AI Based Damage Detection and Localization in Steel Beams Using Hybrid Time–Frequency Domain Features and Transfer Learning

Supervisor: Professor Gian Paolo Cimellaro
Student: Farzad Naseri

2026

Contents

Figures

Abstract

Structural damage detection in steel beams is a key challenge in Structural Health Monitoring (SHM), aiming to improve structural safety, reliability, and maintenance planning. This study presents a vibration-based damage assessment framework that integrates finite element (FE) modeling, feature extraction, machine learning, and transfer learning for damage detection, localization, and severity estimation in steel beams.

A high-fidelity three-dimensional finite element model of an IPE160 steel beam was developed in ANSYS and validated against experimental measurements through comparison of natural frequencies and mode shapes. The validated model was subsequently used to generate a dataset comprising 259 damage scenarios with varying crack locations and severities. From the simulated vibration responses, both frequency-domain features (natural frequencies and mode shapes) and time-domain features (root mean square, peak value, and crest factor) were extracted and used as inputs for machine learning models.

A comprehensive set of machine learning algorithms representing different learning paradigms was evaluated, including tree-based, kernel-based, linear, probabilistic, distance-based, and neural-network-based methods. The results demonstrated that ensemble tree-based algorithms, particularly Extra Trees, Random Forest, and Gradient Boosting, consistently provided the highest accuracy and robustness across damage detection, localization, and severity estimation tasks.

To reduce the discrepancy between numerical simulations and experimental measurements, a single-shot domain adaptation strategy was implemented. Due to the limited availability of experimental data, only one experimentally measured damage scenario was used for model adaptation. For damage detection, the prediction confidence increased from approximately 0.92–0.93 to nearly 1.00 after adaptation. For damage localization, the classification accuracy improved from 0.462 to 0.921, while the mean absolute error (MAE) decreased from 210.8 mm to 45.2 mm. For damage severity estimation, the MAE decreased from 30.47 mm to 0.24 mm. At the same time, the performance degradation on the finite element dataset remained limited.

These results demonstrate that machine learning models trained on finite element simulation data can be successfully adapted to a specific real-world damage scenario using only a single experimental measurement while preserving the knowledge acquired from the numerical domain. Furthermore, the proposed framework provides a foundation for incremental model updating, enabling additional experimental measurements to be incorporated as they become available, thereby progressively improving model performance while retaining previously learned structural knowledge. Overall, the findings highlight the potential of transfer learning as a practical pathway for bridging the simulation-to-experiment gap and facilitating the deployment of data-driven SHM systems in situations where experimental data are scarce.

Keywords: Structural Health Monitoring (SHM), Damage Detection, Hybrid Time and Frequency Domain Features, Machine Learning, Transfer Learning,

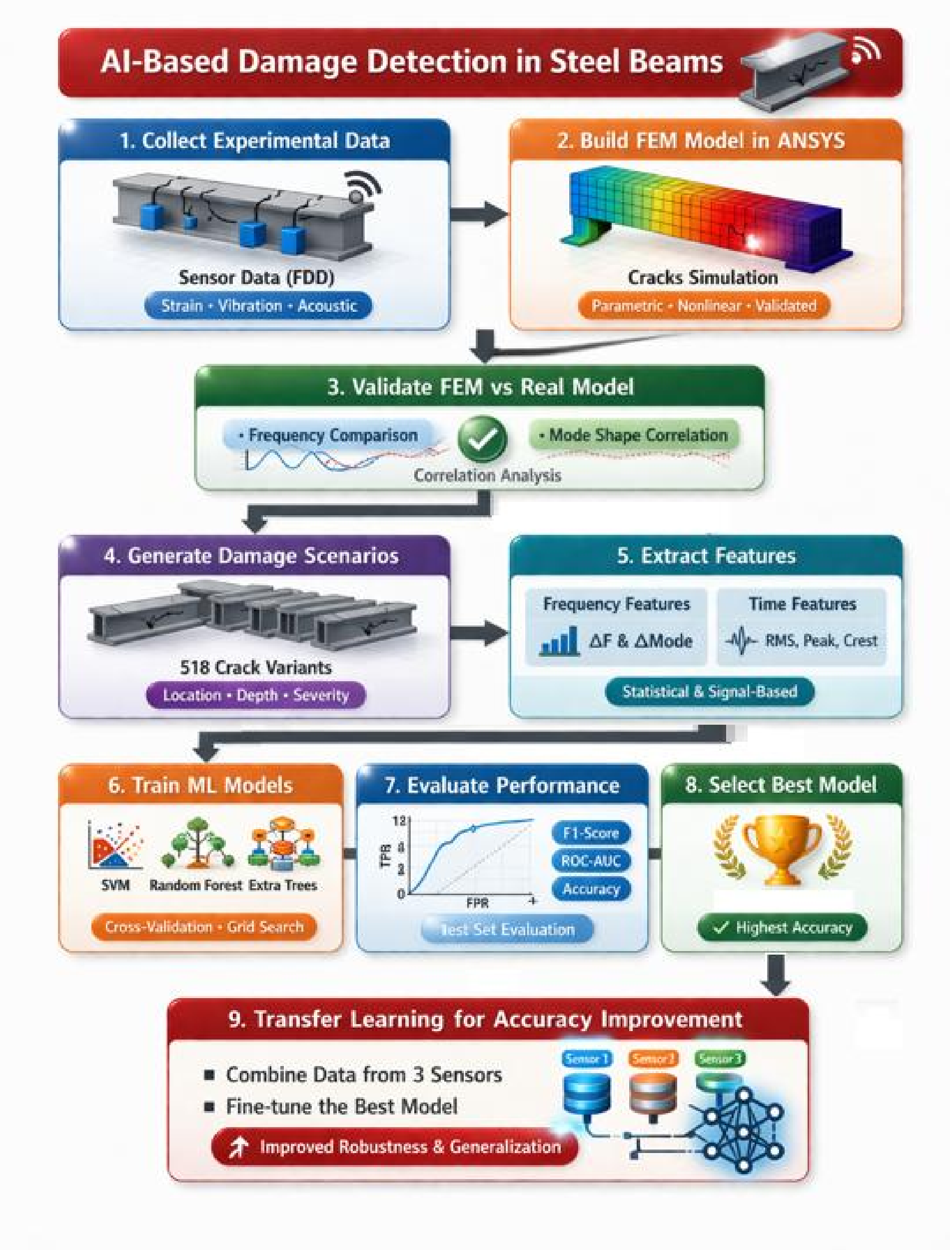

Figure 1. Graphical Workflow

List of Abbreviations

Abbreviation	Definition
SHM	Structural Health Monitoring
FE	Finite Element
FEM	Finite Element Method
AI	Artificial Intelligence
ML	Machine Learning
TL	Transfer Learning
SS-TL	Single-Shot Transfer Learning
DD	Damage Detection
DL	Damage Localization
DS	Damage Severity
RF	Random Forest
ET	Extra Trees
GB	Gradient Boosting
SVM	Support Vector Machine
SVR	Support Vector Regression
KNN	k-Nearest Neighbors
MLP	Multi-Layer Perceptron
ANN	Artificial Neural Network
GPR	Gaussian Process Regression
FDD	Frequency Domain Decomposition
FFT	Fast Fourier Transform
PSD	Power Spectral Density
CSD	Cross Spectral Density
FRF	Frequency Response Function
RMS	Root Mean Square
PV	Peak Value
CF	Crest Factor
STD	Standard Deviation
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
R^2	Coefficient of Determination
ROC-AUC	Receiver Operating Characteristic – Area Under Curve
APDL	ANSYS Parametric Design Language
MAC	Modal Assurance Criterion

1. INTRODUCTION

1.1. Definitions and Fundamental Concepts

This section introduces the fundamental concepts and definitions that underpin the methodology adopted in this study. These concepts span structural health monitoring (SHM), vibration-based feature extraction, machine learning paradigms, and transfer learning strategies. A clear understanding of these foundations is essential for interpreting the proposed framework and its results.

Structural Health Monitoring (SHM)

Structural Health Monitoring (SHM) is a multidisciplinary field concerned with the continuous or periodic evaluation of structural integrity using measured data, analytical models, and decision-making algorithms. The primary goal of SHM is to detect damage at an early stage, thereby improving safety, extending service life, and optimizing maintenance strategies [1-3].

In engineering practice, SHM systems typically consist of sensors, data acquisition units, signal processing techniques, and decision-making algorithms. Among various SHM approaches, vibration-based methods are widely used due to their non-invasive nature and ability to capture global structural behavior [4,5].

Damage in a structure generally leads to a reduction in stiffness, which in turn alters its dynamic properties. These changes can be detected through variations in vibration signals, making SHM a powerful tool for condition assessment of beam-like structures.

Damage Identification Framework

Damage assessment in SHM is commonly structured into hierarchical levels that reflect increasing complexity and information requirements [6,7]:

- **Level 1 – Damage Detection (DD):**
  Determines whether damage exists in the structure. This is typically formulated as a binary classification problem.
- **Level 2 – Damage Localization (DL):**
  Identifies the spatial location of damage within the structure. This may be treated as a classification or regression problem depending on the formulation.
- **Level 3 – Damage Severity (DS):**
  Quantifies the extent of damage, such as crack depth or stiffness reduction. This is typically a regression problem and is highly sensitive to noise and modeling uncertainties.

These levels form a progressive framework, where each stage requires more detailed information and presents greater challenges.

Finite Element Modeling (FEM)

The Finite Element Method (FEM) is a numerical technique used to approximate the behavior of complex structures by discretizing them into smaller elements interconnected at nodes. Each element is governed by physical equations that describe

structural behavior, and the global response is obtained by assembling these local contributions.
In the context of this study, FEM is used to simulate the dynamic response of a steel beam under both healthy and damaged conditions. By introducing parametric variations in crack location and severity, a large dataset of vibration responses can be generated.
This simulation-based approach offers several advantages:

- Controlled generation of diverse damage scenarios,
- Repeatability of experiments,
- Reduced cost compared to physical testing,
- Ability to study extreme or unsafe conditions.

Time-Domain Features

Time-domain features are directly computed from the signal amplitude as a function of time. Common features include:

- **Root Mean Square (RMS):** Measures the energy content of the signal.
- **Peak Value (PV):** Represents the maximum amplitude.
- **Crest Factor (CF):** Ratio of peak value to RMS, indicating impulsiveness.

These features are computationally efficient and suitable for real-time applications. They are particularly sensitive to transient events and sudden changes, but may lack spatial resolution for precise damage localization.

Frequency-Domain Features

Frequency-domain analysis transforms the signal into the frequency spectrum using techniques such as the Fast Fourier Transform (FFT). Key features include:

- **Natural Frequencies:** Fundamental vibration characteristics linked to structural stiffness.
- **Mode Shapes:** Spatial deformation patterns of the structure.

Damage alters stiffness distribution, causing shifts in natural frequencies and distortions in mode shapes. These features are strongly grounded in structural dynamics and provide physically interpretable indicators of damage.

Hybrid Features

Hybrid features combine time-domain and frequency-domain information, capturing both local signal variations and global structural behavior. This integrated representation improves the ability of machine learning models to handle complex tasks such as localization and severity estimation.

Machine Learning Paradigms in SHM

Machine learning enables the development of predictive models that learn relationships between extracted features and structural conditions. In this study, multiple categories of machine learning models are considered to ensure a comprehensive evaluation.

- **Tree-Based Models:**
  Ensemble methods such as Random Forest (RF), Extra Trees (ET), and Gradient

Boosting (GB) are capable of capturing nonlinear relationships and interactions between features. They are robust to noise and provide high predictive performance.

- **Neural Network-Based Models:**
  Multi-Layer Perceptron (MLP) models approximate complex nonlinear mappings and can learn hierarchical feature representations.
- **Kernel and Boundary-Based Models:**
  Support Vector Machines (SVM) construct optimal decision boundaries in high-dimensional feature space, making them effective for classification tasks.
- **Linear Models:**
  Logistic Regression and Ridge Regression provide simple and interpretable baselines, useful for assessing linear separability.
- **Probabilistic Models:**
  Gaussian Process Regression (GPR) offers a probabilistic framework, providing both predictions and associated uncertainty.
- **Distance-Based Models:**
  k-Nearest Neighbors (KNN) relies on similarity between data points and serves as a non-parametric baseline method.

The inclusion of diverse model families allows for a systematic comparison of different learning mechanisms and their suitability for SHM tasks.

Transfer Learning and Domain Adaptation
Transfer Learning (TL) is a technique in which knowledge gained from a source domain is leveraged to improve learning in a target domain. In this study:

- **Source Domain:** Simulated data generated from FEM models
- **Target Domain:** Experimental vibration data

A major challenge in SHM is the **domain gap**, which arises due to differences between simulated and real-world conditions. These differences may be caused by:

- Modeling simplifications,
- Measurement noise,
- Environmental variability,
- Boundary condition uncertainties.

Without adaptation, models trained on simulated data may perform poorly on real measurements.

Single-Shot Transfer Learning
In many SHM applications, experimental data is extremely limited. To address this, a Single-Shot Transfer Learning (SS-TL) strategy is adopted, where model adaptation is performed using a very small number of real samples.
This approach is particularly suitable for engineering applications where data collection is expensive or impractical. The objectives of SS-TL include:

- Reducing overfitting under data scarcity,
- Mitigating bias toward simulation data,
- Improving generalization to real-world conditions.

In this study, adaptation is implemented through retraining or fine-tuning of pretrained models using experimental data.

Model Evaluation Metrics

The performance of machine learning models is evaluated using task-specific metrics:

- **Classification Metrics:** Accuracy, Precision, Recall, F1-score, and ROC-AUC.
- **Regression Metrics:** Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

These metrics provide a comprehensive assessment of both predictive accuracy and generalization capability.

1.2. Research Background

Structural Health Monitoring (SHM) has gained prominence as a critical tool for maintaining and ensuring the safety and performance of civil, mechanical, and aerospace structures. As structures age or endure extreme conditions, they become prone to damage that compromises their integrity. Among the various approaches developed for SHM, vibration-based damage detection has emerged as one of the most widely adopted techniques because of its non-invasive nature and sensitivity to structural changes [8].

Time-domain and frequency-domain analyses have long served as the foundation for extracting features sensitive to structural anomalies. However, modern applications require greater robustness and adaptability under varying operational and environmental conditions. With the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML), new opportunities have emerged for automating and improving damage detection accuracy [9].

Despite significant progress, a major gap still exists between simulation-based model training and real-world deployment. This literature review examines previous studies on time-domain and frequency-domain feature extraction, explores AI-driven SHM applications, and highlights the underexplored yet promising role of transfer learning in bridging the simulation-to-real-world gap.

Time-domain analysis examines the raw vibration signal as it varies with time, making it suitable for capturing transient events and sudden anomalies. Common features include Root Mean Square (RMS), peak amplitude, crest factor, skewness, kurtosis, and signal energy. Sohn et al. (2002) provided a comprehensive review of these statistical indicators for damage detection [10]. RMS values measure signal energy, whereas kurtosis and skewness describe the distribution characteristics of the signal amplitude, which may indicate impulsive or asymmetrical events associated with damage [11,12].

Autoregressive (AR) and Autoregressive Moving Average (ARMA) models are also widely used in vibration-based SHM. Worden and Manson (2007) emphasized that AR coefficients are effective in detecting subtle dynamic changes caused by damage, particularly in systems with low-level noise [13]. Time-domain features have been successfully applied for both damage detection and localization in various structures. For example, Hakim et al. (2015) used time-domain vibration signals to train neural networks capable of predicting the location and severity of damage in steel beams [14].

Frequency-domain analysis involves transforming time-series signals into the frequency spectrum to identify modal characteristics such as natural frequencies, damping ratios, and mode shapes. Natural frequencies and mode shapes are among the most widely used features because changes in these properties are strong indicators of structural damage [15]. Modal curvature, defined as the second derivative of mode shapes, has also been shown to be highly sensitive to localized stiffness reduction.

Fourier analysis, particularly the spatial Fourier transform of mode shapes, has also been effectively applied in SHM. Pawar et al. (2007) demonstrated that damage in a fixed–fixed beam causes significant variations in the Fourier coefficients of its mode shapes. These coefficients were subsequently used as inputs for neural networks to predict damage location and severity accurately [16]. Similarly, Power Spectral Density (PSD), Cross Spectral Density (CSD), and Frequency Response Functions (FRFs) have been extensively used for damage assessment. Lee and Kim (2007) introduced a Signal Anomaly Index (SAI) based on deviations in FRFs and used a neural network for damage localization [17].

Time–frequency analysis methods, such as the Short-Time Fourier Transform (STFT) and Wavelet Transform, address the limitations of purely time-domain or frequency-domain approaches, particularly for non-stationary signals. Wavelet analysis is especially effective in detecting transient anomalies [18]. Chang and Chen (2005) applied wavelet decomposition to mode shapes for crack detection in beams [19]. More recently, studies have applied deep wavelet scattering and hybrid Convolutional Neural Network (CNN) architectures for multi-scale feature extraction, reinforcing the growing relevance of time–frequency fusion in SHM [20].

Machine Learning (ML) has significantly advanced structural damage detection by enabling intelligent models to learn complex relationships directly from data. Common ML methods include Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Trees, and Artificial Neural Networks (ANNs). Neural-network-based methods, in particular, have been extensively investigated. Abdeljaber et al. (2016) proposed a non-parametric structural damage detection framework that combined Self-Organizing Maps (SOMs) with feedforward neural networks to process time-domain acceleration data [21]. Their study demonstrated the effectiveness of combining unsupervised feature extraction with supervised classification.

Supervised learning approaches generally rely on labeled datasets representing both healthy and damaged structural states. Although effective, these methods are often constrained by the limited availability of experimental data. Creating representative labeled datasets for real-world damage scenarios can be expensive and time-consuming. Consequently, recent studies have explored autoencoders, Convolutional Neural Networks (CNNs), and graph-based learning techniques to better capture spatial and temporal features in complex structures [22].

Recent research has also investigated the integration of features from multiple domains to improve the robustness and accuracy of damage detection models. Liu et al. (2020) demonstrated that combining time-domain statistical features with frequency-domain modal parameters improved fault classification performance in rotating machinery systems [23]. Although this approach shows strong potential for civil SHM applications, its adoption remains relatively limited.

Hybrid learning frameworks that combine supervised and unsupervised learning phases are also emerging. In such approaches, a Self-Organizing Map (SOM) may first be used for anomaly detection, followed by an ANN for quantifying damage location and severity. This dual-phase architecture reduces reliance on labeled data while improving robustness [21]. These findings reinforce the idea that structural damage manifests across multiple domains and dimensionalities, requiring multi-domain feature representations for optimal performance.

Most previous studies have focused either on purely simulated datasets or entirely experimental datasets. Numerical simulations, commonly performed using the Finite Element Method (FEM), enable the generation of controlled and repeatable damage scenarios. For example, Chang and Chen (2005) simulated crack propagation in beams [19], whereas Pawar et al. (2007) investigated different crack severities and locations in beam models [16].

Experimental investigations, on the other hand, often involve laboratory-scale structures or full-scale structural components. Abdeljaber et al. (2016) validated their SOM–ANN framework using a steel grid structure subjected to ambient vibrations [21]. However, models trained exclusively on FEM-generated data frequently exhibit limited performance when applied directly to real experimental measurements.

Despite substantial advances in vibration-based damage detection and machine-learning-based SHM, transferring models trained on simulated data to real-world applications remains a major challenge. This limitation primarily arises from the domain gap between numerical simulations and experimental measurements, which is caused by modeling assumptions, measurement noise, environmental variability, and uncertainty in boundary conditions.

Although transfer learning has been extensively explored in fields such as computer vision and speech recognition, its application in vibration-based SHM remains relatively limited, particularly for adapting simulation-trained models to experimental structural data.

In this study, a transfer-learning-based adaptation strategy is investigated to improve the compatibility between simulation-trained models and experimental measurements. The objective is not to eliminate the simulation-to-real-world gap completely, but rather to evaluate whether model performance can be improved using limited experimental data. This approach represents a preliminary step toward more practical and deployable SHM systems.

Overall, time-domain and frequency-domain analyses continue to form the foundation of vibration-based structural damage detection. Numerous studies have demonstrated the effectiveness of these approaches, especially when combined with supervised machine learning algorithms. Nevertheless, the transition from simulated environments to real-world applications remains a critical challenge. Current literature lacks systematic investigations into transfer learning for SHM, despite its strong potential to improve the adaptability and robustness of ML models in practical engineering applications. Future research should therefore focus on developing transferable models that can be pretrained on simulated datasets and efficiently adapted to real structures, thereby addressing one of the most significant challenges in AI-driven SHM.

1.3 Research Gap and Problem Statement

Although vibration-based Structural Health Monitoring (SHM) and machine learning techniques have demonstrated promising results for damage detection, localization, and severity estimation, most existing studies rely either on numerical simulations or large experimental datasets. As a result, machine learning models often suffer from a simulation-to-real domain gap, leading to reduced performance when applied to real structures. Furthermore, transfer learning approaches in SHM typically require substantial experimental data, which is rarely available in practice.

Therefore, a need exists for a data-efficient framework capable of transferring knowledge from simulation-based models to real-world measurements using minimal experimental data. This study addresses this gap by combining finite element modelling, vibration-based feature extraction, machine learning, and a Single-Shot Domain Adaption (SS-DA) strategy to investigate whether reliable damage assessment can be achieved using only a single experimentally measured damage scenario.

2. METHODOLOGY

Figure 2 presents the overall methodology adopted in this study. The framework begins with the definition of the structural health monitoring problem, including damage detection, localization, and severity estimation tasks. Numerical and experimental vibration data are then generated through finite element modelling and laboratory testing of an instrumented steel beam. Subsequently, time-domain, frequency-domain are extracted and used to train and evaluate a range of machine learning models. To improve the applicability of simulation-trained models to real-world conditions, a simulation-to-real domain adaptation strategy is implemented using experimental measurements. Finally, the developed models are assessed and compared, followed by a discussion of the key findings, limitations, and future research directions.

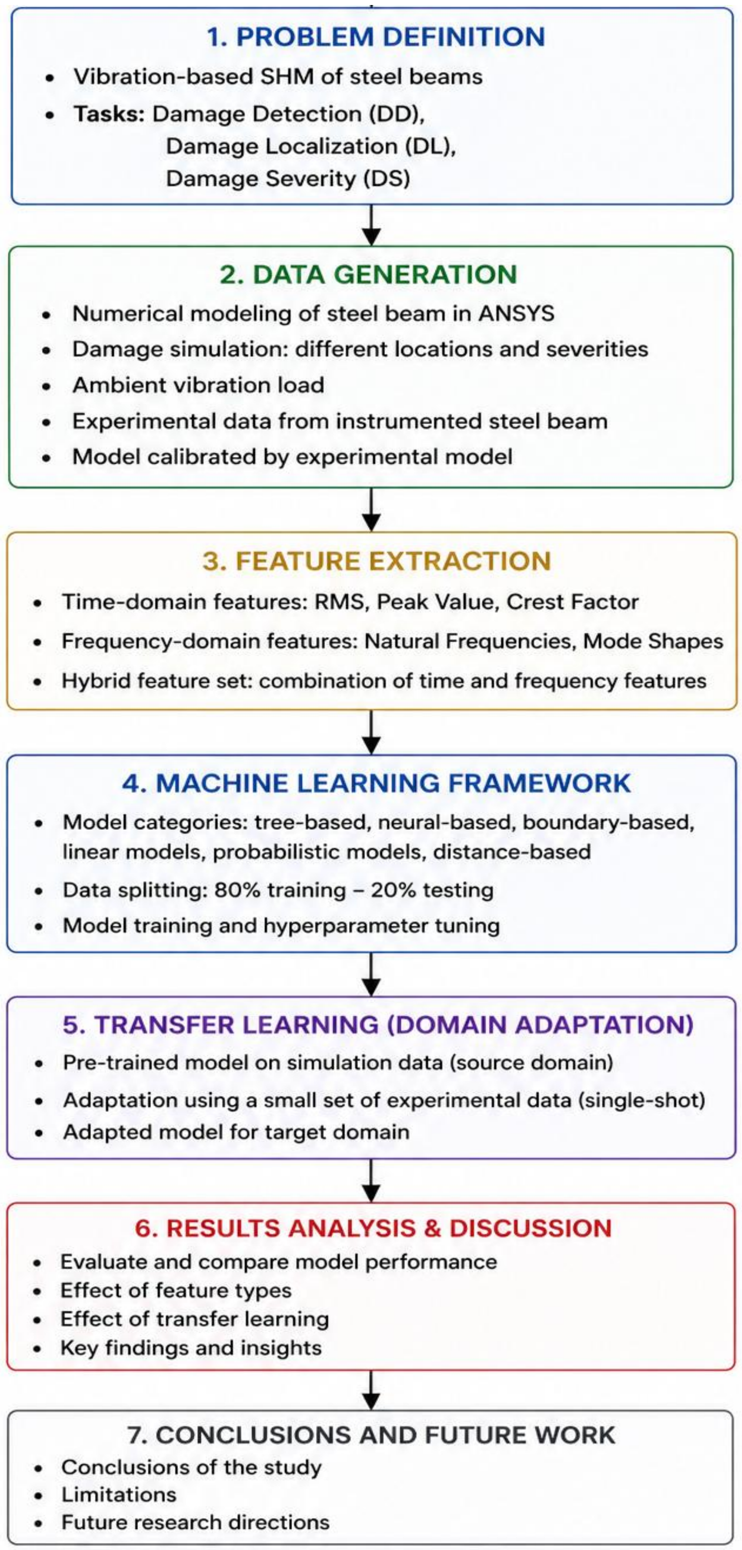

Figure 2. Research Methodology Framework for Vibration-Based Structural Health Monitoring of Steel Beams

2.1. Lunitek Sensor Data Processing and Modal values

This report summarizes the steps carried out to prepare and process Lunitek sensor data for identifying the first ten natural frequencies and corresponding mode shapes of a beam using Frequency Domain Decomposition (FDD). The analysis is based on multiple documents and experimental data provided in CSV format.

2.1.1. Sensor Setup and Orientation

Lunitek Force Balance (FB) sensors were mounted above the top flange of the beam at positions L/4, L/2 and 3L/4 m of left-hand side of beam (**Figure 3**).

- X axis: Transverse direction, perpendicular to the beam's longitudinal axis (parallel to E axis).
- Y axis: Longitudinal (axial) direction, parallel to the beam's length (parallel to N axis).
- Z axis: Vertical direction, pointing upward for Lunitek sensors.

This axis mapping is essential for correctly interpreting the recorded acceleration data.

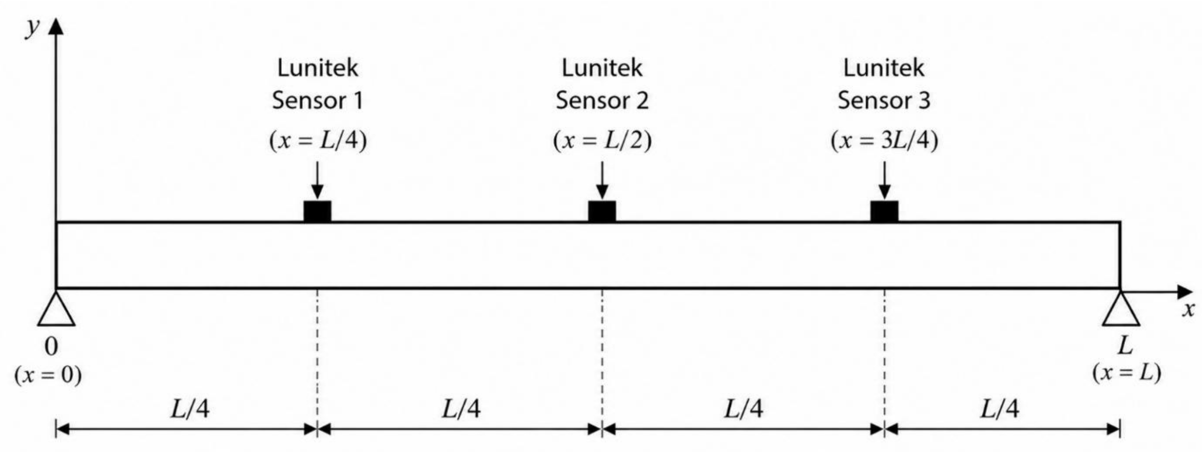

Figure 3. Sensors position on beam

2.1.2. Nature of the Lunitek Data

The Lunitek data consists of acceleration measurements in three directions (X, Y, Z) along with date, time, and timestep or sample index information (**Figure 4**). The acceleration values are small in magnitude, consistent with ambient vibration recordings. While the exact units are not explicitly stated, they are most likely in m/s^2 or g, depending on the Lunitek sensor configuration. In our case the data collected by sensors was in mseed format and should multiply by the correction factor equal to 0.0000031185394639978348 To get the acceleration values.

Field	Value	Size	Class
network	'LK'	1×2	char
station	'EB050'	1×5	char
location	'00'	1×2	char
channel	'HNN'	1×3	char
dataquality	''	0×0	char
type	''	0×0	char
startTime	1.6893e+09	1×1	double
endTime	0	1×1	double
sampleRate	250	1×1	double
sampleCount	6044520	1×1	double
numberOfSamples	6044520	1×1	double
sampleType	'i'	1×1	char
data	*6044520×1 double*	6044520×1	double
dateTime	*1×1 struct*	1×1	struct
dateTimeString	'2023/07/14 06:45:12.876...	1×26	char
matlabTimeVector	*6044520×1 double*	6044520×1	double

Figure 4. Lunitek sensors Information

2.1.3. Data Processing in MATLAB

The MATLAB data processing steps for Lunitek sensors:

1. Read the mseed files containing raw data.
2. Multiply by correction factor and parse the raw text into structured columns: Date, Time, Sensor ID (or timestep), and X, Y, Z acceleration values.
3. Split combined timestamps into separate Date and Time columns for easier filtering and analysis.

2.1.4. Modal Analysis Preparation

To perform modal analysis using FDD, multi-channel acceleration data from all Lunitek sensors must be combined into a single matrix, with each column representing one sensor's acceleration time series in the chosen direction (X, Y, or Z). This multi sensor approach is required to extract both natural frequencies and spatial mode shapes. Single sensor data can provide frequency estimates but cannot produce complete mode shapes.

2.1.5. Planned FDD Methodology

The planned Frequency Domain Decomposition procedure is as follows:

1. Compute the Power Spectral Density (PSD) matrix from the multi-channel time series.
2. Apply Singular Value Decomposition (SVD) to the PSD matrix at each frequency.
3. Identify peaks in the first singular value spectrum, representing the natural frequencies.
4. Extract singular vectors correspond to these peaks to obtain mode shapes.

This will be used to determine the first nine modes and their associated natural frequencies.

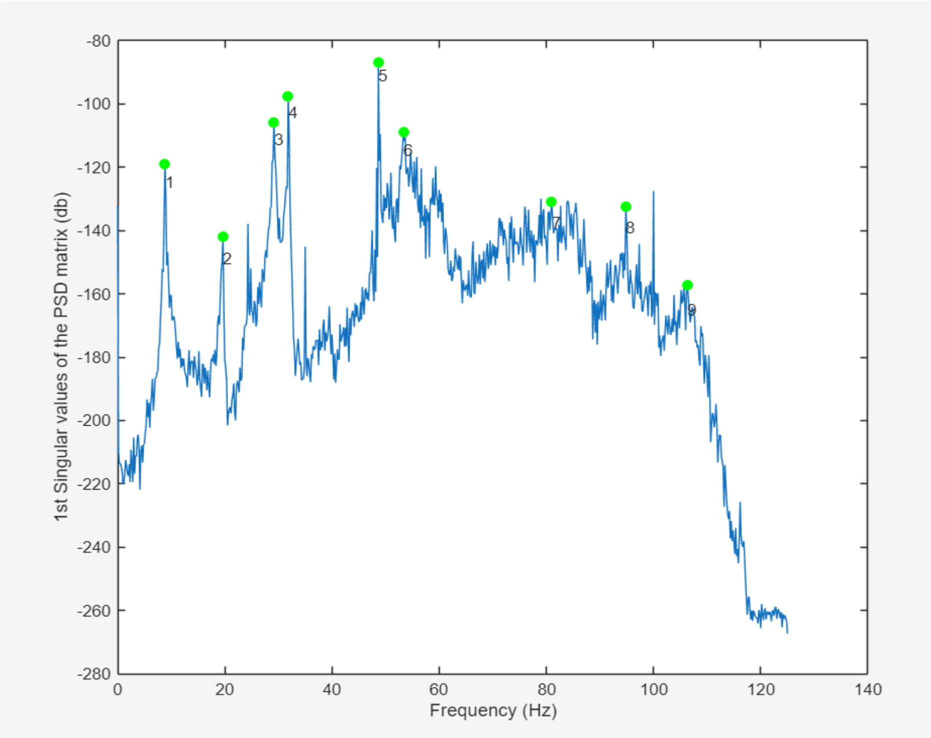

Figure 5. FDD on Experimental data

2.1.6. Conclusion

We have clarified the Lunitek sensor orientations, understood the structure of their acceleration data, and prepared MATLAB scripts for data reading and structuring. Then merge data from all Lunitek sensors and apply the FDD method to determine the first nine natural frequencies and mode shapes under ambient vibration.

2.2. Numerical Modelling and Dataset Generation for AI Training

To develop a robust AI based damage detection framework, it is essential to have a representative dataset encompassing various damage scenarios. Experimental data alone may not cover the full range of possible damage severities and locations, particularly when testing is constrained by time, cost, and safety considerations. For this reason, a numerical modeling approach was adopted to simulate the dynamic response of an IPE160 steel beam under both healthy and damaged conditions. The resulting synthetic data was combined with experimental measurements to form a comprehensive database for AI training.

2.2.1. Beam Geometry and Material Properties

The structure under investigation was modeled as an **IPE160 steel beam**, consistent with the physical specimen used in experimental measurements (**Figure 6**). The geometric parameters were taken from standard IPE profiles, while material properties (Young's modulus, density, and Poisson's ratio) were defined according to manufacturer specifications and verified against engineering steel standards.

- **Length:** *3.8m*
- **Material:** Structural steel 450
- **Boundary Conditions:** Simply supported at both ends, consistent with laboratory conditions

Figure 6. Steel beams under investigation

2.2.2. Crack Modelling

Damage scenarios were introduced in the model by simulating transverse cracks of varying depth and location along the beam span. The crack was represented as a notch in ANSYS Workbench model.

Table 1. Damage Scenario Parameters for the Beam Model

Parameter	Value
Beam length (L)	3.8 m
Crack locations	x = 0.10 m – 3.70 m
Location increment	0.10 m
Number of locations	37
Crack depths (mm)	7.4, 22.0, 36.44, 43.7, 51.0, 65.5, 80.0

Number of severity levels	7
Total damage scenarios	259

Table 1 summarizes the crack modeling parameters considered in the numerical study. Cracks were introduced at 37 locations along the 3.8 m beam span, starting from 0.10 m and ending at 3.70 m with a spacing of 0.10 m. Seven crack severity levels were defined based on crack depth measured from the bottom flange of the IPE beam, ranging from 7.4 mm to 80 mm. Combining all crack locations and severity levels resulted in a total of 259 damage scenarios used for model training and evaluation.

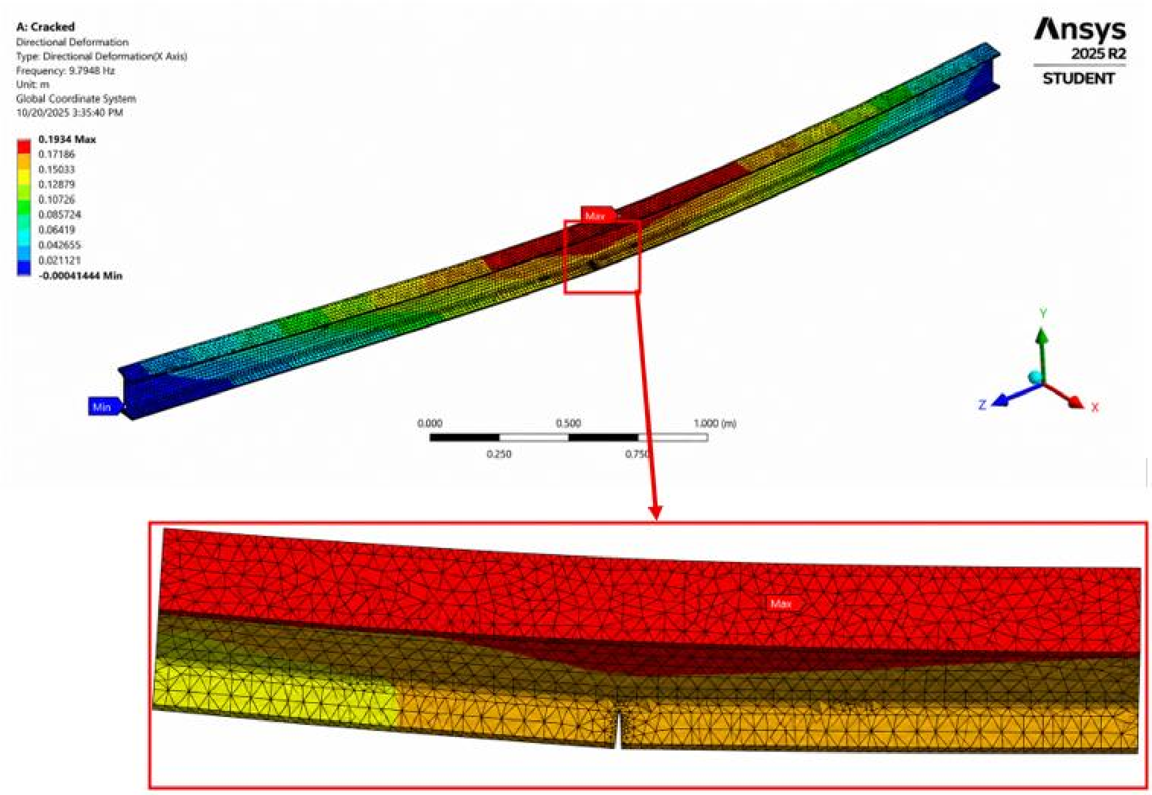

Figure 7. Ansys Model and crack visualization

2.2.3. Finite Element Analysis Setup

The dynamic behavior of the beam was analyzed using modal analysis and Transient Analysis within ANSYS Workbench and Ansys Mechanical APDL:

1. **Modal Analysis:** Extracted natural frequencies and mode shapes for each damage case.
2. **Harmonic/Transient Analysis:** Simulated vibration responses under ambient or base excitation and collecting 5 modeled sensors data.
3. **Meshing:** A fine mesh consisting of SOLID186 elements with an element size of 8 mm was employed. The resulting finite element model comprised approximately 55,000 nodes and 23,000 elements, ensuring sufficient accuracy for capturing the local stiffness variations induced by crack damage.

2.2.4. Transient Dynamic Analysis and Acceleration Response Extraction in ANSYS

Transient dynamic analysis was performed to obtain acceleration time histories at selected points along the beam corresponding to sensor locations. This analysis type allows the structural response to be evaluated as a function of time under time dependent loading and is suitable for simulating impact and short duration dynamic excitations. The finite element model was developed using linear elastic material properties, appropriate element formulations, and validated boundary conditions based on the literature.

The beam geometry was modelled in accordance with the specified dimensions and material properties. A linear elastic, isotropic material model was assumed, as the analysis focused on vibration response within the elastic range of the structure. This assumption is consistent with classical structural dynamics and modal testing studies, where excitation amplitudes are sufficiently small to avoid material nonlinearity. The beam was discretized using appropriate beam or three-dimensional solid elements depending on the required level of modelling fidelity. Boundary conditions were applied to represent the intended support configuration, such as simply supported or cantilever conditions, following classical beam theory and established finite element modelling practice. Particular care was taken to ensure that the numerical boundary conditions accurately reflected the physical constraints, To accurately reproduce the experimental support conditions, displacement supports were used at both beam ends. The left support was fully restrained in translation, while the right support allowed axial displacement along the beam length. Rotations were left free at both supports, thereby representing a simply supported beam configuration. Different support configurations were evaluated and compared against the experimental results. The adopted boundary condition model yielded the best agreement with the measured natural frequencies and mode shapes and was therefore selected for the final finite element model.

- **Sensor Location Definition**

Sensor locations were defined numerically by selecting nodes at specified positions along the beam length, corresponding to the physical locations of accelerometers used in experimental measurements. In finite element dynamic analysis, accelerometers are commonly represented as response points rather than explicitly modelled components, and acceleration time histories are extracted directly from nodal degrees of freedom. The mass of the physical sensors was neglected in the numerical model, as accelerometer mass is typically small relative to the structural mass of slender beam structures and has a negligible influence on global dynamic response. This modelling approach is widely adopted in experimental–numerical correlation studies and has been shown to provide reliable agreement when sensor locations are consistently defined between the model and the experiment.

- **Transient Dynamic Analysis Setup**

A transient dynamic analysis was performed to compute the time dependent response of the beam under dynamic excitation. The governing equations of motion were solved using standard time integration procedures implemented in ANSYS, allowing acceleration responses to be obtained directly at the selected sensor locations. A time step of **0.004 s (1/250 s)** was adopted in the transient analysis to accurately capture the dominant vibration response of the structure while ensuring numerical stability and computational efficiency.

Depending on the analysis requirements, Rayleigh damping or modal damping may be introduced to represent energy dissipation; however, undamped or lightly damped conditions are commonly assumed in impact type studies to avoid obscuring the inherent dynamic characteristics of the structure [24]. The resulting acceleration time histories were post processed for comparison with experimental measurements.

- **Summary of Finite Element Modeling Assumptions and Numerical Parameters**

Table 2 summarizes the main modelling assumptions and numerical considerations adopted in the finite element simulation of the beam. A three-dimensional continuum approach was implemented using SOLID186 elements to capture local effects around the crack more accurately than classical beam elements. A structured hexahedral mesh was generated, and a mesh convergence study based on the first natural frequencies led to an optimal element size of 8 mm. The beam was modeled with pin–roller boundary conditions, where rotations were free and axial displacement was restrained at only one end to avoid over-constraint. Crack damage was simulated by removing material at the specified location. Sensors were numerically represented by selecting nodal response points, from which acceleration time histories were extracted. The material behavior was assumed to be linear, elastic and isotropic. A transient dynamic analysis was performed under base excitation to simulate ambient vibration conditions, with the time step selected according to the highest frequency of interest (1/250 rule). Structural damping was neglected, and an undamped model was considered for simplicity.

Table 2. Key modelling considerations for accurate IPE 160 beam analysis in ANSYS

Aspect	What Considered
Element type	Used SOLID186 (3D continuum)
Beam vs solid modelling	solid elements over beam elements
Mesh size	Perform mesh convergence based on natural frequencies (8mm)
Mesh type	Hexahedral elements
Simply supported beam BCs	Rotation Free, Restrained axial displacement at one end only (pin–roller)
Crack Modelling	Removing Material in crack location
Sensor modelling	Represent sensors as nodal response points (Selecting Nodes at sensors points)
Acceleration extraction	Extract acceleration time histories from selected nodes

Material behaviour	Assume linear elastic isotropic material
Transient analysis	Use transient dynamic analysis for base acceleration loading (Ambient Vibration)
Time step selection	Choose time step based on highest frequency of interest (1/250)
Number of Nodes and Elements	Approximately 55000 nodes and 23000 elements

2.2.5. Automation via Scripting

To efficiently generate large datasets, the entire damage modeling process was automated using ANSYS APDL scripting (**Figure 8**). The script iteratively updated crack parameters (position, depth), executed the analysis, and exported the results in a structured format for further processing. This automation significantly reduced the time required for dataset generation, enabling the creation of hundreds of simulated scenarios.

```
/CWD,"E:\Farzad\AI for Damage Detection of Beams\Prepared Data\Time Domain\crack1\1"
! ========================================================
! PARAMETERS
! ========================================================
dt = 1.0/250
g = 9.81
! ========================================================
! SENSOR WINDOW (unchanged)
! ========================================================
x_s = 0.0
y_s = 0.16
dx_s = 0.01
dy_s = 0.01
! ========================================================
! GET SENSOR NODES (unchanged)
! ========================================================
/PREP7
! ---- find sensor nodes again
NSEL, S, LOC, Z,0.945,0.955
NSEL, R, LOC, X, x_s-dx_s, x_s + dx_s
NSEL, R, LOC, Y, y_s-dy_s, y_s+dy_s
*GET,n1, NODE,0,NUM,MIN
ALLSEL,ALL
NSEL,S,LOC,Z,1.420,1.430
NSEL,R,LOC,X,x_s-dx_s,x_s+dx_s
NSEL,R,LOC,Y,y_s-dy_s,y_s+dy_s
*GET,n2,NODE,0,NUM,MIN
ALLSEL,ALL
NSEL, S, LOC, Z,1.895,1.905
NSEL,R,LOC,X,x_s-dx_s,x_s+dx_s
NSEL,R,LOC,Y,y_s-dy_s,y_s+dy_s
*GET,n3,NODE,0,NUM,MIN
ALLSEL,ALL
NSEL,S,LOC,Z,2.370,2.380
NSEL,R,LOC,X,x_s-dx_s,x_s+dx_s
NSEL, R,LOC,Y,y_s-dy_s,y_s+dy_s
*GET,n4,NODE,0,NUM,MIN
ALLSEL,ALL
NSEL,S,LOC,Z,2.845,2.855
NSEL,R,LOC,X,x_s-dx_s,x_s+dx_s
NSEL,R,LOC,Y,y_s-dy_s,y_s+dy_s
*GET, n5, NODE,0, NUM,MIN
ALLSEL,ALL
FINISH
```

Figure 8. APDL script for dynamic analysis

2.2.6. Data Post Processing in MATLAB

The ANSYS simulation outputs, including acceleration time histories and modal parameters, were imported into MATLAB for further processing.

- **Separation of Directional Data:** X, Y, and Z accelerations were extracted and stored in independent arrays.

- **Feature Extraction:** Frequency Domain Decomposition (FDD) and statistical metrics were computed.
- **Dataset Labeling:** Each case was labeled according to its damage location and severity for supervised AI training.

2.2.7. Integration into AI Framework

The resulting dataset comprised both **healthy** and **damaged** cases, covering a wide range of possible scenarios. This balanced database allowed machine learning models to learn the relationship between vibration characteristics and structural condition, enabling accurate classification and damage localization in future real-world measurements.

2.3. Feature Extraction and Selection for Structural Health Monitoring

In vibration based Structural Health Monitoring (SHM), the extraction of meaningful features from raw sensor measurements is a critical step toward assessing the condition of a structure. The selected features must capture the effects of structural damage while remaining robust to environmental and operational variations. An optimal feature set not only enhances the accuracy of damage detection but also improves interpretability for engineering decision making. In this study, acceleration measurements from models and modal parameters were processed to obtain features suitable for detecting stiffness changes and localizing potential damage in a simply supported beam.

2.3.1. Classification of Features

Structural vibration features are generally classified into three categories: time domain, frequency domain, and time–frequency domain.
Time domain features (**Table 2Table 3**) such as Root Mean Square (RMS), kurtosis, skewness, auto regressive (AR) model coefficients, and signal energy are directly computed from the raw vibration signal. They are computationally efficient and well suited for real time monitoring. Their sensitivity to transient events makes them effective for early anomaly detection, although they may be less precise in damage localization.
Frequency domain features including natural frequencies, modal curvature, power spectral density (PSD), frequency response function (FRF), Fourier coefficients of mode shapes, spectral centroid, and bandwidth are obtained by transforming signals into the frequency domain using spectral analysis techniques such as the Fast Fourier Transform (FFT) (**Table *4***). These features are strongly correlated with the physical properties of the structure and are highly sensitive to stiffness changes, making them valuable for both detection and localization of damage.
Time–frequency domain features such as wavelet coefficients, wavelet entropy, and the Hilbert–Huang Transform (HHT) provide localized frequency content over time, allowing for the analysis of nonstationary and transient signals (**Table 5**). These methods are

particularly useful for identifying damage that evolves over time or occurs under varying load conditions.

Table 3. Time domain features

Feature	Description	Usefulness in Damage Detection
RMS (Root Mean Square)	Measure of the energy or intensity of a signal.	Higher RMS may indicate increased vibration due to damage.
Peak Value	Maximum amplitude observed in a signal.	Can detect sudden events like cracks or impacts.
Crest Factor	Ratio of peak value to RMS.	Identifies impulsive or spiky behavior, possibly due to cracks.
Kurtosis	Measures "tailedness" of the signal distribution.	High kurtosis may indicate impulsive damage events.
Skewness	Measures asymmetry in signal amplitude distribution.	Changes can suggest nonlinear or asymmetrical damage effects.
Standard Deviation	Measures signal variability.	Larger deviation may be linked to structural degradation.
Autocorrelation	Correlation of the signal with itself at different time lags.	Can capture repetitive damage related patterns.
AR/ARMA coefficients	Coefficients from Auto Regressive (AR) models.	Useful for capturing system dynamics; sensitive to changes.
Signal Energy	Integral of squared amplitude over time.	Indicates overall signal power; damage can change energy level.
Entropy (Shannon/Signal)	Measures signal complexity or randomness.	Damage may cause more chaotic signal behavior.

Table 4. Frequency domain features

Feature	Description	Usefulness in Damage Detection
Natural Frequencies	Peaks in the frequency spectrum representing resonance.	Changes in stiffness due to damage shift these frequencies.
Mode Shapes	Vibration shape patterns of the structure at resonance.	Altered mode shapes indicate localized damage.
Modal Curvature	Second derivative of mode shape.	Highly sensitive to local stiffness loss.

Frequency Response Function (FRF)	Ratio of output (response) to input (excitation) in frequency domain.	Changes in FRF magnitude/phase reveal damage.
Power Spectral Density (PSD)	Distribution of signal power over frequency.	Shifts or reductions in power at certain frequencies indicate damage.
Cross Spectral Density (CSD)	Similar to PSD but between two signals.	Reveals phase and amplitude relationships between sensors.
Spectral Entropy	Measure of randomness in the frequency domain.	Damage may make the spectrum less concentrated.
Frequency Centroid	"Center of mass" of the frequency spectrum.	Damage can shift energy concentration in spectrum.
Bandwidth (Effective/3dB)	Width of the frequency range contains significant energy.	Damage can cause modal broadening (increased damping).
Fourier Coefficients (esp. of mode shapes)	Coefficients in the sine/cosine series representation.	Sensitive to both damage **location** and **severity** (as used by Pawar et al.).
Resonance Peak Amplitude	Height of the peak in frequency spectrum.	Damage can dampen or amplify certain modes.

Table 5. Time Frequency domain features

Feature	Description	Usefulness
Wavelet Coefficients	Localized frequency content over time.	Captures transient effects and local damage.
Wavelet Entropy	Complexity of signal in time frequency space.	High sensitivity to sudden or local changes.
Hilbert Huang Transform (HHT)	Empirical method to extract instantaneous frequency.	Good for nonlinear/non stationary systems.

Table 6. Each feature and its application in structural health monitoring (SHM)

Feature Type	Feature	Description & Application	Reference

Time domain	RMS (Root Mean Square)	Measures the energy of a vibration signal. Increases in RMS often indicate structural looseness or cracks.	(Sohn et al, 2003)
Time domain	Kurtosis	Describes the "tailedness" of the signal distribution. High kurtosis indicates impulsive events or localized damage.	(Sohn et al, 2003)
Time domain	Skewness	Indicates asymmetry in the signal amplitude distribution. May suggest unbalanced damage or nonlinear effects.	(Sohn et al, 2003)
Time domain	AR/ARMA coefficients	Parameters from time series models representing the dynamic behavior of the structure. Sensitive to even minor changes in stiffness or mass.	(Pandey et al, 1991)
Frequency domain	Natural frequencies	Key modal property. Shifts in natural frequencies indicate changes in stiffness or boundary conditions due to damage.	(Hakim et al, 2015)
Frequency domain	Modal Curvature	Second derivative of mode shapes. Highly sensitive to local stiffness loss, enabling precise damage localization.	(Hakim et al, 2015)
Frequency domain	Fourier coefficients of mode shapes	Coefficients from spatial Fourier series expansion of mode shapes. Provide a compact and sensitive damage index for detection and localization.	(Fiandaca et al, 2022)
Frequency domain	FRF (Frequency Response Function)	Ratio of output to input vibration in frequency domain. Changes in amplitude or phase suggest degradation or material cracking.	(Pawar et al, 2007)
Frequency domain	PSD (Power Spectral Density)	Describes power distribution over frequency. Damaged structures may show power reduction at certain frequencies or shifting peaks.	(Pawar et al, 2007)
Time Frequency	Wavelet coefficients	Captures signal behavior across both time and frequency. Useful for detecting transient damage events and crack initiations.	(Staszewski et al, 1997)
Time Frequency	Wavelet entropy	Quantifies complexity of the time frequency representation. High entropy may indicate chaotic or damaged behavior in the structure.	(Pawar et al, 2007)
Time Frequency	Continuous Wavelet Transform (CWT)	Provides high resolution localization of energy at different frequency bands and time instances. Useful for identifying mode shape distortions.	(Lee and Kim, 2007)
Time domain	Signal energy	Total energy contained in the vibration signal. Can increase with sudden impacts or decrease due to damping caused by cracks.	(Sohn et al, 2003)
Frequency domain	Spectral centroid	The center of gravity of the spectrum. Indicates where the energy of the frequency spectrum is concentrated and may shift due to damage.	(Chaphalkar et al, 2015)
Frequency domain	Bandwidth (Effective/3dB)	The spread of frequency content around modal peaks. Broadening indicates increased damping, possibly from microcracks or delamination.	(Kralovec et al, 2020)

2.3.2. Criteria for Selecting Optimal Features

The selection of the most relevant features in SHM is guided by the following criteria:

1. **Sensitivity to Damage** – The feature should exhibit measurable and consistent variation in response to structural degradation.
2. **Robustness to Noise** – The feature should remain stable under normal operational and environmental fluctuations.
3. **Interpretability** – The physical meaning of the feature should be clear to facilitate practical engineering conclusions.
4. **Compatibility with Analysis Methods** – The feature should be suited to the intended signal processing or machine learning algorithms.
5. **Low Redundancy** – Highly correlated features should be avoided to reduce computational complexity and prevent overfitting in data driven models.

2.3.3. Evaluation and Robustness Analysis

In this work, feature sensitivity was assessed by comparing measurements from healthy and artificially damaged states of the beam. The percentage change in each feature was calculated, and features with the largest relative variations were prioritized. Robustness was evaluated by introducing simulated measurement noise into the data and observing the stability of feature values. Frequency domain features, particularly natural frequencies and modal curvature were found to be less affected by noise than higher order statistical measures.

2.3.4. Selected Features for the Present Study

Considering the use of the Frequency Domain features method in this project, the primary selected features are:

- **Natural Frequencies** – Identified from FDD peaks, directly related to structural stiffness.
- **Mode Shapes** – Spatial deformation patterns, used for damage localization.

Secondary features such as **RMS**, **Peak Value**, and **Crest Factor** Obtained from time domain analysis were retained for fast anomaly detection and operational checks.

2.3.5. Applications for AI Based Damage Detection

The extracted features can be used as inputs to machine learning (ML) and artificial intelligence (AI) algorithms for automated SHM. Classical ML methods such as Support Vector Machines (SVM) and Random Forests benefit from low dimensional, high quality feature sets. Deep learning models, including Convolutional Neural Networks (CNN), can process richer representations such as mode shape images or time–frequency scalograms. Unsupervised learning approaches, including Self Organizing Maps (SOM) and Autoencoders, can model the undamaged baseline and detect deviations as potential damage events.

2.4. Machine Learning Framework for Damage Detection

2.4.1. Overview of the ML Strategy

To develop a reliable structural damage detection framework for the IPE160 beam, a supervised machine learning approach was adopted. The objective was to construct predictive models capable of:

1. **Damage detection** (Healthy vs Damaged — classification)
2. **Damage localization** (Crack position —Multi class classification/Regression)
3. **Damage severity estimation** (Crack depth — Regression)

The overall machine learning workflow consisted of (**Figure 9**):

- Data preparation and labeling
- Feature extraction (time and frequency domain)
- Feature preprocessing and scaling
- Model selection
- Cross validation strategy to prevent data leakage
- Hyperparameter tuning
- Final model training

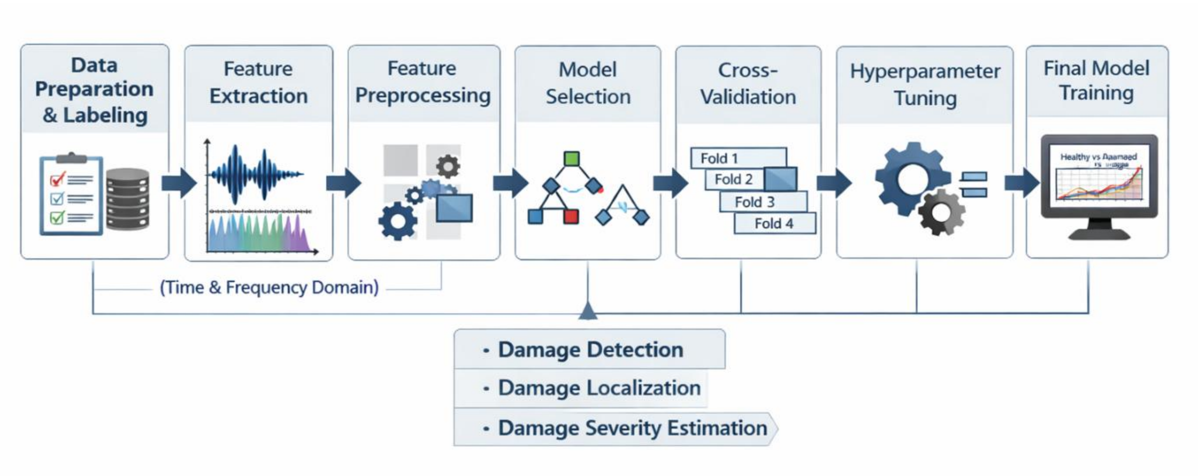

Figure 9. Machine learning framework for damage detection

2.4.2. Feature Engineering

Features were extracted from modal and vibration responses of the beam in both:

- **Time domain** (RMS, Peak value, Crest factor)
- **Frequency domain** (Natural Frequencies, Mode shapes)

These features were selected based on previous literature indicating their sensitivity to stiffness reduction and crack-induced modal changes.

2.4.3. Data Splitting and Validation Strategy

To ensure generalization and avoid data leakage, the dataset was divided using:

- **Stratified K Fold Cross Validation** for classification tasks
- **Group K Fold Cross Validation** where necessary to prevent leakage between crack scenarios

This approach ensures that samples from the same crack configuration are not simultaneously used in training and testing phases (**Figure *10***).

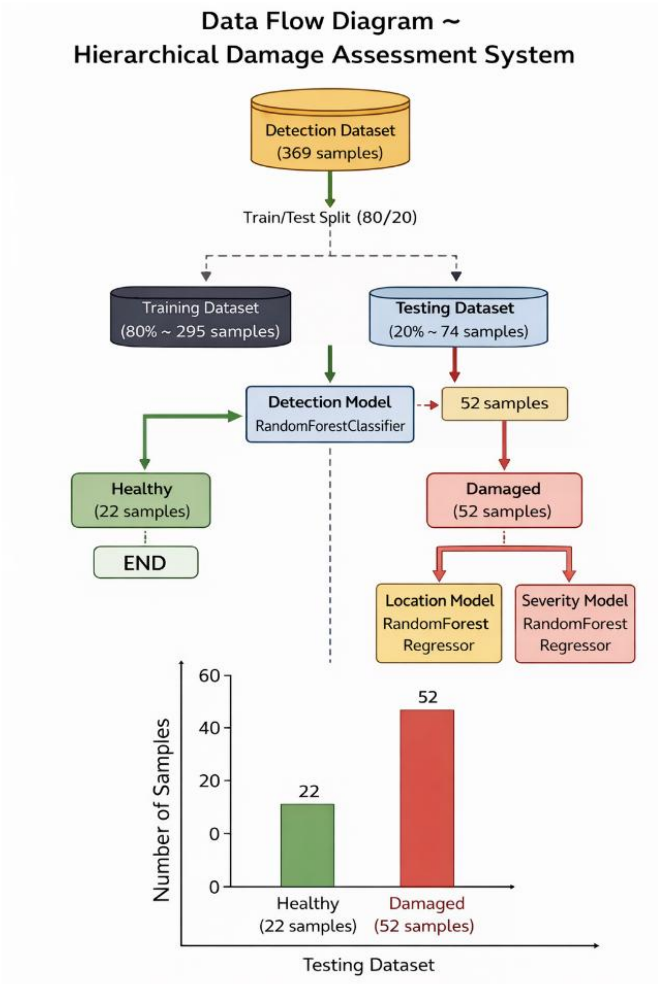

Figure 10. Hierarchical machine learning framework for damage assessment

To assess whether the proposed model suffers from overfitting or data leakage, three complementary evaluation strategies were employed. First, k-fold cross-validation was performed, and the consistency of the coefficient of determination (R^2) scores across folds was examined. Stable performance across folds indicates that the model generalizes well and is less likely to suffer from overfitting [27,28]. Second, a permutation test was conducted by randomly shuffling the target labels and retraining the model. The substantial decrease in performance observed after permutation suggests that the model captures meaningful relationships in the data rather than spurious correlations or data leakage [29]. Third, learning curves were analyzed by comparing training and validation performance. The relatively small gap between training and validation scores indicates that the model maintains good generalization capability without exhibiting severe overfitting [27]. Furthermore, achieving a perfect training score (R^2 = 1 or F1-score = 1) does not necessarily imply overfitting or data leakage. Such results may arise from several factors, including limited dataset size, the use of highly expressive machine learning models, and the presence of highly informative features that strongly correlate with the target variables [27]. Overall, the stable cross-validation performance, significant performance degradation under label permutation, and the small gap between training and validation scores collectively indicate that the model generalizes well and does not exhibit severe overfitting. Nevertheless, although these findings reduce the likelihood of data leakage, they cannot completely eliminate the possibility. Therefore, careful data preprocessing, feature engineering, and dataset splitting procedures must still be ensured throughout the modeling process.

2.4.4. Selected Machine Learning Models

In this study, a diverse set of machine learning algorithms was employed to evaluate their effectiveness in vibration-based structural damage assessment tasks, including damage detection, localization, and severity estimation. The selected models represent different machine learning paradigms, including instance-based learning, kernel-based methods, tree-based algorithms, ensemble learning approaches, boosting techniques, and neural network models, enabling a comprehensive comparison of their predictive capabilities.

To ensure reliable and unbiased model comparison, all algorithms were optimized through hyperparameter tuning and cross-validation. The evaluated models include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), and Multi-Layer Perceptron (MLP). These models were selected due to their widespread application in structural health monitoring (SHM) and their ability to capture nonlinear relationships in vibration-based structural response data [31,32].

Mathematically, the employed models can be summarized as follows.

The KNN algorithm predicts a target value based on the nearest neighboring samples using a distance metric, typically the Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \qquad (2.1)$$

Support Vector Machines determine an optimal separating hyperplane:

$$f(x) = w^T x + b \qquad (2.2)$$

while nonlinear relationships are modeled using the radial basis function (RBF) kernel:

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \qquad (2.3)$$

Random Forest and Extra Trees combine multiple decision trees through ensemble averaging:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \qquad (2.4)$$

where B represents the number of trees and $f_b(x)$ denotes the prediction of the b^{th} tree.

Gradient Boosting constructs the final model sequentially by minimizing prediction error iteratively:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \qquad (2.5)$$

where $h_m(x)$ is the weak learner added at iteration m, and η is the learning rate.

The Multi-Layer Perceptron (MLP) model computes nonlinear transformations through weighted neuron activations:

$$a_j^{(l)} = \sigma(\sum_i w_{ij}^{(l)} x_i + b_j^l) \qquad (2.6)$$

where $w_{ij}^{(l)}$ and b_j^l are the weights and biases, respectively, and σ is the activation function.

The inclusion of these diverse machine learning approaches enables a comprehensive evaluation of different predictive mechanisms for vibration-based structural health monitoring and facilitates identification of the most suitable algorithms for each damage assessment task.

Table 7. Comparative Summary of Selected Machine Learning Models

Model	Main Characteristics	Advantages	Limitations	SHM Application
RF	Bagging-based ensemble trees	Robust, interpretable, resistant to overfitting	Moderate computational cost	Damage localization
ET	Highly randomized ensemble trees	Strong generalization, fast training	Reduced interpretability	Damage detection
GB	Sequential boosting ensemble	High predictive accuracy	Longer training time, risk of overfitting	Severity estimation
SVM / SVR	Kernel-based learning	Effective nonlinear regression with strong generalization	Computationally intensive for large dataset	Damage severity and localization regression
KNN	Distance-based learning	Simple and nonparametric	Sensitive to scaling and noise	Baseline comparison
LR	Linear statistical models	Interpretable and computationally efficient	Limited nonlinear capability	Baseline analysis
RR	Regularized linear regression	Improved stability with correlated features	Limited nonlinear representation	Localization baseline

MLP (ANN)	Neural-network-based learning	Captures complex nonlinear relationships	Requires larger datasets and tuning	Nonlinear pattern learning
GB	Bayesian probabilistic learning	Uncertainty estimation, strong small-data performance	Computationally expensive	Regression and uncertainty analysis

2.4.5. Model Configuration

To ensure consistency and fair comparison, all machine learning models were implemented using optimized hyperparameter settings (**Table 8**). The hyperparameters were selected through preliminary testing and performance evaluation to achieve stable training behavior, balanced model complexity, improved generalization capability, and reduced risk of overfitting and data leakage. Tree-based ensemble models were configured to enhance robustness and minimize variance, while kernel-based methods employed RBF kernels to capture nonlinear relationships in the vibration features. MLP neural networks utilized multiple hidden layers and adaptive learning strategies to model complex patterns while limiting overfitting during training. KNN models relied on standardized distance-based learning, and linear models incorporated regularization to improve stability and predictive performance.

Table 8. Hyperparameter Configuration of the Selected Machine Learning Models

Model	Main Hyperparameters	Selected Values
Support Vector Machine (SVM)	Kernel, C, Gamma	RBF, 10, 0.01
Support Vector Regression (SVR)	Kernel, C, Gamma, Epsilon	RBF, 10, 0.01, 0.01
Random Forest (RF)	n_estimators, max_depth, max_features	200, 20, sqrt
Extra Trees (ET)	n_estimators, max_depth, max_features	200, 20, sqrt
Gradient Boosting (GB)	n_estimators, learning_rate, max_depth	200, 0.05, 4
k-Nearest Neighbors (KNN)	n_neighbors, metric, weights	5, Euclidean, uniform
Logistic Regression	Solver, Regularization (C)	lbfgs, 1.0
Ridge Regression	Alpha	1.0
Multi-Layer Perceptron (MLP)	Hidden Layers, Activation, Learning Rate	(128,64), ReLU, 0.001
Gaussian Process (GP)	Kernel, Length Scale, Alpha	RBF, 1.0, (10^{-6})

2.4.6. Rationale for Model Diversity

The selected models represent a broad range of machine learning paradigms:

- **linear models:** Logistic Regression, Ridge Regression,
- **distance-based models:** KNN classifier and regressor,
- **kernel-based models:** SVM, SVR, Gaussian Process Classifier, Gaussian Process Regressor,
- **neural networks:** MLP classifier and regressor,

- **ensemble tree methods:** Random Forest, Extra Trees, and Gradient Boosting in both classifier and regressor forms.

This diversity makes the comparison more comprehensive and enables a clearer understanding of which modeling strategy is most suitable for vibration-based damage detection and damage localization.

Table 9. AI Training Methods in SHM

AI Method	Type	Typical Use in SHM	Advantages	Reference
Artificial Neural Network (ANN)	Supervised	Predict damage location and severity from modal features (frequencies, mode shapes, etc.)	Nonlinear modeling, good for complex systems	(Fiandaca et al, 2022)
Support Vector Machine (SVM)	Supervised	Classify damaged vs. undamaged states using extracted features	High accuracy with small datasets, good generalization	(Pawar et al, 2007)
k Nearest Neighbors (k NN)	Supervised	Classify damage levels based on similarity to known cases	Simple to implement, intuitive	(Pandey et al, 1991)
Decision Trees	Supervised	Classify types of damage or structural states	Interpretable rules, fast classification	(Lie et al, 2018)
Random Forest (RF)	Supervised	Ensemble approach to improve classification robustness	Handles overfitting better than single trees	(Lie et al, 2018)
Self-Organizing Map (SOM)	Unsupervised	Detect anomalies and cluster modal patterns from ambient vibration signals	Reduces dimensionality, visualizes damage clusters	(Adeljaber and Avci , 2016)
Autoencoder (AE)	Unsupervised	Learn latent representation of undamaged state; reconstruction error indicates damage	Works well with unlabeled data, anomaly detection	(Puntambekar et al, 2025)
Gaussian Mixture Model (GMM)	Unsupervised	Probabilistically cluster feature distributions to infer damage	Probabilistic reasoning, useful when damage follows a known distribution	(Dervilis et al, 2014)
Convolutional Neural Network (CNN)	Supervised/Deep	Used on images of mode shapes or scalograms (time frequency plots) to	Captures spatial and time frequency patterns, effective for wavelet-based input	(Altabey et al, 2020)

		classify damage locations		
Recurrent Neural Network (RNN/LSTM)	Supervised/Deep	Capture temporal sequences in vibration signals for long term monitoring	Excellent for time series; models sequential dependencies	(Barzegar et al, 2022)
Bayesian Neural Network (BNN)	Supervised	Estimate damage and uncertainty in predictions	Produces confidence intervals, helpful for risk-based decision making	(Gal and Ghahremani, 2016)
Hybrid SOM + ANN	Hybrid (Unsupervised + Supervised)	SOM detects anomaly; ANN classifies severity and location	Combines anomaly detection with accurate quantification	(Farrar and Worden, 2012)
Transfer Learning (Fine Tuning)	Supervised + Transfer	Pretrain on FEM simulation data, fine tune on real experimental beam data	Reduces need for large labeled real-world datasets	(Coletta et al, 2020; Fidma et al,2025; Lim et al, 2025; chen et al, 2025; Ahmed M, 2025)

2.4.7. Model Evaluation Strategy

Different evaluation metrics were defined depending on the task:

For Classification:

- Accuracy
- Precision
- Recall
- F1 score
- Confusion Matrix
- Roc_Auc
- Run Time

For Regression:

- Mean Squared Error (MSE)
- Mean Absolut Error (MAE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)
- Run Time

These metrics allow comprehensive assessment of both detection and quantification capabilities.

2.5. Transfer Learning Framework

2.5.1. Motivation

In structural health monitoring applications, acquiring large quantities of labeled experimental data is often expensive, time consuming, and sometimes impractical. In contrast, numerical simulations based on calibrated FE models can generate extensive datasets covering multiple crack locations and severities.
However, even after calibration, differences remain between numerical and experimental data due to:

- Material uncertainties
- Boundary condition imperfections
- Sensor noise
- Modeling simplifications
- Manufacturing tolerances

These differences create a domain gap between simulated (source domain) and real experimental data (target domain). Training a model solely on numerical data may therefore reduce its generalization capability when applied to real measurements.
To address this limitation, a transfer learning strategy was implemented to transfer knowledge learned from the numerical dataset to the experimental dataset.

2.5.2. Transfer Learning Strategy

The adopted approach is based on simulation-to-real domain adaptation using a limited amount of experimental data (Single model). The objective is to investigate whether knowledge learned from a large finite element (FE) dataset can improve prediction performance under real experimental conditions while reducing the discrepancy between simulated and measured structural responses.
It should be noted that the adopted approach does not correspond to conventional transfer learning or full model fine-tuning. Instead, it represents a case-specific simulation-to-real domain adaptation strategy developed for situations where experimental data are severely limited. The machine learning models were initially trained using the FEM-generated dataset containing 259 simulated damage scenarios with different crack locations and severities. Model adaptation was then performed using measurements obtained from a single experimental damage configuration corresponding to a crack located at 1900 mm with a severity of 7.4 mm. To preserve previously learned structural knowledge, a subset of FEM samples was retained during adaptation and combined with the experimental samples. For the tree-based models, the experimental samples were assigned a higher importance during retraining through weighted adaptation (λ=30). Consequently, the reported improvements should be interpreted as evidence of successful adaptation to the available experimental configuration rather than proof of generalization to arbitrary real-world damage scenarios. Further validation using multiple experimental damage locations, severities, and environmental conditions is required before broader conclusions regarding real-world deployment can be made.

To reduce the simulation-to-experiment domain gap, a single-shot domain adaptation framework was developed. The source domain is defined as:

$$D_s = \{X_s, Y_s\} \qquad (2.7)$$

where X_s denotes the feature matrix extracted from FEM simulations and Y_s represents the corresponding damage labels. Similarly, the target domain is defined as:

$$D_t = \{X_t, Y_t\} \qquad (2.8)$$

where X_t contains the experimentally extracted vibration features.
Although both domains share the same feature space, their underlying probability distributions differ:

$$P(X_s) \neq P(X_t) \qquad (2.9)$$

The machine learning models were first pretrained on the source-domain dataset to learn the mapping:

$$f_s\colon X_s \rightarrow Y_s \qquad (2.10)$$

To preserve previously learned knowledge during adaptation, a subset of FEM samples was retained as simulation memory. The adaptation dataset was therefore constructed as:

$$D_{adapt} = D_{memory} \cup D_t \qquad (2.11)$$

where D_{memory} denotes the retained FEM samples and D_t represents the experimental adaptation samples.

For damage detection and localization tasks, the adapted classifier predicts:

$$\hat{y} = f_t(x) \qquad (2.12)$$

where $\hat{y}$ is the predicted class label corresponding to either damage state or crack location.

For crack severity estimation, the adapted regression model predicts:

$$\hat{s} = f_t(x) \qquad (2.13)$$

where ($\hat{S}$) represents the estimated crack severity.

The adaptation process was performed using weighted retraining, where greater importance was assigned to the experimental samples. The objective function can be expressed as:

$$L = \sum_{i=1}^{N_s} L_s(x_i, y_i) + \lambda \sum_{j=1}^{N_t} L_t(x_j, y_j) \qquad (2.14)$$

Where L_s and L_t denote the source-domain and target-domain loss functions, respectively, and λ is the weighting factor assigned to experimental samples. This strategy enables the model to adapt to real-world measurements while maintaining the predictive knowledge acquired from the FEM-generated dataset.

2.5.3. Physical Justification for Transferability

The use of transfer learning in this study is supported by the fact that both numerical and experimental systems are governed by the same underlying structural dynamics. In particular, damage-induced stiffness reduction affects modal properties (natural frequencies and mode shapes) in a physically consistent manner across both domains. This shared physics provides a strong foundation for knowledge transfer between simulated and real data.

However, it is important to note that the experimental validation in this study is limited to a relatively small dataset corresponding to a one damage configuration. As a result, the proposed approach follows a data-constrained (single-shot) adaptation framework, and the observed improvements should be interpreted as evidence of effective model adaptation.

3. Results and Discussion

3.1. Validation of ANSYS Model and Experimental Results

To ensure the reliability of the FE model, its dynamic behavior was validated against experimental measurements obtained from the Lunitek sensors datasets. The validation process focused on comparing the first 9 natural frequencies and corresponding mode shapes derived from experimental modal analysis with those simulated numerically.

3.1.1. Comparison of Natural Frequencies

Modal analyses were performed for the undamaged (healthy) beam configuration using both experimental Frequency Domain Decomposition (FDD) results and the ANSYS FE model. **Table *10*** summarizes the comparison between the first nine natural frequencies. The relative errors between experimental and numerical frequencies remained within an acceptable range (typically below 10%), indicating strong correlation and adequate modeling accuracy.

Table 10. Comparison of frequencies of Experimental analysis and Model

Mode	Bema 3.8 m _ Cracked in middle of Bottom Flange		Differences %
	Experimental	Ansys	
	Frequencies (HZ)	Frequencies (HZ)	
1	8.9	9.7	8.99
2	19.6	20.6	5.1
3	29.4	29.2	0.68
4	31.8	31.2	1.89
5	48.7	49.015	0.65
6	53.5	55.533	3.8
7	81.9	80.287	1.97

8	95	87.948	7.42
9	110.3	112.52	2.01
Mean			3.61

3.1.2. Mode Shape Correlation

Mode shapes were visually compared and quantitatively assessed using the Modal Assurance Criterion (MAC). MAC values for the first few modes exceeded 0.9, confirming a high degree of similarity between the simulated and experimental mode shapes. Minor discrepancies in higher modes were attributed to sensor placement limitations and boundary condition simplifications in the model (

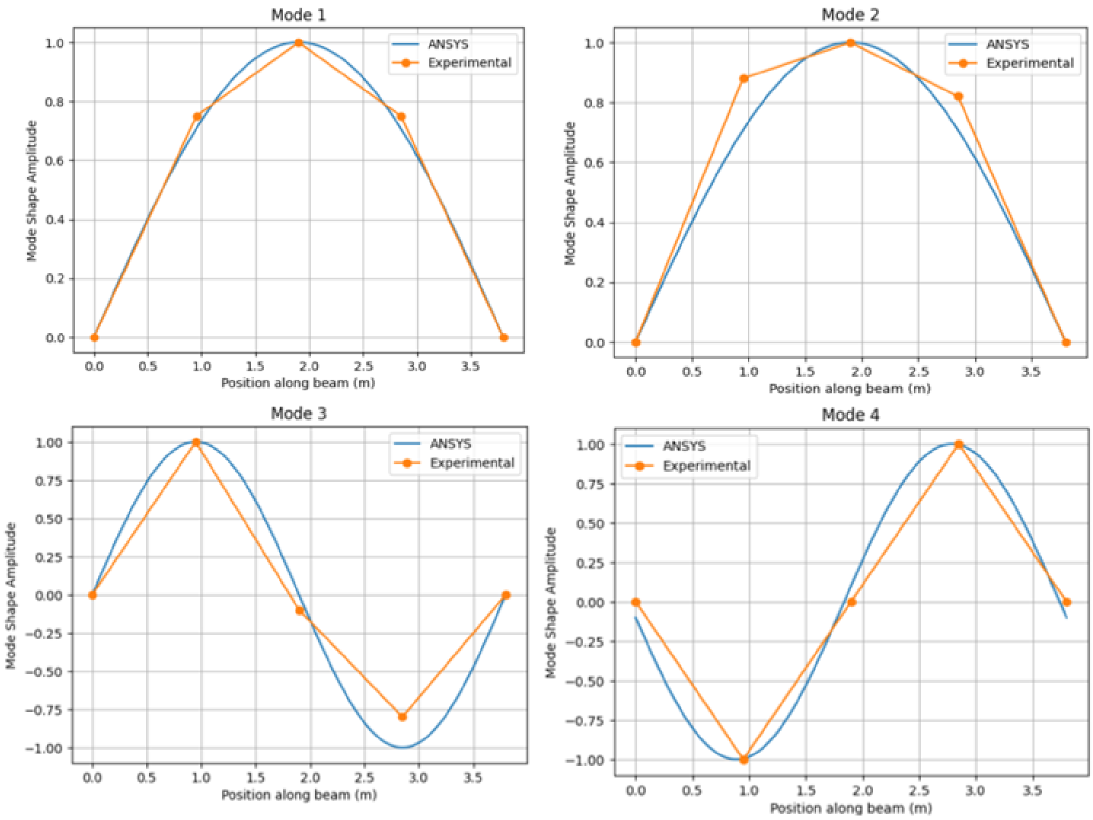

Figure 11 and **Table *11*.** Modal Assurance Criterion).

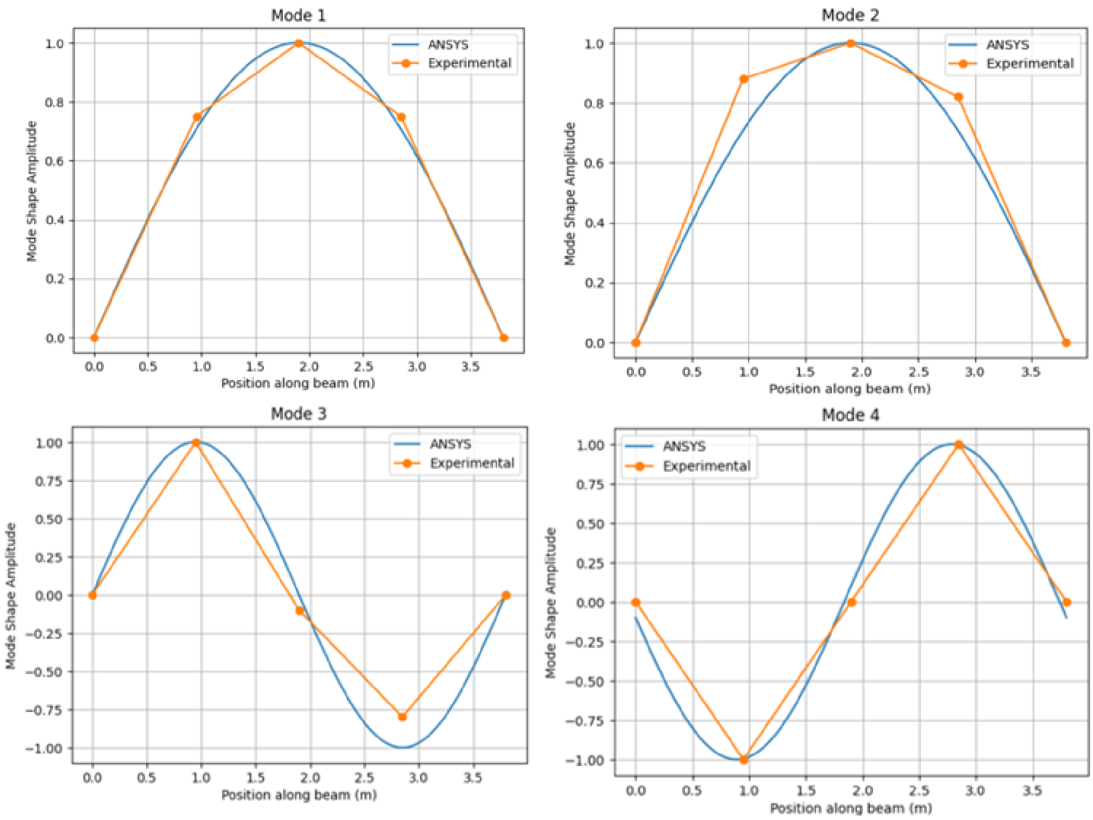

Figure 11. Mode shapes (Experimental and Model) overlays

Table 11. Modal Assurance Criterion

Mode	type	Dir	MAC
1st	Es. Flexural 1st order	X	0.998445
2nd	Es. Flexural 2nd order	X	0.966514
3rd	Es. Flexural 3rd order	X	0.966241
4th	Es. Flexural 4th order	X	0.923659
1st	Es. Flexural 1st order	Y	0.923451

The close agreement observed between the experimentally identified modal parameters and those predicted by the FE model confirms the validity of the numerical representation. In particular, the comparison of natural frequencies and mode shapes demonstrated strong correlation, indicating that the FE model reliably captures the dynamic behavior of the IPE160 beam. Consequently, the numerical model can be considered sufficiently accurate for subsequent analyses. This validation provides confidence that the damage scenarios simulated within the FE framework can serve as a dependable basis for feature extraction, dataset generation, and machine learning model development. Therefore, the results obtained from the FE model are deemed representative of the physical system and suitable for structural damage detection studies.

3.2. Machine learning Results

Machine learning models were trained using the obtained features from Frequency and Time domain analysis. The objective was to evaluate the ability of Models and features to detect damage, localize crack position, and estimate crack severity.

3.2.1. Machine Learning Algorithms on Frequency Features

In this section, machine learning models were trained using the first four natural frequencies (f_1–f_4) as input features. The objective was to evaluate the ability of frequency domain features to detect damage, localize crack position, and estimate crack severity.

- **Damage Detection (Only Frequencies)**

The performance of the machine learning models trained using only the first four natural frequencies (f_1–f_4) clearly demonstrates that raw modal properties contain sufficient information for accurate damage detection. As illustrated in **Figure 12**, ensemble-based models including Random Forest, Extra Trees, and Gradient Boosting achieved perfect classification performance, with Accuracy, F1-score, and ROC-AUC all equal to 1.000. This indicates that these models can effectively capture the relationship between changes in structural stiffness (reflected in natural frequencies) and the presence of damage.

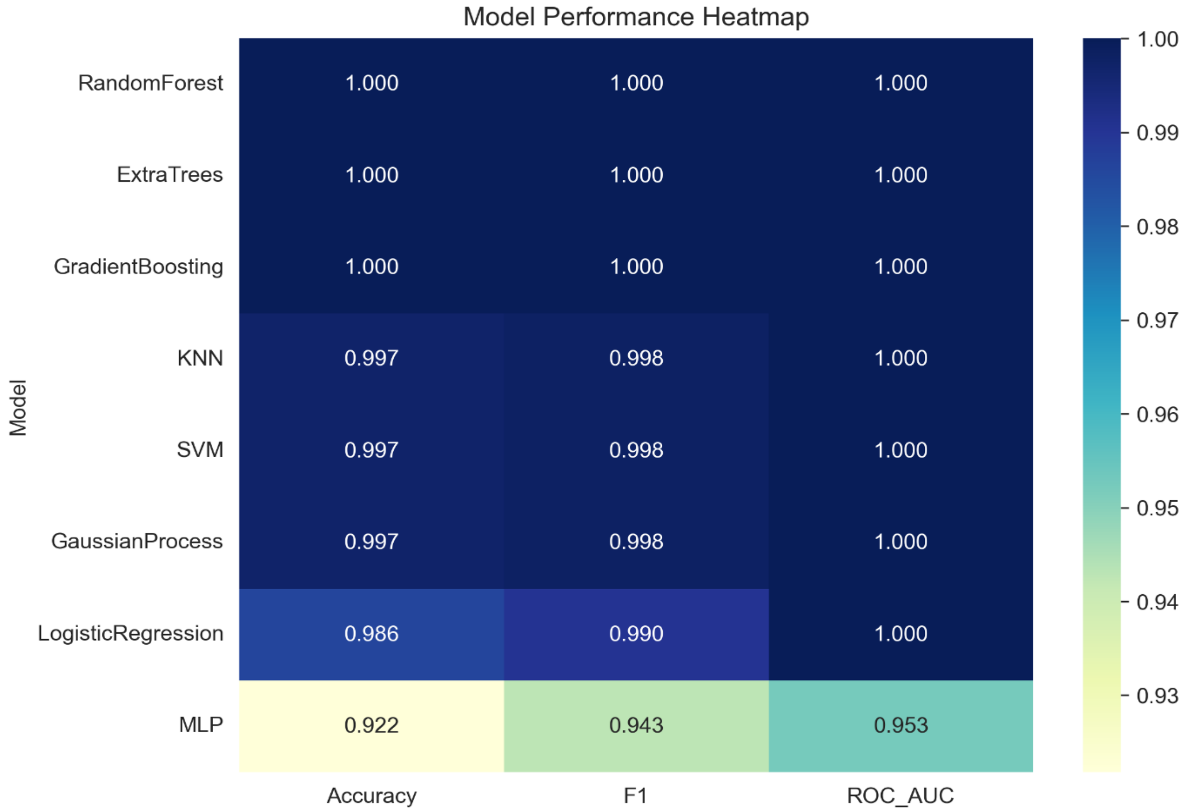

Figure 12. Heatmap of Model Performance for Damage Detection

Other models such as KNN, SVM, and Gaussian Process also showed near-perfect performance, with accuracy values around 0.997 and ROC-AUC equal to 1.000. This

confirms that even without complex feature engineering, the natural frequencies alone provide highly discriminative information for distinguishing between healthy and damaged states.
Logistic Regression achieved slightly lower but still strong performance (Accuracy = 0.986, F1 = 0.990), suggesting that the relationship between frequencies and damage is largely linearly separable. On the other hand, the MLP model showed comparatively lower performance (Accuracy = 0.922, F1 = 0.943, ROC-AUC = 0.953), which may be attributed to the limited complexity of the input features or insufficient tuning, indicating that deep learning approaches are not necessarily advantageous for this type of data.
Overall, the results demonstrate that natural frequencies (f_1–f_4) alone are highly reliable features for damage detection, enabling most machine learning models to achieve near-perfect performance. Ensemble learning methods, in particular, provide the most accurate and stable predictions, making them the most suitable choice for this task.

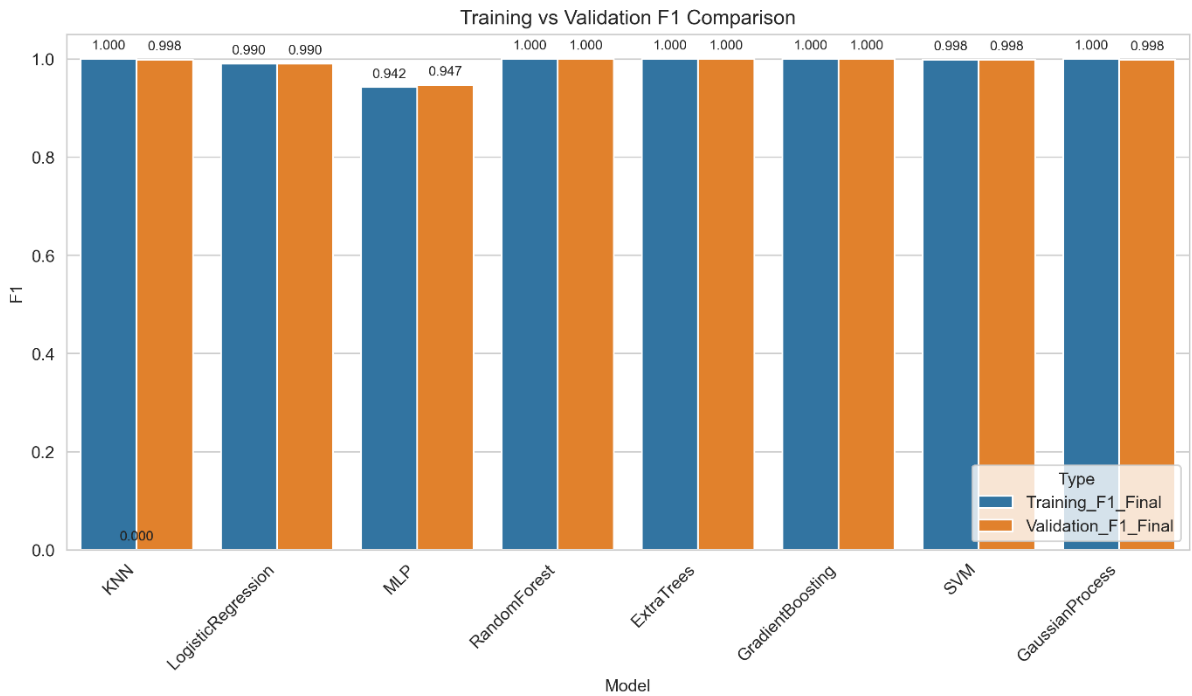

Figure 13. Training vs Validation F1-Score Comparison for ML Models

The comparison between training and validation F1-scores shows that most models exhibit excellent generalization performance, with minimal differences between training and validation results. Ensemble models (Random Forest, Extra Trees, Gradient Boosting) achieved perfect or near-perfect F1-scores on both datasets, indicating no overfitting. Similarly, KNN, SVM, and Gaussian Process maintain very high and consistent performance. Logistic Regression shows a slight drop, while MLP has the lowest scores but still remains stable between training and validation. Overall, the small gaps confirm that the models are well-trained and robust, with negligible overfitting.

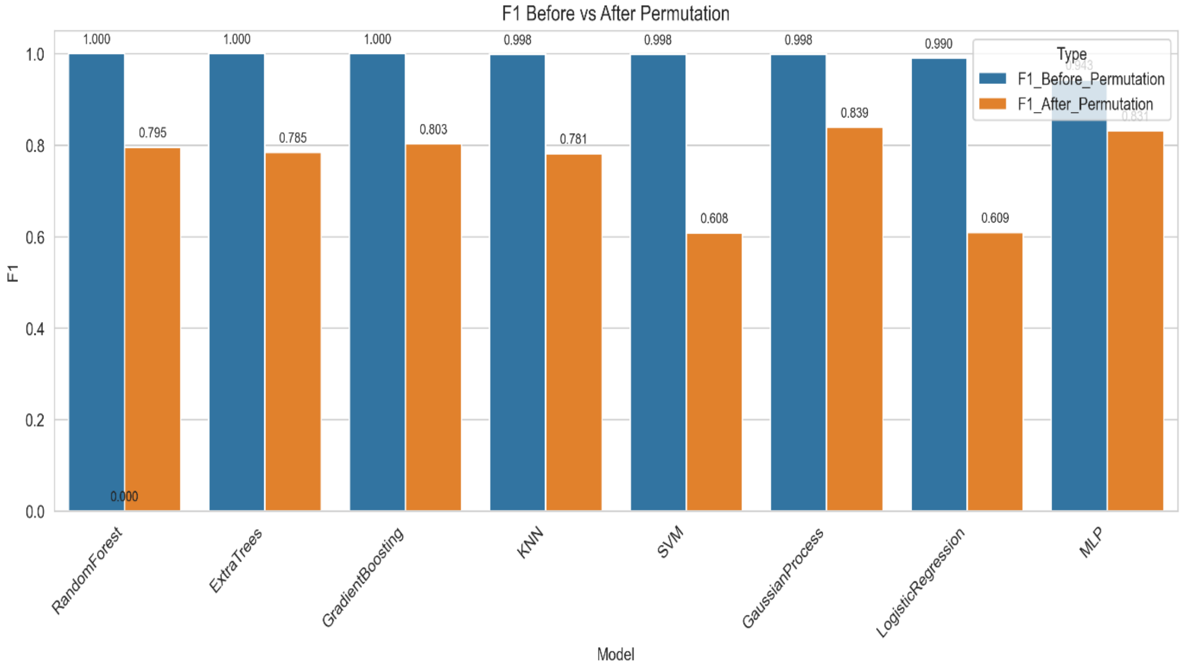

Figure 14. Impact of Feature Permutation on Model Performance (F1-Score Comparison)

The results show a significant decrease in F1-score after feature permutation for all models (**Figure 14**), indicating that the input features (natural frequencies) carry strong and meaningful information for damage detection. Ensemble models such as Random Forest, Extra Trees, and Gradient Boosting experience notable performance drops (from 1.000 to 0.78–0.80), while SVM and Logistic Regression show even larger declines, reaching approximately 0.60. Gaussian Process and MLP retain relatively higher performance after permutation but still exhibit clear degradation. Overall, this confirms that model predictions are highly dependent on the original feature structure, and the features used are both informative and essential rather than redundant.

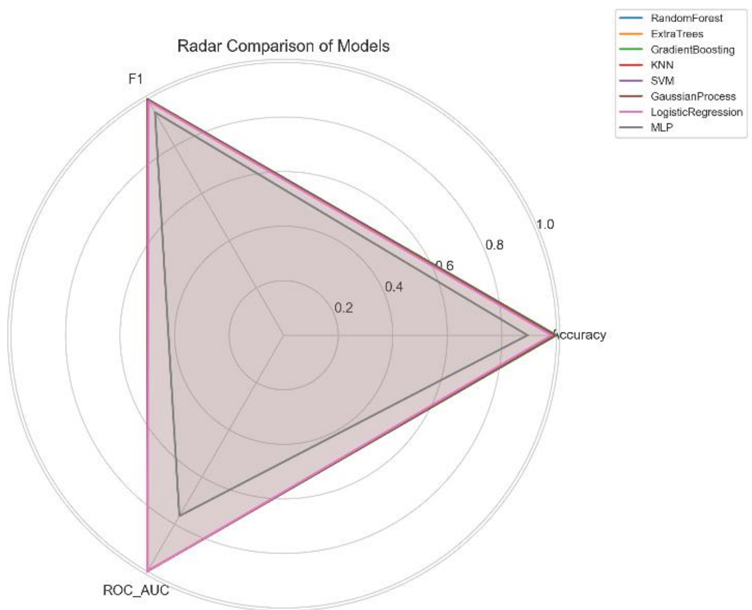

Figure 15. Radar Chart Comparison of Machine Learning Model Performance

The radar chart illustrates the comparative performance of all models across Accuracy, F1-score, and ROC-AUC. Most models, particularly ensemble methods (Random Forest, Extra Trees, and Gradient Boosting), achieve near-perfect and overlapping performance, forming almost identical outer boundaries on the chart. KNN, SVM, and Gaussian Process also show very high and consistent results. In contrast, MLP displays a noticeably smaller area, indicating relatively lower performance across all metrics. Overall, the chart highlights the strong and consistent performance of most models, while also clearly showing the slight underperformance of MLP.

- **Crack Localization (Only Frequencies)**

The results indicate that crack localization is significantly more challenging than damage detection when using only natural frequencies. Tree-based ensemble models, particularly Extra Trees and Random Forest, achieved the best performance with accuracies around 0.67 and strong ROC-AUC values above 0.92, confirming their superior capability in capturing complex relationships for localization tasks. Gradient Boosting shows moderate performance, while KNN, SVM, and Logistic Regression perform relatively poorly with accuracies around 0.30. Gaussian Process performs slightly better but remains limited. The MLP model exhibits very weak performance, indicating that frequency-only features are insufficient for deep learning-based localization. Overall, while frequencies are effective for detection, their ability to precisely localize cracks is limited, with ensemble models providing the most reliable results.

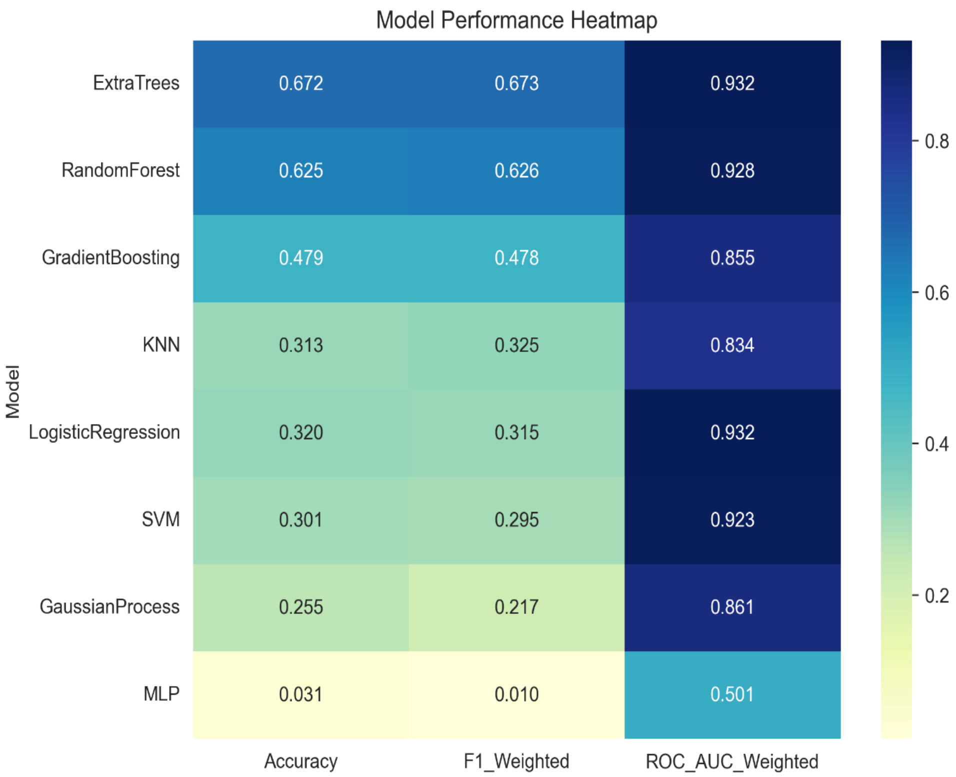

Figure 16. Heatmap of Model Performance for Crack Localization Using Natural Frequencies (f_1–f_4)

The comparison between training and validation F1-weighted scores reveals a noticeable performance gap for all models, indicating the presence of overfitting in the crack localization task. Ensemble models (Random Forest, Extra Trees, and Gradient Boosting) achieve the highest training performance; however, their validation scores drop significantly, suggesting limited generalization to unseen data. Simpler models such as KNN, Logistic Regression, and SVM show consistently low performance on both training and validation sets, indicating underfitting. The MLP model performs poorly overall. These results confirm that crack localization using only natural frequencies is a challenging problem, with reduced generalization capability compared to damage detection.

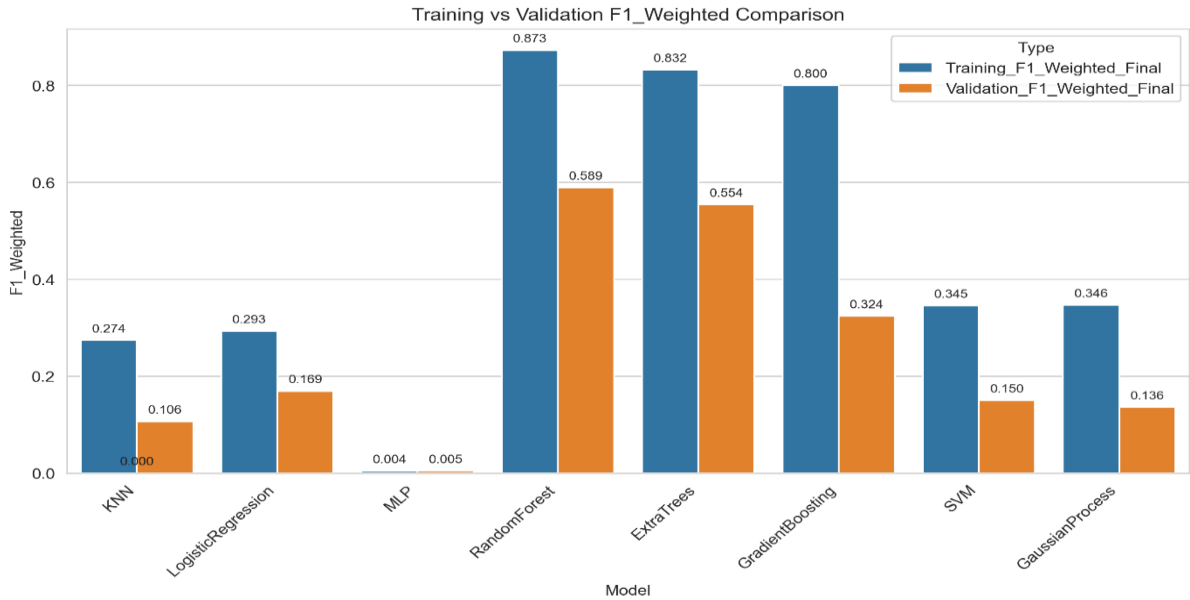

Figure 17. Training vs Validation F1-Weighted Comparison for Crack Localization Models

To ensure the validity and reliability of the developed models, careful measures were taken to prevent data leakage throughout the training and evaluation process. The dataset was properly split into training and validation sets prior to model development, ensuring that no information from the validation data was used during training. Additionally, all features were derived solely from the input variables without incorporating any target-related information. The permutation analysis further supports this, as a significant drop in model performance was observed when the feature structure was randomly disrupted, indicating that the models rely on meaningful relationships within the data rather than memorization or unintended data leakage. While these results strongly suggest that data leakage is unlikely, the adopted data handling and validation procedures provide additional confidence in the robustness and generalizability of the models.

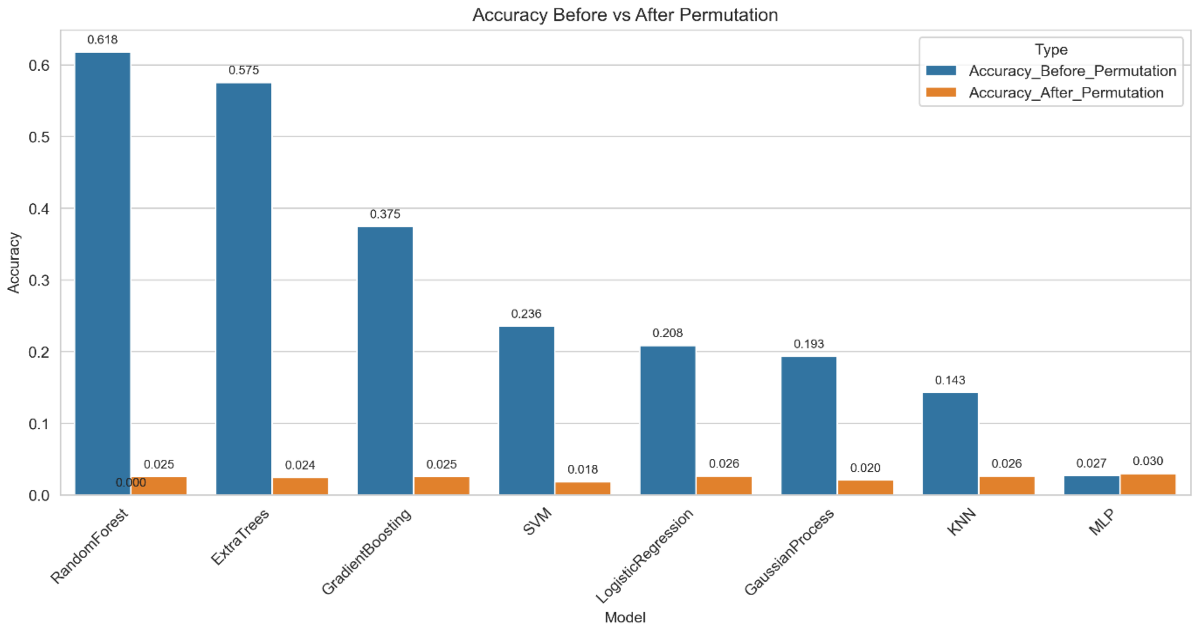

Figure 18. Impact of Feature Permutation on Accuracy for Crack Localization Models

The radar chart highlights clear performance differences among models in the crack localization task. Ensemble methods, particularly Extra Trees and Random Forest, achieve the highest overall performance, forming the largest areas across all metrics. Gradient Boosting shows moderate results, while KNN, SVM, and Logistic Regression demonstrate limited performance. Gaussian Process performs slightly better than these models but remains below ensemble methods. The MLP model exhibits the weakest performance across all metrics. Overall, the chart emphasizes that tree-based ensemble models are the most effective for crack localization, although performance remains lower compared to damage detection.

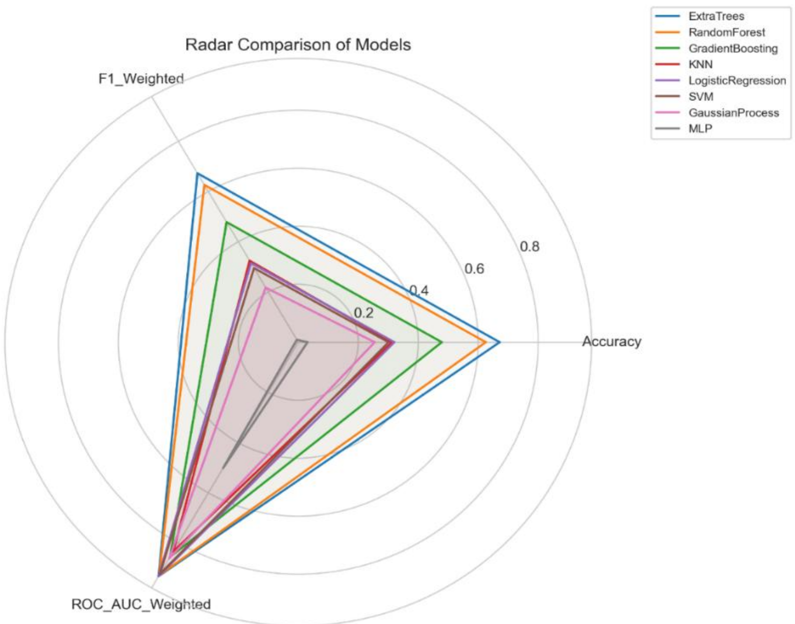

Figure 19. Radar Chart Comparison of Model Performance for Crack Localization Using Natural Frequencies (f_1–f_4)

➢ Crack Severity Estimation (Only Frequencies)

The results show that crack severity estimation using only natural frequencies is moderately successful, with notable differences between models. Extra Trees achieves the best overall performance, with the highest R^2 (0.916) and the lowest error values (RMSE ≈ 6.64, MAE ≈ 4.24), indicating strong predictive capability. KNN and Gradient Boosting also perform well, while MLP and Random Forest show slightly lower but acceptable accuracy. Linear Regression and Ridge exhibit weaker performance, suggesting that the relationship between frequencies and severity is not purely linear. SVM performs poorly, and Gaussian Process fails significantly, with negative R^2 and very high errors. Overall, the results indicate that while natural frequencies can be used for severity estimation, the task is more complex than damage detection, and ensemble models provide the most reliable predictions.

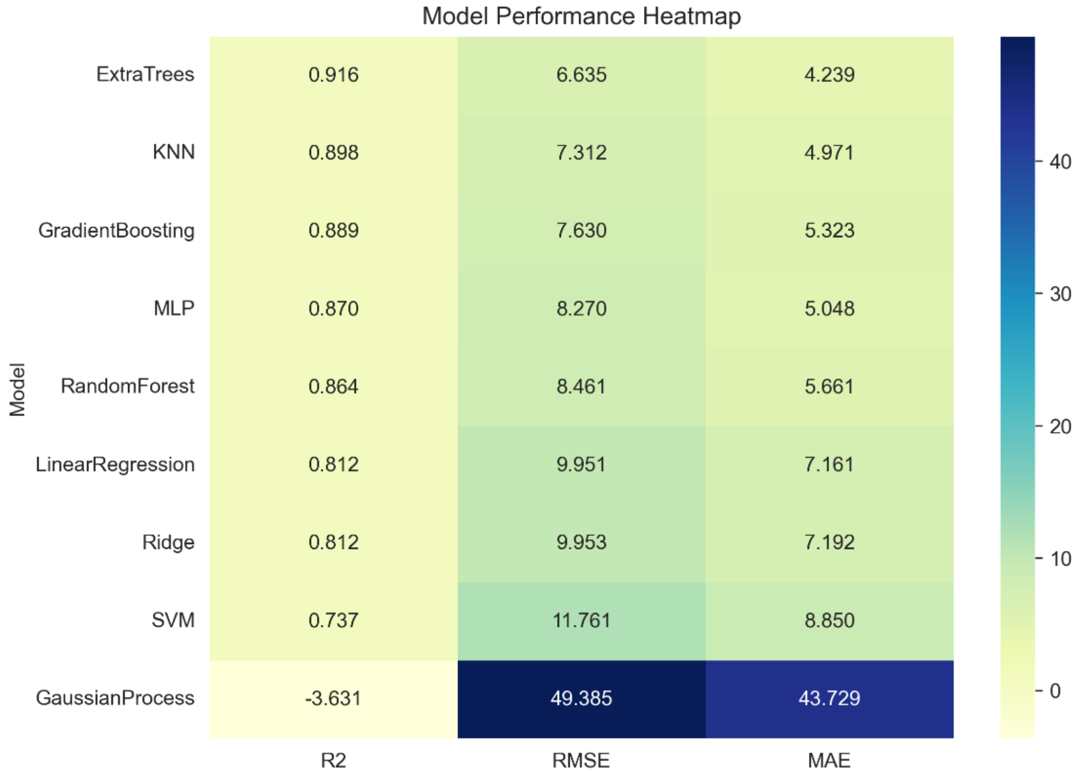

Figure 20. Heatmap of Model Performance for Crack Severity Estimation Using Natural Frequencies (f_1–f_4)

The comparison of training and validation R^2 values indicates that most models achieve strong predictive performance with relatively good generalization. Ensemble models such as Extra Trees, Random Forest, and Gradient Boosting show high R^2 values on both training and validation sets, although a slight drop suggests mild overfitting. Gaussian Process demonstrates the most consistent performance with nearly identical training and validation scores, indicating excellent generalization. KNN and MLP also perform well with moderate gaps, while Ridge and SVM show lower performance overall. These results suggest that severity estimation is feasible using natural frequencies, but some models exhibit limited generalization compared to their training performance.

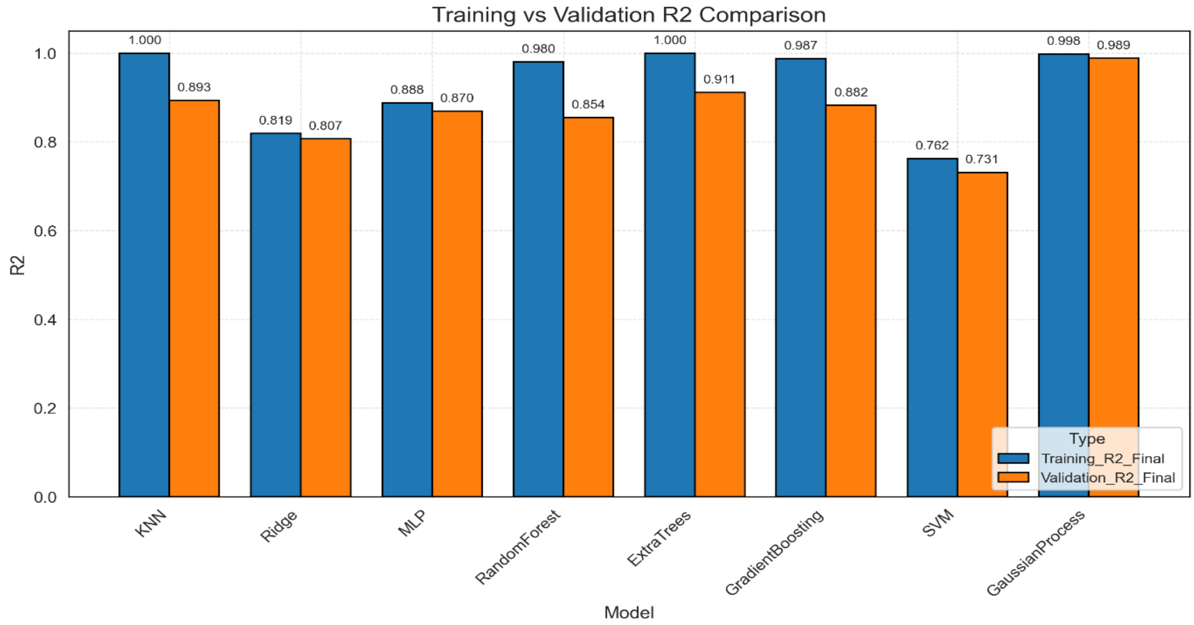

Figure 21. Training vs Validation R^2 Comparison for Crack Severity Estimation Models

The results show a dramatic decline in R^2 values after feature permutation, with all models dropping to near-zero or negative values. This indicates that the models lose their predictive capability when the relationship between input features and target values is disrupted. Before permutation, all models demonstrate strong performance, particularly Gaussian Process and ensemble methods. However, after permutation, negative R^2 values confirm that the models perform worse than a simple mean prediction. This clearly demonstrates that the natural frequency features contain essential information for severity estimation and that model predictions are strongly dependent on the original feature structure.

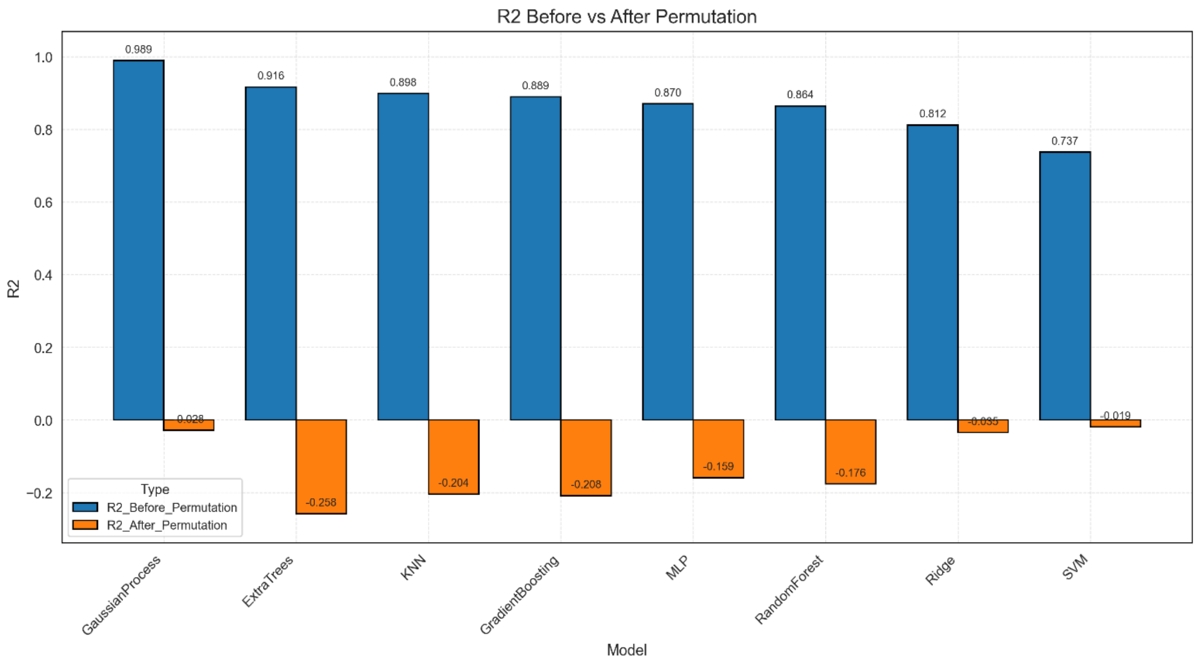

Figure 22. Impact of Feature Permutation on R^2 for Crack Severity Estimation Models

The scatter plots compare predicted versus true crack severity values for each model, illustrating their regression performance. Ensemble models such as Random Forest and Extra Trees show predictions closely aligned with the ideal diagonal line, indicating higher accuracy and consistency. Gradient Boosting and Gaussian Process also perform reasonably well, with moderate dispersion. In contrast, KNN, Ridge, and SVM exhibit larger scatter and deviation from the reference line, indicating lower predictive accuracy. The MLP model shows noticeable variability and less stable predictions. Overall, the plots confirm that ensemble models provide the most reliable severity estimation, while other models struggle to accurately capture the relationship between natural frequencies and crack severity.

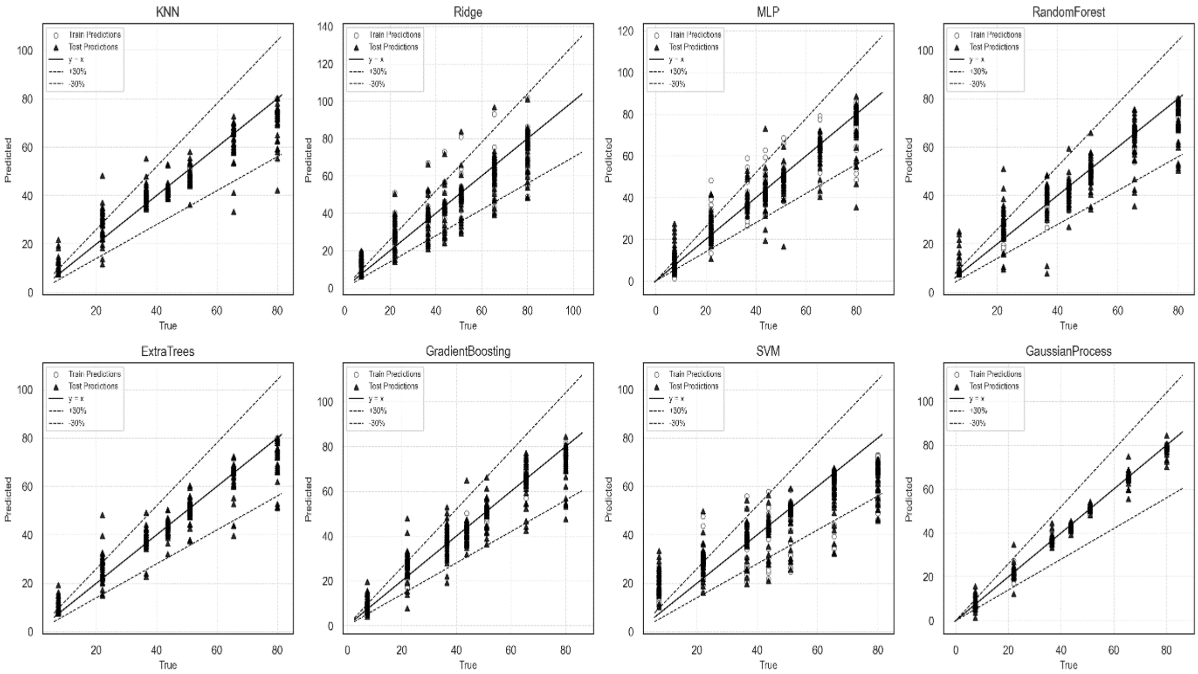

Figure 23. Predicted vs True Crack Severity for Different Machine Learning Models

- **Results and Discussion for Frequency Features**

The results demonstrate that natural frequency features (f_1–f_4) are highly effective for structural damage detection because changes in structural stiffness directly influence the global dynamic properties of the beam. Damage-induced stiffness reduction produces measurable shifts in the natural frequencies, allowing most machine learning models to achieve near-perfect classification performance with strong generalization capability. The clear separation between healthy and damaged states indicates that frequency features contain highly discriminative information for binary classification tasks.

However, the effectiveness of frequency-only features decreases for more complex tasks such as crack localization and severity estimation. Natural frequencies primarily describe global structural behavior and therefore provide limited spatial information regarding the exact crack position. As a result, different damage locations may generate similar frequency variations, reducing feature separability and increasing prediction ambiguity during localization tasks. This explains why localization accuracy remains only moderate despite the strong detection performance.

Ensemble learning methods, particularly Extra Trees and Random Forest, consistently outperform other approaches across all tasks because they effectively capture nonlinear feature interactions while remaining robust against noise, correlated inputs, and local data variations. Their averaging mechanisms also help reduce variance and improve generalization stability. Gradient Boosting models similarly achieve strong predictive performance because the sequential boosting strategy progressively corrects residual prediction errors. However, boosting-based methods exhibit a greater tendency toward overfitting, particularly for localization and regression tasks where the feature relationships become more complex.

In contrast, linear models such as Logistic Regression and Linear Regression show comparatively lower performance for localization and severity estimation because the relationship between crack characteristics and frequency variations is inherently nonlinear. Similarly, KNN models demonstrate fewer stable predictions because distance-based learning becomes sensitive to overlapping feature distributions and local clustering effects. MLP neural networks also do not consistently outperform ensemble methods, likely because the available dataset size is relatively limited for deep nonlinear learning and may not fully support stable neural-network generalization.

For crack severity estimation, predictive performance remains acceptable but exhibits larger variability and increased prediction errors compared with binary damage detection. Small changes in crack severity often produce relatively subtle frequency shifts, making regression tasks more sensitive to measurement noise, modeling uncertainty, and feature overlap. Although ensemble regression methods achieve the strongest performance, the limited sensitivity of global frequency features restricts precise damage quantification.

The permutation analysis further confirms that model performance strongly depends on the physical relevance and inherent structure of the frequency features. Significant performance degradation after feature permutation indicates that the models rely on meaningful structural information rather than random statistical correlations. Overall, the findings indicate that natural frequencies provide a reliable and computationally efficient basis for structural damage detection, whereas improved localization and severity estimation require additional damage-sensitive features such as mode shapes, frequency response functions, or time-frequency indicators that contain richer spatial and local structural information.

3.2.2. Machine Learning Algorithms on Mode Shape Features

In this section, machine learning models were trained using four transverse mode shapes (M_1–M_4). Each sample contained 28 features (4 modes × 28 measurement points).

The objective was to evaluate the effectiveness of high dimensional mode shape features for damage detection and quantification.

- **Damage Detection (Only Mode Shapes)**

The results (**Figure 24**) demonstrate that mode shape features provide exceptionally strong performance for damage detection. Several models, including Random Forest, SVM, Gradient Boosting, and Extra Trees, achieve perfect classification across all metrics (Accuracy, F1, and ROC-AUC = 1.000), indicating that mode shapes contain highly discriminative information. Logistic Regression and Gaussian Process also show near-perfect results, while KNN and MLP exhibit only slightly lower performance. Overall, these findings confirm that mode shape features are highly effective for damage detection, offering performance comparable to or slightly more stable than frequency-based features.

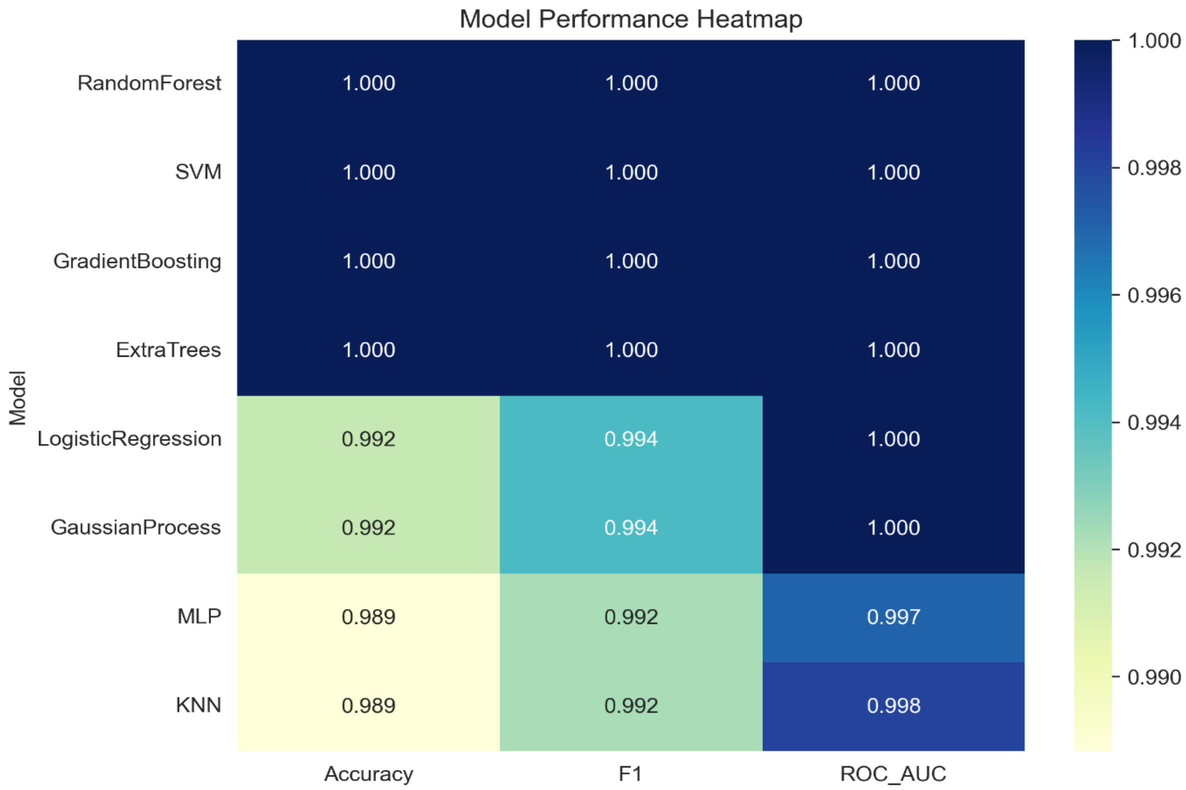

Figure 24. Heatmap of Model Performance for Damage Detection Using Mode Shape Features (M_1–M_4)

The comparison between training and validation F1-scores shows that most models achieve perfect or near-perfect performance with minimal differences between training and validation results. Ensemble models, SVM, and Random Forest exhibit identical scores (F1 = 1.000), indicating excellent generalization and no signs of overfitting. Logistic Regression, KNN, MLP, and Gaussian Process show only slight drops in validation performance, but remain very close to perfect. Overall, the results confirm that mode shape features enable highly robust and stable damage detection across all models.

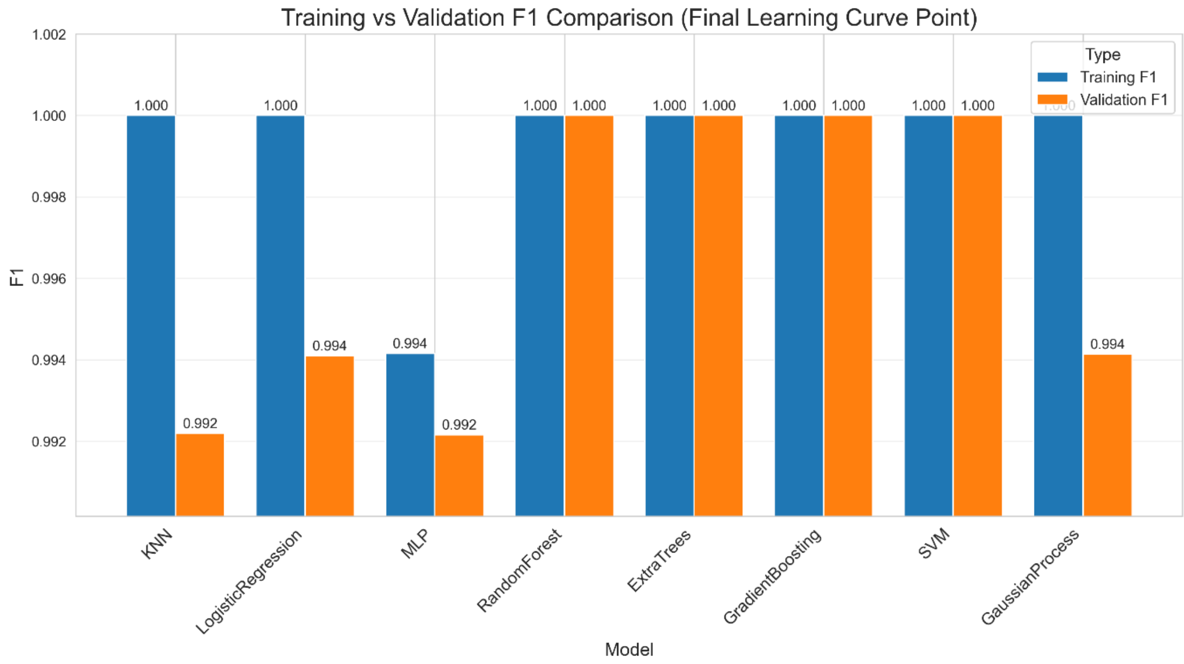

Figure 25. Training vs Validation F1-Score Comparison for Damage Detection Using Mode Shape Features

The results show a clear reduction in F1-score after feature permutation for all models, confirming that mode shape features contain meaningful and informative patterns for damage detection. Although performance decreases, most models still retain relatively high F1-scores (around 0.78–0.84), indicating some robustness of the models even when feature structure is disrupted. SVM and Logistic Regression experience the largest performance drops, suggesting higher sensitivity to feature relationships. Overall, the results demonstrate that model predictions strongly depend on the original mode shape features, further validating their importance for accurate damage detection.

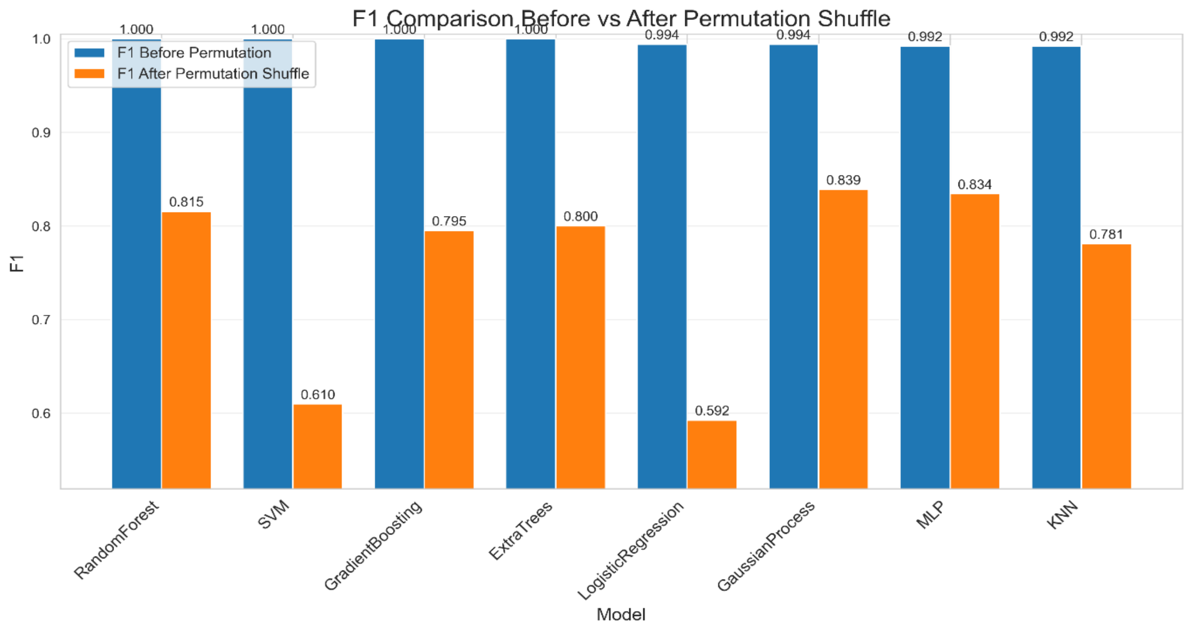

Figure 26. Impact of Feature Permutation on F1-Score for Damage Detection Using Mode Shape Features

The radar chart shows that all models achieve nearly identical and perfect performance across Accuracy, F1-score, and ROC-AUC. The overlapping shapes indicate that mode shape features enable consistently high classification performance regardless of the chosen algorithm. Unlike frequency-based features, there is almost no visible difference between models, highlighting the strong discriminative power of mode shapes. Overall, this confirms that mode shape features provide highly robust and model-independent performance for damage detection.

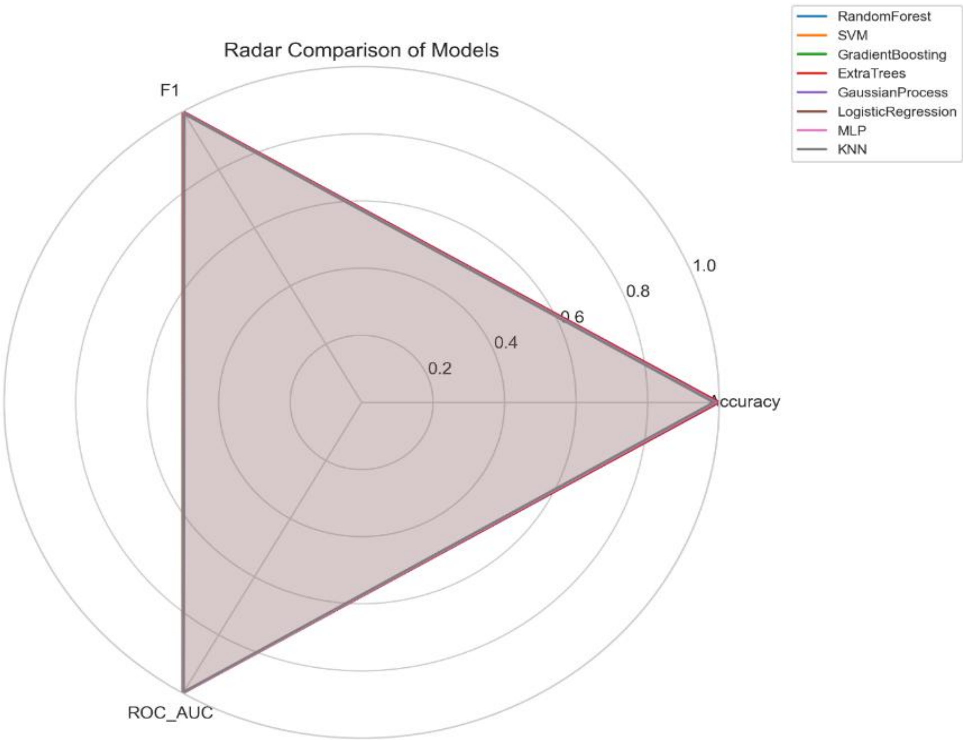

Figure 27. Radar Chart Comparison of Model Performance for Damage Detection Using Mode Shape Features

➢ Crack Localization Performance (Only Mode Shapes)

The results indicate that mode shape features improve crack localization performance compared to frequency-only features, although the task remains challenging. Tree-based ensemble models, particularly Random Forest and Extra Trees, achieve the best results with accuracies above 0.60 and high ROC-AUC values (~0.95), confirming their strong capability in capturing spatial damage information. Gradient Boosting shows moderate performance, while Gaussian Process and Logistic Regression provide limited accuracy. SVM and KNN perform poorly, and MLP exhibits very weak results. Overall, mode shape features enhance localization capability, but ensemble models remain the most effective approach.

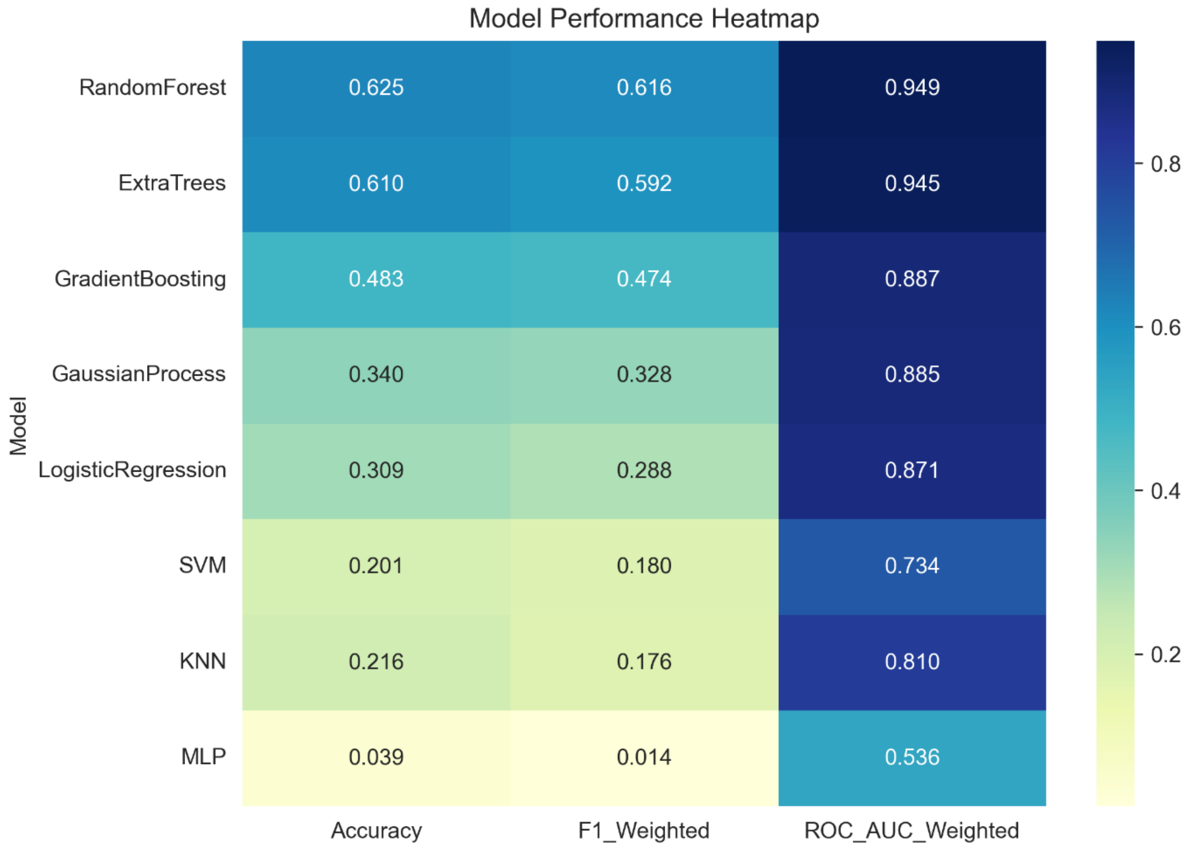

Figure 28. Heatmap of Model Performance for Crack Localization Using Mode Shape Features (M_1–M_4)

The comparison between training and validation F1-weighted scores reveals a noticeable performance gap for most models, indicating overfitting in the crack localization task. Ensemble models (Random Forest, Extra Trees, and Gradient Boosting) achieve high training performance but show a significant drop in validation scores, suggesting limited generalization. Simpler models such as KNN and SVM perform poorly on both training and validation sets, indicating underfitting. Gaussian Process and Logistic Regression show moderate performance, while MLP remains weak overall. These results confirm that, despite improvements from mode shape features, crack localization remains a challenging problem with reduced generalization capability.

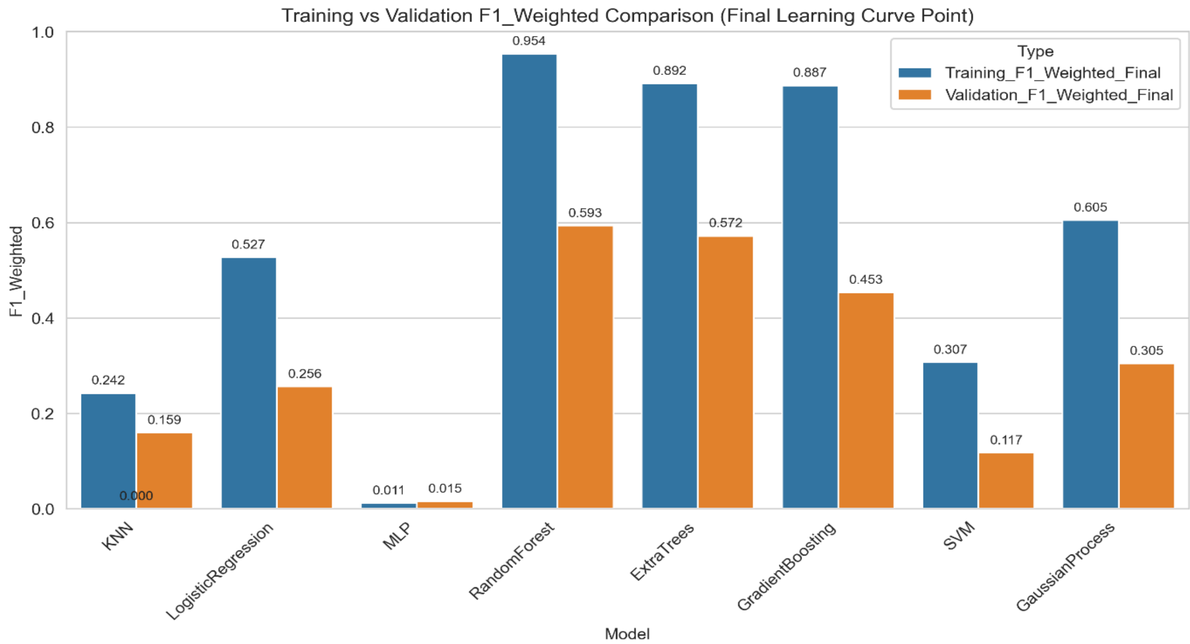

Figure 29. Training vs Validation F1-Weighted Comparison for Crack Localization Using Mode Shape Features

The results show a substantial decrease in accuracy after feature permutation for all models, with performance dropping to near-random levels (~0.02–0.03). This indicates that the models strongly depend on the original mode shape features for crack localization. Ensemble models (Random Forest and Extra Trees) achieve the highest accuracy before permutation but experience a complete loss of predictive capability after permutation. Overall, this confirms that mode shape features contain critical information for localization, although the task itself remains complex and challenging.

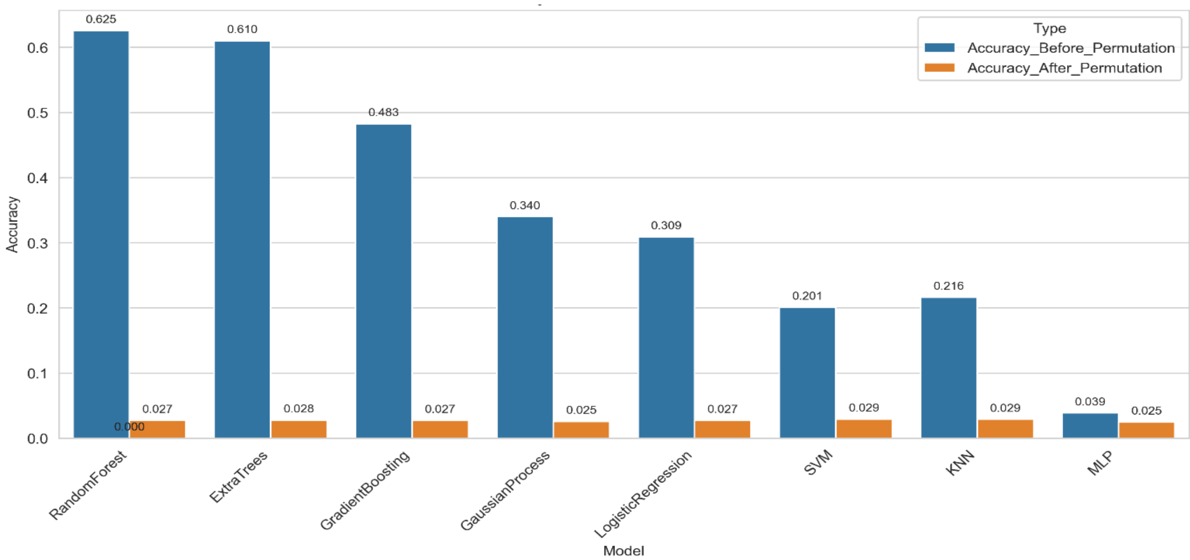

Figure 30. Impact of Feature Permutation on Accuracy for Crack Localization Using Mode Shape Features

The radar chart highlights clear performance differences among models in the crack localization task. Ensemble models, particularly Random Forest and Extra Trees, achieve the highest overall performance, forming the largest areas across all metrics. Gradient Boosting shows moderate performance, while Gaussian Process and Logistic Regression perform at a lower level. KNN and SVM exhibit weaker results, and MLP shows the lowest performance across all metrics. Overall, the chart emphasizes that although mode shape features improve localization, ensemble methods remain the most effective, while other models struggle to achieve comparable accuracy.

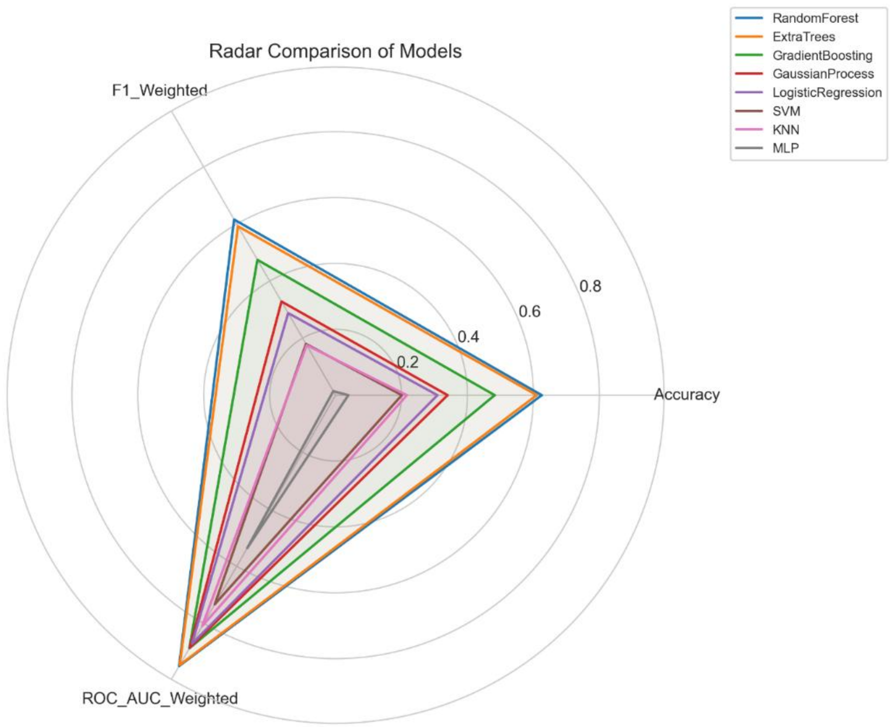

Figure 31. Radar Chart Comparison of Model Performance for Crack Localization Using Mode Shape Features

- **Crack Severity Estimation (Only Mode Shapes)**

The results indicate that crack severity estimation using only mode shape features is relatively challenging, with moderate predictive performance across models. Ensemble methods such as Random Forest, Extra Trees, and Gradient Boosting achieve the best results, with R^2 values around 0.46–0.50 and lower error metrics compared to other models. SVM, KNN, and Gaussian Process show weaker performance, while MLP and Ridge exhibit the poorest results with low R^2 and high error values. Overall, although mode shape features contain useful information for severity estimation, their effectiveness is limited, and the models struggle to achieve high accuracy compared to damage detection tasks.

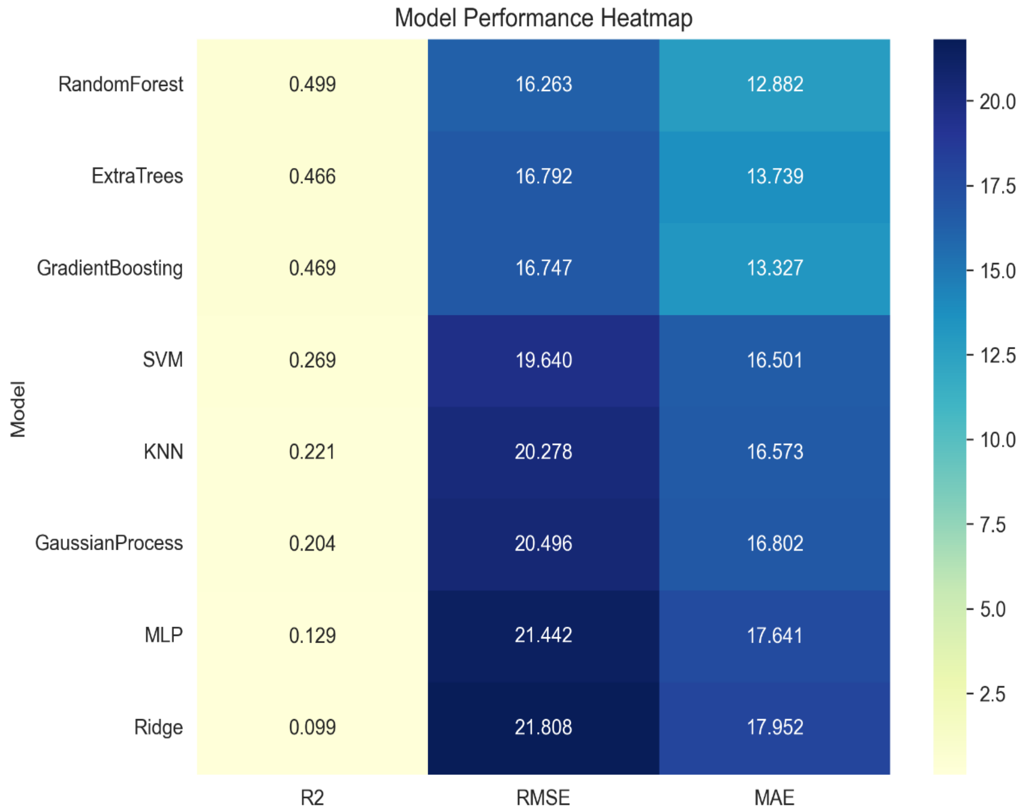

Figure 32. Heatmap of Model Performance for Crack Severity Estimation Using Mode Shape Features (M_1–M_4)

The comparison between training and validation R^2 values reveals a significant performance gap for most models, indicating strong overfitting in the crack severity estimation task. Ensemble models (Random Forest, Extra Trees, and Gradient Boosting) achieve high training R^2 values but show a considerable drop in validation performance, suggesting limited generalization. Gaussian Process demonstrates very high training performance but fails to generalize effectively. Simpler models such as Ridge and MLP show low performance on both training and validation sets, indicating underfitting. Overall, these results confirm that severity estimation using mode shape features is a challenging problem with reduced generalization capability compared to other tasks.

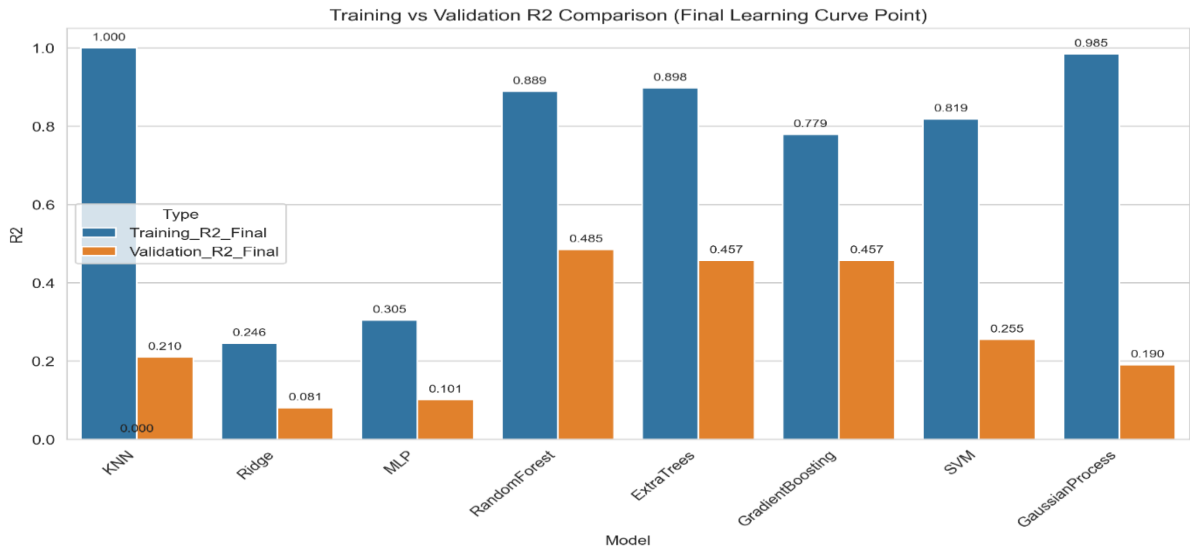

Figure 33. Training vs Validation R^2 Comparison for Crack Severity Estimation Using Mode Shape Features

The results show a clear drop in R^2 values after feature permutation, with all models falling to near-zero or negative values. This indicates that the predictive relationships between mode shape features and crack severity are completely disrupted when feature structure is randomized. Ensemble models initially achieve the highest R^2 values, but their performance collapses after permutation, similar to other models. The presence of negative R^2 values further confirms that the models perform worse than a baseline mean prediction when meaningful feature information is removed. Overall, this demonstrates that mode shape features contain essential information for severity estimation, although the task itself remains difficult.

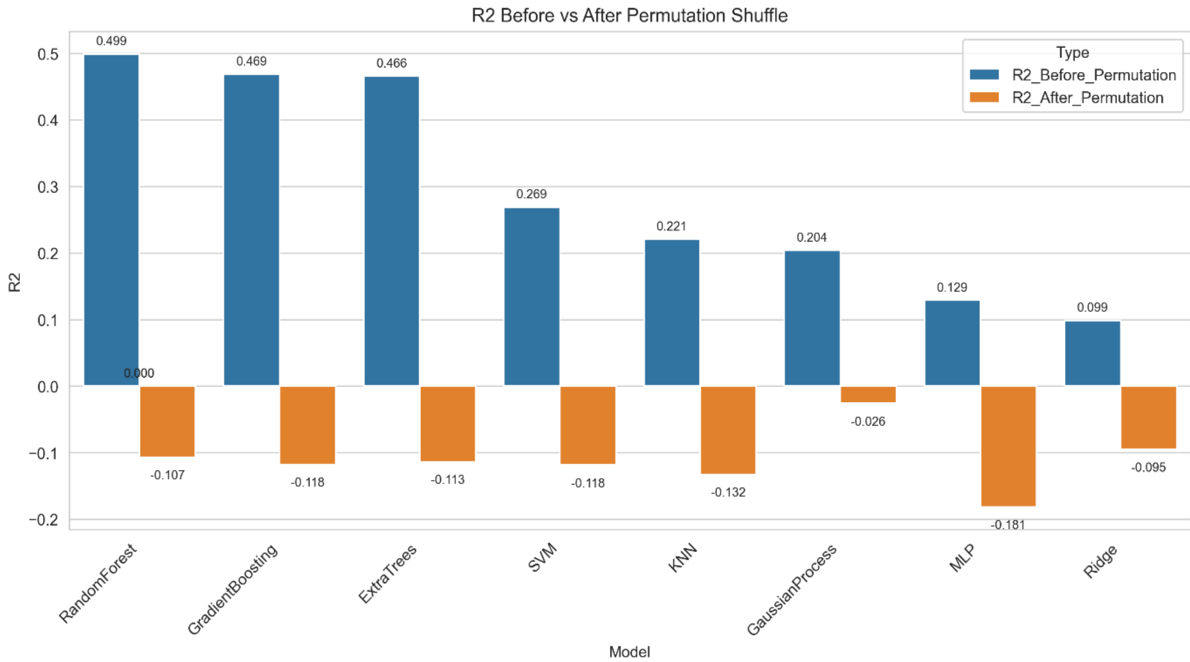

Figure 34. Impact of Feature Permutation on R^2 for Crack Severity Estimation Using Mode Shape Features

The scatter plots illustrate the relationship between predicted and true crack severity for each model. Ensemble models, particularly Random Forest and Extra Trees, show predictions that are more closely aligned with the ideal diagonal line, indicating better accuracy and consistency. Gradient Boosting also performs reasonably well, while Gaussian Process shows moderate agreement with some dispersion. In contrast, KNN, Ridge, and SVM exhibit larger scatter and deviation, reflecting weaker predictive capability. The MLP model demonstrates high variability and less reliable predictions. Overall, the results confirm that ensemble models provide the most accurate severity estimation when using mode shape features, although noticeable prediction errors remain across all models.

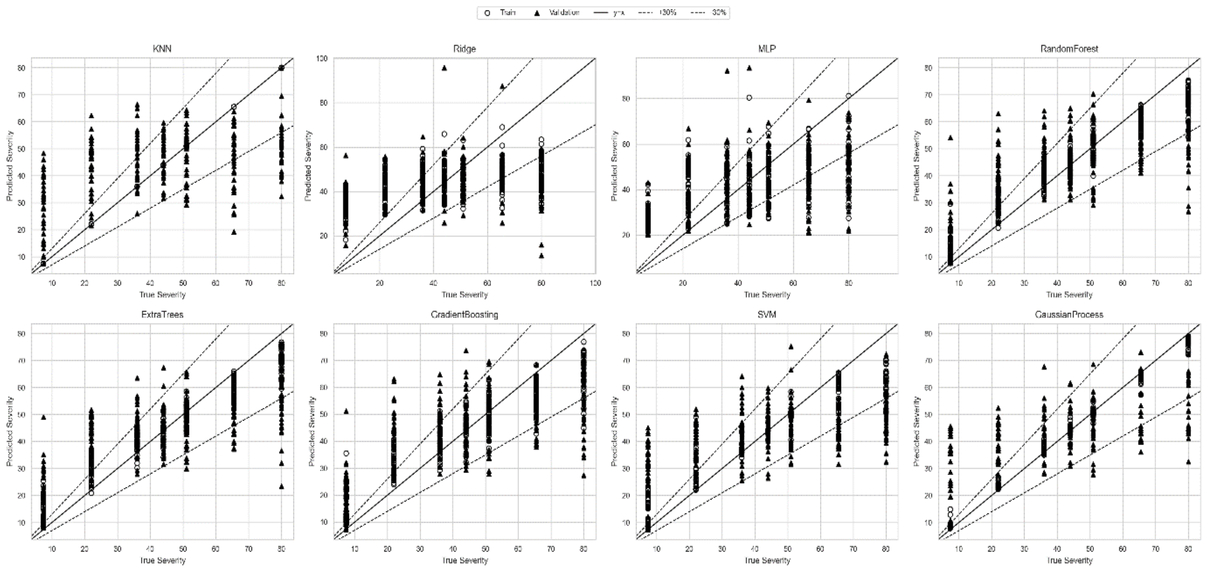

Figure 35. Predicted vs True Crack Severity Using Mode Shape Features for Different Models

- **Results and Discussion for Mode Shape Features**

The results demonstrate that mode shape features (M1–M4) provide highly informative and discriminative representations for structural damage assessment because they contain detailed spatial information associated with structural deformation behavior. Unlike natural frequencies, which mainly describe global dynamic characteristics, mode shapes capture localized changes in structural response, making them more sensitive to crack position and deformation patterns.

For damage detection, all machine learning models achieved near-perfect performance, indicating that the differences between healthy and damaged mode shapes are sufficiently distinguishable in the extracted feature space. Even relatively simple models achieved excellent results because the damage-induced modal changes produce strong feature separability. However, ensemble learning methods such as Random Forest and Extra Trees still provided the most stable and consistent performance due to their robustness against feature correlations and their ability to model nonlinear interactions between modal components. In contrast, linear models showed slightly reduced flexibility because the relationship between structural damage and modal deformation is not entirely linear.

Compared with frequency-only features, mode shapes significantly improve crack localization performance because they directly contain spatial deformation information related to crack position. Nevertheless, localization remains considerably more challenging than binary damage detection. Damage scenarios located at neighboring positions often generate similar modal deformation patterns, increasing overlap between feature distributions and reducing class separability. Ensemble models again achieved the strongest localization performance because tree-based learning methods can effectively partition complex nonlinear feature spaces and identify localized modal variations associated with crack position. Gradient Boosting also demonstrated strong predictive capability, although it exhibited a higher tendency toward overfitting because sequential boosting methods can become sensitive to noise and localized data variations. Distance-based methods such as KNN showed less stable performance because localization results depend strongly on local feature clustering and dataset distribution.

For crack severity estimation, predictive performance remains more limited, with only moderate (R^2) values and increased prediction errors compared with detection and localization tasks. This suggests that mode shapes alone are insufficient for precise damage quantification because changes in crack severity often produce relatively subtle variations in modal deformation patterns. Regression tasks are therefore more sensitive to noise, modeling uncertainty, and feature overlap than classification tasks. Ensemble regression methods still outperform linear and distance-based approaches because they better capture nonlinear relationships between modal characteristics and stiffness degradation. In contrast, linear regression models underperform because severity evolution is not linearly proportional to modal changes. Similarly, MLP neural networks do not consistently surpass ensemble models, likely because the available dataset size is relatively limited for deep nonlinear learning and may not fully support stable neural-network generalization.

The permutation analysis further confirms that model performance strongly depends on the inherent structure and physical relevance of the mode shape features. The substantial performance degradation observed after feature permutation demonstrates that the models rely on physically meaningful modal information rather than random correlations. Overall, mode shape features are highly effective for damage detection and substantially improve localization capability due to their spatial sensitivity. However, additional or hybrid features, such as frequency-domain indicators or time-frequency representations, are required to achieve more reliable and accurate crack severity estimation.

3.2.3. Machine learning Algorithms on Combined Frequency Domain features

In this section, the performance of the proposed machine learning models is evaluated using the combined feature set obtained from both frequency domain variations and modal shape variations. The objective is to investigate whether integrating these complementary sources of structural information enhances the accuracy and robustness of damage detection, localization, and severity estimation. Eight machine learning algorithms, including ensemble methods, kernel-based models, neural networks, and Gaussian process regression, are trained and tested on the fused feature space. The results are compared in terms of classification accuracy, F1 score, balanced accuracy, regression error metrics (MAE and R^2), and computational cost to assess the effectiveness of the hybrid feature representation.

- **Damage Detection**

The heatmap (**Figure 36**) demonstrates that combining frequency domain features results in consistently high and near-perfect damage detection performance across all models. Ensemble methods (Random Forest, Extra Trees, and Gradient Boosting) as well as SVM achieve perfect scores in all evaluation metrics, indicating their strong capability to exploit the enriched feature set. Other models such as KNN, Logistic Regression, and Gaussian Process also perform very close to optimal, with only minor reductions in accuracy and F1-score. The MLP model shows comparatively lower performance but still maintains high reliability.
Overall, the results suggest that combining frequency features enhances the separability between damaged and undamaged states, leading to improved model robustness and reduced sensitivity to algorithm choice.

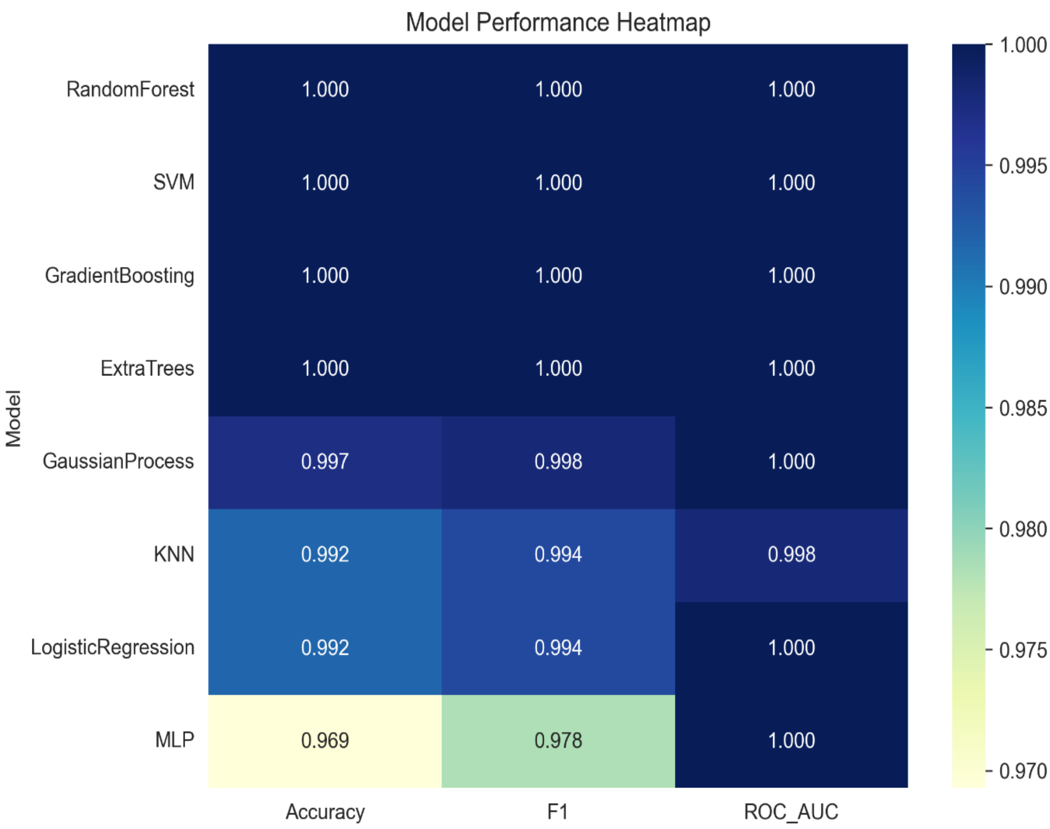

Figure 36. Damage Detection Performance (Combined Frequency Features)

The comparison shows that most models achieve identical or near-identical training and validation F1-scores, indicating excellent generalization and no signs of overfitting. Ensemble models, SVM, and Random Forest reach perfect performance (F1 = 1.000) on both datasets, demonstrating strong stability. KNN, Logistic Regression, MLP, and Gaussian Process show only minimal drops in validation scores, remaining very close to optimal. Overall, the results confirm that combined frequency features enable highly robust and consistent damage detection across different models.

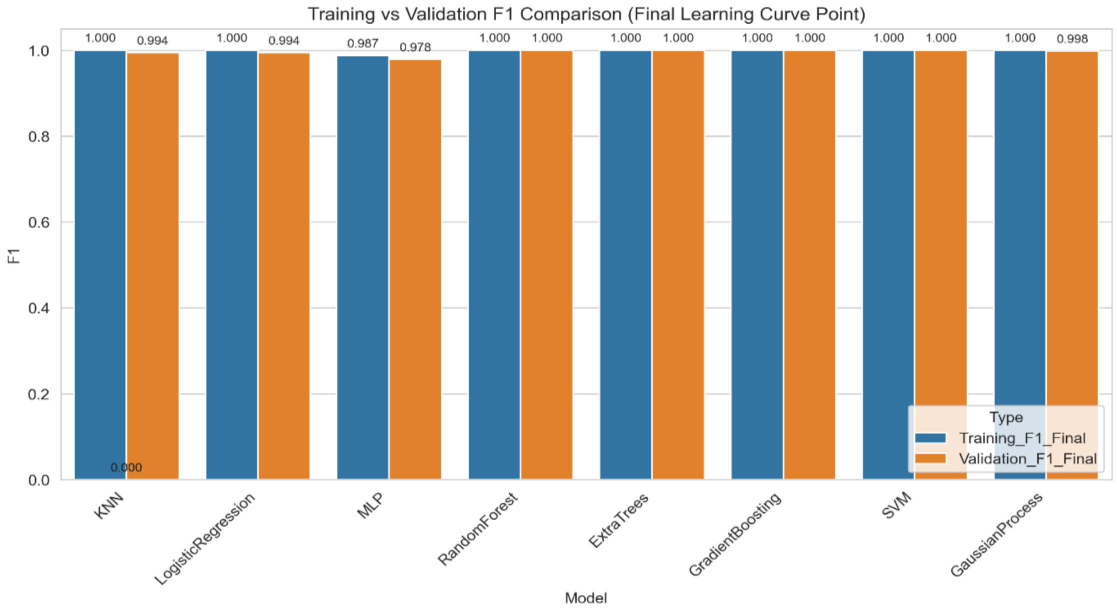

Figure 37. Training vs Validation F1 (Combined Frequency Features)

The results (**Figure 38**) show a clear reduction in F1-score after feature permutation for all models, confirming that the combined frequency features contain meaningful and informative patterns. Despite the drop, most models still retain relatively high performance (around 0.75–0.85), indicating some robustness to feature disruption.

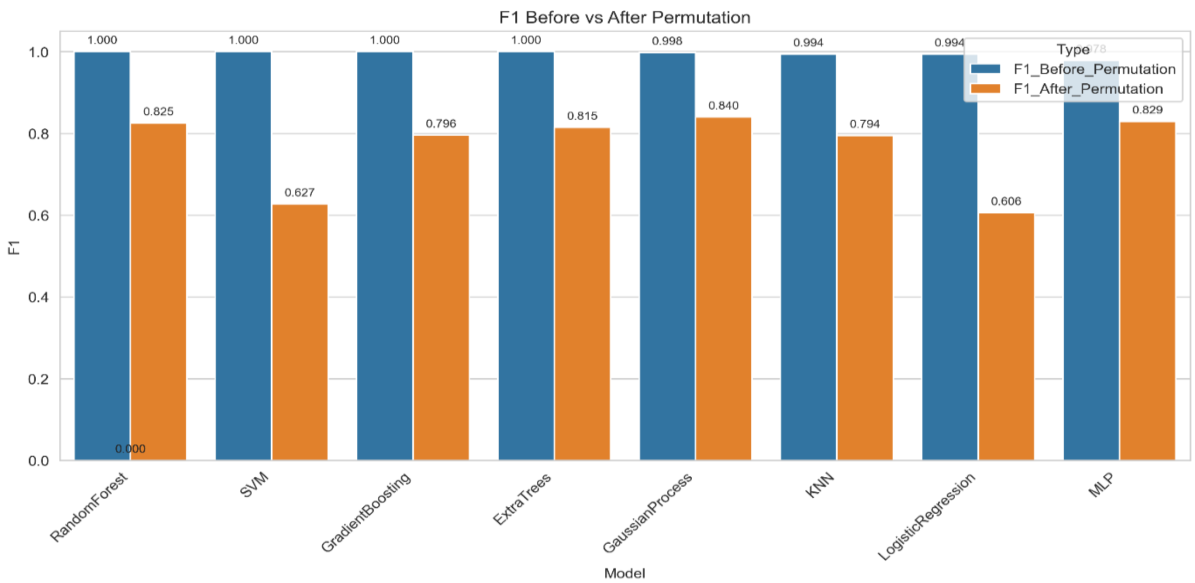

Figure 38. F1 Before vs After Permutation (Combined Frequency Features)

SVM and Logistic Regression exhibit the largest decreases, suggesting higher sensitivity to feature structure. Overall, the results demonstrate that model performance strongly depends on the original feature relationships, validating the importance of the combined frequency features.

The radar chart shows that all models achieve nearly identical and perfect performance across Accuracy, F1-score, and ROC-AUC. The complete overlap of the curves indicates that combined frequency features provide highly discriminative information, making the choice of model less critical. This consistency highlights the robustness of the feature set, as all algorithms are able to achieve optimal damage detection performance.

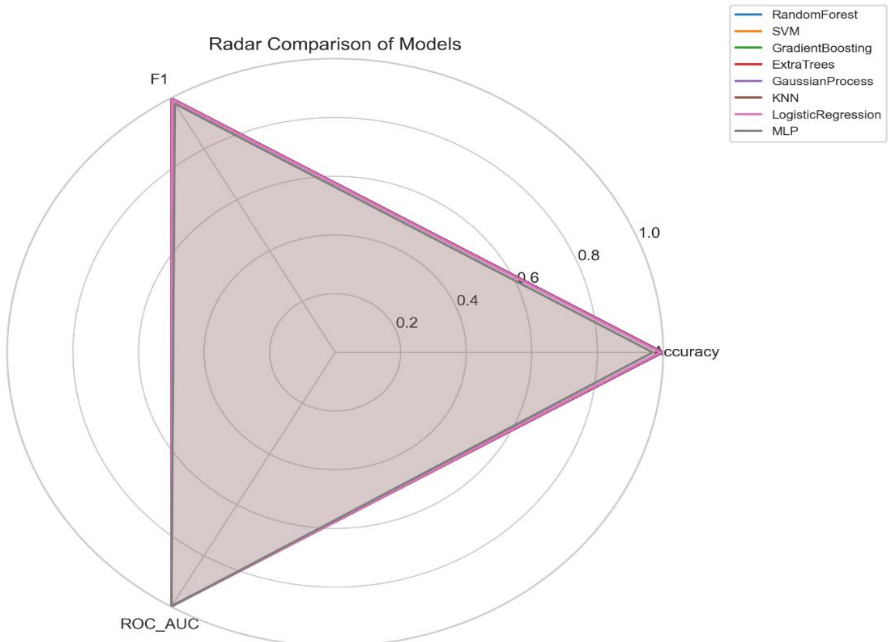

Figure 39. Model Comparison (Combined Frequency Features)

➢ Damage Localization

The heatmap shows that damage localization performance improves when using combined frequency features, but the task remains moderately challenging. Ensemble models, particularly Random Forest and Extra Trees, achieve the best results with accuracies around 0.65 and high ROC-AUC values (~0.95), indicating strong discrimination capability. Gradient Boosting provides moderate performance, while Logistic Regression, SVM, and KNN show lower accuracy. Gaussian Process performs similarly but with reduced consistency. The MLP model exhibits poor performance across all metrics. Overall, combined frequency features enhance localization compared to simpler features, but accuracy remains limited, with ensemble methods providing the most reliable results.

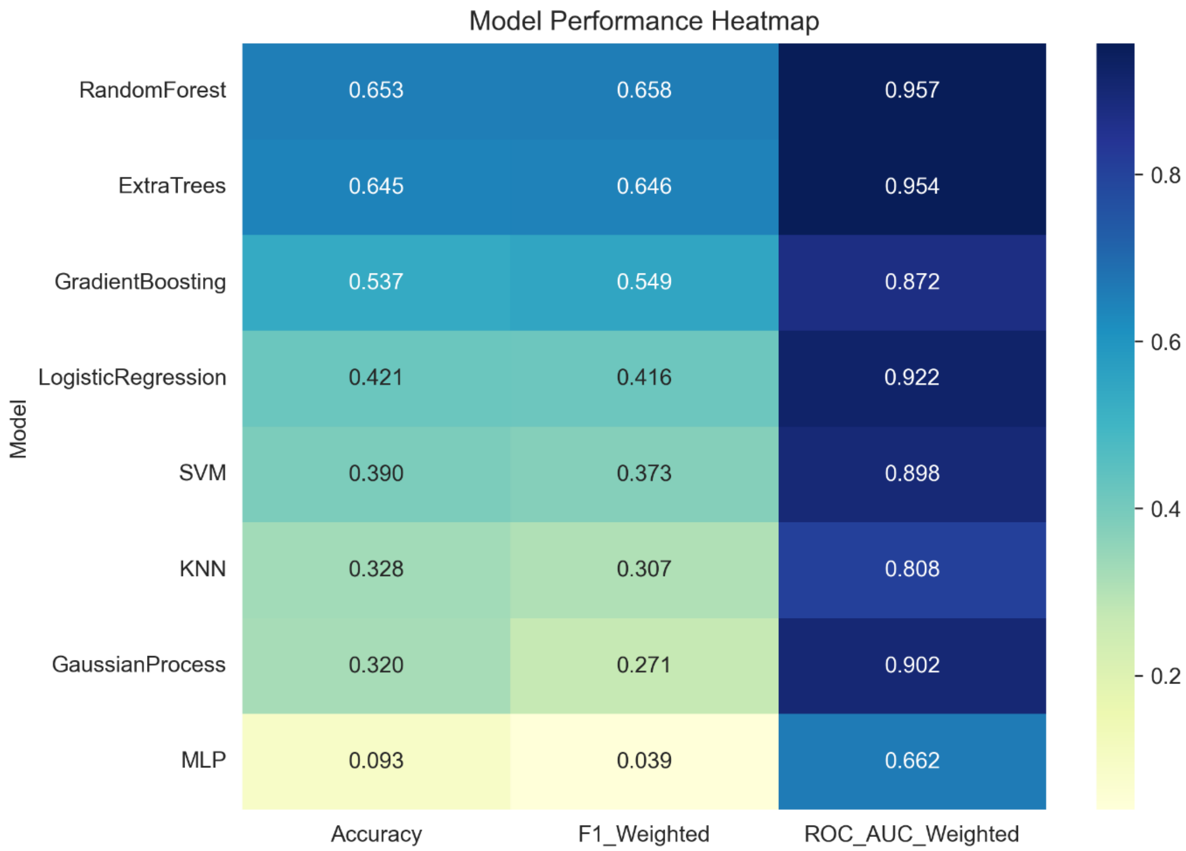

Figure 40. Damage Localization Performance (Combined Frequency Features)

The comparison shows a significant gap between training and validation F1-scores for most models, indicating overfitting in the localization task. Ensemble models (Random Forest, Extra Trees, and Gradient Boosting) achieve perfect training performance but show notable drops in validation scores, suggesting limited generalization. Logistic Regression and SVM exhibit moderate performance, while KNN and Gaussian Process perform poorly. The MLP model shows very low performance on both datasets, indicating underfitting. Overall, despite improvements from combined features, damage localization remains a challenging task with reduced generalization capability.

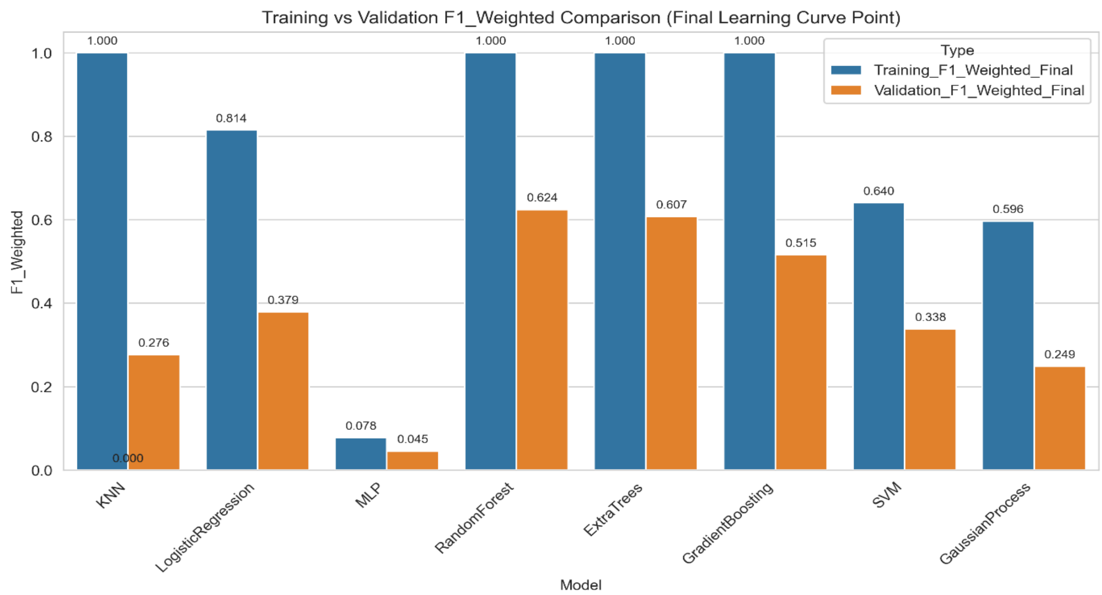

Figure 41. Training vs Validation F1 (Localization – Combined Frequency Features)

The results show a sharp drop in accuracy after feature permutation for all models, with values falling to near-random levels (~0.02–0.03). This indicates that the models strongly depend on the original combined frequency features for damage localization. Ensemble models achieve the highest accuracy before permutation but lose nearly all predictive capability after permutation. Overall, this confirms that the features contain essential information, although localization remains a complex and sensitive task.

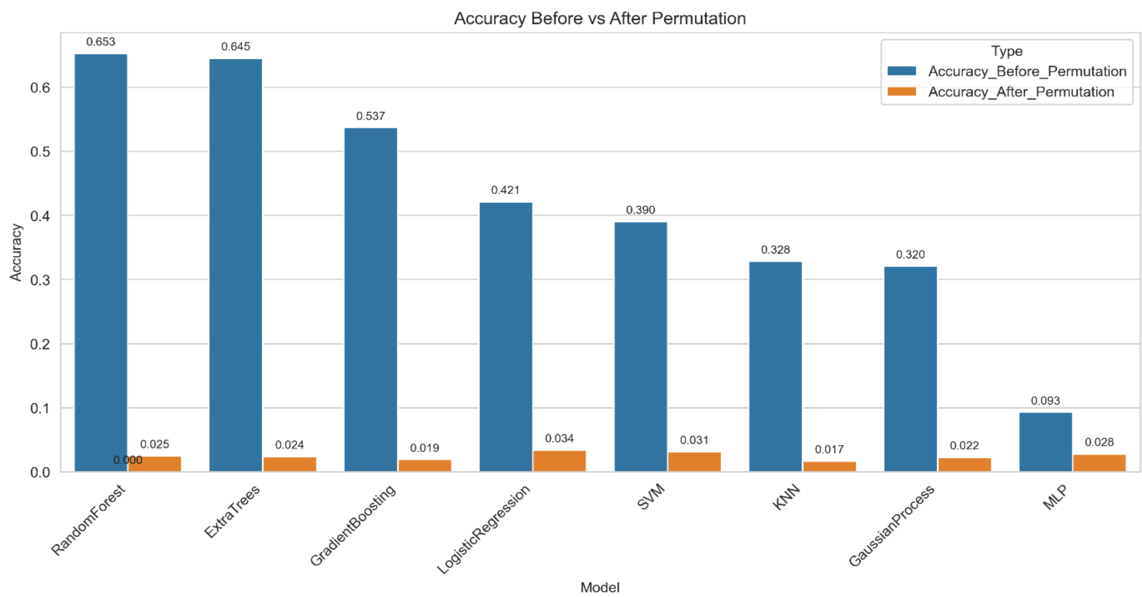

Figure 42. Accuracy Before vs After Permutation (Localization – Combined Frequency Features)

The radar chart highlights clear performance differences among models in the localization task. Ensemble models, particularly Random Forest and Extra Trees, achieve the highest overall performance, forming the largest areas across all metrics. Gradient Boosting shows moderate results, while Logistic Regression and SVM perform at a lower level. KNN and Gaussian Process exhibit weaker performance, and MLP shows the poorest results. Overall, the chart confirms that combined frequency features improve localization, but performance still varies significantly across models, with ensemble methods remaining the most effective.

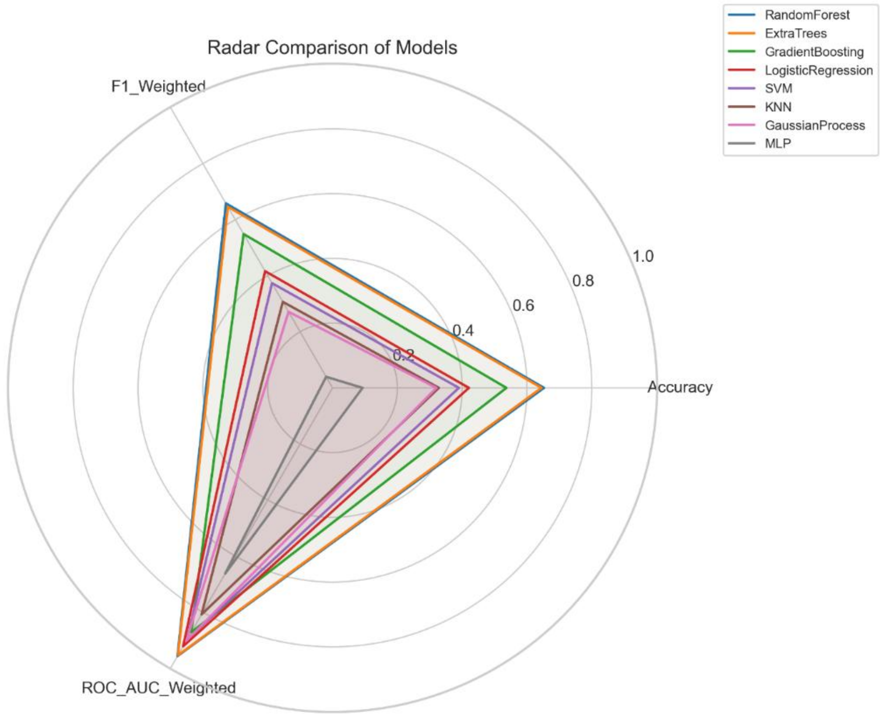

Figure 43. Model Comparison (Localization – Combined Frequency Features)

➢ Damage Severity

The heatmap shows that crack severity estimation improves significantly when using combined frequency features. Ensemble models, particularly Extra Trees and Gradient Boosting, achieve the best performance with high R^2 values (~0.88–0.90) and lower error metrics, indicating strong predictive capability. MLP and Random Forest also perform well, while SVM and Ridge show moderate results. Gaussian Process provides mixed performance with higher errors, and KNN performs the worst with low R^2 and very high error values. Overall, combined frequency features enhance severity estimation, with ensemble models offering the most accurate and reliable predictions.

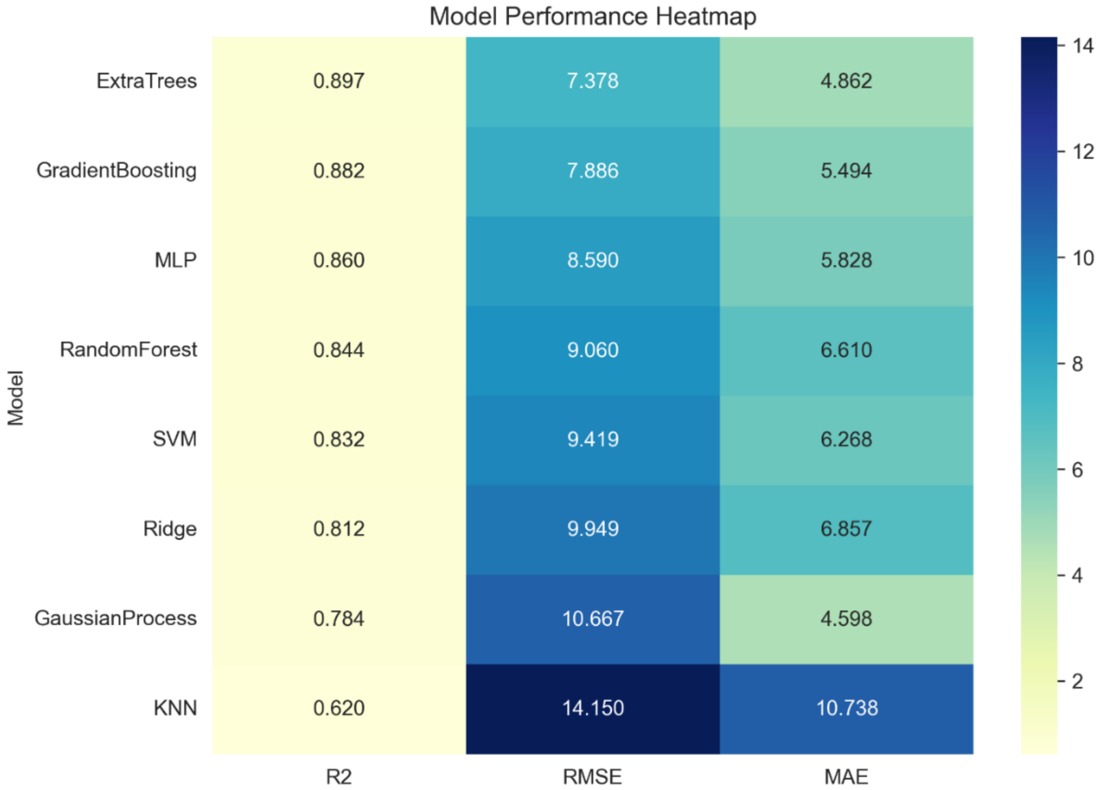

Figure 44. Severity Estimation Performance (Combined Frequency Features)

The comparison shows that most models achieve high R^2 values on both training and validation sets, indicating strong predictive performance with improved generalization compared to previous feature sets. Ensemble models (Extra Trees, Random Forest, and Gradient Boosting) maintain high validation scores with relatively small gaps, suggesting good stability. SVM, Ridge, and Gaussian Process also perform well with moderate differences. KNN shows the largest gap, indicating overfitting, while MLP demonstrates slightly lower but consistent performance. Overall, combined frequency features significantly improve severity estimation and generalization across models.

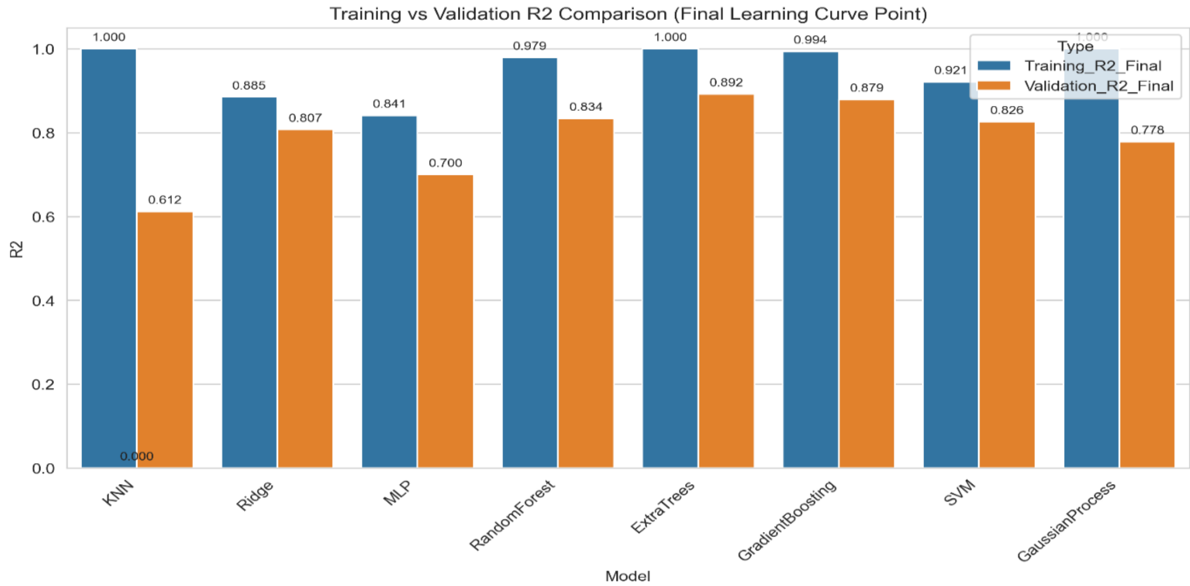

Figure 45. Training vs Validation R^2 (Severity – Combined Frequency Features)

The results show a substantial increase in MAE after feature permutation for all models, indicating a significant loss in prediction accuracy. Before permutation, ensemble models achieve the lowest errors, while KNN shows the highest MAE. After permutation, errors increase dramatically across all models, with Gaussian Process showing the most severe degradation. This confirms that the models strongly depend on the original combined frequency features for accurate severity estimation and that the feature structure contains essential predictive information.

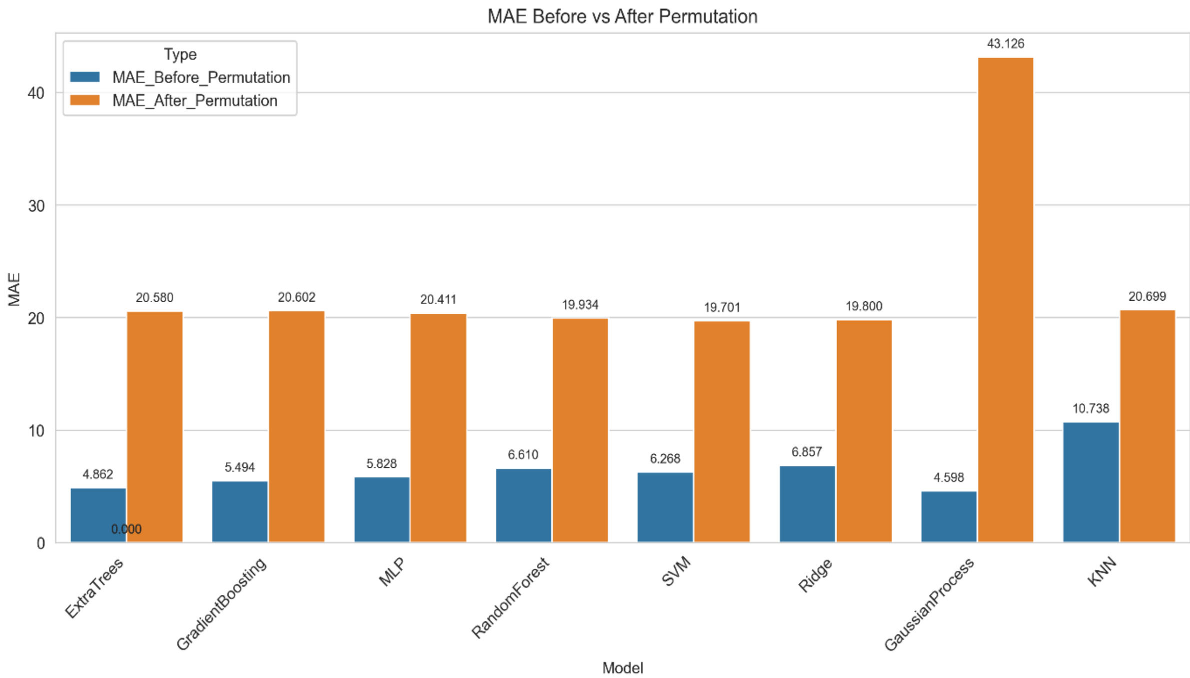

Figure 46. MAE Before vs After Permutation (Severity – Combined Frequency Features)

The scatter plots show that ensemble models, particularly Random Forest and Extra Trees, produce predictions that closely follow the ideal diagonal line, indicating higher accuracy and consistency. Gradient Boosting and SVM also show reasonable alignment with moderate dispersion. In contrast, KNN and Ridge exhibit larger scatter, reflecting lower accuracy, while Gaussian Process shows instability with significant deviations. MLP demonstrates moderate performance with noticeable variability. Overall, the results confirm that combined frequency features improve severity estimation, with ensemble models providing the most reliable predictions.

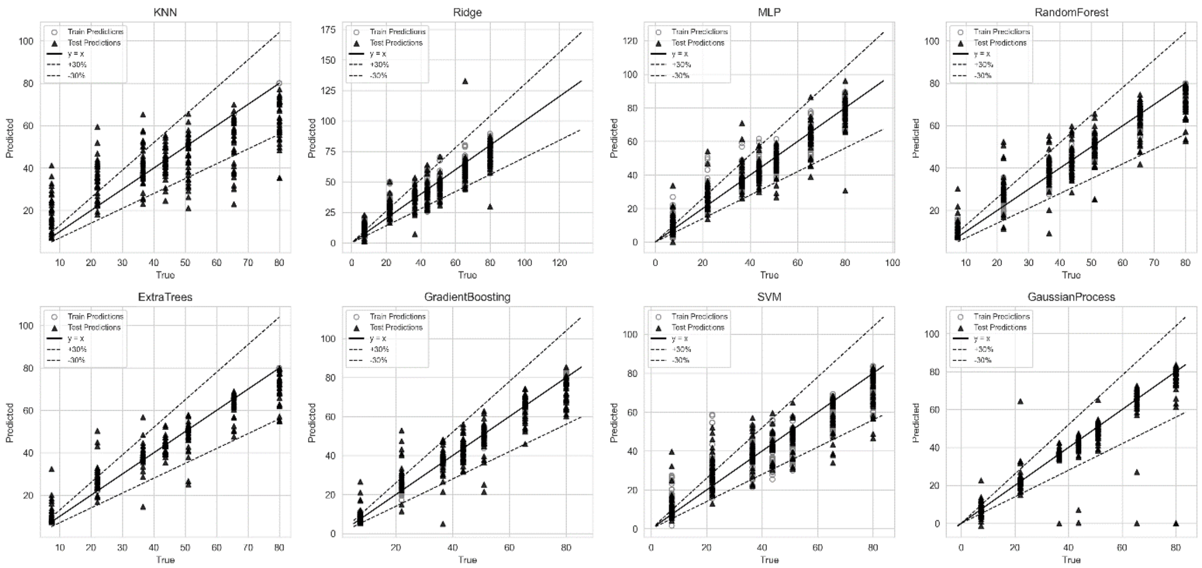

Figure 47. Predicted vs True Severity (Combined Frequency Features)

- **Results and Discussion of Combined features (Frequencies and Mode shapes)**

The results demonstrate that combining frequency-domain features significantly enhances model performance across all structural damage assessment tasks by providing complementary information about the structural dynamic response. The integration of natural frequencies, mode shapes, and other frequency-based indicators improves the representation of both global and local structural behavior, allowing the models to capture more complex damage-sensitive patterns than individual feature sets alone.

For damage detection, all machine learning models achieved near-perfect classification performance with strong generalization capability. This indicates that the combined features produce clear separability between healthy and damaged structural states. Linear models such as Logistic Regression already achieve very high accuracy because the distinction between healthy and damaged conditions becomes sufficiently separable in the enriched feature space. However, ensemble learning methods, particularly Random Forest and Extra Trees, consistently provide the most stable and robust performance due to their ability to model nonlinear feature interactions while remaining resistant to noise and feature correlations. In contrast, models such as KNN show slightly less stable behavior because distance-based methods are more sensitive to feature distribution and local variations within the dataset.

For damage localization, the combined features substantially improve predictive capability compared with individual feature sets because they contain additional spatial and modal information related to crack position. Nevertheless, localization remains significantly more challenging than binary damage detection. Damage scenarios located at nearby positions often generate similar vibration responses, reducing class separability and increasing prediction ambiguity. Ensemble methods again outperform other approaches because tree-based models can better capture complex nonlinear relationships between modal characteristics and crack location. Gradient Boosting also performs strongly due to its sequential error-correction mechanism, although it exhibits a greater tendency toward overfitting because boosting models are highly sensitive to local variations and noise in the training data. Linear models show comparatively lower localization accuracy because the relationship between crack position and vibration response is inherently nonlinear and cannot be fully represented using linear decision boundaries.

For crack severity estimation, the combined features produce the largest improvement among all tasks, achieving higher (R^2) values and lower prediction errors than individual feature groups. This indicates that combining multiple vibration characteristics increases sensitivity to stiffness degradation and crack propagation effects. However, regression performance remains lower than classification performance because severity estimation requires continuous prediction of damage magnitude rather than discrete state separation. Small changes in crack severity may generate only subtle variations in the vibration response, making regression tasks more sensitive to noise, modeling uncertainty, and feature overlap. Ensemble regression methods again provide the strongest performance because they effectively capture nonlinear relationships and reduce variance through averaging mechanisms. In contrast, linear regression models underperform because damage severity evolution is not strictly linear with respect to modal parameters. Similarly, MLP neural networks do not consistently outperform ensemble methods, likely because the available dataset size is relatively limited for deep nonlinear learning, increasing the risk of unstable training and overfitting.

The permutation analysis further confirms that model performance strongly depends on the inherent structure and physical relevance of the combined frequency-domain features. Significant performance degradation after feature permutation indicates that the models rely on meaningful structural information rather than random correlations. Overall, the combined feature set provides the most robust and consistent performance among all investigated feature groups, while ensemble learning methods consistently achieve superior results due to their robustness, nonlinear modeling capability, resistance to overfitting, and ability to handle complex feature interactions in vibration-based structural health monitoring applications.

3.2.4. Machine learning Algorithms on Hybrid features (Time and Frequency Domain Features)

In this section, machine learning models were trained using a combination of time-domain and frequency-domain features. The time-domain features include RMS, peak value, and crest factor, while the frequency-domain features consist of natural frequencies and mode shapes. The objective was to evaluate whether integrating complementary information from both domains can improve the performance of damage detection, localization, and severity estimation. The results show that hybrid features provide the most comprehensive representation of structural behavior, capturing both global dynamic characteristics (frequencies and mode shapes) and local signal variations (time-domain features). This combination leads to improved performance across all tasks compared to using individual feature sets. For damage detection, models achieve consistently perfect or near-perfect performance with strong generalization. In damage localization, hybrid features enhance accuracy and reduce uncertainty, although the task remains challenging. For crack severity estimation, the inclusion of time-domain features significantly improves predictive capability, resulting in higher R^2 values and lower error metrics. Overall, hybrid features offer the most robust and reliable framework for structural damage assessment, with ensemble models continuing to provide the best overall performance.

- **Damage Detection**

The heatmap shows that hybrid features (time and frequency domain) achieve near-perfect damage detection performance across all models. Ensemble methods (Random Forest, Extra Trees, and Gradient Boosting) and SVM reach perfect scores in all metrics, indicating exceptional discriminative capability.

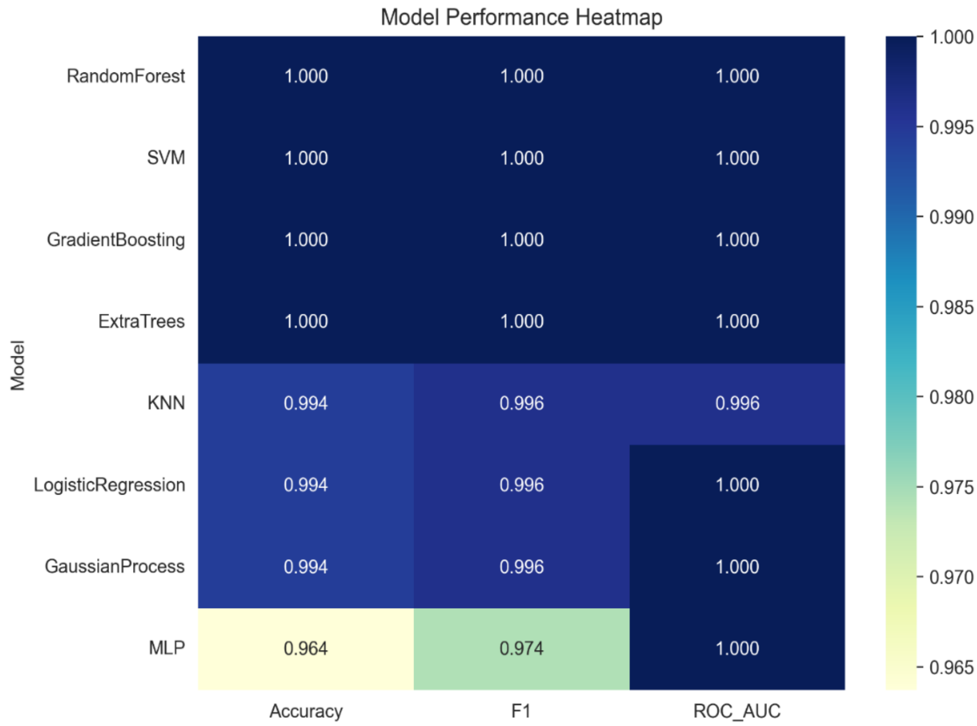

Figure 48. Damage Detection Performance (Hybrid Features)

KNN, Logistic Regression, and Gaussian Process also perform almost perfectly, with only minor deviations. MLP shows slightly lower accuracy and F1-score but still maintains very high performance. Overall, the results confirm that hybrid features provide the most powerful and robust representation for damage detection, making model choice less critical.

The comparison shows that all models achieve nearly identical training and validation F1-scores, indicating excellent generalization and no signs of overfitting. Ensemble models, SVM, and Random Forest reach perfect performance on both datasets, while other models show only minimal reductions in validation scores. MLP exhibits slightly lower values but remains highly accurate. Overall, the results confirm that hybrid features enable extremely robust and stable damage detection across all models.

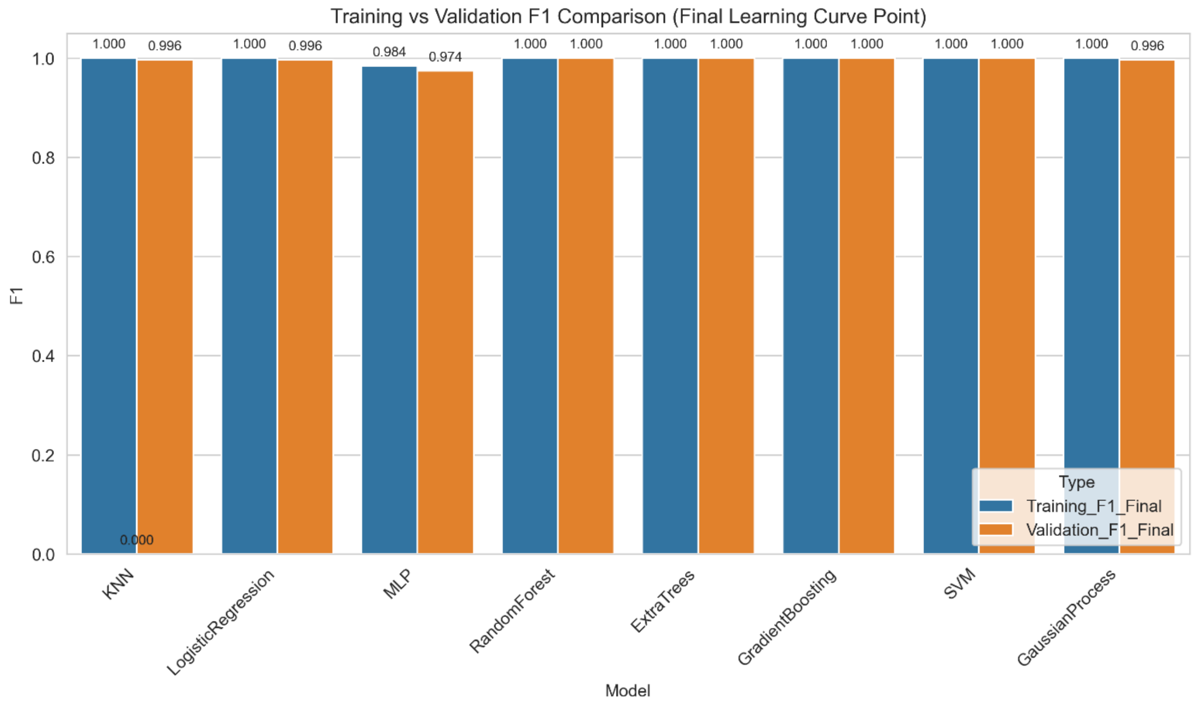

Figure 49. Training vs Validation F1 (Detection – Hybrid Features)

The results show a noticeable decrease in F1-score after feature permutation for all models, confirming that hybrid features contain meaningful and essential information for damage detection. Despite the drop, most models still maintain relatively high performance, indicating robustness to feature disruption. SVM and Logistic Regression experience the largest reductions, showing higher sensitivity to feature structure. Overall, the results validate that model performance strongly depends on the combined time and frequency features.

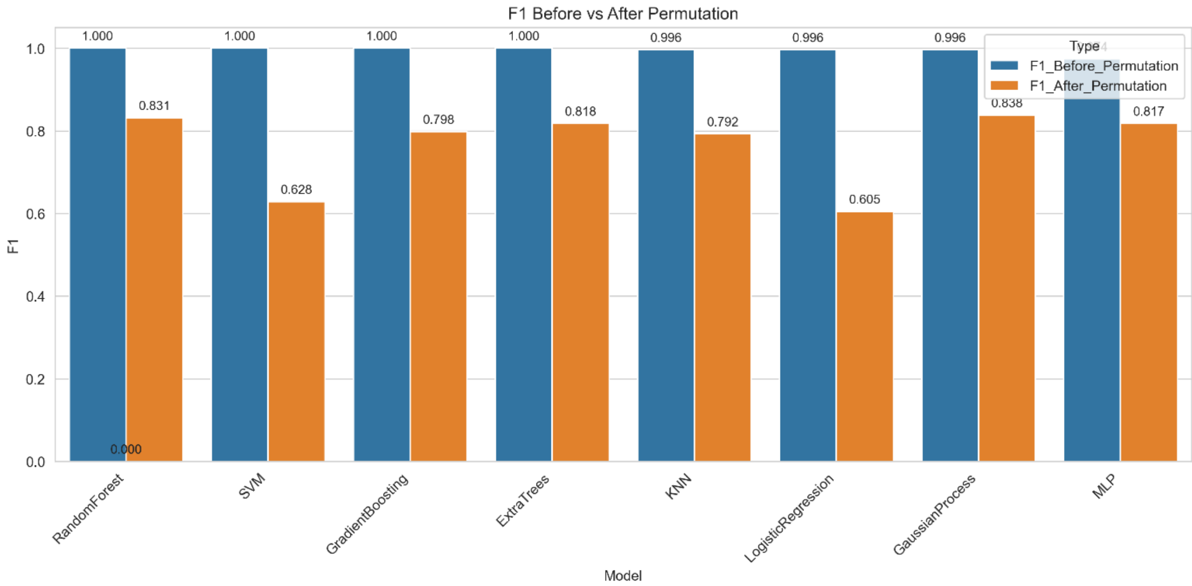

Figure 50. F1 Before vs After Permutation (Detection – Hybrid Features)

The radar chart shows that all models achieve nearly identical and perfect performance across Accuracy, F1-score, and ROC-AUC. The complete overlap of the curves indicates that hybrid features provide extremely strong discriminative power, making the choice of model less critical. This highlights the robustness and effectiveness of combining time and frequency features for damage detection.

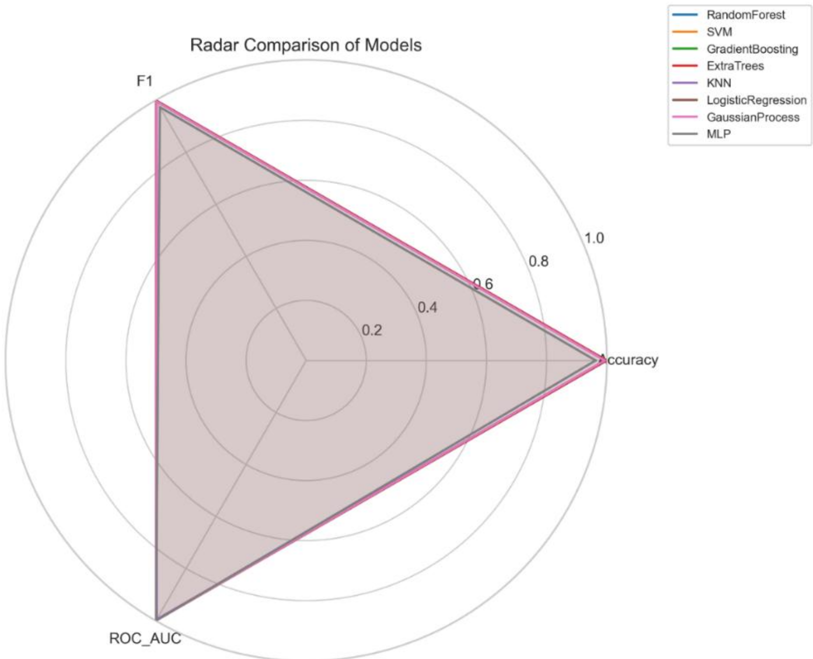

Figure 51. Model Comparison (Detection – Hybrid Features)

➢ **Damage Location**

The heatmap shows that hybrid features improve crack localization performance compared to previous feature sets, but the task remains challenging. Ensemble models, particularly Random Forest and Extra Trees, achieve the best results with the highest accuracy (~0.60–0.67) and lowest error values. Gradient Boosting shows moderate performance, while MLP and KNN perform at a lower level. Linear models and SVM exhibit weaker results, and Gaussian Process performs the worst with high errors and low accuracy. Overall, hybrid features enhance localization capability, especially for ensemble methods, but significant prediction errors still remain.

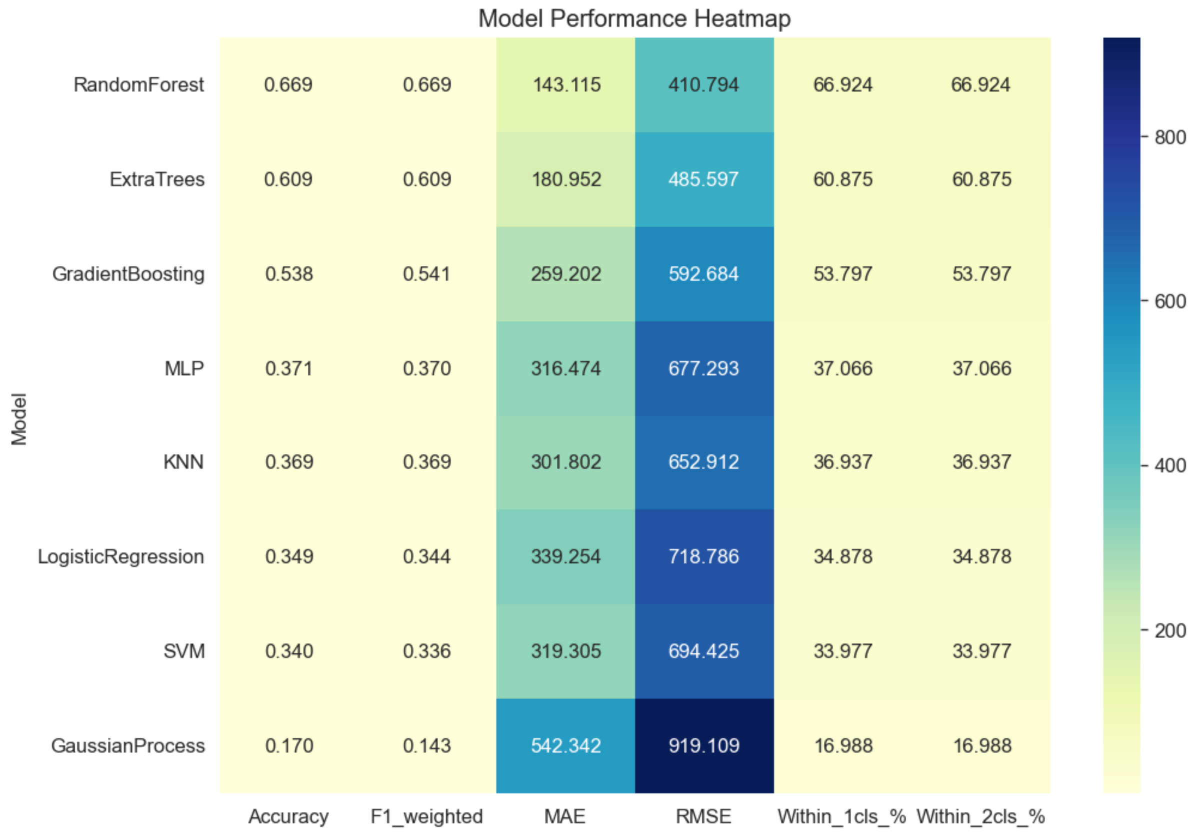

Figure 52. Localization Performance (Hybrid Features)

The comparison shows a clear gap between training and validation F1-scores for most models, indicating overfitting in the localization task. Ensemble models (Random Forest, Extra Trees, and Gradient Boosting) achieve high training performance but show noticeable drops in validation scores, suggesting limited generalization. MLP performs moderately, while KNN, SVM, Logistic Regression, and Gaussian Process exhibit lower performance on both datasets. Overall, despite improvements from hybrid features, damage localization remains a challenging problem with reduced generalization capability.

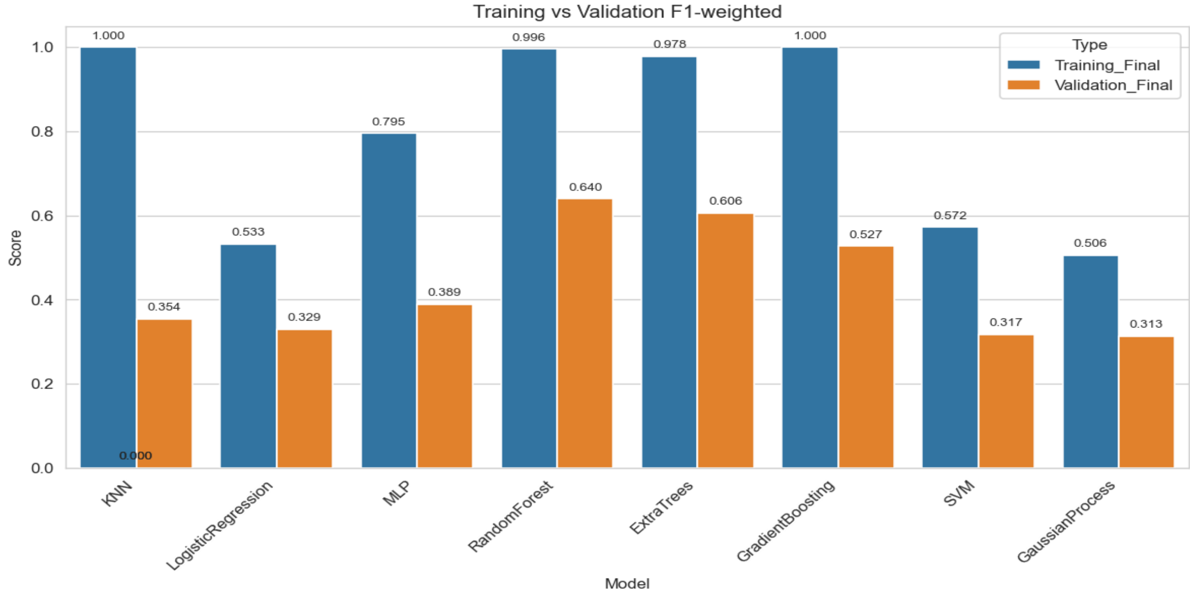

Figure 53. Training vs Validation F1 (Localization – Hybrid Features)

The results show a significant drop in accuracy after feature permutation for all models, with values falling to near-random levels (0.02–0.04). This indicates that the models strongly rely on the original hybrid features for damage localization. Ensemble models achieve the highest accuracy before permutation but lose most of their predictive capability after permutation. Overall, this confirms that the combined time and frequency features contain essential information, although localization remains a complex and sensitive task.

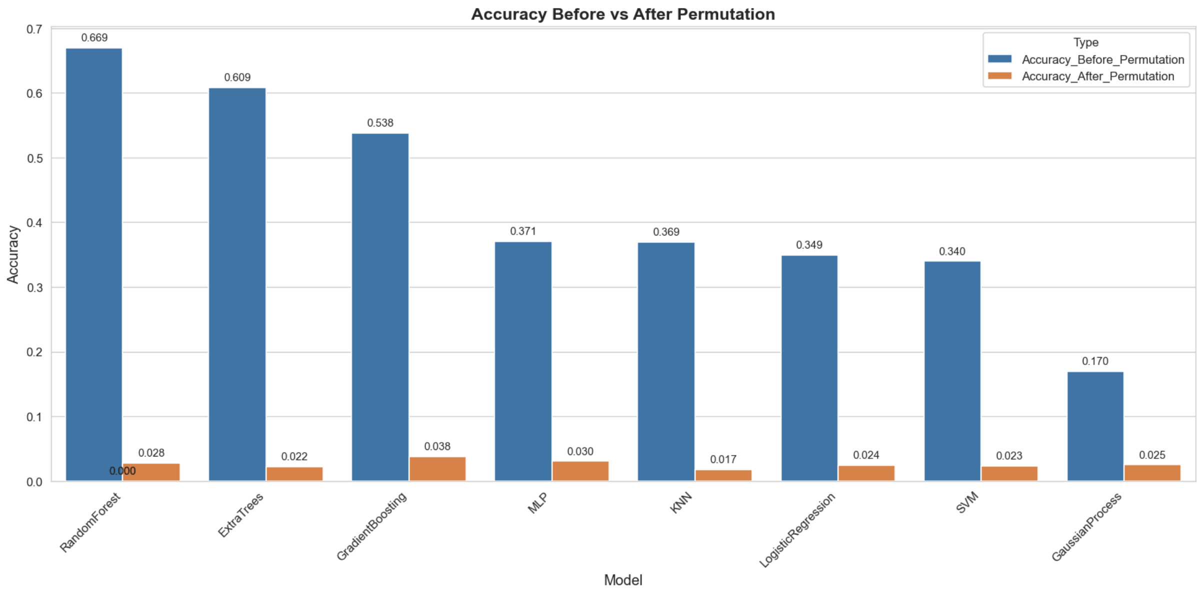

Figure 54. Accuracy Before vs After Permutation (Localization – Hybrid Features)

The confusion matrices show that ensemble models, particularly Random Forest and Extra Trees, achieve the most concentrated diagonal patterns, indicating more accurate localization predictions. Gradient Boosting also performs relatively well with a clear diagonal trend. In contrast, KNN, Logistic Regression, and SVM exhibit more scattered predictions, reflecting higher misclassification across locations. Gaussian Process shows the most dispersed pattern, indicating poor localization performance. Overall, hybrid features improve class separation, but significant off-diagonal elements confirm that crack localization remains a challenging task.

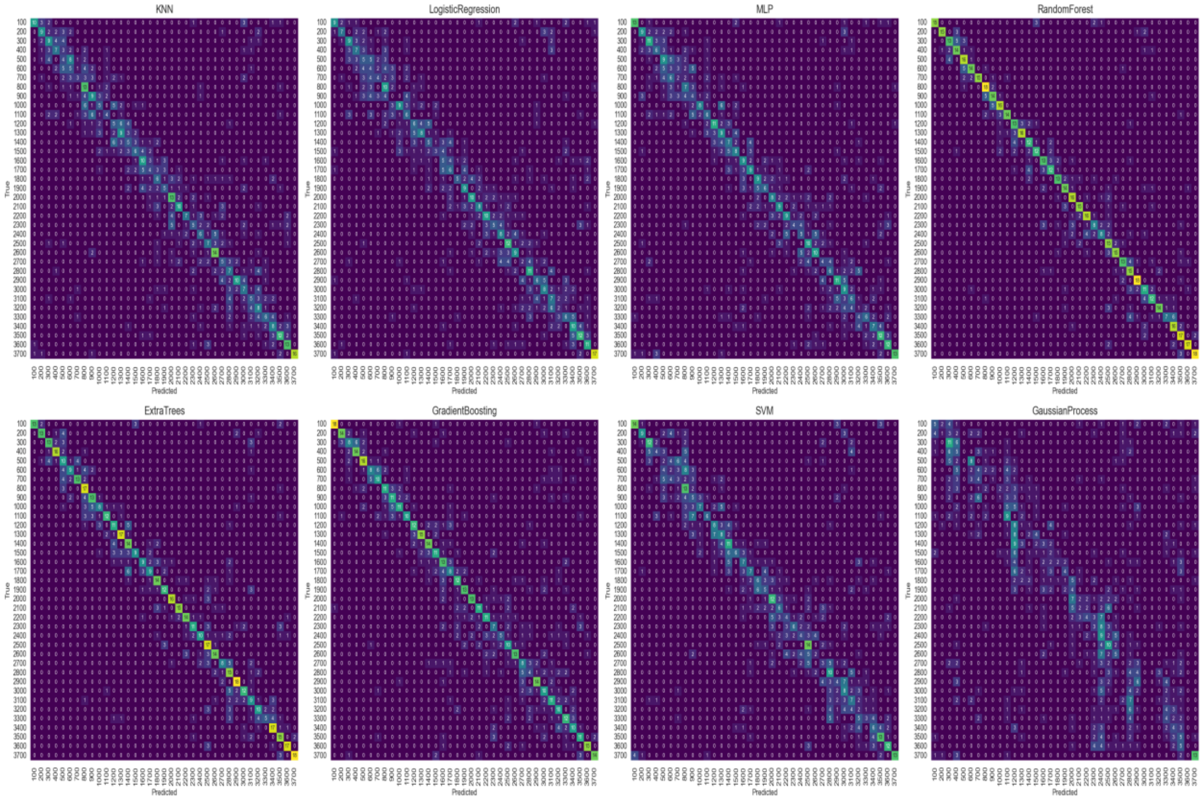

Figure 55. Confusion Matrix Comparison (Localization – Hybrid Features)

The scatter plots show that ensemble models, particularly Random Forest and Extra Trees, produce predictions that are more closely aligned with the ideal diagonal line, indicating better localization accuracy. Gradient Boosting also demonstrates reasonable performance with moderate dispersion. In contrast, KNN, Logistic Regression, and SVM show wider scatter, reflecting higher prediction errors. Gaussian Process exhibits the most variability and least reliable predictions. Overall, hybrid features improve localization trends, but noticeable deviations from the ideal line confirm that precise crack location prediction remains challenging.

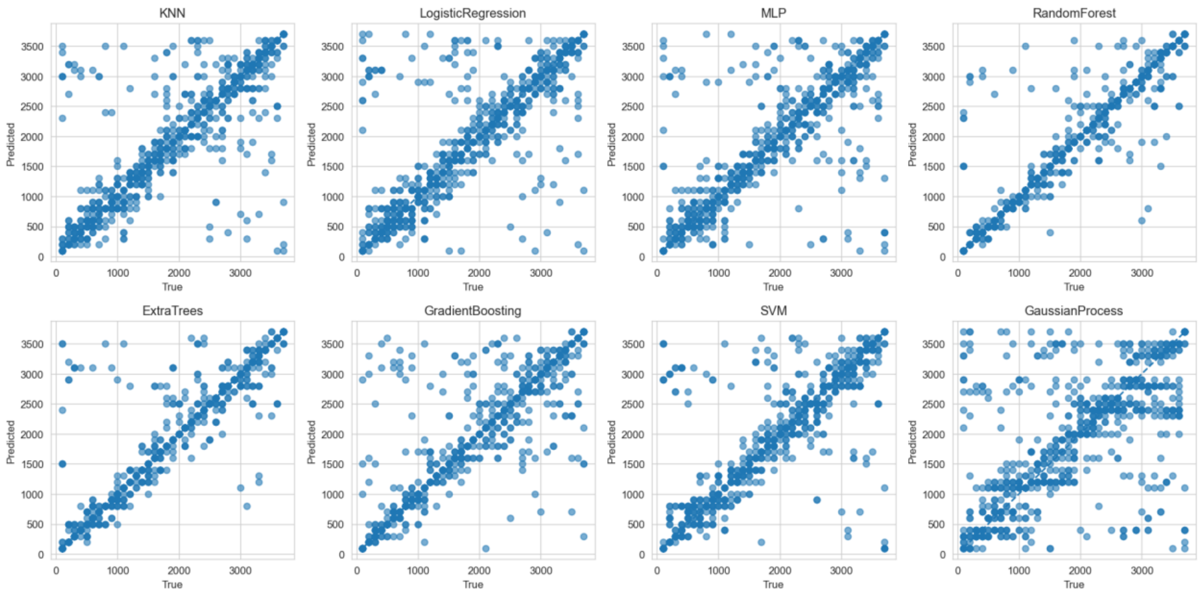

Figure 56. Predicted vs True Location (Hybrid Features)

The radar chart highlights that ensemble models, particularly Random Forest and Extra Trees, achieve the highest overall performance across all metrics, forming the largest areas. Gradient Boosting shows moderate performance, while MLP, KNN, Logistic Regression, and SVM perform at a lower level with similar patterns. Gaussian Process exhibits the weakest performance with significantly smaller values across metrics. Overall, the chart confirms that hybrid features improve localization performance, but clear differences remain between models, with ensemble methods providing the most reliable results.

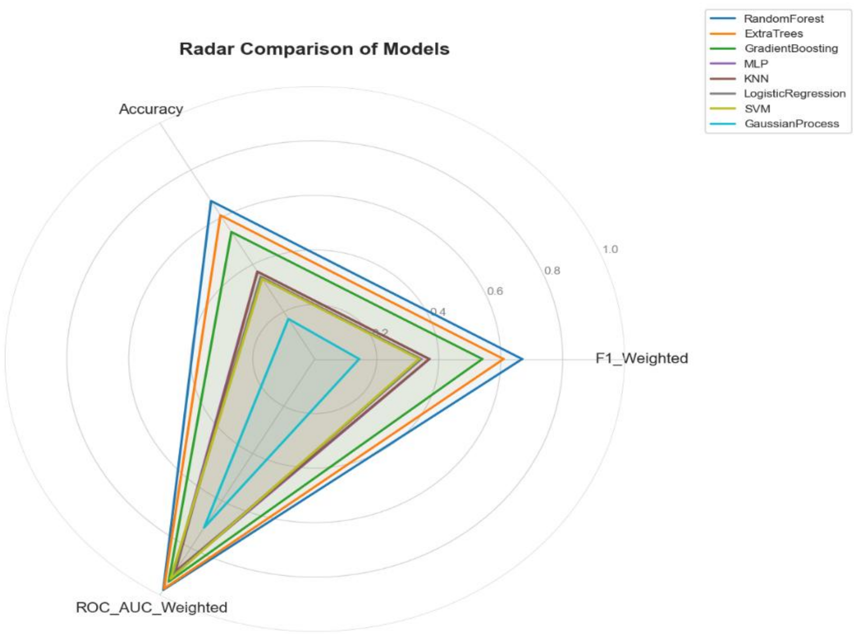

Figure 57. Model Comparison (Localization – Hybrid Features)

The results show significant differences in training time across models. Gradient Boosting and Gaussian Process require substantially longer training times (in minutes), indicating higher computational cost. In contrast, most models—including Random Forest, Extra Trees, SVM, MLP, and KNN—train very quickly within seconds. Logistic Regression is the fastest model. Overall, ensemble models offer a good balance between performance and computational efficiency, while Gradient Boosting and Gaussian Process are more computationally expensive despite their competitive performance.

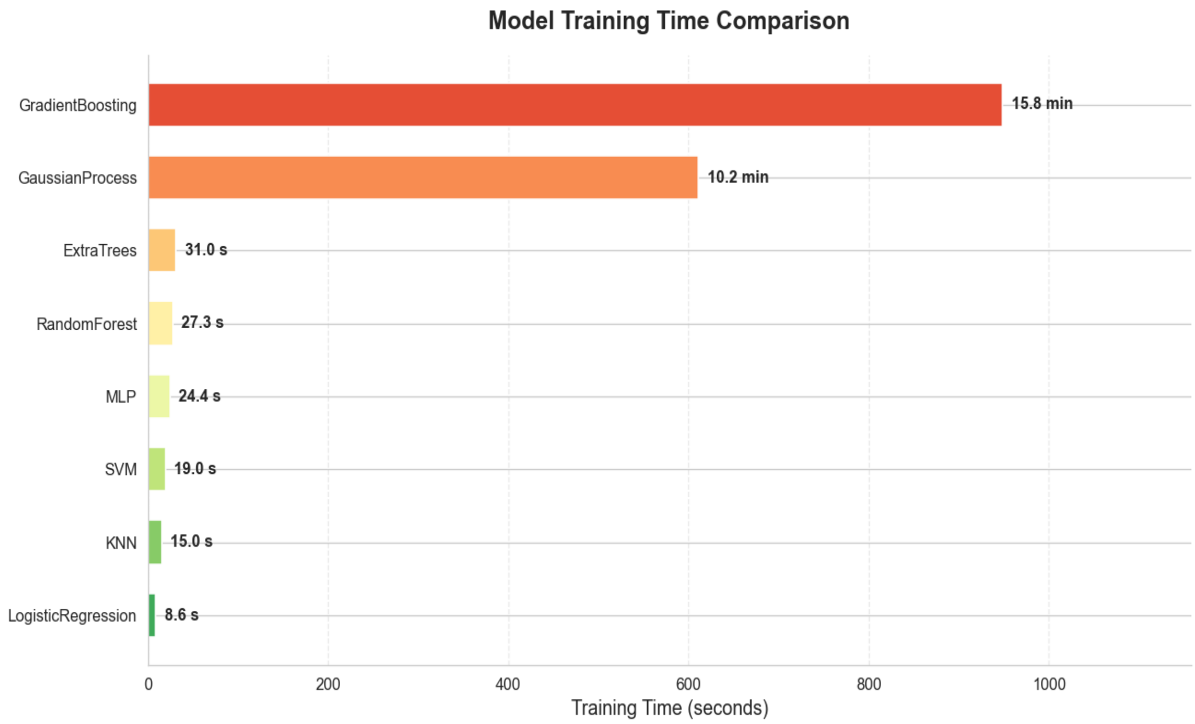

Figure 58. Model Training Time Comparison (Hybrid Features)

The Taylor diagram indicates that ensemble models, particularly Random Forest and Extra Trees, are closest to the reference point, showing the highest correlation and best agreement with true damage locations. Gradient Boosting also performs strongly, maintaining a good balance between correlation and standard deviation. KNN, Logistic Regression, and SVM exhibit moderate performance with slightly lower correlation and higher deviation from the reference. Gaussian Process shows the weakest performance, with lower correlation and poor variance representation. Overall, hybrid features significantly enhance localization accuracy, with tree-based ensemble models providing the most reliable predictions.

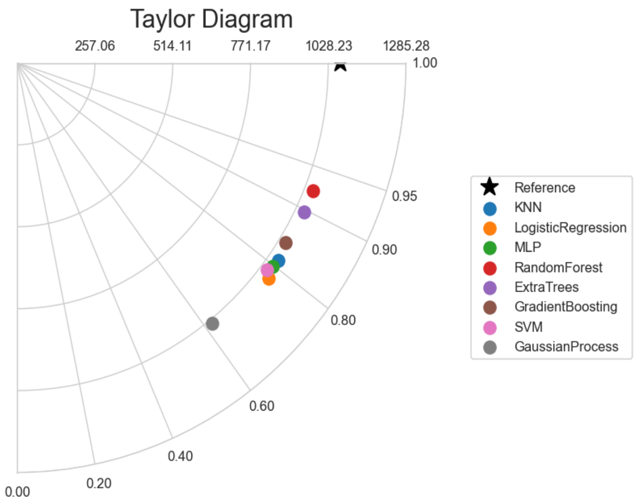

Figure 59. Taylor Diagram of Model Performance (Damage Localization – Hybrid Features)

➢ **Damage Severity**

The heatmap shows that hybrid features provide improved performance for crack severity estimation, with Gradient Boosting and Extra Trees achieving the best results, characterized by lower MAE and RMSE and higher R^2 values. Random Forest and Gaussian Process show moderate performance, while SVM and KNN exhibit higher errors and lower predictive accuracy. MLP performs slightly worse, and Ridge shows the poorest performance with the highest errors and lowest R^2. Overall, hybrid features enhance severity estimation compared to individual feature sets, with ensemble models delivering the most reliable predictions.

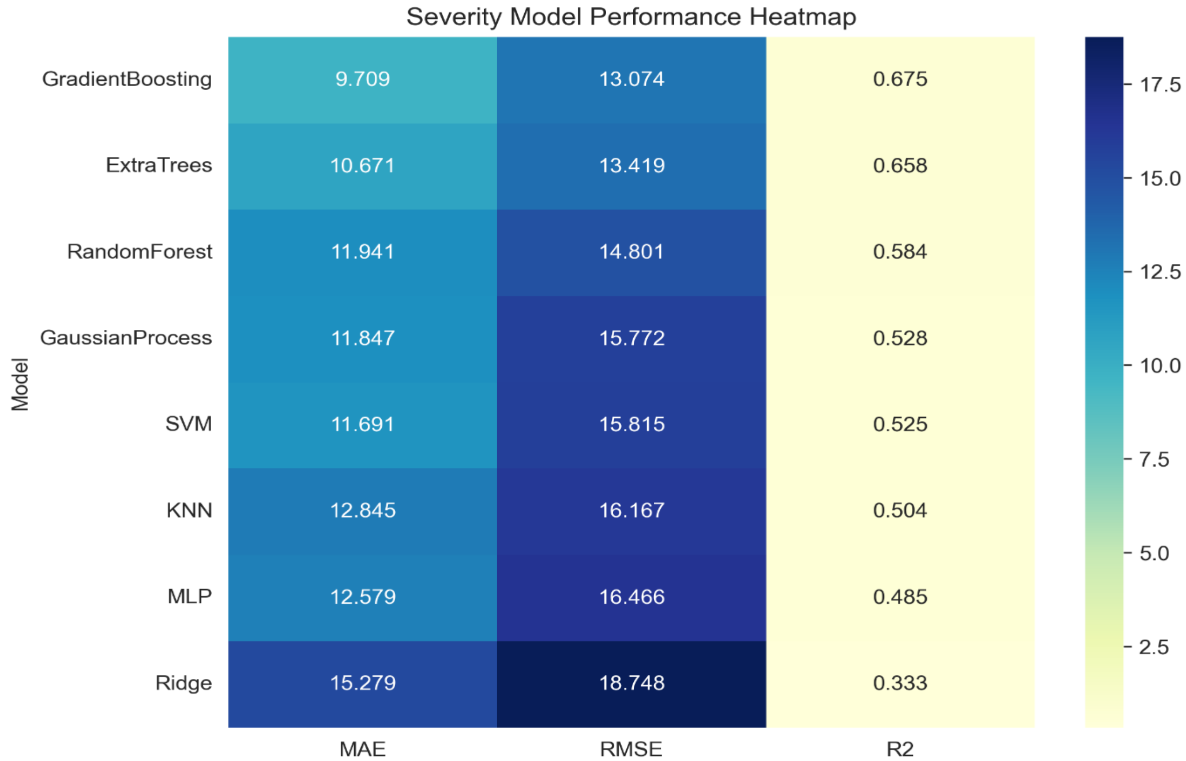

Figure 60. Severity Estimation Performance (Hybrid Features)

The comparison shows that hybrid features provide strong performance for severity estimation, but a noticeable gap remains between training and validation R^2 values, indicating overfitting. Ensemble models, particularly Extra Trees and Gradient Boosting, achieve the highest validation scores, demonstrating better generalization. Random Forest also performs well, while MLP and SVM show moderate results. Ridge performs poorly with low R^2 values. KNN and Gaussian Process exhibit high training performance but reduced validation accuracy, suggesting instability. Overall, hybrid features improve severity prediction, but generalization remains a key challenge.

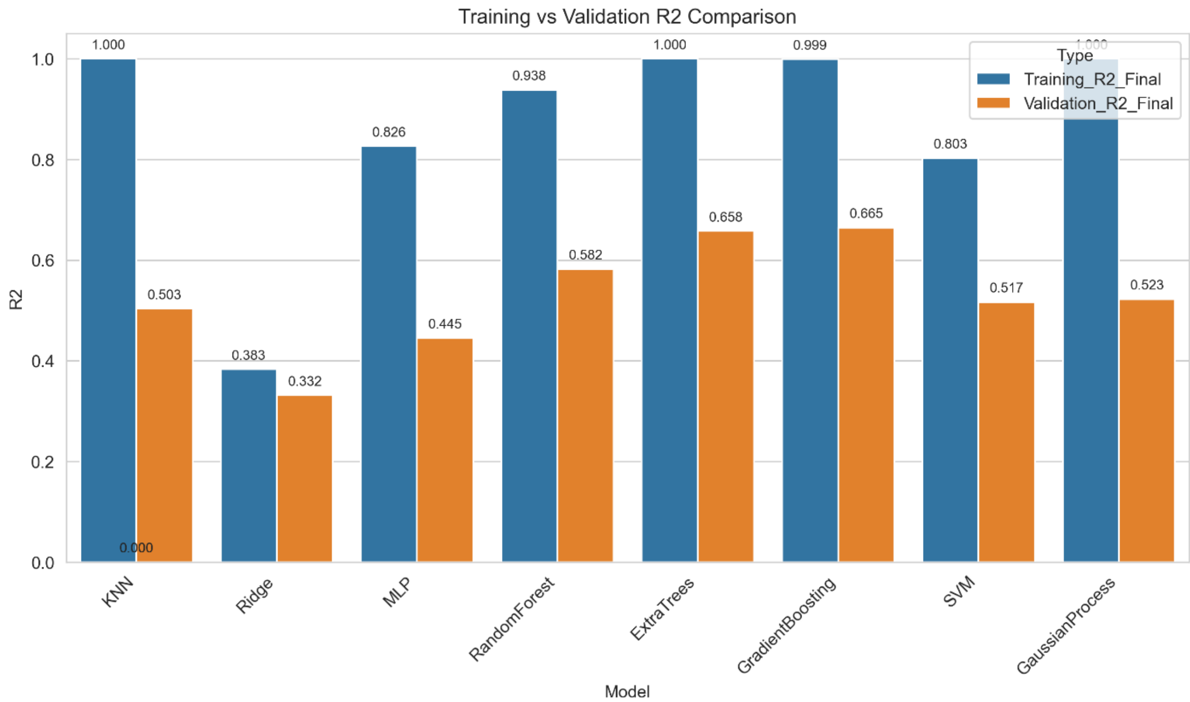

Figure 61. Training vs Validation R^2 (Severity – Hybrid Features)

The results show a significant drop in R^2 values after feature permutation, with all models falling to near-zero or negative values. This indicates that the predictive relationships between hybrid features and crack severity are strongly disrupted when feature structure is randomized. Ensemble models initially achieve the highest R^2 values but lose their predictive capability after permutation, similar to other models. The presence of negative R^2 values confirms that models perform worse than a baseline prediction when meaningful feature information is removed. Overall, this demonstrates that hybrid features contain essential information for accurate severity estimation.

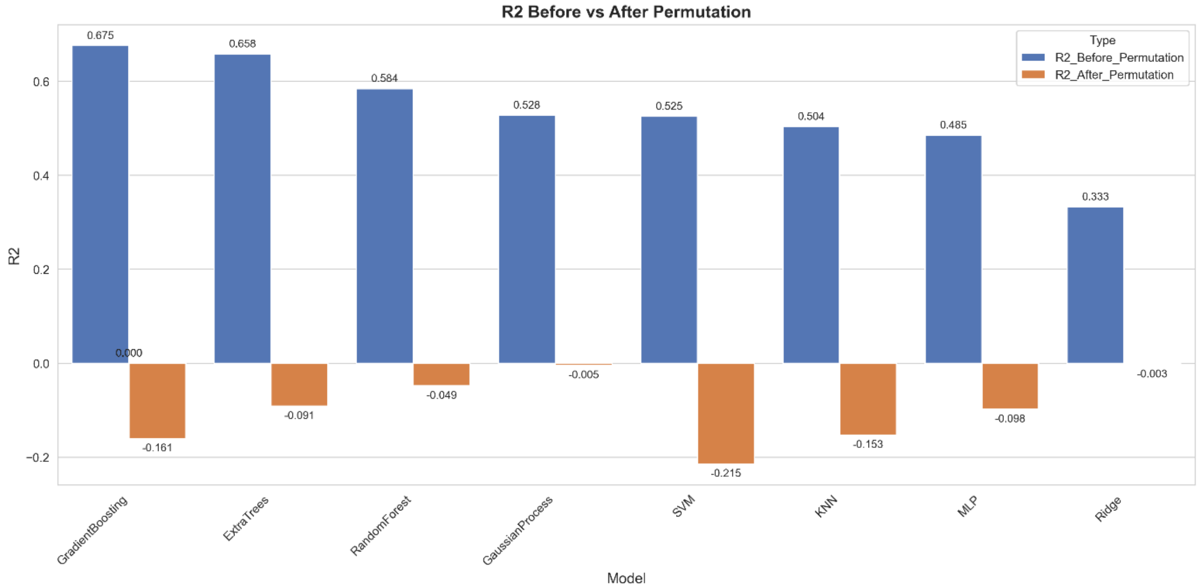

Figure 62. R^2 Before vs After Permutation (Severity – Hybrid Features)

The scatter plots show that ensemble models, particularly Random Forest and Extra Trees, produce predictions that align more closely with the ideal diagonal line, indicating better accuracy and consistency in severity estimation. Gradient Boosting also demonstrates good performance with moderate dispersion. In contrast, KNN, Ridge, and SVM show wider scatter, reflecting higher prediction errors. Gaussian Process exhibits instability with noticeable deviations, while MLP shows moderate but less consistent predictions. Overall, hybrid features improve severity estimation, with ensemble models providing the most reliable results, although some prediction variability remains.

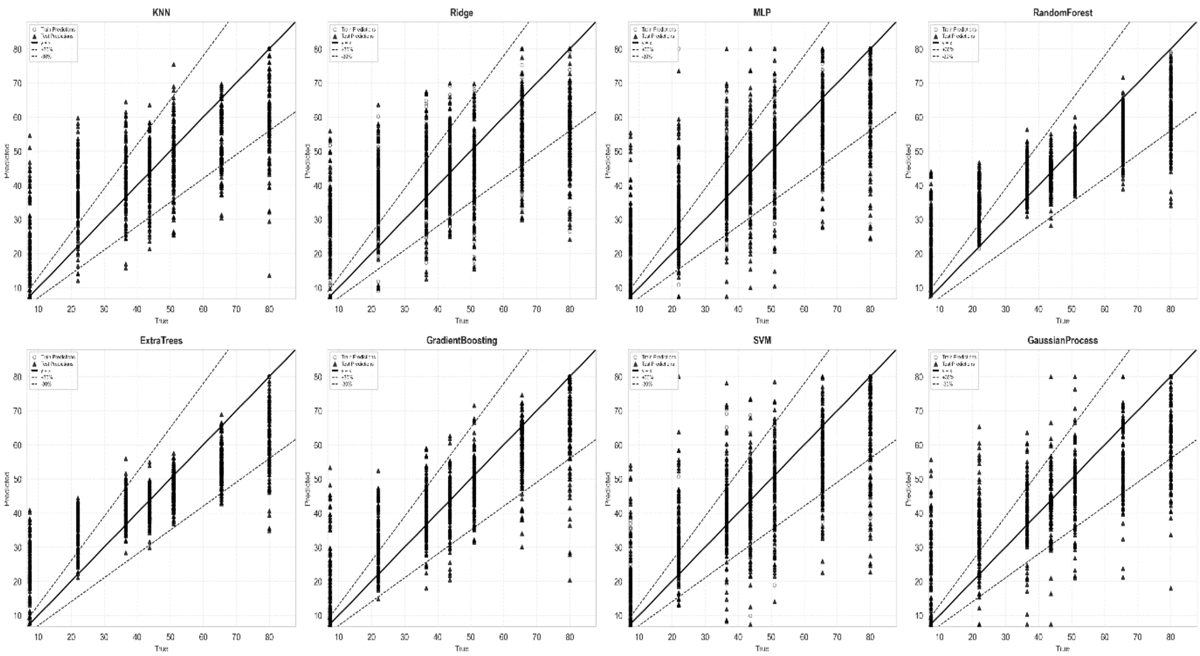

Figure 63. Predicted vs True Severity (Hybrid Features)

The results show a clear trade-off between model complexity and computational cost. Gradient Boosting and Gaussian Process require the longest training times (around 9 minutes), indicating higher computational demand. Random Forest, Extra Trees, and MLP show moderate training times, balancing performance and efficiency. SVM has relatively low training time, while KNN and Ridge are the fastest models, requiring only seconds. Overall, ensemble models provide strong performance at a moderate computational cost, whereas simpler models are computationally efficient but less accurate.

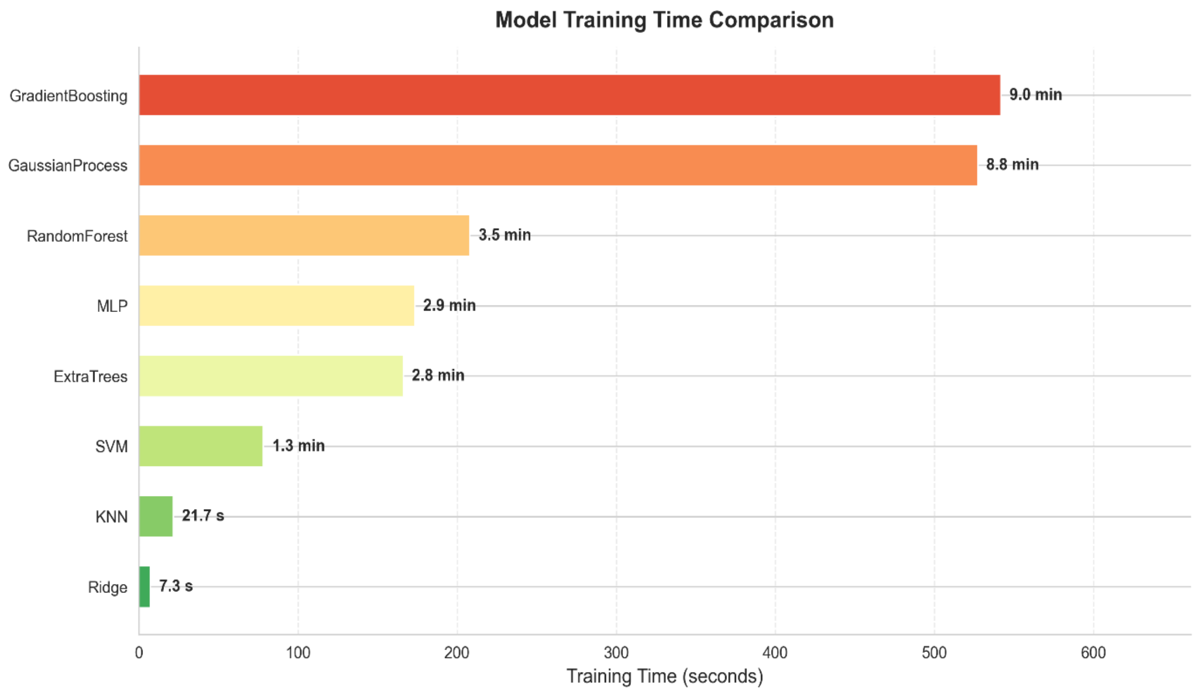

Figure 64. Training Time Comparison (Hybrid Features)

The Taylor diagram shows that ensemble models, particularly Random Forest, Extra Trees, and Gradient Boosting, are closest to the reference point, indicating higher correlation and better agreement with true severity values. These models also exhibit more appropriate standard deviation, reflecting good variability capture. In contrast, simpler models such as KNN, Ridge, and SVM are farther from the reference, indicating lower correlation and reduced accuracy. Gaussian Process and MLP show moderate performance but still deviate from the ideal point. Overall, ensemble models provide the most reliable and consistent severity predictions when using hybrid features.

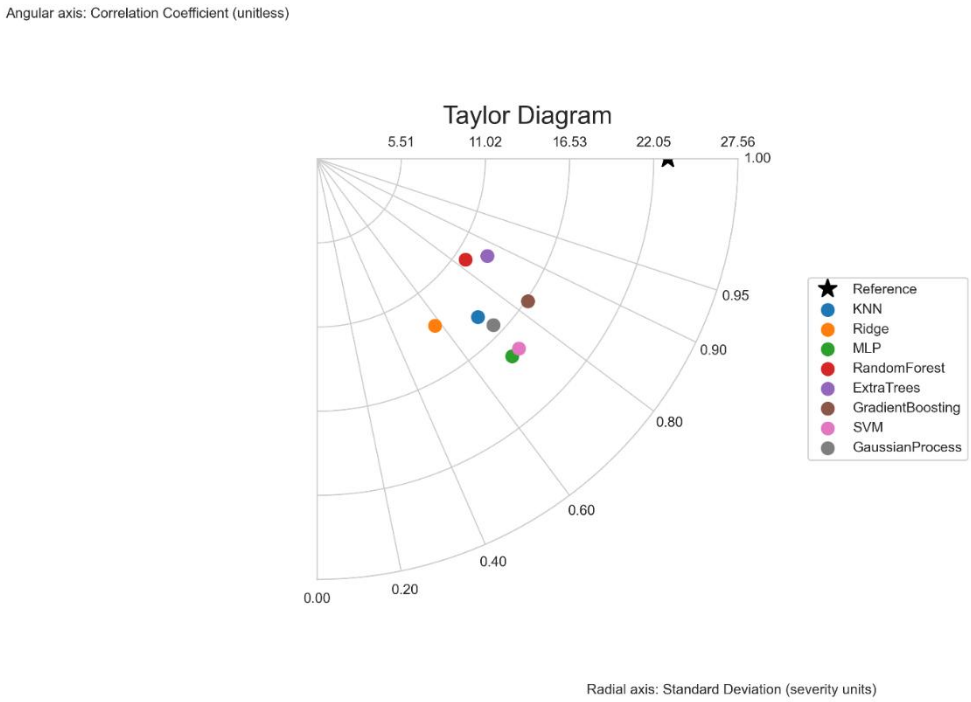

Figure 65. Taylor Diagram of Model Performance (Severity – Hybrid Features)

- **Results and Discussion for Hybrid Features (All features)**

The results demonstrate that hybrid features combining time-domain and frequency-domain information provide the most comprehensive and informative representation for structural damage assessment. By integrating statistical time-domain indicators with modal and spectral characteristics, the hybrid feature set captures both global dynamic behavior and localized transient structural responses. This complementary information significantly enhances the sensitivity of the machine learning models to different damage conditions and improves overall predictive capability across all structural health monitoring tasks.

For damage detection, nearly all machine learning models achieved near-perfect classification accuracy with excellent generalization performance, indicating that the hybrid features provide strong separability between healthy and damaged structural states. The integration of multiple feature domains improves robustness against noise and measurement variability while increasing the discriminative capability of the feature space. Ensemble learning methods, particularly Random Forest and Extra Trees, consistently produced the most stable and reliable results because they effectively exploit nonlinear interactions between time-domain and frequency-domain features while remaining resistant to overfitting and feature correlations.

For crack localization, the hybrid feature set substantially improves performance compared with individual feature groups because it combines the spatial sensitivity of frequency-domain information with the transient and statistical characteristics captured in the time domain. Ensemble models again outperform other approaches due to their ability to partition complex nonlinear feature spaces and identify subtle relationships associated with crack position. Gradient Boosting also demonstrates strong localization

capability because the sequential boosting process progressively refines prediction accuracy. However, moderate overfitting and reduced generalization for certain damage cases indicate that localization remains inherently challenging, particularly for structurally similar damage scenarios that produce overlapping vibration responses.
For crack severity estimation, hybrid features provide the strongest predictive performance among all investigated feature sets, achieving higher (R^2) values and lower prediction errors. The improved regression capability suggests that combining multiple feature domains increases sensitivity to stiffness degradation and crack propagation effects while reducing the limitations associated with individual feature types. Ensemble regression methods achieve the best overall performance because they effectively capture nonlinear relationships and reduce prediction variance through averaging mechanisms. In contrast, linear models remain limited because the relationship between damage severity and vibration response is strongly nonlinear. Similarly, neural-network-based models such as MLP do not consistently outperform ensemble methods, likely because the available dataset size is relatively limited for deep nonlinear learning and may increase the risk of unstable convergence and overfitting.
The permutation analysis further confirms that model performance strongly depends on the inherent structure and physical relevance of the hybrid features. Significant performance degradation after feature permutation demonstrates that the models rely on meaningful structural information rather than random statistical patterns. Overall, the hybrid time-frequency feature set provides the most robust, consistent, and physically informative representation for structural health monitoring, while ensemble learning methods consistently deliver the highest predictive capability due to their robustness, nonlinear modeling strength, and ability to exploit complementary information from multiple feature domains.

3.2.5. Best Performance Models

This section aims to identify the most suitable machine learning models for each structural health monitoring task, including damage detection, localization, and severity estimation. Since each task has different levels of complexity and prediction requirements, model performance is evaluated based on appropriate classification and regression metrics. By comparing multiple algorithms across these tasks, the goal is to highlight which models are most reliable, robust, and effective for practical implementation.

- **Damage Detection**

The results show that tree-based ensemble models, particularly Extra Trees, Random Forest and Gradient Boosting, consistently achieve perfect or near-perfect performance (Accuracy ≈ 1.000, F1 ≈ 1.000, ROC-AUC ≈ 1.000). These models demonstrate excellent capability in distinguishing between damaged and undamaged states with strong generalization and robustness.

The learning curve shows that the Extra Trees model achieves perfect training performance (F1 = 1.000) across all sample sizes, while validation performance steadily improves and converges to 1.00 as the number of training samples increases. The small gap between training and validation curves indicates excellent generalization with no signs of overfitting. Additionally, the narrow-shaded region at higher sample sizes reflects reduced variance and stable model behavior. Overall, this confirms that Extra Trees is a highly reliable and well-generalized model for this task.

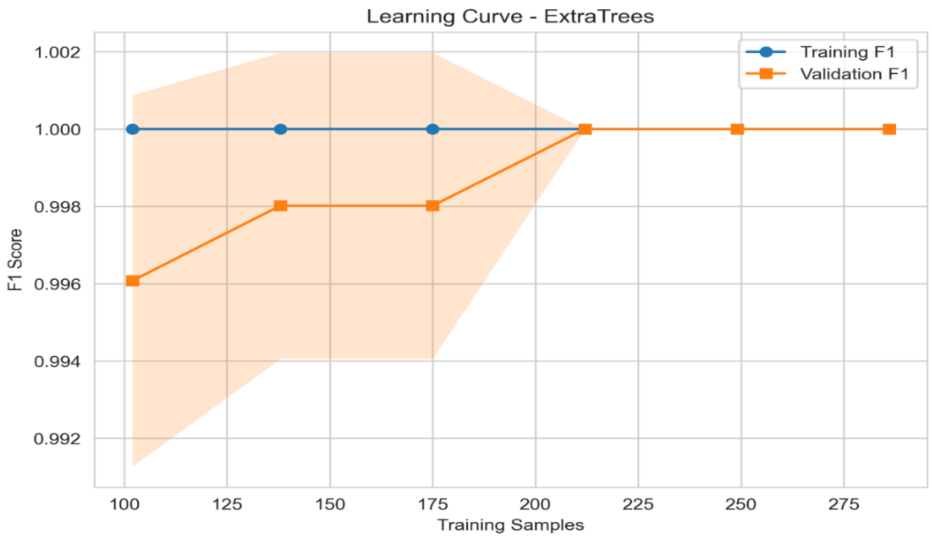

Figure 66. Learning Curve of Extra Trees (Best Model)

The permutation test shows that model performance significantly drops and fluctuates when feature relationships are randomized, particularly in ROC-AUC and accuracy. Despite F1 remaining relatively high, the instability across runs indicates that the model relies on meaningful feature patterns rather than memorization. The absence of consistently high performance after permutation suggests that there is no data leakage with respect to permutated features, and the model’s strong results are based on genuine learned relationships.

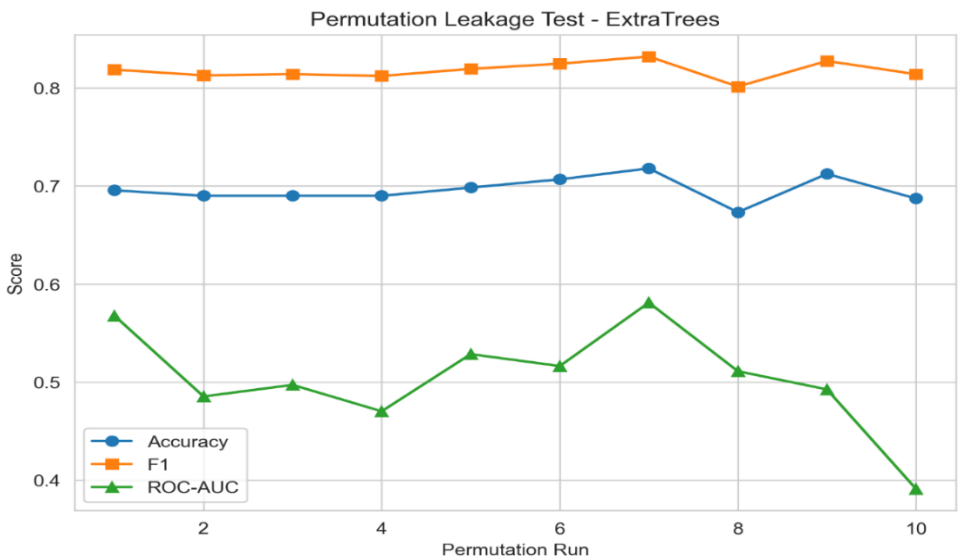

Figure 67. Permutation Leakage Test (Extra Trees – Best Model)

The confusion matrix shows perfect classification performance, with all samples correctly identified. The model accurately classifies all 99 no-damage cases and all 259 damage cases, with zero false positives and zero false negatives. This indicates that the Extra Trees model achieves 100% accuracy, precision, and recall, demonstrating exceptional reliability for damage detection.

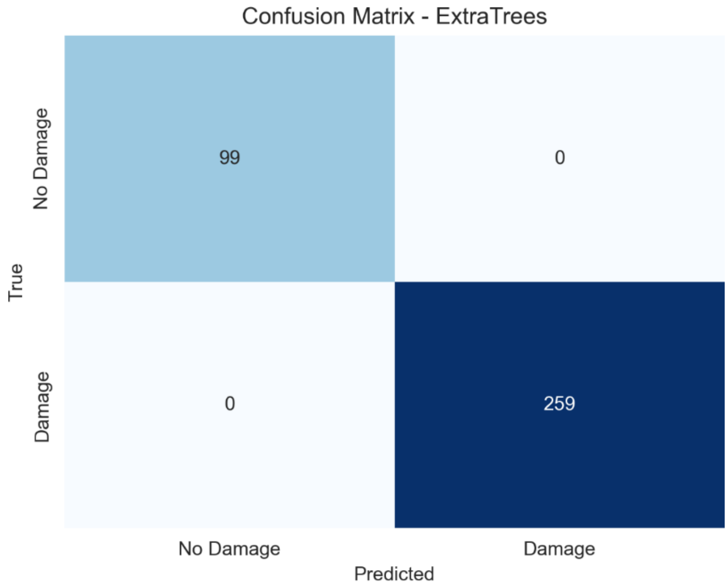

Figure 68. Confusion Matrix of Extra Trees (Damage Detection)

The ROC curve shows perfect classification performance, with the curve reaching the top-left corner and an AUC of 1.000. This indicates that the model can completely distinguish between damaged and undamaged cases without any misclassification. The sharp deviation from the diagonal baseline confirms excellent sensitivity and specificity, demonstrating that the Extra Trees model is highly reliable for damage detection.

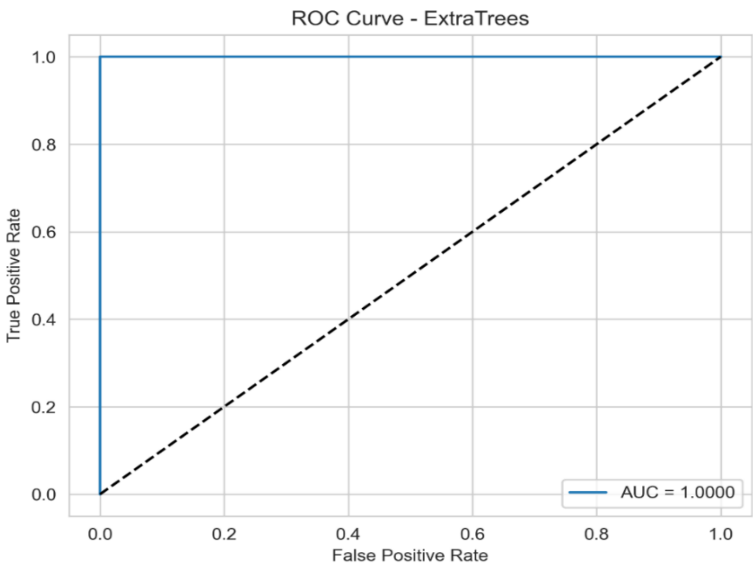

Figure 69. ROC Curve of Extra Trees (Damage Detection)

The feature importance results (**Figure 70**) indicate that mode shape features are the dominant contributors to the model's performance, especially those from higher modes such as Mode M4 and M3. In particular, Mode_M4-P2 stands out as the most influential feature, suggesting that higher-order vibration modes are more sensitive to structural changes and damage characteristics. This is expected, as higher modes capture more localized and subtle variations in the structure.

Features from Mode M2 and M1 also contribute, but with gradually decreasing importance, indicating that lower modes contain useful but less detailed information about damage.

On the other hand, frequency-based features (FR_F1–FR_F4) have comparatively lower importance. While they do not dominate the predictions, they still provide valuable complementary information that helps improve overall model performance when combined with mode shapes.

Overall, the model relies primarily on spatial information from mode shapes, while frequency features enhance robustness, confirming that hybrid features effectively capture both global and local damage characteristics.

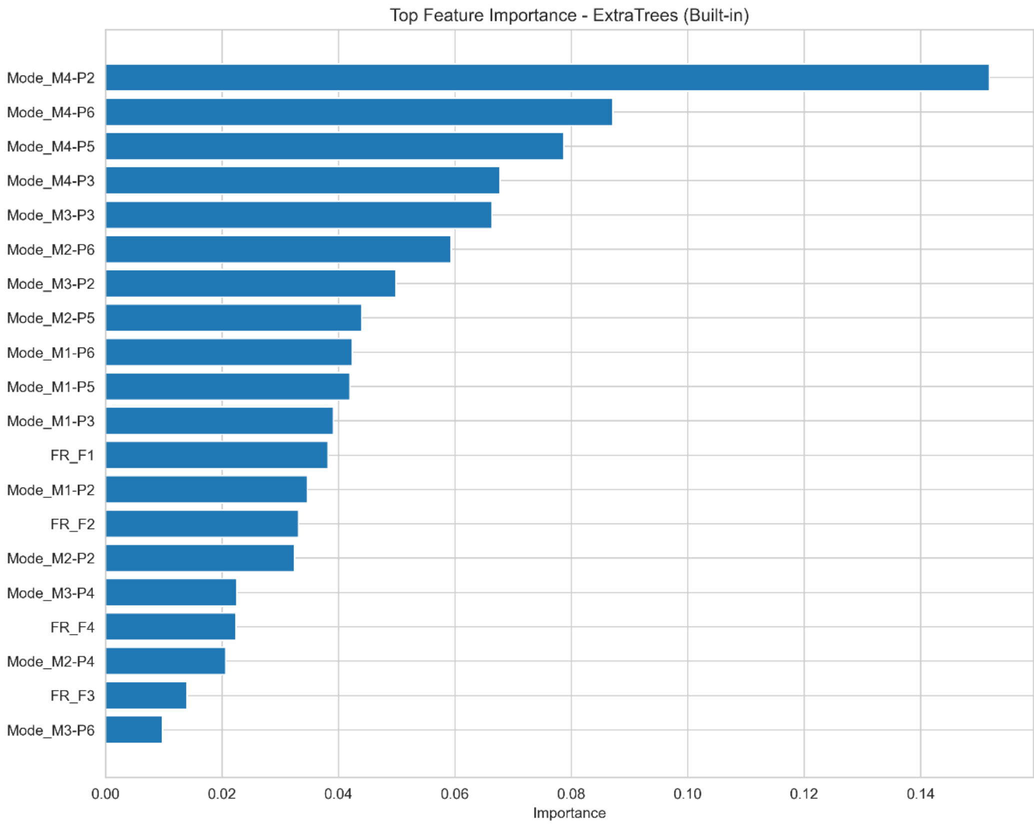

Figure 70. Feature Importance Analysis (Extra Trees – Hybrid Features)

➢ Damage Location

This section focuses on identifying the most effective model for predicting the location of damage within the structure. Unlike detection, localization requires capturing spatial variations in structural behavior, making it a more challenging task. The models are evaluated based on their ability to accurately map damage positions while maintaining consistency and robustness. The aim is to determine which approach best captures the relationship between input features and damage location for reliable structural assessment.

The confusion matrix shows (**Figure 71**) that most predictions lie along the diagonal, indicating that the model accurately identifies the correct damage locations in the majority of cases. The presence of small off-diagonal values suggests that misclassifications occur mainly between neighboring locations, which is expected due to the similarity of structural responses at close positions. Overall, the Random Forest model demonstrates strong localization capability with high accuracy and reasonable spatial precision.

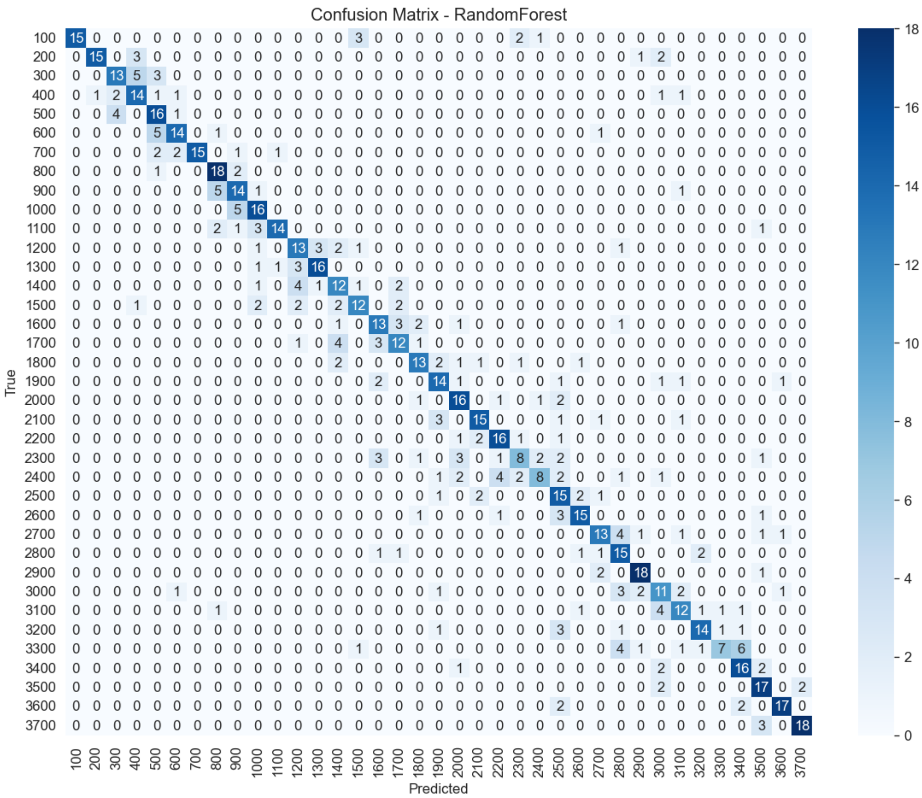

Figure 71. Confusion Matrix of Random Forest (Damage Localization)

The permutation test shows that both accuracy and F1-weighted scores drop to very low values when the input features are randomized. This indicates that the model loses its predictive capability without the true feature relationships. The consistently low performance across runs confirms that the model is not relying on any data leakage and that its original performance is based on meaningful structural patterns.

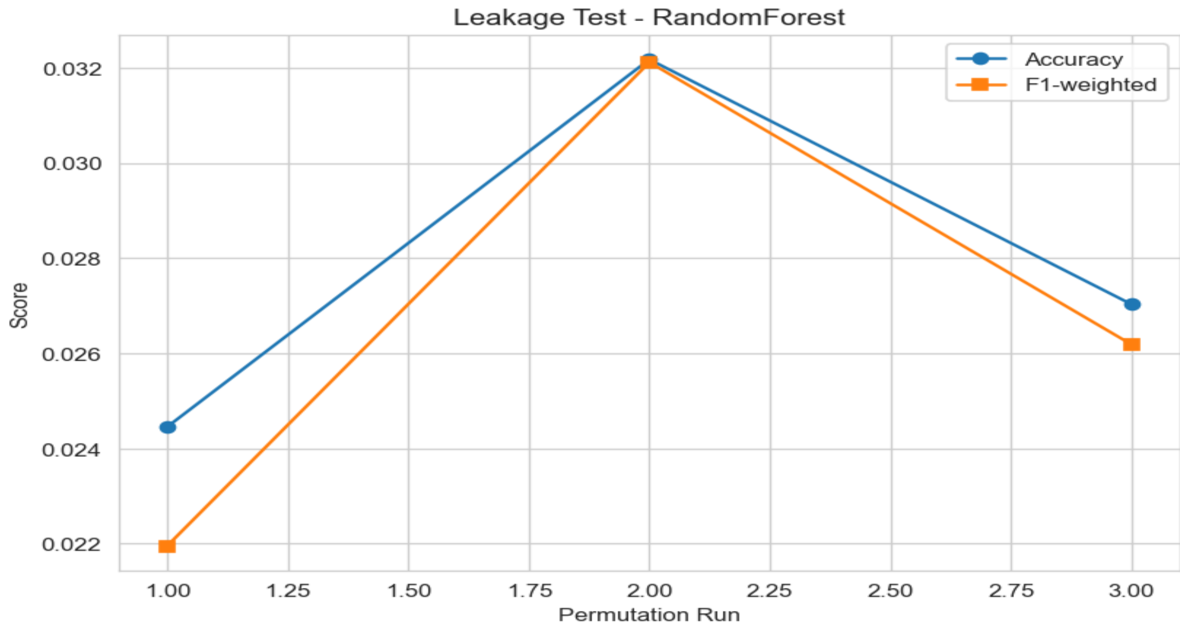

Figure 72. Permutation Leakage Test (Random Forest – Damage Localization)

The scatter plot shows that most predictions align closely with the diagonal line, indicating good agreement between predicted and true damage locations. The spread of points around the line suggests some prediction error, particularly at higher locations, but the overall trend remains strong. This indicates that the Random Forest model captures the general spatial pattern effectively, with most errors occurring near neighboring locations rather than large deviations.

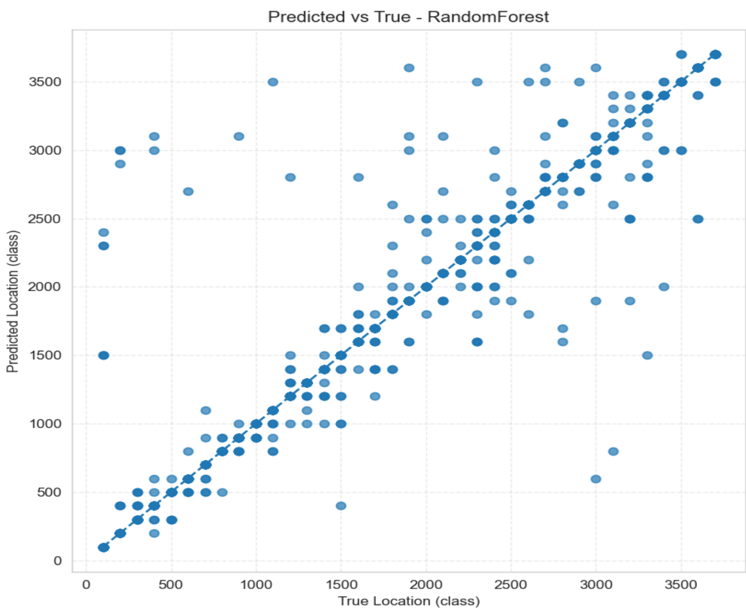

Figure 73. Predicted vs True Location (Random Forest)

The histogram shows that most prediction errors are concentrated at low values, indicating that the model accurately predicts damage locations in the majority of cases. A small number of larger errors are present, forming a long tail, which suggests occasional mispredictions at distant locations. Overall, the distribution confirms that the Random Forest model achieves high localization accuracy with errors mainly limited to small deviations.

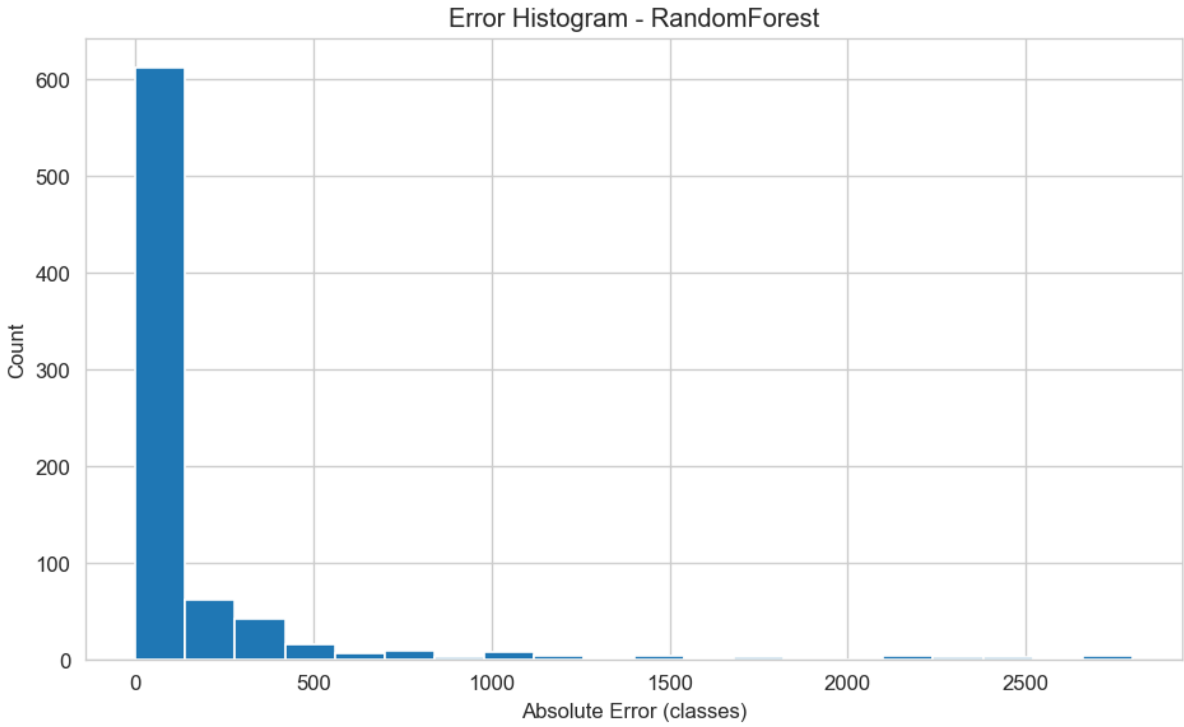

Figure 74. Error Distribution (Random Forest – Damage Localization)

The feature importance results show that both mode shape features and frequency-based features play a key role in damage localization. Features from Mode M2 and M4, particularly at specific measurement points, have the highest importance, indicating strong sensitivity to damage position. Additionally, frequency features (FR_F1) and time-domain features (Peak and RMS) also contribute, enhancing the model's ability to capture both global and local structural changes. Overall, the results highlight the effectiveness of hybrid features, with mode shapes being the dominant contributors and other features providing complementary information for improved localization accuracy.

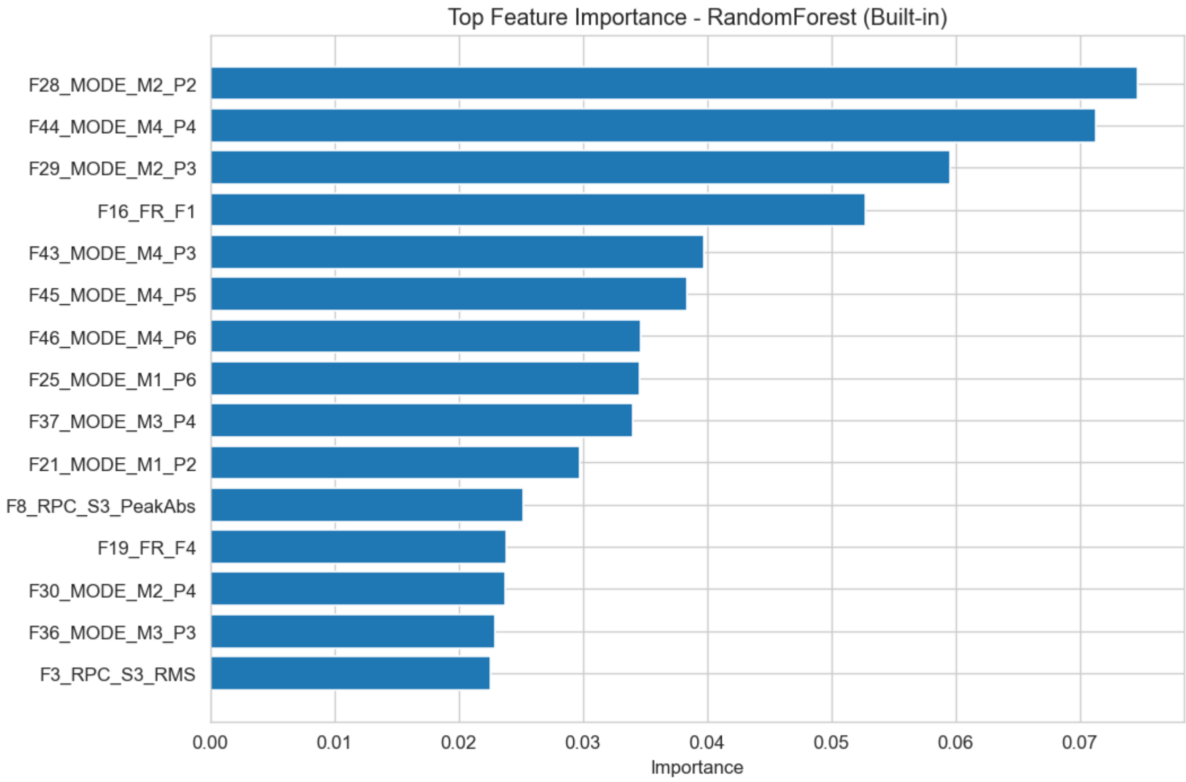

Figure 75. Feature Importance (Random Forest – Damage Localization)

- **Damage Severity**

This section focuses on identifying the most effective model for estimating the severity of damage, which is inherently more complex than detection and localization. Unlike classification tasks, severity estimation requires accurate regression of continuous values, making it more sensitive to noise and model generalization. The models are evaluated based on their ability to minimize prediction error while maintaining stable and reliable performance. The objective is to determine which model can best capture the relationship between input features and the extent of damage for precise quantification. The learning curve shows (**Figure 76**) that the model achieves perfect training performance ($R^2 \approx 1.000$), while validation performance improves steadily as the number of samples increases. Initially, validation R^2 is low and unstable, but it gradually rises and stabilizes, indicating improved generalization with more data. The gap between training and validation suggests some overfitting, but this reduces at higher sample sizes. Overall, the model demonstrates strong learning capability, with performance becoming more reliable as data increases.

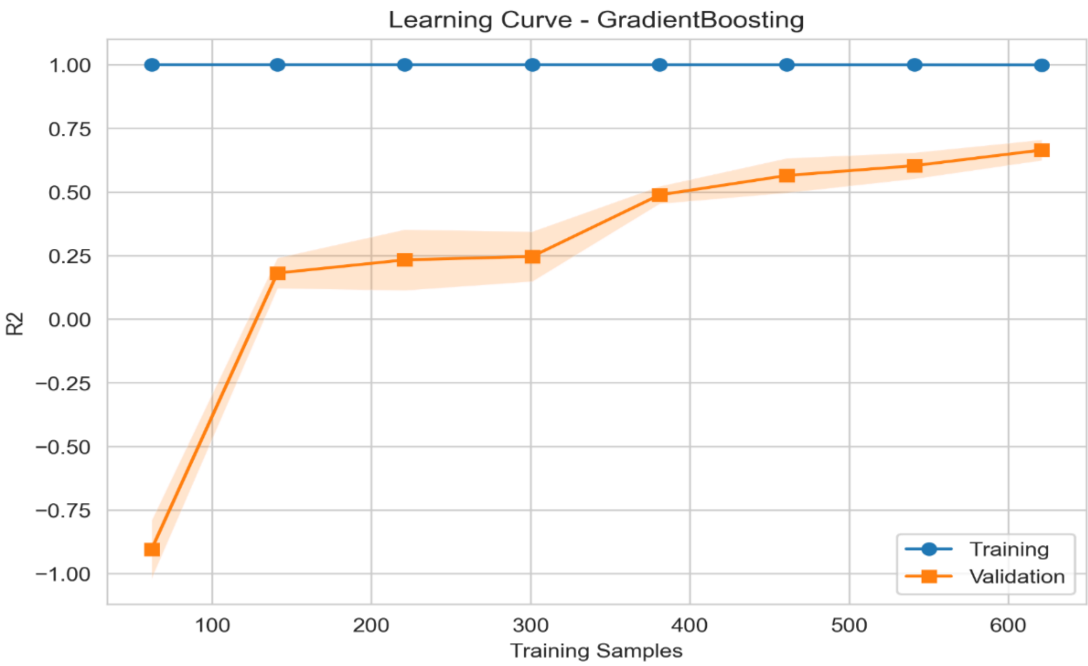

Figure 76. Learning Curve of Gradient Boosting (Damage Severity)

The cross-validation results show consistent performance across all folds, with relatively stable MAE and RMSE values and only minor variations. The R^2 values remain moderately high and stable, indicating reliable predictive capability across different data splits. This consistency suggests that the model generalizes well and is not overly sensitive to specific training data, confirming its robustness for severity estimation.

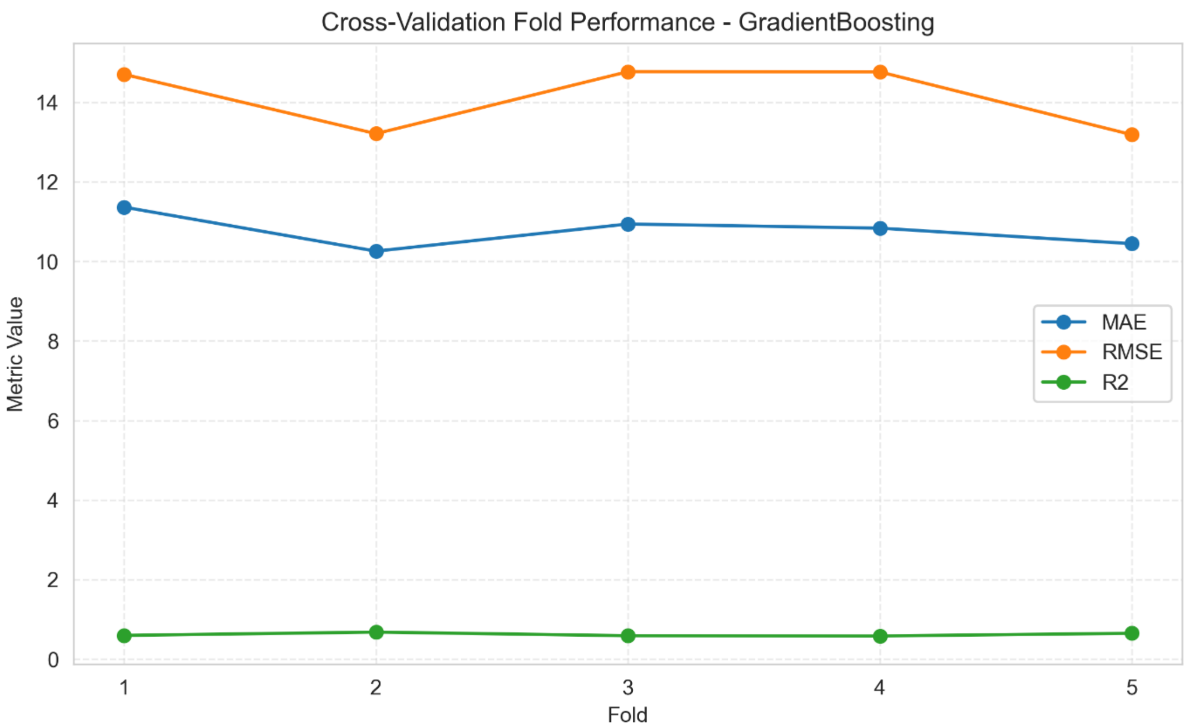

Figure 77. Cross-Validation Performance (Gradient Boosting – Damage Severity)

The results of the permutation leakage test indicate that the Gradient Boosting model does not suffer from data leakage. When the target variable is randomly shuffled, the model's predictive performance significantly degrades, as shown by the near-zero or negative R^2 values across all runs. Simultaneously, the MAE and RMSE remain consistently high, reflecting large prediction errors and the model's inability to learn from randomized data. The stability of these metrics over multiple permutation runs further confirms that the model is not exploiting hidden correlations or memorized patterns. Instead, its original performance is based on meaningful relationships within the dataset, demonstrating the validity and robustness of the severity prediction model.

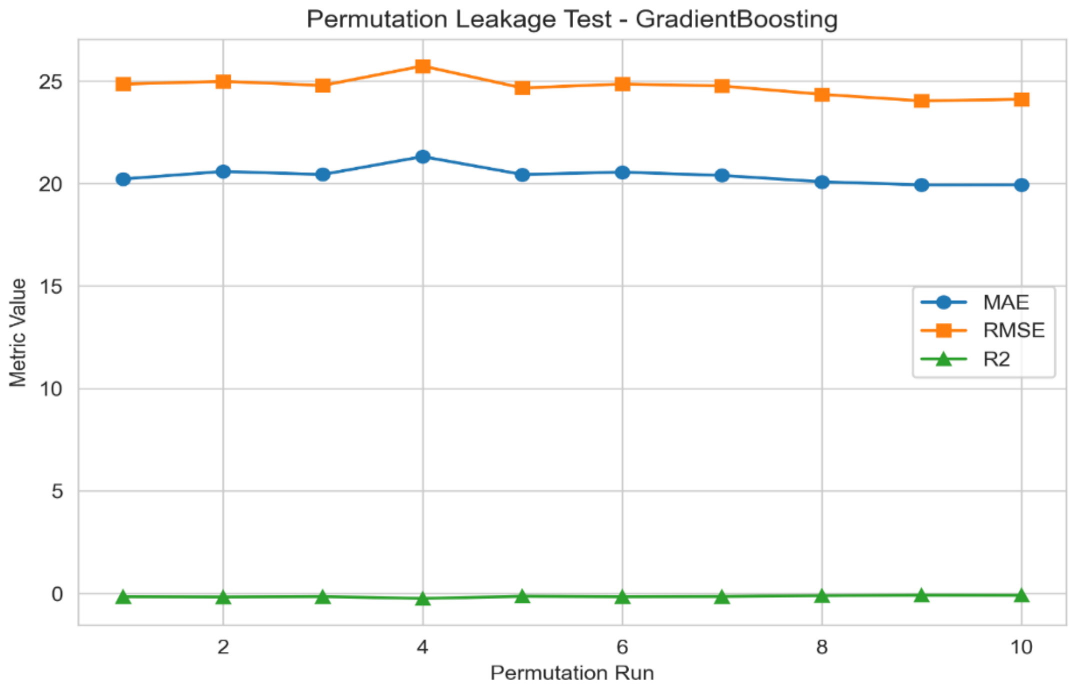

Figure 78. Permutation Leakage Test for Damage Severity Prediction using Gradient Boosting

The predicted versus true severity plot demonstrates that the Gradient Boosting model is capable of capturing the overall trend of damage severity with reasonable accuracy. Most data points are distributed around the diagonal reference line, indicating a good agreement between predicted and actual values. However, a noticeable spread is observed, particularly at higher severity levels, suggesting increased prediction uncertainty as damage severity increases. While many predictions fall within the ±30% error bounds, some outliers indicate occasional overestimation or underestimation. Overall, the model shows a strong ability to approximate severity values, but with moderate dispersion that reflects the inherent complexity of the regression task.

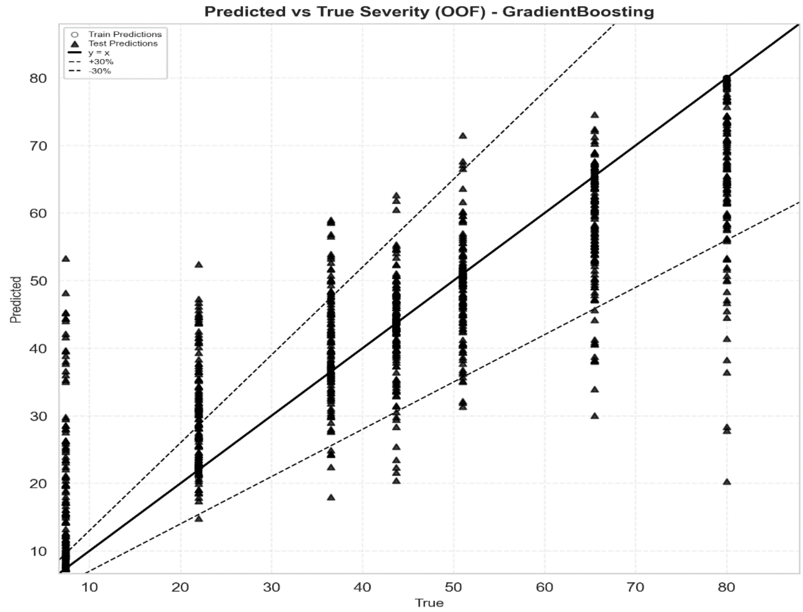

Figure 79. Predicted vs. True Damage Severity (Out-of-Fold) – Gradient Boosting Model

The error histogram shows that the majority of prediction errors are concentrated at low values, indicating that the Gradient Boosting model achieves good accuracy for most samples. A high frequency of small absolute errors suggests reliable performance in estimating damage severity. However, the distribution exhibits a right-skewed tail, meaning that a limited number of cases result in larger errors.

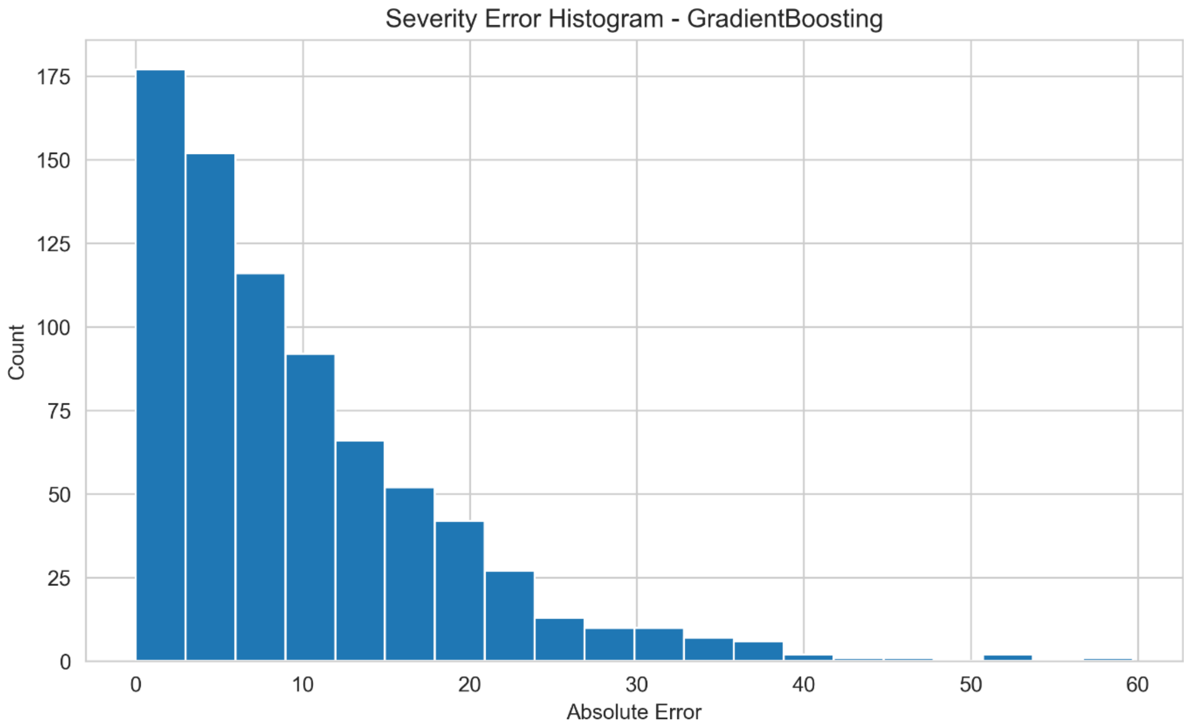

Figure 80. Error Distribution of Damage Severity Prediction – Gradient Boosting Model

The feature importance analysis reveals that frequency-domain features, particularly FR_F3, FR_F2, and FR_F1, play a dominant role in predicting damage severity, with FR_F3 having a significantly higher contribution than all other features. This indicates that higher-order frequency responses are highly sensitive to changes in damage magnitude. Additionally, several mode shape features (Mode_M4-P6, Mode_M4-P4, Mode_M3-P4) also contribute meaningfully, highlighting the importance of spatial deformation characteristics in severity estimation. Time-domain features such as RMS and peak-related indicators show relatively lower importance, suggesting that they provide supplementary rather than primary information. Overall, the results confirm that a combination of frequency and mode shape features is essential for accurate damage severity prediction, with frequency features being the most influential.

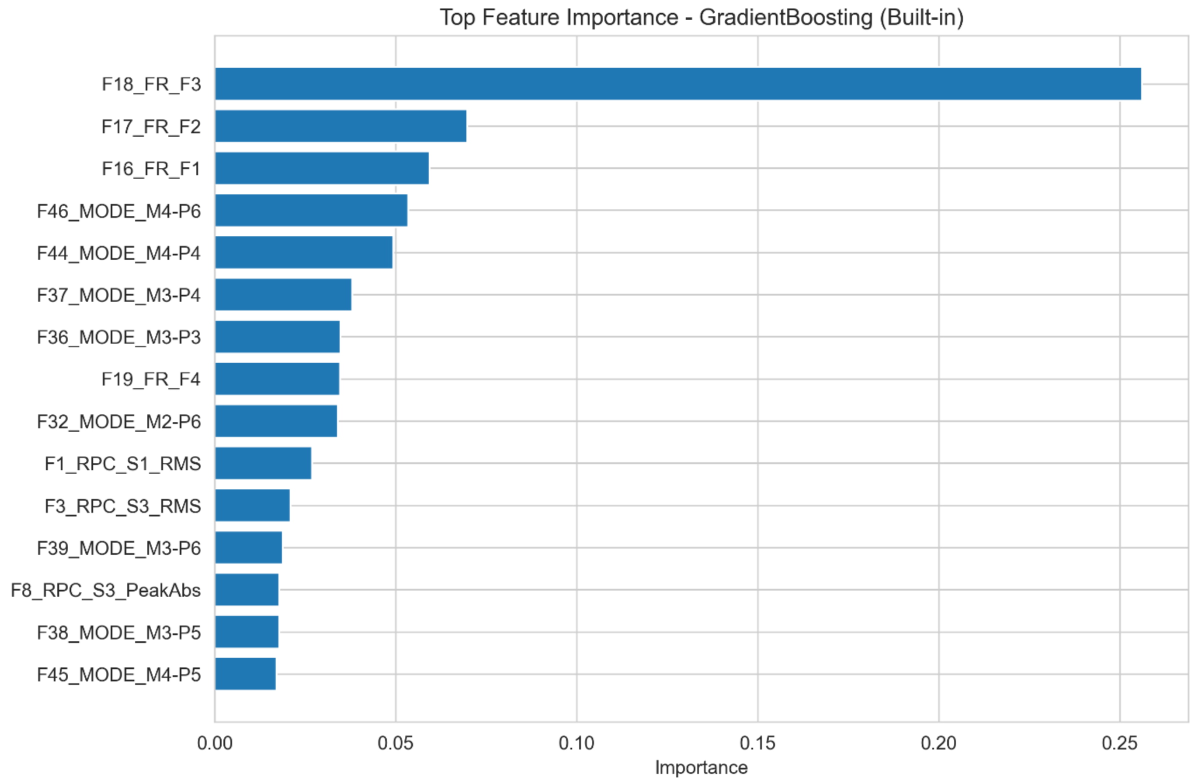

Figure 81. Top Feature Importance for Damage Severity Prediction – Gradient Boosting Model

3.2.6. Generalization Performance on Unseen Models

To evaluate the generalization capability of the developed machine learning models, five new structural models were created in ANSYS that were not previously used during the training phase. These unseen models were used to assess how well the trained algorithms can predict damage detection, location, and severity under new conditions. The input data for these models are presented in the following tables.
The evaluation was performed in two stages. First, the models were tested using raw features directly extracted from the structural responses. In the second stage, Δ features

were used, defined as the difference between damaged and intact states. This approach aims to enhance sensitivity to damage-related changes. By comparing the prediction results from both feature types, the effectiveness of each representation in improving model performance on unseen data is analyzed.

- **Test Models**

Table 12Table 12 presents the specifications of the unseen ANSYS models used for validation, including the damage condition, crack location, and severity for each test case. These models were not included in the training phase and are used to evaluate the generalization capability of the developed machine learning algorithms under new structural conditions.

Table 12. Test Models Information

Models	Damage (0= Undamaged, 1=Damaged)	Location [mm]	Severity [mm]
Test1	0	0	0
Test2	1	140	70
Test3	1	1170	35
Test4	1	2005	10
Test 5	1	3150	45

- **Test Results**

In this section, the prediction performance of the developed machine learning models is evaluated using five different feature configurations, including frequency-domain features, mode shape and Frequencies, their combination, time-domain features (RMS, peak value, and crest factor), and a comprehensive combination of all features. The analysis is conducted for all three tasks: damage detection, localization, and severity estimation, to determine the most effective feature set for each case. Overall, 240 predictions for assessment of the models performed and best model results are represented in the next tables.

Table 13. Best Model Prediction Results on Unseen Models Using Selected Feature Sets (Raw Data)

		Real	Predicted Status	Prob_No_Damage	Prob_Damage
Damage Detection (Extra Trees)	Test 1	No Damage	No Damage	83.67	16.33
	Test 2	Damage	Damage	0	100
	Test 3	Damage	Damage	6	94
	Test 4	Damage	Damage	0	100
	Test 5	Damage	Damage	28.33	71.67

Damage Location (Random Forest)		**Real (mm)**	**Predicted Location**	**Error (mm)**	**Error (cm)**
	Test 2	140	100	40	4
	Test 3	1170	1200	30	3
	Test 4	2005	2000	5	0.5
	Test 5	3150	3100	50	5
Damage Severity (Gradient Boosting)		**Real (mm)**	**Predicted Severity**	**Error (mm)**	**Error (cm)**
	Test 2	70	66.13	3.87	0.40
	Test 3	35	28.81	6.19	0.60
	Test 4	10	21.34	11.34	1.13
	Test 5	45	37.98	7.02	0.70

Table 13 presents the prediction results of the best-performing machine learning models for each task when applied to unseen models using raw features for training the models. For damage detection, the Extra Trees classifier using frequency-domain features correctly classifies all test cases, achieving good accuracy. The model shows high confidence in most predictions, with damage probabilities reaching up to 100% (Tests 2 and 4), while even the lowest confidence case (Test 5) still correctly identifies damage with approximately 71.67% probability.

For damage localization, the Random Forest model using combined frequency and mode shape features provides accurate predictions with relatively small errors ranging from 5 mm to 50 mm (0.5 cm to 5 cm). The most precise prediction is observed in Test 4 with only 5 mm error, while larger deviations occur in Tests 2 and 5, though still within acceptable engineering limits.

For damage severity estimation, the Gradient Boosting model using combined time and frequency features achieves reasonable accuracy, with errors ranging from approximately 3.87 mm to 11.34 mm (0.40 cm to 1.13 cm). The model performs best for moderate severity levels (e.g., Test 2), while slightly larger errors are observed in lower severity cases such as Test 4.

Overall, the table demonstrates that the selected models achieve perfect detection performance, highly accurate localization, and acceptable severity prediction on unseen data, confirming their strong generalization capability.

Table 14. Best Model Prediction Results on Unseen Models Using Δ Features

	Combined All ΔFeatures (Freq and time Domain)				
		Real	**Predicted Status**	**Prob_No_Damage**	**Prob_Damage**
Damage Detection (MLP)	Test 1	No Damage	No Damage	53.92	46.08
	Test 2	Damage	Damage	44.71	55.29
	Test 3	Damage	Damage	25.57	74.43
	Test 4	Damage	Damage	35.5	64.5
	Test 5	Damage	Damage	30.48	69.52
Location Detection (Random Forest)	**Combined All ΔFeatures (Freq and time Domain)**				
		Real (mm)	**Predicted Location**	**Error (mm)**	**Error (cm)**
	Test 2	140	100	40	4
	Test 3	1170	1200	30	3
	Test 4	2005	2000	5	0.5
	Test 5	3150	3200	50	5
Damage Severity (Gradient Boosting)	**Combined All ΔFeatures (Freq and time Domain)**				
		Real (mm)	**Predicted Severity**	**Error (mm)**	**Error (cm)**
	Test 2	70	67.5	2.5	0.25
	Test 3	35	26.92	8.08	0.81
	Test 4	10	23.48	13.48	1.35
	Test 5	45	46.54	1.54	0.15

Table 14 presents the prediction results of the best-performing machine learning models when trained and tested using Δ features, defined as the difference between damaged and intact structural responses. These features enhance sensitivity to damage-related changes and are evaluated on unseen ANSYS models.

For damage detection, the MLP model using combined Δ features correctly classifies all test cases, achieving 100% accuracy. However, compared to raw feature results, the prediction confidence is lower and more uncertain. The probability of damage ranges from 55.29% to 74.43%, while even the intact case (Test 1) shows a relatively balanced probability (53.92% vs 46.08%), indicating reduced confidence despite correct classification.

For damage localization, the Random Forest model maintains consistent performance, with errors ranging from 5 mm to 50 mm (0.5 cm to 5 cm), similar to the raw feature case. The most accurate prediction is observed in Test 4 with only 5 mm error, while other cases show moderate deviations. This consistency confirms that Δ features do not significantly alter localization performance but preserve model stability.

For damage severity estimation, the Gradient Boosting model shows improved performance in some cases compared to raw features. The prediction errors range from 1.54 mm to 13.48 mm (0.15 cm to 1.35 cm). Notably, Test 2 and Test 5 achieve very low errors (2.5 mm and 1.54 mm), indicating enhanced sensitivity of Δ features to damage magnitude. However, larger deviations are still observed in low-severity cases such as Test 4.

Overall, the results demonstrate that Δ features improve sensitivity to damage changes, particularly in severity estimation, while maintaining stable localization performance. However, they introduce increased uncertainty in classification confidence for damage detection compared to raw features.

3.3. Transfer Learning Implementation for Model Adaptation

The main objective of this stage is to apply transfer learning to the previously selected best-performing models in order to adapt them from simulated data to real experimental conditions. This process aims to reduce the discrepancy between numerical (ANSYS) data and real structural responses, thereby improving prediction accuracy in practical applications.

Methodology / Processing Steps

To achieve this, the following procedure is implemented:

- **Experimental Data Acquisition:**
  Real experimental vibration data corresponding to a damaged structure with a crack located at **1900 mm** and a severity of **7.4 mm** are utilized.
- **Data Segmentation (Windowing):**
  The continuous signal (approximately **900 seconds**) is divided into multiple smaller samples (**20-second windows**) to increase the number of training instances and improve learning efficiency.
- **Feature Extraction:**
  The same feature extraction process used for simulated data is applied, including frequency-domain, mode shape, and time-domain features. This ensures consistency between source (simulation) and target (experimental) domains.
- **Model Fine-Tuning:**
  The pre-trained models are fine-tuned using the experimental dataset. Instead of training from scratch, the models leverage previously learned representations and adapt their parameters to better match real-world data, following a data-constrained (single-shot) transfer learning strategy.
- **Evaluation:**
  The updated models are then evaluated to assess improvements in damage detection, localization, and severity estimation performance.

➢ **Damage Detection**

The performance of the selected Extra Trees model under single-shot transfer learning is illustrated in Figures 81 and 82, which present the predicted damage probability across multiple signal windows before and after domain adaptation.

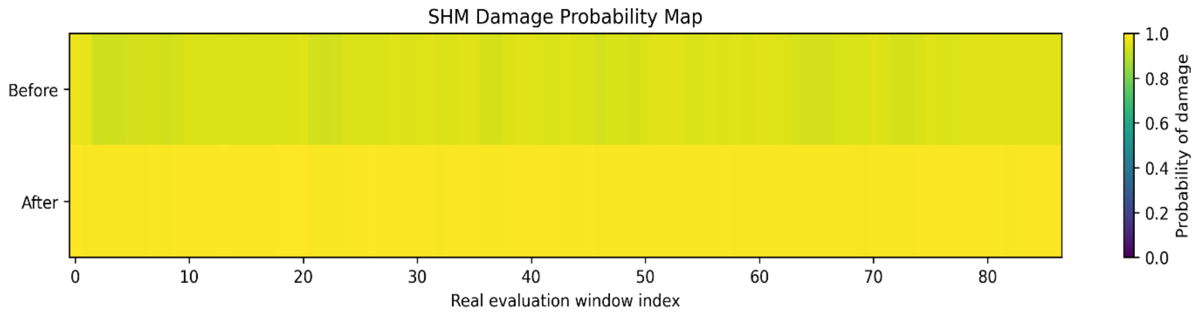

Figure 82. Damage Probability Trend Before and After Transfer Learning Across Signal Windows

The heatmap representation (**Figure 82**) shows a clear shift toward higher and more uniform probability values after adaptation. Prior to transfer learning, the predictions exhibit noticeable variability across windows, indicating sensitivity to the distribution mismatch between simulated and experimental data. In contrast, after single-shot adaptation, the predicted probabilities become more homogeneous and consistently close to unity, suggesting improved alignment with the real data distribution.

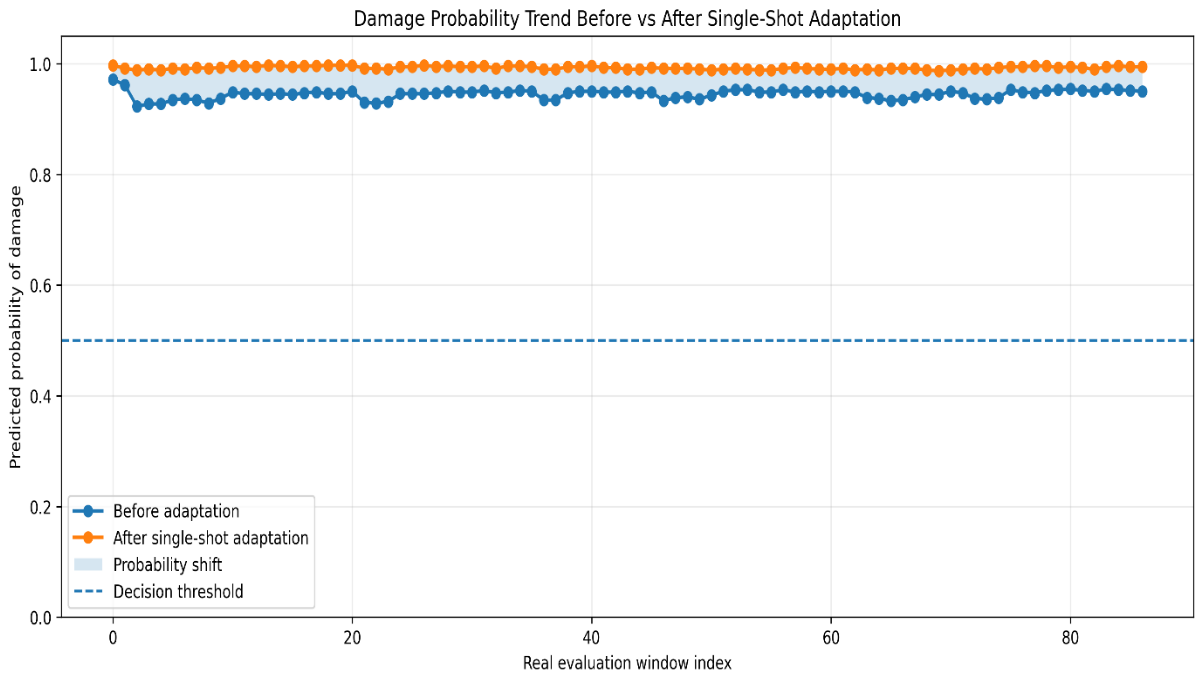

Figure 83. Distribution of Predicted Damage Probability Before and After Transfer Learning

This behavior is further confirmed in the probability trend plot (**Figure 83**). Before adaptation, the Extra Trees model produces predictions within a slightly lower and more variable range (approximately 0.92–0.93), reflecting residual uncertainty. After adaptation, the predictions increase systematically and stabilize near 1.00 across all evaluation windows. The reduction in variance and the upward shift in predicted probabilities indicate a substantial improvement in model confidence and robustness.

It is worth noting that the Extra Trees model, as a tree-based ensemble method, does not support classical parameter fine-tuning. Therefore, the observed improvement is attributed to effective model adaptation through retraining on limited experimental data, which enables better alignment with the target domain.

Importantly, all predictions remain well above the decision threshold both before and after adaptation, confirming correct damage detection. However, the post-adaptation results demonstrate enhanced reliability, as evidenced by reduced fluctuations and a narrower prediction band.

Overall, these results confirm that the proposed single-shot transfer learning strategy effectively reduces the simulation-to-real domain gap, enabling the Extra Trees model to achieve more stable, confident, and consistent damage detection under real experimental conditions.

- **Damage Location**

The performance of the Random Forest model for damage localization before and after single-shot transfer learning is presented in Figures 83 and 84.

Figure 84 shows the predicted probability assigned to the true damage location across multiple evaluation windows. Before domain adaptation, the model assigns very low confidence to the correct location, with probability values remaining close to zero (typically below 0.1). This indicates a significant domain mismatch between the simulated training data and real experimental measurements, resulting in poor localization capability.

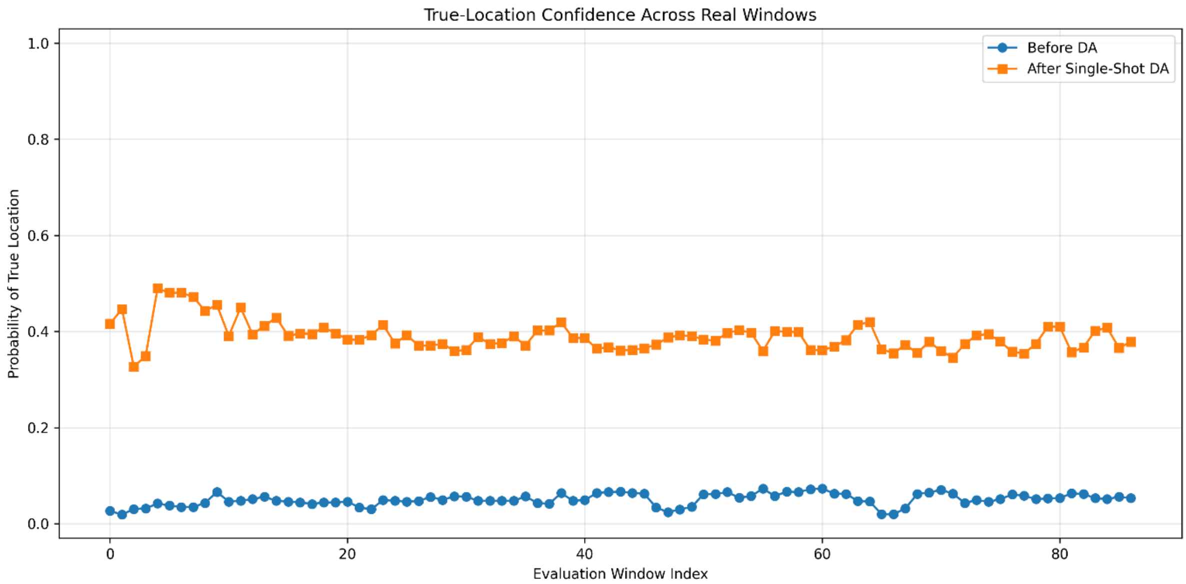

Figure 84. Classification Performance Improvement Before and After Transfer Learning for Damage Location

After applying single-shot transfer learning, a substantial improvement is observed. The predicted probability of the true location increases consistently to a significantly higher range (approximately 0.35–0.54), with reduced variability across windows. This shift indicates that the model has successfully adapted its decision boundaries to better capture the underlying patterns in the experimental data. The increase in confidence and stability demonstrates improved alignment with the target domain.

Table 15. Quantitative Comparison of Damage Localization Performance Before and After TL

Metric	Before TL	After TL	Improvement
Accuracy	0.462	0.921	0.46
F1_weighted	0.422	0.951	0.53
MAE [mm]	210.769	45.222	165.548
RMSE [mm]	531.616	97.125	434.404

Table 15 presents a quantitative comparison of the Random Forest model performance for damage localization before and after applying single-shot transfer learning.
A substantial improvement is observed across all evaluation metrics. The classification accuracy increases significantly from 0.462 to 0.921, while the F1-weighted score improves from 0.422 to 0.951, indicating a major enhancement in the model's ability to correctly identify damage locations with balanced performance across classes.
In terms of regression accuracy, the localization errors are dramatically reduced. The mean absolute error (MAE) decreases from approximately 210.8 mm to 45.2 mm, while the root mean square error (RMSE) drops from 531.6 mm to 97.2 mm. These reductions correspond to improvements of over 165 mm in MAE and 434 mm in RMSE, highlighting a significant increase in spatial prediction accuracy.
Overall, the results demonstrate that the proposed single-shot transfer learning strategy substantially improves both classification and localization precision, effectively reducing the simulation-to-real domain gap and enabling more accurate damage localization under real experimental conditions.

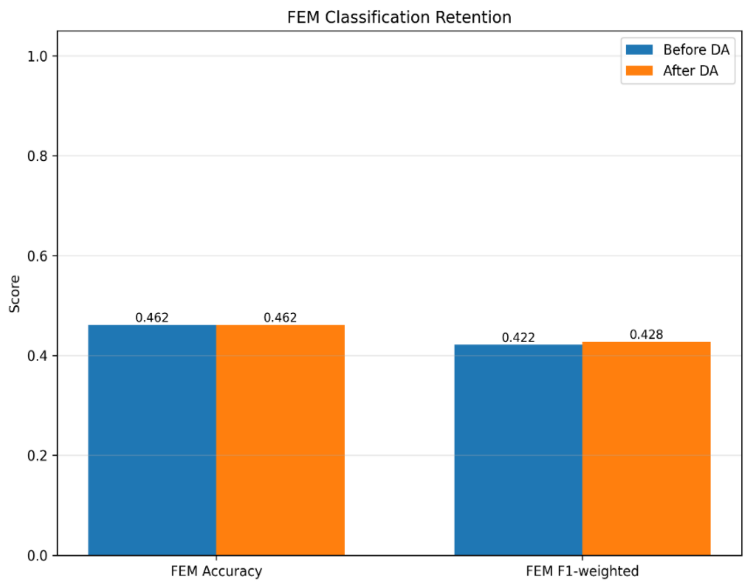

Figure 85. Retention of FEM-Based Classification Performance Before and After TL

Figure 85 shows that the Random Forest model retains its FEM-based classification performance after transfer learning. The accuracy remains unchanged (0.462), while the F1-weighted score shows a slight improvement (from 0.422 to 0.428). This indicates that incorporating experimental data does not degrade performance on the simulation domain. Combined with the improvements observed in localization on real data, these results confirm that the proposed single-shot transfer learning approach enhances model adaptability without introducing catastrophic forgetting.

- **Damage Severity**

Figure 86 presents the distribution of predicted crack severity on real test data using the Gradient Boosting model before and after single-shot transfer learning.

Before adaptation, the model exhibits a strong bias toward overestimation, with predicted severity values concentrated around approximately 35–40 mm, far from the true value of 7.4 mm. This indicates a significant domain mismatch between the simulated training data and real experimental conditions.

After applying transfer learning, the predicted severity distribution shifts dramatically toward the true value, with predictions tightly clustered around 7–8 mm. The mean prediction improves from 37.87 mm to 7.21 mm, closely matching the ground truth and indicating a substantial reduction in systematic error.

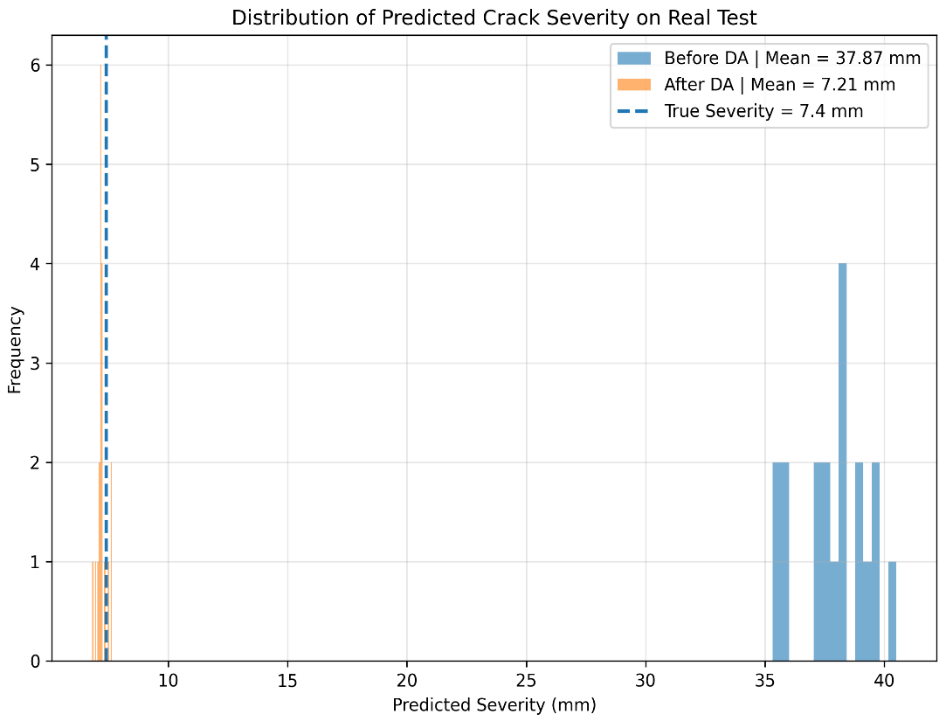

Figure 86. Improvement in Crack Severity Prediction After Single-Shot Transfer Learning

This shift demonstrates that the proposed single-shot transfer learning approach effectively corrects the bias in severity estimation and enables the model to capture the true damage scale in the experimental domain. In addition, the reduced spread of the predictions suggests improved consistency and robustness.

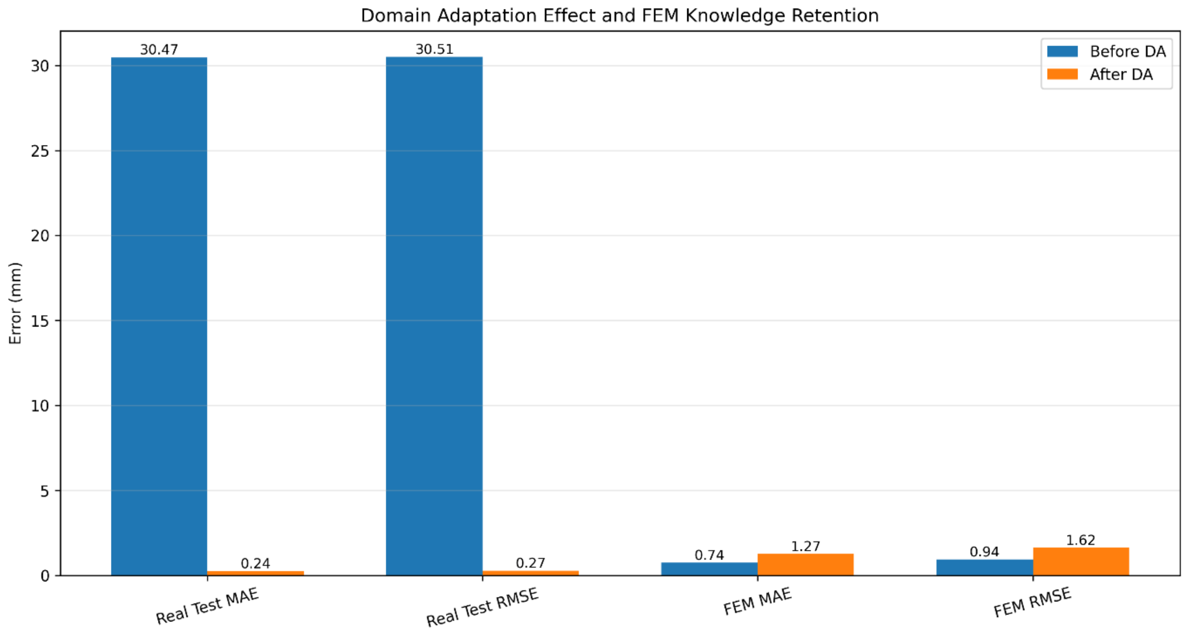

Figure 87. Trade-Off Between Real-World Improvement and FEM Knowledge Retention in Crack Severity Prediction

Figure 87 compares the error metrics for crack severity prediction on both real test data and the FEM domain before and after transfer learning using the Gradient Boosting model.

A substantial improvement is observed on real data, where the MAE decreases dramatically from approximately 30.47 mm to 0.24 mm, and the RMSE drops from 30.51 mm to 0.27 mm, indicating a near-perfect alignment with the true severity values after adaptation. This confirms that transfer learning effectively corrects the strong overestimation bias observed before adaptation.

On the FEM domain, a slight increase in error is observed (MAE from 0.74 mm to 1.27 mm, RMSE from 0.94 mm to 1.62 mm). However, these values remain very low, indicating that the model largely preserves its performance on simulated data despite adaptation to the experimental domain.

Overall, the results demonstrate that the proposed single-shot transfer learning strategy achieves a favorable trade-off: it significantly improves prediction accuracy on real data while maintaining acceptable performance on the source domain. This highlights the effectiveness of the approach in reducing domain discrepancy without severe loss of previously learned knowledge.

3.3.1. Transfer Learning Results and Discussion

The performance of the simulation-trained machine learning models was further evaluated on experimental data using a transfer learning-inspired adaptation strategy. Due to the nature of the selected algorithms (Extra Trees, Random Forest, and Gradient Boosting), adaptation was implemented through retraining using features extracted from experimental measurements (Single Shot Transfer Domain Transfer Learning). This approach enables the models to adjust to the target domain while preserving the structural patterns learned from simulation data.

- **Damage Detection**

For damage detection, the adapted models produced more stable and consistent predictions across different signal windows compared to the pre-adaptation results. The predicted damage probabilities exhibited reduced variability and increased confidence after incorporating experimental data, indicating improved alignment between the model and the experimental feature distribution.
Despite these improvements, it should be noted that the evaluation is based on a limited experimental dataset corresponding to a specific damage condition.

- **Damage Localization**

A substantial improvement was observed in damage localization performance following model adaptation. The classification accuracy and F1-weighted scores increased significantly, while the mean absolute error (MAE) of the predicted damage location was markedly reduced.
These results indicate that incorporating experimental data enables the model to better capture spatial variations in structural response. In particular, the adaptation process effectively refines the decision boundaries initially learned from simulation, allowing improved localization accuracy in the target domain.
However, given that the adaptation is based on a limited experimental dataset, the reported performance should be interpreted as effective domain adaptation for the specific experimental condition rather than evidence of robust generalization across varying damage locations or structural configurations.

- **Damage Severity**

For damage severity estimation, model adaptation resulted in a pronounced reduction in prediction error and improved agreement between predicted and true severity values. The post-adaptation predictions exhibit significantly reduced bias and lower dispersion compared to the pre-adaptation results.
This behavior indicates that retraining with experimental data enhances the model’s ability to capture the relationship between vibration-based features and damage magnitude. In particular, the reduction in absolute error suggests improved sensitivity to real-world structural characteristics.
Nevertheless, the performance remains dependent on the available experimental data, and variability may increase under different damage levels or structural conditions.

4. Conclusion

This study presented a vibration-based structural health monitoring framework for damage detection, localization, and severity estimation in an IPE160 steel beam by combining finite element simulation, feature extraction, machine learning, and transfer learning. A calibrated three-dimensional ANSYS model was developed and validated against experimental modal data through comparison of natural frequencies and mode shapes. The close agreement between numerical and experimental results confirmed that the finite element model provided a reasonable representation of the beam's dynamic behavior and could therefore be used to generate systematic damage scenarios.

Using the validated numerical model, a dataset covering multiple crack locations and severities was generated and processed to extract frequency-domain, mode-shape, and selected time-domain features. These features were then used to train and compare a diverse set of machine learning algorithms for damage detection, localization, and severity estimation. The results showed that feature selection strongly influenced model performance. Natural frequencies were highly effective for damage detection, while mode shapes improved localization performance. The combination of multiple feature types produced the most informative representation and achieved the most balanced performance across all tasks.

Among the evaluated models, tree-based ensemble methods demonstrated the most reliable and consistent behavior. Extra Trees achieved near-perfect damage detection performance, Random Forest provided the best localization capability, and Gradient Boosting achieved the strongest severity estimation results. The evaluation on unseen numerical models further confirmed good generalization capability, although lower-severity damage remained more difficult to estimate accurately. The comparison between raw features and Δ features also showed that the most effective feature representation depends on the specific prediction task.

In addition to predictive performance, computational efficiency was evaluated through model training time analysis. The results showed noticeable differences among the considered algorithms. Ensemble tree-based models achieved an effective balance between accuracy and computational cost, whereas more complex models (GB,BP) generally required longer training times without providing a proportional improvement in prediction performance. These findings indicate that ensemble methods are particularly suitable for practical SHM applications where both accuracy and computational efficiency are important considerations.

Feature importance analysis was performed for the best-performing models in each task to identify the vibration features contributing most to the predictions. The results revealed that frequency-domain features, particularly natural frequencies and mode-shape-related parameters, consistently exhibited the highest importance scores. This observation confirms the physical relationship between structural damage and changes in modal characteristics. Time-domain features contributed additional information,

especially when combined with frequency-domain features, leading to improved performance in damage localization and severity estimation.
A key contribution of this work is the application of transfer learning to bridge the gap between simulation-based and experimental data. Using a single-shot domain adaptation strategy, models trained on 259 finite element damage scenarios were adapted using only one experimentally measured damage scenario. The adaptation process resulted in substantial improvements in damage localization and severity estimation while preserving most of the knowledge acquired from the numerical domain. This demonstrates a favorable trade-off between adaptation and knowledge retention, where the model becomes significantly more accurate for the target experimental condition without severe degradation of its simulation-based predictive capability. The results therefore indicate that transfer learning can provide a practical mechanism for incorporating limited experimental information into simulation-trained SHM models.
Despite these promising findings, several limitations should be acknowledged. The transfer learning validation was performed using a single experimental damage configuration, and therefore the observed improvements should be interpreted as successful adaptation to a specific target scenario rather than evidence of broad real-world generalization. In addition, the experimental dataset was limited to a single beam configuration, and environmental and operational variability effects were not explicitly considered. Consequently, further validation using a wider range of damage conditions, structural configurations, and environmental scenarios is required to fully assess the robustness and scalability of the proposed framework.
Overall, the study demonstrates that the combination of calibrated finite element modeling, machine learning, and transfer learning provides an effective approach for vibration-based damage assessment in steel beams. The findings show that simulation-trained models can be successfully adapted using minimal experimental data while retaining previously learned structural knowledge. Moreover, the proposed framework establishes a foundation for incremental model updating, enabling future experimental measurements to be incorporated progressively as they become available and thereby supporting the long-term deployment of data-driven SHM systems in real-world structures.

Key Results

- The calibrated finite element (FE) model developed in ANSYS showed good agreement with the experimental modal properties, providing a reliable basis for generating 259 damage scenarios covering multiple crack locations and severity levels.
- Features extracted from the FE simulations enabled accurate damage detection, localization, and severity estimation on previously unseen numerical cases. The combination of frequency, mode-shape, RMS, PV and CV provided the most balanced overall performance.
- Among the evaluated machine learning algorithms, ensemble tree-based methods demonstrated the most consistent and robust behavior. Extra Trees achieved the best

damage detection performance, Random Forest provided the highest localization accuracy, and Gradient Boosting produced the most accurate severity predictions.

• Feature importance analysis confirmed that natural frequencies and mode-shape-related features contain the most relevant information for vibration-based damage assessment, providing physical interpretability in addition to predictive accuracy.

• Training time analysis showed that ensemble tree-based models provide a favorable balance between prediction accuracy and computational efficiency, making them attractive candidates for practical SHM deployment.

• Single-shot domain adaptation was successfully performed using only one experimentally measured damage scenario, demonstrating the feasibility of adapting simulation-trained models under severe experimental data scarcity.

• For damage detection, model adaptation increased prediction confidence from approximately 0.92–0.93 to nearly 1.00 while reducing variability across signal windows.

• For damage localization, the classification accuracy improved from 0.462 to 0.921 and the F1-weighted score increased from 0.422 to 0.951. The mean absolute error (MAE) was reduced from 210.8 mm to 45.2 mm, indicating a substantial improvement in spatial prediction accuracy.

• For damage severity estimation, the MAE decreased from 30.47 mm to 0.24 mm, demonstrating a significant improvement in agreement between predicted and actual damage severity.

• The results demonstrate a favorable trade-off between adaptation and knowledge retention, whereby the model becomes highly accurate for the target experimental scenario while maintaining acceptable performance across the FEM-generated damage scenarios.

• The proposed transfer learning framework provides a foundation for incremental model updating, enabling new experimental measurements to be incorporated progressively while preserving previously acquired structural knowledge.

• Since the experimental validation was performed using a single damage configuration, the reported improvements should be interpreted as successful case-specific domain adaptation rather than evidence of full generalization to arbitrary real-world damage scenarios.

5. References

[1] Wang, G., & Ke, J. (2024). Literature review on the structural health monitoring (SHM) of sustainable civil infrastructure: an analysis of influencing factors in the implementation. *Buildings, 14*(2), 402.

[2] Zhang, B., Ren, Y., He, S., Gao, Z., Li, B., & Song, J. (2025). A review of methods and applications in structural health monitoring (SHM) for bridges. *Measurement, 245*, 116575.

[3] Li, J., Weng, S., Wu, Q., Meng, M., Li, X., & Zhu, H. (2025). A Review of Vibration-Based Civil Structural Damage Detection Using Deep Learning: From 2016 to 2025. *International Journal of Structural Stability and Dynamics*, 2730001.

[4] Sonbul, O. S., & Rashid, M. (2023). Algorithms and techniques for the structural health monitoring of bridges: Systematic literature review. *Sensors, 23*(9), 4230.

[5] Wang, Q. A., Dai, Y., Ma, Z. G., Wang, J. F., Lin, J. F., Ni, Y. Q., ... & Yan, J. R. (2024). Towards high-precision data modeling of SHM measurements using an improved sparse Bayesian learning scheme with strong generalization ability. *Structural Health Monitoring, 23*(1), 588–604.

[6] Zhang, C. D., & Xu, Y. L. (2016). Structural damage identification via multi-type sensors and response reconstruction. Structural Health Monitoring, 15(6), 715-729.

[7] Rytter, A. (1993). *Vibrational based inspection of civil engineering structures*.

[8] Aktan, E., Bartoli, I., Glišić, B., & Rainieri, C. (2024). Lessons from bridge structural health monitoring (SHM) and their implications for the development of cyber-physical systems. *Infrastructures, 9*(2), 30.

[9] He, Y., Yang, B., Jin, W., Zhang, L., & Chen, H. (2025). A parallel neural network based structural anomaly detection: Leveraging time-frequency domain features. *Neurocomputing, 634*, 129907.

[10] Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemates, D. W., Nadler, B. R., & Czarnecki, J. J. (2003). A review of structural health monitoring literature: 1996–2001. *Los Alamos National Laboratory, USA, 1*(16), 10–12989.

[11] Benfenati, L., Pagliari, D. J., Zanatta, L., Velez, Y. A. B., Acquaviva, A., Poncino, M., ... & Burrello, A. (2025). Foundation models for structural health monitoring. *IEEE Transactions on Sustainable Computing*.

[12] Pandey, A. K., Biswas, M., & Samman, M. M. (1991). Damage detection from changes in curvature mode shapes. *Journal of Sound and Vibration, 145*(2), 321–332.

[13] Worden, K., & Manson, G. (2007). The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365*(1851), 515–537.

[14] Hakim, S. J. S., Razak, H. A., & Ravanfar, S. A. (2015). Fault diagnosis on beam-like structures from modal parameters using artificial neural networks. *Measurement, 76*, 45–61.

[15] Fiandaca, D., Di Matteo, A., Patella, B., Moukri, N., Inguanta, R., Llort, D., ... & Pirrotta, A. (2022). An integrated approach for structural health monitoring and damage detection of bridges: An experimental assessment. *Applied Sciences, 12*(24), 13018.

[16] Pawar, P. M., Venkatesulu Reddy, K., & Ganguli, R. (2007). Damage detection in beams using spatial Fourier analysis and neural networks. *Journal of Intelligent Material Systems and Structures, 18*(4), 347–359.

[17] Lee, J., & Kim, S. (2007). Structural damage detection in the frequency domain using neural networks. *Journal of Intelligent Material Systems and Structures, 18*(8), 785–792.

[18] Staszewski, W. J., Worden, K., & Tomlinson, G. R. (1997). Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition. *Mechanical Systems and Signal Processing, 11*(5), 673–692.

[19] Chang, C. C., & Chen, L. W. (2005). Detection of the location and size of cracks in the multiple cracked beams by spatial wavelet-based approach. *Mechanical Systems and Signal Processing, 19*(1), 139–155.

[20] Ye, X., Cao, Y., Liu, A., Wang, X., Zhao, Y., & Hu, N. (2023). Parallel convolutional neural network toward high efficiency and robust structural damage identification. *Structural Health Monitoring, 22*(6), 3805–3826.

[21] Abdeljaber, O., & Avci, O. (2016). Nonparametric structural damage detection algorithm for ambient vibration response: utilizing artificial neural networks and self-organizing maps. *Journal of Architectural Engineering, 22*(2), 04016004.

[22] Zhao, M., Sadhu, A., & Capretz, M. (2022). Multiclass anomaly detection in imbalanced structural health monitoring data using convolutional neural network. *Journal of Infrastructure Preservation and Resilience, 3*(1), 10.

[23] Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing, 108*, 33–47.

[24] García-Pérez, A., Coppotelli, G., & Alonso, G. (2021). Impact dynamic response of low damped structures subjected to impulse base accelerations. *Mechanical Systems and Signal Processing, 157*, 107740.

[25] Chaphalkar, S. P., Khetre, S. N., & Meshram, A. M. (2015). Modal analysis of cantilever beam structure using finite element analysis and experimental analysis. *American Journal of Engineering Research, 4*(10), 178–185.

[26] Kralovec, C., & Schagerl, M. (2020). Review of structural health monitoring methods regarding a multi-sensor approach for damage assessment of metal and composite structures. *Sensors, 20*(3), 826.

[27] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*.

[28] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (Vol. 14, No. 2, pp. 1137–1145).

[29] Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of Machine Learning Research, 11*(6).

[30] Bao, Y., & Li, H. (2021). Machine learning paradigm for structural health monitoring. *Structural Health Monitoring, 20*(4), 1353–1372.

[31] Gomez-Cabrera, A., & Escamilla-Ambrosio, P. J. (2022). Review of machine-learning techniques applied to structural health monitoring systems for building and bridge structures. *Applied Sciences, 12*(21), 10754.

[32] Bandara, R. P., Chan, T. H., & Thambiratnam, D. P. (2014). Structural damage detection method using frequency response functions. Structural Health Monitoring, 13(4), 418-429.

[33] Puntambekar, R., Vyas, P., Thakkar, A., & Patel, D. (2025). A survey of machine learning and deep learning methods for vibration-based bearing fault diagnosis: The need, challenges, and potential future research directions. *Neurocomputing*, 131628.

[34] Dervilis, N., Choi, M., Taylor, S. G., Barthorpe, R. J., Park, G., Farrar, C. R., & Worden, K. (2014). On damage diagnosis for a wind turbine blade using pattern recognition. *Journal of Sound and Vibration, 333*(6), 1833–1850.

[35] Altabey, W. A., & Noori, M. (2020). Deep Learning for Automated Damage Detection: A Novel Algorithm in CNN Family for Faster and Accurate Damage Identification. In *Experimental Vibration Analysis for Civil Structures* (pp. 13–25). CRC Press.

[36] Barzegar, V., Laflamme, S., Hu, C., & Dodson, J. (2022). Ensemble of recurrent neural networks with long short-term memory cells for high-rate structural health monitoring. *Mechanical Systems and Signal Processing, 164*, 108201.

[37] Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning* (pp. 1050–1059). PMLR.

[38] Farrar, C. R., & Worden, K. (2012). *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons.

[39] Coletta, G., Miraglia, G., Gardner, P., Ceravolo, R., Surace, C., & Worden, K. (2020, July). A transfer learning application to FEM and monitoring data for supporting the classification of structural condition states. In *European Workshop on Structural Health Monitoring* (pp. 947–957). Springer International Publishing.

[40] Fidma, M. A., Schmidt, F., & Bercher, J. F. (2025). Deep Learning Techniques for Damage Detection in Structural Health Monitoring: A Review. *Structural Engineering International*, 1–17.

[41] Lim, S. Y., Seo, H. B., & Lee, S. Y. (2025). Simulation-Driven Deep Transfer Learning Framework for Data-Efficient Prediction of Physical Experiments. *Mathematics, 13*(23), 3884.

[42] Chen, H., Chen, J., Li, Z., & Liu, Y. A. (2025). Deep Net-based Transfer Learning Framework for Structural Digital Twins: Bridging the Sim-to-Real Gap with Delta-Correction Strategy.

[43] Ahmed, M. Z. (2025, December). Revolutionizing Infrastructure Resilience: AI-Driven Approaches to Structural Health Monitoring. In *2025 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)* (pp. 1–4). IEEE.