] M. Dimitrova, A. Aminzadeh, M. S. Meiabadi, S. Sattarpanah Karganroudi, H. Taheri, H. Ibrahim. A Survey on Non-Destructive Smart Inspection of Wind Turbine Blades Based on Industry 4.0 Strategy. Appl. Mech. 3 (2022) 1299–1326. doi:10.3390/applmech3040075.

[6] R. Katam, P. Kalapatapu, V. D. K. Pasupuleti, in: P. Rizzo, A. Milazzo (Eds.), European Workshop on Structural Health Monitoring, Springer International Publishing, Cham, 2023, pp. 371–380.

[7] S. C. Mukhopadhyay, I. Ihara, Sensors and Technologies for Structural Health Monitoring: A Review, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1–14. doi:10.1007/978-3-642-21099-0_1.

[8] P. Pandey, A. Rai, M. Mitra. Explainable 1-d convolutional neural network for damage detection using lamb wave. Mechanical Systems and Signal Processing 164 (2022). doi:10.1016/j.ymssp.2021.108220.

[9] F. Angeletti, P. Iannelli, P. Gasbarri, M. Panella, A. Rosato. A study on structural health monitoring of a large space antenna via distributed sensors and deep learning. SENSORS 23 (2023).

[10] M. Moradi, P. Komninos, R. Benedictus, D. Zarouchas, in: Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM, volume 14.

[11] Y. W. Ou, E. N. Chatzi, V. K. Dertimanis, M. D. Spiridonakos. Vibration-based experimental damage detection of a small-scale wind turbine blade. STRUCTURAL HEALTH MONITORING-AN INTERNATIONAL JOURNAL 16 (2017) 79–96. doi:10.1177/1475921716663876.

[12] S. Carone, G. Pappalettera, C. Casavola, S. De Carolis, L. Soria. A support vector machine-based approach for bolt loosening monitoring in industrial customized vehicles. SENSORS 23 (2023). doi:10.3390/s23115345.

[13] W. Y. Hsiung, Y. T. Huang, C. H. Loh, K. J. Loh, R. J. Kamisky, D. Nip, C. van Dam, Analyzing the dynamic response of rotating blades in small-scale wind turbines, 2014. doi:10.1117/12.2044351.

[14] D. Goyal, A. Choudhary, B. S. Pabla, S. S. Dhami. Support vector machines based non-contact fault diagnosis system for bearings. Journal of Intelligent Manufacturing 31 (2020) 1275–1289. doi:10.1007/s10845-019-01511-x.

[15] M. Parziale, L. Lomazzi, M. Giglio, F. Cadini, in: Lecture Notes in Civil Engineering, volume 433 LNCE, pp. 540–549. doi:10.1007/978-3-031-39117-0_55.

[16] J. Baquerizo, C. Tutivén, B. Puruncajas, Y. Vidal, J. Sampietro. Siamese neural networks for damage detection and diagnosis of jacket-type offshore wind turbine platforms. Mathematics 10 (2022). doi:10.3390/math10071131.

[17] J. X. Leon-Medina, M. Anaya, N. Parés, D. A. Tibaduiza, F. Pozo. Structural damage classification in a jacket-type wind-turbine foundation using principal component analysis and extreme gradient boosting. Sensors 21 (2021). doi:10.3390/s21082748.

[18] J. X. Leon-Medina, N. Parés, M. Anaya, D. A. Tibaduiza, F. Pozo. Ensemble of feature extraction methods to improve the structural damage classification in a wind turbine foundation. Bulletin of the Polish Academy of Sciences: Technical Sciences 71 (2023).

[19] D. García, D. Tcherniak. An experimental study on the data-driven structural health monitoring of large wind turbine blades using a single accelerometer and actuator. Mechanical Systems and Signal Processing 127 (2019) 102–119. doi:10.1016/j.ymssp.2019.02.062.

[20] Y. Vidal, G. Aquino, F. Pozo, J. E. M. Gutiérrez-Arias. Structural health monitoring for jacket-type offshore wind turbines: Experimental proof of concept. Sensors (Switzerland) 20 (2020). doi:10.3390/s20071835.

[21] C. Tutivén, H. Triviño, Y. Vidal, J. Sampietro, Structural health monitoring of offshore jacket platforms via transformers, 2023. doi:10.1007/978-3-031-07254-3_17.

[22] B. Pavlack, J. Paixao, S. da Silva, A. Cunha, D. G. Cava. Polynomial chaos-kriging metamodel for quantification of the debonding area in large wind turbine blades. STRUCTURAL HEALTH MONITORING-AN INTERNATIONAL JOURNAL 21 (2022) 666–682. doi:10.1177/14759217211007956.

[23] F. D. N. Santos, N. Noppe, W. Weijtjens, C. Devriendt, in: International Conference on Structural Health Monitoring of Intelligent Infrastructure: Transferring Research into Practice, SHMII, volume 2021-June, pp. 1261–1265.

[24] J. Xu, X. Ding, Y. Gong, N. Wu, H. Yan. Rotor imbalance detection and quantification in wind turbines via vibration analysis. Wind Engineering 46 (2022) 3–11. doi:10.1177/0309524X21999841.

[25] J. Peng, A. Kimmig, Z. Niu, J. Wang, X. Liu, D. Wang, J. Ovtcharova. Wind turbine failure prediction and health assessment based on adaptive maximum mean discrepancy. Int. J. Electr. Power Energy Syst. 134 (2022) 107391. doi:10.1016/j.ijepes.2021.107391.

[26] A. Movsessian, D. G. Cava, D. Tcherniak. Interpretable machine learning in damage detection using Shapley Additive Explanations. Engineering Archive (2021). doi:10.31224/osf.io/96yf5.

[27] E. J. Cross, S. J. Gibson, M. R. Jones, D. J. Pitchforth, S. Zhang, T. J. Rogers, Physics-Informed Machine Learning for Structural Health Monitoring, Springer International Publishing, Cham, 2022, pp. 347–367. doi:10.1007/978-3-030-81716-9_17.

[28] S. D. Fassois, F. P. Kopsaftopoulos, in: New trends in structural health monitoring, Springer, 2013, pp. 209–264.

[29] L. Lomazzi, S. Fabiano, M. Parziale, M. Giglio, F. Cadini. On the explainability of convolutional neural networks processing ultrasonic guided waves for damage diagnosis. Mechanical Systems and Signal Processing 183 (2023). doi:10.1016/j.ymssp.2022.109642.

[30] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, K.-R. Müller, Layer-Wise Relevance Propagation: An Overview, 2019. doi:10.1007/978-3-030-28954-6_10.

[31] J. J. P. Abadía, H. Fritz, G. Dadoulis, K. Dragos, K. Smarsly, in: EG-ICE 2021 Workshop on Intelligent Computing in Engineering, Proceedings, pp. 432–441.

[32] D. Katsaprakakis, N. Papadakis, I. Ntintakis. A Comprehensive Analysis of Wind Turbine Blade Damage. Energies 14 (2021) 5974. doi:10.3390/en14185974.

[33] S.-L. Lee, S. Shin. Structural design optimization of a wind turbine blade using the genetic algorithm. Eng. Optim. 54 (2021) 1–18. doi:10.1080/0305215X.2021.1973450.

[34] D. Luckey, H. Fritz, D. Legatiuk, J. J. Peralta Abadía, C. Walther, K. Smarsly, Explainable Artificial Intelligence to Advance Structural Health Monitoring, volume 21, 2022, pp. 331–346. doi:10.1007/978-3-030-81716-9_16.

[35] M. Mudabbiruddin, A. Mosavi, in: SACI 2023 - IEEE 17th International Symposium on Applied Computational Intelligence and Informatics, Proceedings, pp. 405–414. doi:10.1109/SACI58269.2023.10158652.

APPENDICES

Table 1: Techniques and motivations observed in selected papaers

Author	Year	Journal	Settings	Main Purpose	Data-Driven Technique[1]
Ou, YW; Chatzi, EN; Dertimanis, VK; Spiridonakos, MD [11]	2017	STRUCTURAL HEALTH MONITORING-AN INTERNATIONAL JOURNAL	Experimental setup of wind turbine blade.	Damage detection in wind turbines blades. Data is collected from healthy and induced damages.	Statistical and Modal features are compared with healthy and various induced damages.
Angeletti, F; Iannelli, P;Gasbarri, P; Panella, M; Rosato, A [9]	2023	SENSORS	-	Develop a deep learning based damage detection system, damage type detection and localisation	Neural networks LSTM. bi-directional LSTM
Carone, S; Pappalettera, G; Casavola, C; De Carolis, S; Soria, L [12]	2023	SENSORS	Experimental prototype	vibration-based detection of bolt loosening in a rotating joint of a custom sewer cleaning vehicle transmission was performed using support vector machines (SVM)	Statistical features of vibration signal classified using SVM
Hsiung, WY; Huang, YT; Loh, CH; Loh, KJ; Kamisky, RJ; Nip, D; van Dam, C [13]	2014	SENSORS AND SMART STRUCTURES TECHNOLOGIES FOR CIVIL, MECHANICAL, AND AEROSPACE SYSTEMS 2014	Experimental setup of prototype and induced damage	Identify changing modal features with the induced damage	Modal feature extraction
Movsessian, Artur and Cava, David Garcia and Tcherniak, Dmitri[26]	2021	Engineering Archive	Simulated data	The study involves using DSF to calculate a Mahalanobis Distance-based Damage Index (MD-based DI), training an XGBoost regression model with these DSFs and environmental variabilities as predictors, and employing SHAP analysis to interpret the model's outputs, specifically the impact of each DSF and environmental variable on the DI.	XGBoost regression, SHAP
Goyal, D.; Choudhary, A.; Pabla, B.S.; Dhami, S.S. [14]	2020	Journal of Intelligent Manufacturing	Experimental prototype	Comparing classification results of non-contact sensor and accelerometer.	extract statistical features from vibration singal;SVM, KNN classifiers
AbadÃa, J.J.P.; Fritz, H.;Dadoulis, G.; Dragos, K.; Smarsly, K.[31]	2021	EG-ICE 2021 Workshop on Intelligent Computing in Engineering, Proceedings	Review Literature	XAI in SHM, comparing damaged and healthy states	SHAP
Pandey, P.; Rai, A.; Mitra, M.[8]	2022	Mechanical Systems and Signal Processing	FE model simulations	XAI in SHM to explain classification of damaged and healthy signal	CNN, LIME, SHAP
Parziale, M.; Lomazzi, L.; Giglio, M.; Cadini, F.[15]	2023	Lecture Notes in Civil Engineering	Experimental prototype.	Transmissibility Function based engineered features are used to generate matrix which are used as gray-scale image data to train CNN classifier for various damaged scenarios.	CNN; LRP used to explain results

[1]Note: KNN refers to K-Nearest Neighbour, SVM is Support Vector Machine, CNN is Convolution Neural Networks, GB is Gradient Boosting, SVDD is Support Vector Domain Description, LSTM is Long Short-Term Memory

Lomazzi, L.; Fabiano, S.; Parziale, M.; Giglio, M.; Cadini, F.[29]	2023	Mechanical Systems and Signal Processing	Lamb wave, acutators	GSI images generated from Lamb waves which are trained on CNN to classify	LRP used to explain results
Luckey, D.; Fritz, H.; Legatiuk, D.; Peralta AbadÃa, J.J.; Walther, C.; Smarsly, K.[34]	2022	Structural Integrity	Review paper	Framework for implementing XAI in SHM done using ML algorithms	Conceptual XAI framework
Moradi, M.; Komninos, P.; Benedictus, R.; Zarouchas, D.[10]	2022	Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM	Simulated data	Formulate a HI(Health Index) based on PCA derived features and classify HI using neural networks with custom weights, additive and multiplicative weights to make model predictive.	PCA, neural networks
Jin Xu and Xian Ding and Yongli Gong and Ning Wu and Huihuang Yan [24]	2022	Wind Engineering	Real-time monitoring	Demodulated raw data using CWT, complex wavelet transforms. PSD of these transformed signals are used to formulate a health indicator based on amplitudes of harmonics of rotational frequency. Compared this health indiacators on different scenario.	Complex wavelet transform(CWT)
Mudabbiruddin, M.; Mosavi, A. [35]	2023	SACI 2023 - IEEE 17th International Symposium on Applied Computational Intelligence and Informatics, Proceedings	Review paper	Frame work for modeling structural aging process.	neural networks
Baquerizo, J.; TutivÃ©n, C.; Puruncajas, B.; Vidal, Y.; Sampietro, J.[16]	2022	Mathematics	Experimental WT jack	256 values from each column were reshaped into 16 × 16 matrices to train and test classification neural network model. Various classes of induced damage data is generated.	convolutional neural networks
Leon-Medina, J.X.; Anaya, M.; ParÃ©s, N.; Tibaduiza, D.A.; Pozo, F. [17]	2021	Sensors	Experimental WT jack	data from sensor network is reduced in dimensions by PCA, this data used to train, test and validate an XGBoost multi class model.	XGBoost, PCA
GarcÃa, D.; Tcherniak, D. [19]	2019	Mechanical Systems and Signal Processing	Experimental setup of WT blade	Data driven vibration SHM, SSA is used to decompose raw signal and construct feature vectors for statistical analysis with mahanabolis distance as criterion.	Singular Spectrum Analysis (SSA)
Vidal, Y.; Aquino, G.; Pozo, F.; GutiÃ©rrez-Arias, J.E.M. [20]	2020	Sensors (Switzerland)	Experimental setup of WT jack	Signals decomposed using PCA and trained to classify damaged and healthy states.	KNN, SVM
Santos, F.D.N.; Noppe, N.; Weijtjens, W.; Devriendt, C. [23]	2021	International Conference on Structural Health Monitoring of Intelligent Infrastructure: Transferring Research into Practice, SHMII	Real time monitoring	Fatigue assessment, using acceleration and SCADA data. Features selection using Wrapper type feature selection and trained on NN classifier.	neural Networks(NN)

Leon-Medina, J.X.; ParÃ©s, N.; Anaya, M.; Tibaduiza, D.A.; Pozo, F. [18]	2023	Bulletin of the Polish Academy of Sciences: Technical Sciences	Experimental setup of WT jack	Classifying healthy and damaged scenarios.	XGBoost, PCA
TutivÃ©n, C; TriviÃ±o, H; Vidal, Y; Sampietro, J [21]	2023	EUROPEAN WORKSHOP ON STRUCTURAL HEALTH MONITORING (EWSHM 2022), VOL 1	Experimental setup of WT jack	Transformer network to train damage detection classifier.	Transformer network
Peng, Jieyang and Kimmig, Andreas and Niu, Zhibin and Wang, Jiahai and Liu, Xiufeng and Wang, Dongkun and Ovtcharova, Jivka [25]	2022	International Journal of Electrical Power & Energy Systems	Real-time monitoring	The study involves collecting SCADA data from wind turbines, preprocessing it for failure classification, and using an algorithm to compare sensor data with standard health benchmarks. This approach helps in real-time health assessment of turbines, employing a CNN classifier for accurate failure prediction and enabling effective maintenance decisions.	CNN classfier, Max-Mean discrepancy(MMD)
Pavlack, B; Paixao, J; da Silva, S; Cunha, A; Cava, DG [22]	2022	STRUCTURAL HEALTH MONITORING-AN INTERNATIONAL JOURNAL	Experimental setup	Damage detection and quantification of Damage Index using PC-Kriging.	PC-Kriging