

Ensemble Feature Selection for Clustering Damage Modes in Carbon Fiber-Reinforced Polymer Sandwich Composites Using Acoustic Emission

Abdulkadir Gulsen,* Burak Kolukisa, Umut Caliskan, Burcu Bakir-Gungor, and Vehbi Cagri Gungor

Acoustic emission (AE) serves as a noninvasive technique for real-time structural health monitoring, capturing the stress waves produced by the formation and growth of cracks within a material. This study presents a novel ensemble feature selection methodology to rank features highly relevant with damage modes in AE signals gathered from edgewise compression tests on honeycomb-core carbon fiber-reinforced polymer. Two distinct features, amplitude and peak frequency, are selected for labeling the AE signals. An ensemble-supervised feature selection method ranks feature importance according to these labels. Using the ranking list, unsupervised clustering models are then applied to identify damage modes. The comparative results reveal a robust correlation between the damage modes and the features of counts and energy when amplitude is selected. Similarly, when peak frequency is chosen, a significant association is observed between the damage modes and the features of partial powers 1 and 2. These findings demonstrate that, in addition to the commonly used features, other features, such as partial powers, exhibit a correlation with damage modes.

materials in mechanical applications has often required the incorporation of fibers, whiskers, particles, and layers. The specific type of matrix that requires reinforcement depends on its intended application.^[1] Fiber-reinforced composites have found widespread application across various industries, including aerospace, energy, mechanical manufacturing, and automotive, owing to their superior mechanical properties, lightweight nature, and ease of manufacturing. These advanced properties come from their complex hierarchical and/or hybrid microstructures. However, this complexity also makes them harder to use safely. Mechanical and thermal stresses can compromise the integrity of the more vulnerable elements within a composite, leading to potential damage during operational use. Therefore, detecting such damage is crucial to preventing

premature deterioration of composite structures.^[2] A key aspect of understanding and predicting the behavior of composites under stress involves examining their response to lateral loading.^[3]

Lateral loading significantly impacts the durability of carbon fiber composites by causing various types of damage, such as matrix cracking, fiber-matrix debonding, delamination, and fiber breakage.^[4] As lateral pressures are applied, microcracks develop and propagate within the resin matrix, leading to a reduction in the composite's stiffness and strength. The failure of the bond between the fibers and the matrix, often associated with matrix cracking, further reduces the load-carrying capacity of the composite. Delamination, which involves the separation of layers along with fiber breakage, further compromises the stability and load-bearing capability of the composite, particularly under high stress or extensive damage.^[5] These complex damage dynamics underscore the critical need for rigorous health monitoring to ensure the structural integrity and longevity of composites.

Nondestructive testing methods such as computer tomography, visual testing, microscopy, ultrasonic testing, and acoustic emission (AE) are available for monitoring the health of composite materials.^[6] AE is known for its capability to detect damage at the exact moment it happens, including on very small scales.^[7] This method monitors acoustic waves that result from


1. Introduction

The main objective of composite materials is to combine two or more materials to produce unique characteristics. Building up

A. Gulsen, B. Kolukisa, B. Bakir-Gungor
Department of Computer Engineering
Abdullah Gül University
38080 Kayseri, Turkey
E-mail: abdulcadir.gulsen@agu.edu.tr

U. Caliskan
Department of Mechanical Engineering
Erciyes University
38280 Kayseri, Turkey

V. C. Gungor
Network Technologies
Turkcell
34854 Istanbul, Turkey

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adem.202400317>.

© 2024 The Author(s). Advanced Engineering Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/adem.202400317

microstructural changes, offering important insights into the sources of damage.^[8] Interpreting these signals, especially in materials with complex structures, is challenging due to the intricate nature of the data and structural variations. This highlights the necessity of precise signal classification to improve the performance of monitoring systems.^[9]

Machine learning methods are rapidly evolving in the field of damage characterization in composite structures and can be broadly categorized into supervised and unsupervised learning approaches.^[8] Unsupervised learning is particularly useful when the damage classes are unknown, as it groups similar AE signals to help characterize damage modes. However, significant challenges include linking these groups to specific damage modes and determining the optimal number of clusters.^[8] Adding more in situ observations, like computer tomography,^[10] microscopy,^[11] digital image correlation,^[12] or simulating AE patterns^[13] to make a labeled dataset for supervised learning can often help with these problems. In contrast, supervised learning requires prior knowledge of damage classes and involves training models on labeled data. Despite its effectiveness, this method faces the additional obstacle of the complex task of collecting data related to the cracking of individual components within a composite.

In this study, the AE technique is employed to investigate damage modes in carbon fiber-reinforced polymer (CFRP) composites. During compression testing, both time and frequency features of AE signals are extracted. An ensemble-supervised feature selection approach is then utilized to identify relevant AE features for damage characterization, with a focus on the amplitude and peak frequency (PF) reference intervals associated with specific damage modes. These selected AE features are subsequently classified using clustering models, and the performance of these models is comparatively analyzed. This methodology provides a complete overview of how to accurately describe damage in CFRP composites. It showed how AE techniques can be combined with advanced machine learning methods to improve damage characterization and description. The novelty of this study lies in three primary areas: 1) The examination of crack formation and propagation in carbon fiber composite sandwiches under edgewise loading through AE measurements. 2) The application of an ensemble supervised feature selection method that utilizes reference intervals for amplitude and PF features as established in existing literature. 3) A comprehensive analysis of the most effective features through the evaluation of clustering model outcomes.

The article's main contributions are summarized as follows: 1) Four individual and one ensemble supervised feature selection methods are employed to identify key AE features, investigating the reliability and generalizability of the ranges defined in the literature. 2) Five clustering models are applied to classify AE features, and their results are aggregated to more accurately define the damage characterization associated with AE signals. 3) Our experimental findings reveal that, in addition to commonly used features, partial powers also correlate with damage modes. Employing clustering models based on selected features and evaluating their efficacy provides profound insights into the damage mechanisms associated with AE signals.

This study is organized as follows: Section 2 reviews the literature on AE techniques for classifying damage characterization. Section 3 describes the experimental setup and the composite

structures used. Section 4 outlines ensemble-supervised feature selection and clustering methods. Section 5 offers a comparative analysis of the results and discusses the findings. Finally, Section 6 concludes the article with a concise summary and conclusion.

2. Related Work

The literature contains a significant amount of research on the mathematical, analytical, and experimental assessment of damage mechanisms in CFRP materials subjected to tensile, bending, and compression testing.^[14] Numerous studies have investigated the various damage modes associated with AE waves generation.^[15] The extraction of AE features from waveforms is a widely used method for damage classification. Within the existing research, many efforts focus on either a single-parameter approach or employ classifiers to link specific damage modes to AE features.^[16] In composites, particularly CFRP, common AE features for damage assessment include amplitude, PF, counts, rise time, energy, and average frequency (AF).^[14,17] Most studies utilize unsupervised learning models such as K-Means,^[18,19] fuzzy C-means (FCM),^[20,21] K-Means++,^[22,23] and Gaussian mixture models (GMMs)^[24,25] for clustering these AE signals.

Before performing clustering analysis, it is essential to select the relevant AE features to accurately distinguish damage characterization. In the literature, various feature selection methods, including the Laplacian score,^[26,27] correlation coefficient,^[23,26] Davies–Bouldin index,^[28,29] and principal component analysis (PCA),^[30,31] are employed to identify the essential features. In **Table 1**, the most recent studies that use AE for damage mode classification of composites are compared and summarized. In **trg. [27]**, authors used Laplacian scores for feature selection and K-Means clustering to discover feature relationships. Researchers discovered a robust association between the amplitude and the frequency centroid (FC). The study in **ref. [19]** utilized AE features, which include count, signal strength (SS), duration, rise time, and energy, as training data, and used amplitude as the target in a regression process, selecting features based on shapley additive explanations (SHAP) values. In **ref. [26]**, authors applied correlation analysis and Laplacian scores to select optimal features for K-Means++ clustering. Despite these advancements, selecting the optimal feature remains challenging due to the incomplete utilization of all AE features and various clustering models.

The literature indicates that different damage mechanisms in composite materials typically generate AE signals with distinct, almost unique features. For example, matrix cracking is typically characterized by low amplitude, low frequency, long duration, long rise time, and large counts. In contrast, we can identify delamination by its intermediate amplitude, low frequency, and extremely long duration. Conversely, fiber breakage is associated with signals that have high amplitude, high frequency, and a short rise time. These results show that amplitude and PF are the main AE features used to group AE signals.^[32,33] There is a consensus in the literature regarding the correlation between these features and specific damage mechanisms in composite materials.^[14,19,25,34–37] This agreement reinforces the reliability

Table 1. A summary of research using AE for damage mode classification of composites.

References	Year	Material	Type of the Test	Feature Selection	Clustering Models	Best Features	Approach
[18]	2012	CFRP	Double cantilever beam	Laplacian score	K-Means	A, RT, E, C, AF, D	Harmony search K-Means clustering was applied to defined features.
[20]	2012	Glass fiber reinforced polyester (GFRP)	Double cantilever beam	PCA	FCM	–	FCM and PCA were employed to find the correlation.
[26]	2014	GFRP	Tensile	Laplacian score, PCA	K-Means++	A, PF, FC, (RT/A)	Laplacian and PCA utilized to find the best features and applied K-Means++ to find the clusters.
[22]	2016	CFRP	Tensile	PCA	K-Means++	A, PF	K-Means++ and PCA utilized to find the clusters.
[49]	2021	CFRP	Tensile	PCA	FCM	A, IF, RF	FCM and PCA applied to find the correlation.
[27]	2022	CFRP	Tensile	Laplacian score	K-Means	A, FC	Laplacian scores were used to identify the best features, and K-Means clustering was employed to establish their correlation.
[19]	2023	GFRP	Tensile	SHAP	K-Means	A, SS, D, CTP	Ensemble learning models and SHAP values are utilized to find the best features and K-Means is applied to find the correlation.
PP	2024	CFRP	Compression	ANOVA F-test, CS, MI, ridge	K-Means, MiniBatch K-Means, balanced iterative reducing and clustering using hierarchies (BIRCH), agglomerative, GMM	C, E, PP1, PP2	Ensemble supervised feature selection is utilized to find the best features and K-Means, MiniBatch K-Means, BIRCH, agglomerative, GMM clustering models are applied to find the correlation.

of using these AE features to effectively identify and characterize various types of damage.

3. Experimental Section

3.1. Materials

Sandwich panel members consist of core material, top and bottom face sheets, and adhesive. CFRP prepreg laminates were prepared with VTP H 300 epoxy matrix and high-strength carbon fibers with a density of 210 g m^{-2} . The nominal thickness of the ply is 0.210 mm. The core material constituting the sandwich structure is polypropylene honeycomb. The number of plies, the stacking sequence, and other dimensional details of the sandwich specimen are given in **Table 2**. Composite face sheets were produced by the hot pressing method by stacking prepreps and placing them in the mold. The production was completed under appropriate pressure and temperature conditions and cut in sandwich structure geometry. After the honeycomb material was cut to the $100 \times 150 \text{ mm}$ dimensions, the composite plates were ready for sandwich-making. The composite plates were cleaned with ethanol. After the surface of the composite plates

Table 2. Sandwich specimen geometry and characteristics.

Length [mm]	Width [mm]	Thickness [mm]	Number of plies
Face-sheets-CFRP150	100	2,1	10 [0/90°]
Core-polypropylene honeycomb 150	100	15	–

was cleaned with ethanol, the upper surfaces of the core materials were cleaned with a brush, respectively. Epoxy was applied to the face-sheets surface, bonded with core material, and passed to the mold stage for curing.

3.2. Testing Methods

The mechanical test was carried out following ASTM C364—Standard Test Method for Edgewise Compressive Strength of Sandwich Constructions.^[38] The compression test system and AE equipment are shown in **Figure 1b**. The compression load was applied under displacement-controlled mode at a constant crosshead displacement of 1 mm min^{-1} . **Figure 1b** shows AE sensor position and deformed specimen. According to the ASTM standard, the test was completed after at least five repetitions. While the success of sandwich structures against loads such as compression and impact is obvious, the mechanical behavior and damage mechanism in the direction of weakness constitute a behavioral limit. In this respect, the damage mechanism and crack modes were investigated by applying an edgewise compression test to the sandwich structure.

4. AE Analysis

In this research, 19 features are derived from the AE waveforms for analysis, with each feature detailed in **Table 3**. **Figure 2** illustrates a typical AE signal with commonly used features, which have been extensively discussed in the literature.^[14,15,39] **Figure 3** provides a visual representation of the methodology.

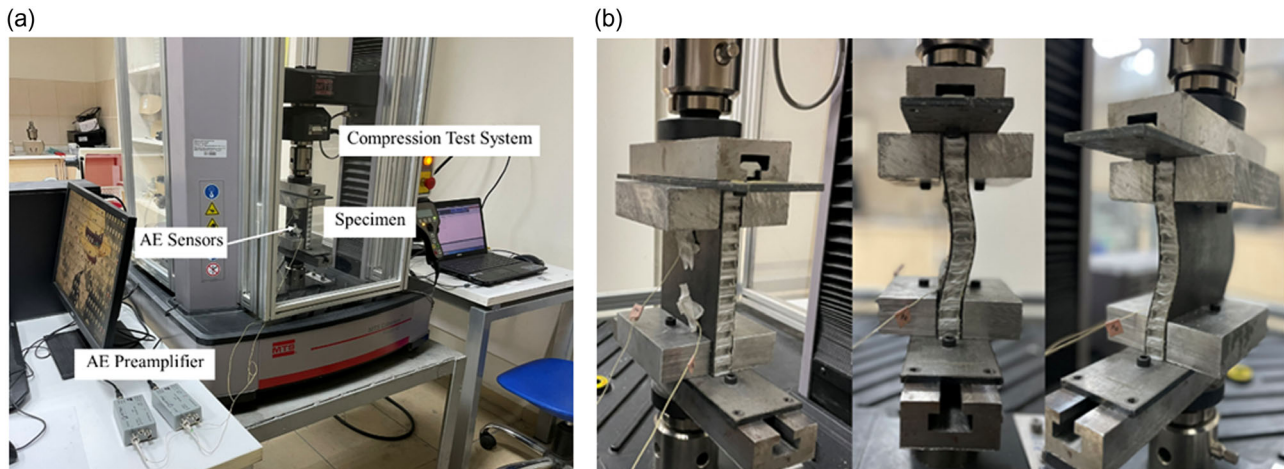


Figure 1. a) Setup of the compression test system and AE equipment, including sensors, preamplifier, and the compression test system with the specimen. b) Position of the AE sensors and the deformed specimen after the compression test.

Table 3. Descriptions of the features extracted from the AE signals.

Feature	Description
Counts [C]	The aggregate number of times the AE signal exceeds the predefined threshold throughout the measurement period [count].
Energy [E]	The total energy of the AE signal from the time it exceeds the detection threshold until the signal ends [aJ].
Duration [D]	The measure is the time from when the signal first crosses the detection threshold to when it last [μs].
Counts To Peak [CTP]	The number of signal counts observed from the initial crossing of the threshold to the point where the signal amplitude is at its maximum [counts].
Rise time [RT]	The moment when the signal exceeds a predefined threshold level [μs].
Average signal level	The average value of the fluctuating AE signal's amplitude over time.
RMS	Root mean square of the constantly fluctuating AE voltage.
RF	$(C - CTP) / (D - RT)$ [kHz].
Initiation frequency [IF]	CTP / RT (kHz).
SS	The overall magnitude of the AE signal, determined by the integral of the rectified voltage signal over the duration of the AE waveform in pVs [picovolt-seconds].
Absolute energy	Calculated by integrating the voltage squared and then dividing by the reference resistance over the entire duration of the waveform [aJ].
AF	The proportion of the total counts and duration of the waveform [kHz].
FC	Weighted mean of frequency components, determined by conducting fast Fourier transform (FFT) and executing calculations on each FFT element [kHz].
PF	The frequency at which the AE signal's amplitude is at its maximum, measured from the FFT data [kHz].
Amplitude [A]	The maximum voltage level of the AE signal, often expressed in decibels relative to a reference voltage [dB].
Partial power 1 [PP1]	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 0 and 200 [kHz].
Partial power 2 [PP2]	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 200 and 400 [kHz].
Partial power 3 [PP3]	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 400 and 600 [kHz].
Partial power 4 [PP4]	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 600 and 800 [kHz].

This study introduces an innovative ensemble feature selection approach to rank features that are highly relevant to damage modes in AE signals from edgewise compression tests on honeycomb-core CFRP. The AE signals are labeled using two key features: amplitude and PF. The ensemble-supervised

feature selection method ranks the features based on their importance to these labels. Subsequently, unsupervised clustering models are used to identify the damage modes based on the ranked features. In the following sections, we explain the experiments and the methods.

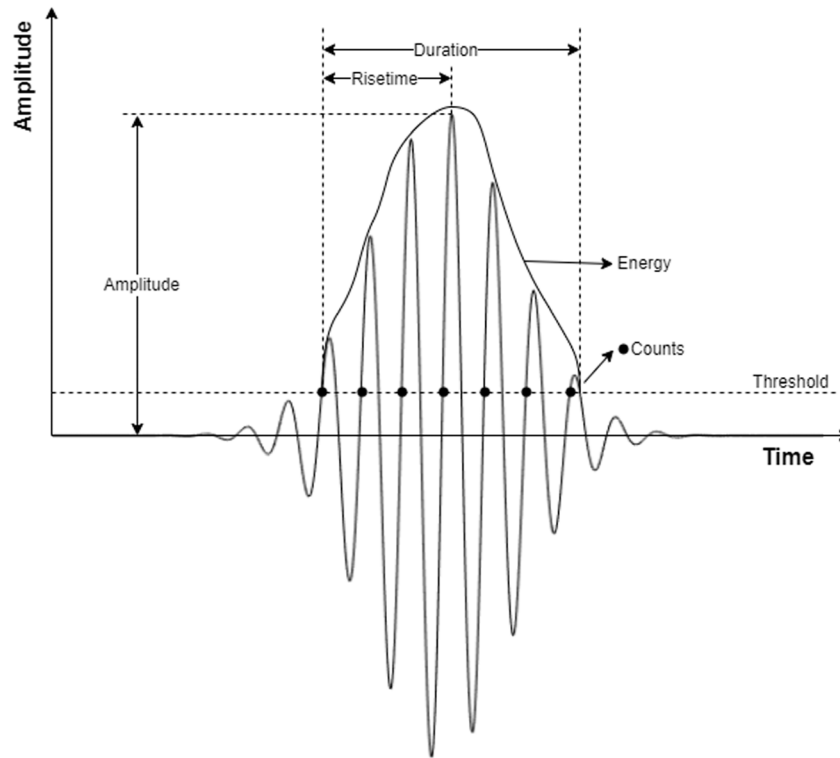


Figure 2. Illustration of key signals: amplitude, rise time, duration, energy, and counts, with a marked threshold.

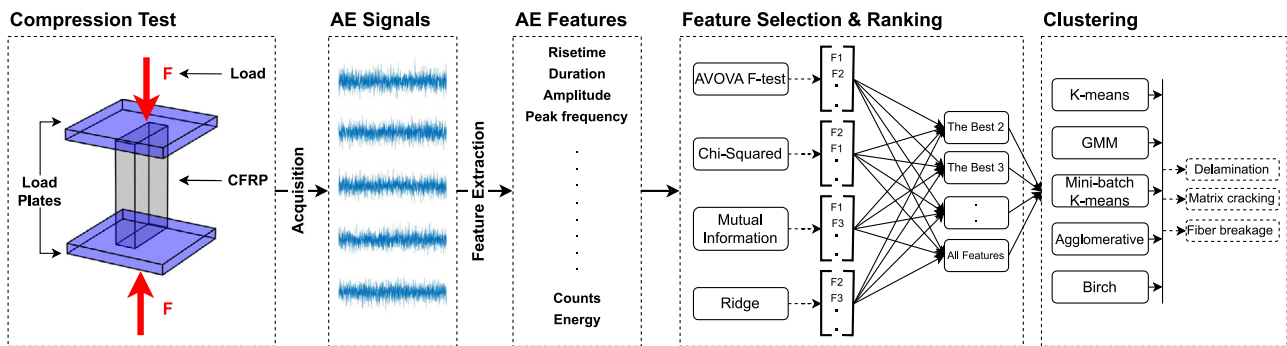


Figure 3. Schematic representation of the proposed methodology, including compression test, AE signal acquisition, feature extraction, feature selection and ranking, and clustering to identify damage characterization.

4.1. Ensemble-Supervised Feature Selection

The objective of feature selection is to enhance model robustness by eliminating unnecessary features that lack relevance to the target variable or possess inadequate predictive capability. In this study, we utilize four feature selection techniques: ANOVA F-test (AFT), Chi-squared (CS), MI, and ridge regression (RD)^[40] aimed at improving clustering performance. Each method assesses the relevance of features according to distinct criteria: AFT considers linear relationships; CS focuses on dependencies between categorical variables; MI evaluates mutual dependencies of information; and RD incorporates regularization to control overfitting.

In this study, PF and amplitude features are utilized as target values respectively. These features have been found to be highly

relevant to the presence of defects in numerous studies, which are discussed in the literature. The reference interval of each feature for damage modes is defined based on the related works. For the amplitude, the reference intervals are selected as 40–60 dB (matrix cracking), 60–80 dB (delamination), and 80–100 dB (fiber breakage) based on the results of this research.^[41] The reference intervals of PF are selected as 60–120 kHz (matrix cracking), 120–210 kHz (delamination), and 200–350 kHz (fiber breakage) according to the outcomes of this study.^[42] The collected dataset is labeled based on these intervals.

In the ensemble-supervised feature selection approach, first, the amplitude is set as the target value, and feature selection is applied to all features except PF during training. Second, PF is chosen as the target value, and feature selection is

conducted on all features except amplitude during training. This approach allows for the evaluation of the relationship between amplitude and PF individually, providing a comprehensive understanding of their predictive capabilities. The best features, which are defined by each feature selection method, are ranked from best to worst, with the best being seventeen and the worst being one. The new rank of each feature is determined by calculating the median of all rank values assigned by the methods. By considering the median rank, we can identify the most consistently valuable features across different methods and make more informed decisions in selecting the optimal set of features for our analysis.

4.2. Clustering Models

In this study, the five widely recognized clustering models (K-Means, MiniBatch K-Means, GMM, Agglomerative, and BIRCH) were chosen for their unique strengths in handling diverse data types. K-Means and MiniBatch K-Means are optimal for large datasets, while GMM excels at modeling complex distributions. Agglomerative clustering offers flexibility in identifying hierarchical structures, and BIRCH provides efficiency for large datasets with many features.

The clustering models are applied to the new rank that is defined for each feature. Initially, clustering models are executed on various subsets of features in the new rank, specifically excluding the target value (amplitude). Three clusters are obtained as a result for each model. To evaluate the clustering performance of the models, the labeled data is compared with clusters, and the v-measure scores are obtained for different numbers of features in the new rank. The v-measure is a measure of the similarity between the predicted clusters and the true labels. It provides an evaluation of how well the clustering models perform in terms of accurately assigning data points to their respective clusters. Additionally, this comparison allows for an assessment of which clustering model yields the highest v-measure, indicating its effectiveness in capturing the underlying patterns and structure in the data. Subsequently, we replicate this entire process using PF as the target variable.

4.2.1. K-Means Clustering

K-Means clustering is a versatile unsupervised machine learning algorithm used for segmenting datasets across various domains. In image processing, it aids in segmenting images based on pixel attributes, while in business analytics, it is utilized for customer segmentation to tailor marketing strategies effectively. Additionally, K-Means facilitates document clustering in text mining, improving information retrieval, and is employed in anomaly detection to identify outliers in datasets such as network security. This clustering technique is pivotal in extracting actionable insights from complex datasets, thereby informing decision-making in both commercial and scientific fields.^[43]

The model aims to divide a dataset into K separate, nonintersecting groups to reveal inherent structures within the data. The algorithm operates on a dataset $X = \{x_1, x_2, \dots, x_n\}$, where each x_i is a d -dimensional vector. It aims to minimize the within-cluster sum of squares, formally represented by the objective function in Equation (1)

$$J(C, \mu) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (1)$$

here, C_k denotes the k th cluster, and μ_k is its centroid, calculated as the mean of all vectors in C_k .

The algorithm initiates with a set of randomly selected centroids or employs more sophisticated models like K-Means++ for initialization. These centroids serve as the starting points for each cluster. In the assignment phase, data points are allocated to the cluster that minimizes the Euclidean distance $\|x_i - \mu_k\|$, effectively updating the cluster assignments as in Equation (2)

$$C_k = \{x_i : \|x_i - \mu_k\| \leq \|x_i - \mu_j\| \forall j, 1 \leq j \leq K\} \quad (2)$$

Subsequently, the centroids μ_k are recalculated in the update phase as the mean of all points in C_k , essentially averaging the data to find the new centroid. The algorithm performs iterative calculations to optimize the positions of the centroids and halts when the centroids have stabilized or when a predefined number of iterations has been reached.

4.2.2. Agglomerative Clustering

Agglomerative clustering, a hierarchical clustering model, is utilized across diverse fields for its ability to build cluster hierarchies and offer insights at various granularities. In the biological sciences, it aids in classifying species and understanding genetic relationships, proving particularly useful in cancer classification through gene expression data. In marketing, it enables nuanced customer segmentation without prespecifying cluster counts, thus enhancing personalized marketing strategies. This model also improves information retrieval by clustering similar documents, benefiting digital libraries, and is employed in social network analysis to identify user communities and understand social dynamics.^[44]

The primary objective of the model is to group similar data points based on their observed characteristics, thereby facilitating the identification of underlying patterns or relationships within the data. The algorithm operates on a dataset $X = \{x_1, x_2, \dots, x_n\}$, where each x_i is a d -dimensional vector. In contrast to partitioning techniques such as K-Means, the agglomerative clustering approach does not necessitate predefining the total number of clusters, K , before beginning the analysis.

The process begins with the algorithm considering each individual data point as its own separate cluster. It then progressively combines these clusters according to a chosen linkage strategy that uses Euclidean distance as a measure, continuing this merging until it consolidates all points into one singular cluster. Common linkage methods include single, complete, average, and ward linkage, each with its own strengths and weaknesses. For instance, single linkage tends to create elongated, chain-like clusters, while complete linkage generally results in more compact, spherical clusters. The linkage criterion is formalized as a dissimilarity measure $D(A, B)$ between sets of observations A and B , and it can vary depending on the chosen method.

The algorithm constructs a dendrogram, a tree-like diagram that illustrates the sequence of cluster merges, with the height of each merge corresponding to the value of the linkage criterion

at that point. The dendrogram can be cut at a desired level to obtain clusters at varying levels of granularity.

4.2.3. Gaussian Mixture Clustering

GMM is particularly effective in applications like speech recognition, where it models the acoustic properties of different phonemes, capturing the variability in speech sounds. In image segmentation, GMM separates objects based on pixel intensities, and in anomaly detection, it identifies abnormal patterns by modeling normal behavior using Gaussian distributions. These diverse applications underscore the versatility and robustness of GMM in capturing complex data structures, although it is sensitive to initial parameter estimates and may converge to a local optimum.^[45]

It operates on the premise that the dataset originates from a blend of K distinct Gaussian distributions, with each one corresponding to a separate cluster. Unlike partition-based models such as K-Means, GMM assigns probabilities to data points, allowing them to belong to multiple clusters simultaneously. This provides a more nuanced understanding of cluster assignment and enables the capture of complex relationships within the data.

Mathematically, given a dataset $X = \{x_1, x_2, \dots, x_n\}$, where each x_i is a d -dimensional vector, the GMM aims to model the data using the likelihood function in Equation (3)

$$p(x_i|\Theta) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \quad (3)$$

Here, Θ represents the model parameters, which include the mixing coefficients π_k , the means μ_k , and the covariance matrices Σ_k . The algorithm employs the Expectation-Maximization (EM) method for parameter estimation, which alternates between the E-step and M-step. In the E-step, the algorithm computes the posterior probabilities w_{ik} that each data point x_i belongs to each Gaussian distribution k in Equation (4). In the M-step, these probabilities are used to update the parameters Θ , thereby maximizing the likelihood of the observed data.

$$w_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)} \quad (4)$$

The algorithm iteratively updates these parameters until convergence is reached, resulting in the final clustering and density estimation.

4.2.4. Birch

The BIRCH algorithm is a scalable and memory-efficient clustering model specifically designed for large datasets. One key application of BIRCH is in customer behavior analysis, where it clusters vast amounts of customer data to identify behavior patterns and trends, aiding businesses in segmenting their market and tailoring marketing strategies.^[46] Additionally, BIRCH is used for anomaly detection in network traffic.^[47] By clustering data points, BIRCH can quickly identify unusual patterns that

deviate from the norm, which are potential indicators of security threats or failures.

The algorithm incrementally constructs a height-balanced tree structure known as the clustering feature tree (CF Tree), which serves as a compact representation of the dataset. This summary, which is stored in memory, strives to condense the dataset's memory footprint by encapsulating dense data clusters into concise CF entries. Mathematically, a CF is a tuple $CF = (N, LS, SS)$, where N is the number of data points in a cluster, LS is the linear sum of the data points, and SS is the square sum of the data points. Given a dataset $X = \{x_1, x_2, \dots, x_n\}$, where each x_i is a d -dimensional vector, the CF for a cluster C containing N data points is defined as in Equation (5)

$$CF = \left(N, \sum_{i=1}^N x_i, \sum_{i=1}^N x_i^2 \right) \quad (5)$$

The CF Tree is parameterized by two key factors: the branching factor B and the threshold T . The branching factor limits the maximum number of child nodes that each internal node can have, while the threshold controls the maximum number of entries in each leaf node, effectively determining the diameter of the subclusters.

The algorithm operates through multiple phases, beginning with the initialization phase where an empty CF Tree is created. This is followed by the loading Phase, where each data point x_i is inserted into the CF Tree while maintaining its CF properties and adhering to the threshold T . Optionally, a condensation phase can be executed if the CF Tree exceeds memory limitations, serving to reduce its size. Finally, a global clustering phase is carried out, where a global clustering algorithm such as K-Means or agglomerative clustering is applied to the leaf nodes to derive the final clusters.

4.2.5. MiniBatch K-Means

The MiniBatch K-Means algorithm is an optimized version of the traditional K-Means clustering algorithm, designed to improve computational efficiency, particularly when dealing with large datasets, such as image or video processing, where it significantly reduces computational costs and accelerates convergence. MiniBatch K-Means is also employed in online learning scenarios, where data arrives in streams, and quick adaptation to new data is crucial, making it suitable for dynamic environments like financial market analysis or real-time sensor data clustering in IoT applications.^[48] Unlike the standard K-Means, which uses all data points to update the centroids of clusters, MiniBatch K-Means employs a subset of randomly selected data points, known as a "MiniBatch," in each iteration. This approximation speeds up the optimization process, reducing computational time while slightly compromising the quality of the clustering.

The algorithm starts with the initialization phase where initial centroids are randomly selected. This is followed by the MiniBatch selection phase, where a subset of data points, known as a "MiniBatch," is randomly chosen from the dataset. In the centroid update phase, these MiniBatches are used to iteratively update the centroids using a learning rate, which is often determined by the reciprocal of the number of data points assigned to

each centroid as the algorithm progresses. This learning rate effectively reduces the number of iterations required for the algorithm to reach convergence. The algorithm continues to iterate through the MiniBatch selection and centroid update phases until no significant changes in the clusters are observed, signaling that convergence has been achieved.

4.3. Performance Metrics

The V-measure is an evaluation metric for clustering algorithms that provides a balanced assessment of clustering quality when ground truth labels are available. It is the harmonic mean of two components: homogeneity and completeness.

Homogeneity quantifies the extent to which a cluster comprises solely elements from one particular class. Complete homogeneity is achieved when each cluster in a clustering solution exclusively consists of elements from a single class. The formula for calculating homogeneity h is presented in Equation (6), where K represents the count of clusters, C signifies the number of actual classes, $n_{k,c}$ is the number of elements in cluster k that are also in class c , n_k denotes the count of all elements within cluster k , and N is the total count of elements across the dataset.

$$h = 1 - \sum_{k=1}^K \sum_{c=1}^C \frac{n_{k,c}}{N} \log\left(\frac{n_{k,c}}{n_k}\right) \quad (6)$$

Completeness c serves to quantify the degree to which all instances of a particular class are grouped into a singular cluster. A clustering outcome is deemed to exhibit completeness if each class is predominantly represented within a single, unique cluster. The formal mathematical representation of completeness is articulated in Equation (7). In this equation, K signifies the aggregate number of clusters, C denotes the whole amount of ground truth classes, $n_{k,c}$ indicates the quantity of data points in cluster k that are also members of class c , n_c is the cumulative count of data points in class c , and N encapsulates the overall data point count in the dataset

$$c = 1 - \sum_{c=1}^C \sum_{k=1}^K \frac{n_{k,c}}{N} \log\left(\frac{n_{k,c}}{n_c}\right) \quad (7)$$

The V-measure is defined as in Equation (8) where h and c are the homogeneity and completeness scores, respectively. The V-measure ranges from 0 to 1, where a value of 1 indicates a perfect clustering that is both homogeneous and complete, and a value of 0 indicates the worst possible clustering

$$V = 2 \times \frac{h \times c}{h + c} \quad (8)$$

5. Results and Discussion

5.1. Ensemble Supervised Feature Selection

In this study, the dataset is labeled separately using two features: amplitude and PF. Four different supervised feature selection methods (AFT, CS, MI, and RD) along with an ensemble feature selection method are used. The results of each feature selection method and the ensemble feature selection for the dataset labeled with amplitude are shown in **Figure 4**. Each feature selection method assigned a score of 17 to the most relevant feature and 1 to the least relevant feature. For the ensemble method, the feature ranking is calculated by taking the median of scores obtained from the four feature selection methods. According to this new ranking, counts, energy, duration, root mean square (RMS), and partial power 2 are found to have a strong correlation with amplitude, while reverberation frequency (RF), AF, partial power 1, initiation frequency, and SS exhibited a weak correlation with amplitude. The same process is applied by selecting PF as the target. The results of the feature selection methods and the ensemble feature selection for PF are shown in **Figure 5**. According to the ensemble feature selection, partial power 1, partial power 2, RF, FC, and partial power 3 exhibited a strong correlation with PF, whereas counts to peak, duration, average signal level, counts, and energy showed a weak correlation with PF. This method offers a more thorough and robust set of features, effectively capturing a broad range of feature interactions and dependencies.

Anova F-Test		Chi-squared		Mutual-info		Ridge		Ensemble Feature Selection	
Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Median of Ranks	Feature
17	Counts	17	PartialPower4	17	Energy	17	Counts	16	Counts
16	ASL	16	ASL	16	SignalStrength	16	Energy	15,5	Energy
15	Energy	15	Energy	15	Counts	15	PartialPower2	13	Duration
14	Duration	14	RMS	14	AbsoluteEnergy	14	Duration	12	RMS
13	RMS	13	Counts	13	CountstoPeak	13	AbsoluteEnergy	11,5	PartialPower2
12	Partial Power 4	12	PartialPower2	12	Duration	12	Risetime	10,5	ASL
11	Partial Power 2	11	Duration	11	RMS	11	SignalStrength	9,5	PartialPower4
10	FrequencyCentroid	10	PartialPower3	10	Risetime	10	FrequencyCentroid	9	AbsoluteEnergy
9	CountstoPeak	9	CountstoPeak	9	ReverberationFrequency	9	PartialPower3	9	FrequencyCentroid
8	PartialPower3	8	FrequencyCentroid	8	FrequencyCentroid	8	RMS	9	CountstoPeak
7	Risetime	7	Risetime	7	InitiationFrequency	7	PartialPower4	8,5	Risetime
6	PartialPower1	6	InitiationFrequency	6	PartialPower2	6	PartialPower1	8,5	PartialPower3
5	InitiationFrequency	5	AbsoluteEnergy	5	ASL	5	ASL	7,5	SignalStrength
4	AverageFrequency	4	SignalStrength	4	AverageFrequency	4	AverageFrequency	5,5	InitiationFrequency
3	AbsoluteEnergy	3	AverageFrequency	3	PartialPower4	3	CountstoPeak	4	PartialPower1
2	SignalStrength	2	PartialPower1	2	PartialPower1	2	ReverberationFrequency	4	AverageFrequency
1	ReverberationFrequency	1	ReverberationFrequency	1	PartialPower3	1	InitiationFrequency	1,5	ReverberationFrequency

Figure 4. Results of four individual and one ensemble supervised feature selection methods labeled with amplitude.

Anova F-Test		Chi-squared		Mutual-info		Ridge		Ensemble Feature Selection	
Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Median of Ranks	Feature
17	PartialPower1	17	PartialPower2	17	PartialPower1	17	PartialPower1	17	PartialPower1
16	PartialPower2	16	PartialPower3	16	PartialPower2	16	PartialPower2	16	PartialPower2
15	ReverberationFrequency	15	PartialPower4	15	AverageFrequency	15	FrequencyCentroid	14	ReverberationFrequency
14	PartialPower3	14	PartialPower1	14	ReverberationFrequency	14	ReverberationFrequency	13	FrequencyCentroid
13	FrequencyCentroid	13	ReverberationFrequency	13	FrequencyCentroid	13	SignalStrength	12,5	PartialPower3
12	PartialPower4	12	AverageFrequency	12	Risetime	12	AbsoluteEnergy	11,5	AverageFrequency
11	AverageFrequency	11	SignalStrength	11	PartialPower3	11	PartialPower3	11	PartialPower4
10	InitiationFrequency	10	AbsoluteEnergy	10	PartialPower4	10	Energy	9	SignalStrength
9	Counts	9	FrequencyCentroid	9	InitiationFrequency	9	AverageFrequency	8,5	InitiationFrequency
8	RMS(16)	8	InitiationFrequency	8	RMS(16)	8	Risetime	8	AbsoluteEnergy
7	SignalStrength	7	RMS(16)	7	Duration	7	PartialPower4	7,5	RMS
6	AbsoluteEnergy	6	ASL	6	SignalStrength	6	CountstoPeak	4,5	Risetime
5	ASL	5	Counts	5	AbsoluteEnergy	5	Duration	4	Energy
4	Energy	4	Energy	4	Energy	4	InitiationFrequency	4	Counts
3	CountstoPeak	3	CountstoPeak	3	ASL	3	Counts	4	ASL
2	Duration	2	Duration	2	CountstoPeak	2	RMS(16)	3,5	Duration
1	Risetime	1	Risetime	1	Counts	1	ASL	3	CountstoPeak

Figure 5. Results of four individual and one ensemble supervised feature selection methods labeled with PF.

5.2. Clustering

In this study, five unsupervised clustering models, including K-Means, MiniBatch K-Means, GMM, agglomerative, and BIRCH, are used to cluster signal data. The clustering results are compared with the labeled dataset based on amplitude and PF. The performance of each model is evaluated using the V-measure. In the clustering performance results for the PF target, BIRCH achieved a V-measure score of 39.7%, outperforming the other models, which suggests its effectiveness in capturing the underlying patterns in the data, as illustrated in **Figure 6**. This notable difference in V-measure scores indicates that the preset labeling intervals for PF in the dataset, as determined by previous studies, may not fully apply to CFRP composites. The high V-measure scores achieved by features partial power 1 and partial power 2 for all clustering models indicate that these features capture essential information related to PF. The inclusion of the third (RF), fourth (FC), and fifth (partial power 3) features results in a decrease in the V-measure score for all clustering models, indicating that these features may introduce noise or irrelevant information. The models consistently show a decline in performance as the number of input features increases, indicating that an abundance of features can lead to



Figure 6. Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the PF target.

complexity rather than clarity, potentially introducing noise or irrelevant variables.

The clustering performance results for the amplitude target show varying trends across different numbers of features, as illustrated in **Figure 7**. The GMM and MiniBatch K-Means clustering models achieve V-measure scores of 48.75% and 40.68%, respectively. The MiniBatch and GMM clustering models demonstrate greater robustness compared to other models with the increase in feature size, whereas K-Means, agglomerative, and BIRCH show a negative impact. As the number of input features increases, the models exhibit a continuous decline in performance, suggesting that an excess of features may introduce complexity instead of clarity, possibly due to the inclusion of noise or irrelevant variables. This observation highlights the challenges posed by a high number of features used in clustering tasks, where the inclusion of additional features can dilute data clusters instead of refining them, resulting in reduced clustering performance.

The GMM consistently achieves high V-measure scores in both targets, indicating its effectiveness as an initial clustering model in AE studies. **Figure 8** shows the GMM clustering results using amplitude as the target for four different feature combinations: a) counts and energy, b) counts, energy, duration, and RMS, c) counts, energy, duration, RMS, and partial power 2,

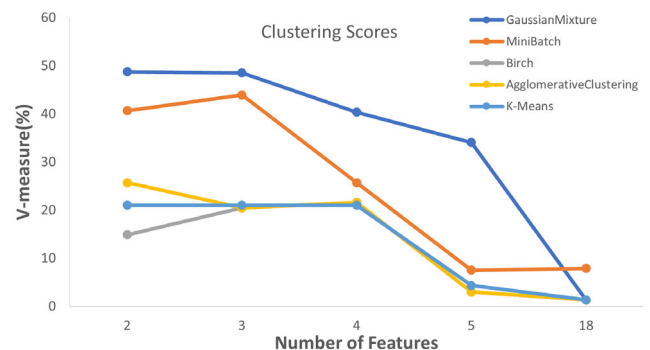


Figure 7. Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the amplitude.

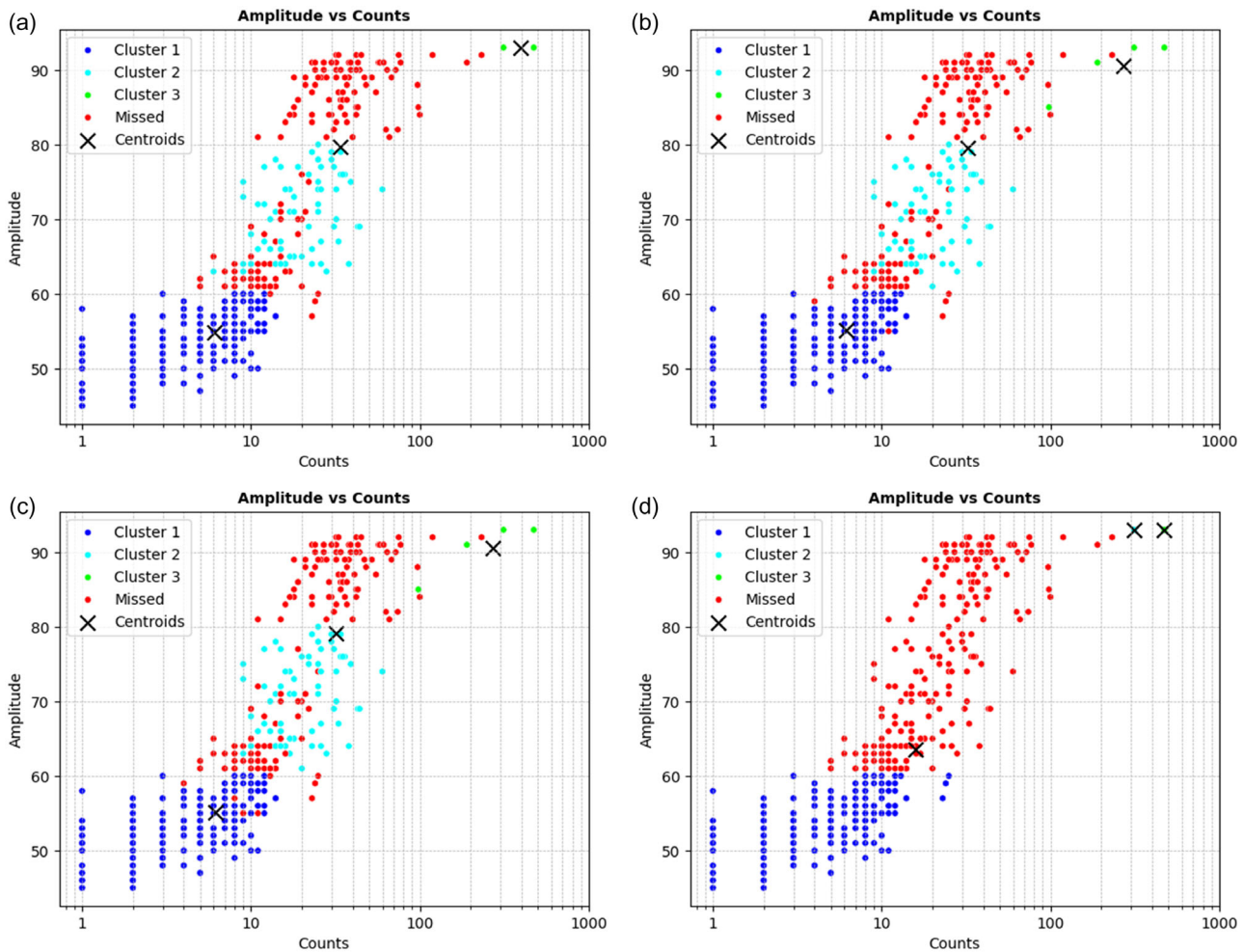


Figure 8. GMM clustering results with different number of features for amplitude target. Missed indicates incorrectly clustered points. The x-axis is scaled logarithmically. a) Amplitude versus counts with counts and energy features, b) amplitude versus counts with counts, energy, duration, and RMS features, c) amplitude versus counts with counts, energy, duration, RMS, and partial power 2 features, and d) amplitude versus counts with all features.

d) all features. In Figure 8a, clusters are defined by specific amplitude ranges: cluster 1 ranges from 45 to 76 dB, cluster 2 from 50 to 92 dB, and cluster 3 maintains a steady amplitude of 93 dB. These clusters correspond to matrix cracking, delamination, and fiber breakage, respectively. The amplitude range of cluster 3 aligns with the values documented in the literature. Nevertheless, cluster 3 experiences frequent misclassification, often leading to erroneous grouping of points into cluster 2. Although the initial range of cluster 2 closely matches the literature, there is a significant misclassification rate for amplitudes between 50 and 60 dB, frequently leading to incorrect assignments to cluster 1. The V-measure, at $\approx 50\%$, indicates potential limitations in the generalizability of the literature's defined amplitude ranges, which may contribute to the observed clustering errors. Moreover, according to Figure 8b,c, the overlap between clusters 1 and 2 contributes to the decrease in V-measure scores. In Figure 8d, cluster 1 shows precise detection, whereas clusters 2 and 3 display notable misclassification.

These findings highlight the significance of ensemble feature selection in improving clustering performance and underscore

its importance in precise damage characterization. The significant decline in performance at 18 features could imply that adding too many features introduces noise or irrelevant information, leading to complications in the clustering process.

6. Conclusion

This study investigates the use of AE features to damage characterization in CFRP composites during compression tests. We utilized supervised ensemble feature selection methods (including AFT, CS, MI, and ridge) and unsupervised clustering models (such as K-Means, MiniBatch K-Means, GMM, agglomerative, and BIRCH) to examine damage mechanisms. In the beginning, the dataset was labeled based on insights from prior studies concerning amplitude and PF. Ensemble feature selection methods were applied to obtain feature ranks, in which the highest correlation has higher scores and the lower correlation has lower scores. Clustering algorithms were then employed on these ranked features, and their performance was rigorously evaluated

using the V-score metric. The results demonstrate a good alignment in the supervised ensemble feature selection methods for both amplitude and PF targets. Furthermore, although all clustering models produced reliable results for PF, GMM and MiniBatch K-Means exhibited significantly better performance in amplitude-based clustering. The study thus provides important perspectives on the utility of AE features for damage characterization in CFRP composites and underscores the efficacy of combining ensemble supervised feature selection methods and unsupervised clustering models.

Acknowledgements

This work was supported by TÜBİTAK ULAKBİM; through its agreement with Wiley, the open access fee for this publication has been covered.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

acoustic emission, carbon fiber-reinforced polymer composites, clustering, damage characterization, ensemble feature selection, industrial innovation, machine learning

Received: February 7, 2024

Revised: June 3, 2024

Published online: July 15, 2024

- [1] A. Kelly, C. Zweben, *Mater. Today* **1999**, 2, 20.
- [2] R. S. Almeida, M. D. Magalhães, M. N. Karim, K. Tushtev, K. Rezwan, *Mater. Des.* **2023**, 227, 111745.
- [3] J. Yu, Y. Z. Liu, R. R. Shi, *Adv. Mater. Res.* **2009**, 79, 1029.
- [4] J. Hwang, C. Lee, W. Hwang, *Appl. Compos. Mater.* **2001**, 8, 411.
- [5] S. M. Lee, *J. Compos. Mater.* **1986**, 20, 185.
- [6] Z. Quiney, E. Weston, P. I. Nicholson, S. Pattison, M. R. Bache, *J. Eur. Ceram. Soc.* **2020**, 40, 3788.
- [7] M. Roth, E. Mojaev, E. Dul'kin, P. Gemeiner, B. Dkhil, *Phys. Rev. Lett.* **2007**, 98, 265701.
- [8] C. Muir, B. Swaminathan, A. Almansour, K. Sevener, C. Smith, M. Presby, J. Kiser, T. Pollock, S. Daly, *Materials* **2021**, 7, 95.
- [9] N. Godin, P. Reynaud, G. Fantozzi, *Appl. Sci.* **2018**, 8, 1267.
- [10] J. P. Forna-Kreutzer, J. Ell, H. Barnard, T. J. Pirezada, R. O. Ritchie, D. Liu, *Mater. Des.* **2021**, 208, 109899.
- [11] W. Zhou, R. Qin, K.-n. Han, Z.-y. Wei, L.-H. Ma, *Polym. Test.* **2021**, 93, 106881.
- [12] E. Özaslan, A. Yetgin, B. Acar, M. A. Güler, *Compos. Struct.* **2021**, 274, 114299.
- [13] M. G. Sause, S. Horn, *J. Nondestruct. Eval.* **2010**, 29, 123.
- [14] M. Saeedifar, D. Zarouchas, *Composites, Part B* **2020**, 195, 108039.
- [15] C. Barile, C. Casavola, G. Pappaletta, V. P. Kannan, *Eng. Fract. Mech.* **2020**, 235, 107083.
- [16] W. Harizi, S. Chaki, G. Bourse, M. Ourak, *Compos. Struct.* **2022**, 289, 115470.
- [17] C. U. Grosse, M. Ohtsu, D. G. Aggelis, T. Shiotani, in *Acoustic Emission Testing: Basics For Research–Applications In Engineering*, Springer Nature, Berlin, Heidelberg **2021**.
- [18] F. Pashmforoush, M. Fotouhi, M. Ahmadi, *J. Reinf. Plast. Compos.* **2012**, 31, 671.
- [19] S. Gholizadeh, Z. Leman, B. Baharudin, *Ultrasonics* **2023**, 132, 106998.
- [20] A. R. Oskouei, H. Heidary, M. Ahmadi, M. Farajpur, *Mater. Des.* **2012**, 37, 416.
- [21] H. Sayar, M. Azadi, A. Ghasemi-Ghalebahman, S. M. Jafari, *Compos. Struct.* **2018**, 204, 1.
- [22] L. Li, Y. Swolfs, I. Straumit, X. Yan, S. V. Lomov, *J. Compos. Mater.* **2016**, 50, 1921.
- [23] F. E. Oz, N. Ersoy, M. Mehdikhani, S. V. Lomov, *Compos. Struct.* **2018**, 196, 163.
- [24] H. A. Sawan, M. E. Walter, B. Marquette, *Compos. Sci. Technol.* **2015**, 107, 89.
- [25] M. Saeedifar, M. A. Najafabadi, D. Zarouchas, H. H. Toudeshky, M. Jalalvand, *Composites, Part B* **2018**, 144, 206.
- [26] L. Li, S. V. Lomov, X. Yan, V. Carvelli, *Compos. Struct.* **2014**, 116, 286.
- [27] C. Barile, C. Casavola, G. Pappaletta, V. P. Kannan, *Appl. Acoust.* **2022**, 185, 108425.
- [28] D. D. Doan, E. Ramasso, V. Placet, S. Zhang, L. Boubakar, N. Zerhouni, *Mech. Syst. Signal Process.* **2015**, 64, 465.
- [29] J. Tang, S. Soua, C. Mares, T.-H. Gan, *Sensors* **2017**, 17, 2507.
- [30] M. Assarar, M. Bentahar, A. El Mahi, R. El Guerjouma, *Int. J. Damage Mech.* **2015**, 24, 787.
- [31] H. Heidary, N. Z. Karimi, M. Ahmadi, A. Rahimi, A. Zucchelli, *J. Compos. Mater.* **2015**, 49, 559.
- [32] J. Yousefi, M. A. Najafabadi, H. H. Toudeshky, M. Akhlaghi, *Appl. Compos. Mater.* **2018**, 25, 1021.
- [33] M. Fotouhi, P. Suwarta, M. Jalalvand, G. Czel, M. R. Wisnom, *Composites, Part A* **2016**, 86, 66.
- [34] R. Gutkin, C. Green, S. Vangrattanachai, S. Pinho, P. Robinson, P. Curtis, *Mech. Syst. Signal Process.* **2011**, 25, 1393.
- [35] S. F. Karimian, M. Modarres, *Compos. Struct.* **2021**, 268, 113987.
- [36] W. Zhou, W.-z. Zhao, Y.-n. Zhang, Z.-j. Ding, *Compos. Struct.* **2018**, 195, 349.
- [37] M. Haggui, A. El Mahi, Z. Jendli, A. Akrouf, M. Haddar, *Appl. Acoust.* **2019**, 147, 100.
- [38] ASTM C364. *Standard Test Method for Edgewise Compressive Strength of Sandwich Constructions*, ASTM, West Conshohocken, PA **2000**.
- [39] M. Hamstad, in *Acoustic Emission Beyond The Millennium*, Elsevier, Amsterdam **2000**, pp. 77–91.
- [40] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, H. Liu, *ACM Comput. Surveys* **2017**, 50, 1.
- [41] P. Liu, J. Chu, Y. Liu, J. Zheng, *Mater. Des.* **2012**, 37, 228.
- [42] R. Boominathan, V. Arumugam, C. Santulli, A. A. P. Sidharth, R. A. Sankar, B. Sridhar, *Composites, Part B* **2014**, 56, 591.
- [43] A. K. Jain, *Pattern Recognit. Lett.* **2010**, 31, 651.
- [44] F. Murtagh, P. Contreras, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **2012**, 2, 86.
- [45] N. Bouguila, W. Fan, in *Mixture Models and Applications*, Springer, New York **2020**.
- [46] A. D. Fontanini, J. Abreu, in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, Piscataway, NJ **2018**, pp. 1–5.
- [47] A. G. Roselin, P. Nanda, S. Nepal, X. He, *IEEE Access* **2021**, 9, 47243.
- [48] Z. Wang, Y. Zhou, G. Li, in *2020 5th IEEE Int. Conf. on Big Data Analytics (ICBDA)*, IEEE, Piscataway, NJ **2020**, pp. 11–17.
- [49] C. Barile, C. Casavola, G. Pappaletta, P. K. Vimalathithan, *Appl. Sci.* **2021**, 11, 393.