Abdulkadir GÜLŞEN

Ph.D. Thesis

AGU 2024

DETECTION AND CLASSIFICATION OF FLAWS FROM ULTRASONIC TOMOGRAPHY IMAGES OF COMPOSITE MATERIALS BASED ON DEEP LEARNING

Ph.D. THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Abdulkadir GÜLŞEN

December 2024

DETECTION AND CLASSIFICATION OF FLAWS FROM ULTRASONIC TOMOGRAPHY IMAGES OF COMPOSITE MATERIALS BASED ON DEEP LEARNING

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Abdulkadir GÜLŞEN
December 2024

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Abdulkadir GÜLŞEN
Signature :

REGULATORY COMPLIANCE

Ph.D. thesis titled “Detection and Classification of Flaws from Ultrasonic Tomography Images of Composite Materials Based on Deep Learning” has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor	Co Advisor
Abdulkadir GÜLŞEN	Assoc. Prof. Burcu BAKIR GÜNGÖR	Dr. Burak KOLUKISA

Head of the Electrical and Computer Engineering Program

Assoc. Prof. Samet GÜLER

ACCEPTANCE AND APPROVAL

Ph.D. thesis titled “Detection and Classification of Flaws from Ultrasonic Tomography Images of Composite Materials Based On Deep Learning” and prepared by Abdulkadir GÜLŞEN has been accepted by the jury in the Electrical and Computer Engineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

…….. /…….. / 2024

JURY:

Advisor : Assoc. Prof. Burcu BAKIR GÜNGÖR

Co-Advisor : Dr. Burak KOLUKISA

Member : Prof. Ahmet Turan ÖZDEMİR

Member : Asst. Prof. Mehmet Gökhan BAKAL

Member : Asst. Prof. Abdulkadir KÖSE

Member : Asst. Prof. Nazlı TEKİN

APPROVAL:

The acceptance of this Ph.D. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated ….. /….. / …….. and numbered ……..……. .

…….. /…….. / ……..

(Date)

Graduate School Dean
Prof. İrfan ALAN

ABSTRACT

DETECTION AND CLASSIFICATION OF FLAWS FROM ULTRASONIC TOMOGRAPHY IMAGES OF COMPOSITE MATERIALS BASED ON DEEP LEARNING

Abdulkadir GÜLŞEN
Ph.D. in Electrical and Computer Engineering
Advisor: Assoc. Prof. Burcu BAKIR GÜNGÖR
Co-advisor: Dr. Burak KOLUKISA
December 2024

This thesis introduces novel methodologies for enhancing defect classification and characterization in advanced composite materials by leveraging state-of-the-art machine learning (ML), deep learning (DL), and federated learning (FL) techniques within ultrasonic and acoustic emission (AE) inspection environments. First, a new ultrasonic dataset (UNDT), comprising 1,150 images from 60 distinct composite materials, is introduced. Applying transfer learning methods to both the UNDT and a publicly available dataset demonstrates the efficacy of advanced neural architectures—such as DenseNet121 and VGG19—achieving accuracy rates up to 98.8% and 98.6%, respectively. Next, the scope is extended to AE-based health monitoring by introducing an ensemble feature selection methodology to identify features strongly correlated with damage modes. By selecting amplitude and peak frequency for labeling and subsequently applying unsupervised clustering, the analysis confirms that both traditional AE features (e.g., counts and energy) and less commonly employed features (e.g., partial powers) correlate with distinct defect types. Finally, a novel FL framework is introduced to address the scarcity of publicly available, real-world ultrasonic datasets. This decentralized approach preserves data privacy while maintaining performance levels comparable to centralized methods, ensuring scalability and confidentiality in diverse data environments. Overall, these contributions significantly advance the field of NDT, offering robust defect classification and characterization. In doing so, the findings not only improve the accuracy and reliability of material integrity assessments but also lay a durable foundation for more secure, collaborative, and efficient NDT systems.

Keywords: Non-destructive Testing (NDT), Defect Classification, Machine Learning, Deep Learning, Federated Learning

ÖZET

DERIN ÖĞRENME TABANLI KOMPOZİT MALZEMELERİN ULTRASONİK TOMOGRAFİ GÖRÜNTÜLERİNDEN KUSURLARIN TESPİTİ VE SINIFLANDIRILMASI

Abdulkadir GÜLŞEN
Elektrik ve Bilgisayar Mühendisliği Anabilim Dalı Doktora
Tez Danışmanı: Doç. Dr. Burcu BAKIR-GÜNGÖR
İkinci Tez Danışmanı: Dr. Burak KOLUKISA
Aralık-2024

Bu tez, gelişmiş kompozit malzemelerde kusur sınıflandırma ve karakterizasyonunu iyileştirmek amacıyla, ultrasonik ve akustik emisyon muayene ortamlarında en son makine öğrenme, derin öğrenme ve federe öğrenme yaklaşımlarını bir araya getiren yenilikçi yöntemler sunmaktadır. İlk olarak, 60 farklı kompozit malzemeden elde edilen 1150 görüntüden oluşan yeni bir ultrasonik veri seti (UNDT) tanıtılmıştır. Hem UNDT hem de halka açık bir başka veri seti üzerinde uygulanan transfer öğrenimi yöntemleri, DenseNet121 ve VGG19 gibi gelişmiş sinir ağı mimarilerinin %98,8 ve %98,6'ya varan doğruluk oranlarıyla etkinliğini kanıtlamaktadır. Bunu takiben, tezin kapsamı akustik emisyon tabanlı yapısal sağlık izleme alanına genişletilerek, hasar modlarıyla güçlü ilişkisi bulunan özellikleri belirlemek amacıyla topluluk temelli bir özellik seçimi metodolojisi sunulmuştur. Genlik ve tepe frekansı etiketleme için seçilmiş, ardından uygulanan denetimsiz kümeleme analizleri, hem geleneksel akustik emisyon özelliklerinin (ör. sayılar ve enerji) hem de daha az kullanılan kısmi güçler gibi özelliklerin farklı kusur tipleriyle güvenilir biçimde ilişkili olduğunu göstermiştir. Son olarak, gerçek dünyaya uygun, halka açık ultrasonik veri setlerinin sınırlılığını gidermek amacıyla yenilikçi bir federe öğrenme yaklaşımı geliştirilmiştir. Bu dağıtık yaklaşım, merkezi yöntemlere kıyasla benzer performans seviyelerini korurken veri mahremiyetini de güvence altına alarak ölçeklenebilir ve gizliliğe duyarlı bir çözüm sunmaktadır. Genel olarak, bu katkılar tahribatsız muayene alanında önemli ilerlemeler sağlamakta, kusur sınıflandırma ve karakterizasyonunun doğruluğunu ve güvenilirliğini artırmaktadır. Böylece, daha güvenli, işbirlikçi ve verimli tahribatsız muayene sistemlerinin geliştirilmesi için kalıcı ve sağlam bir temel oluşturulmaktadır.

Anahtar kelimeler: Tahribatsız Muayene, Kusur Sınıflandırma, Makine Öğrenme, Derin Öğrenme, Federe Öğrenme

Acknowledgements

First and foremost, I extend my deepest gratitude and sincere appreciation to my supervisors, Dr. V. Çağrı GÜNGÖR, Dr. Burcu BAKIR-GÜNGÖR and Dr. Burak KOLUKISA. Their constant support and insightful guidance have been invaluable throughout my Ph.D. journey.

My cordial thanks also extend to Dr. A. Turan ÖZDEMİR, Dr. M. Gökhan BAKAL, Dr. Abdulkadir KÖSE and Dr. Nazlı TEKİN for serving on my dissertation defense committee.

Finally, my sincere appreciation goes to my parents Abdullah and Emine GÜLŞEN, my wife Pervin GÜLŞEN, and my child Elif Sare GÜLŞEN. Their patience, continuous support, and encouragement have been a source of strength in my pursuit of this thesis.

TABLE OF CONTENTS

1. **INTRODUCTION 1**

1.1 RESEARCH OBJECTIVES AND SOLUTIONS 2

1.1.1 Defect classification of composite materials using transfer learning methods 5

1.1.2 Ensemble Feature Selection for Clustering Damage Modes in Carbon Fiber-Reinforced Polymer Sandwich Composites Using Acoustic Emission 6

1.1.3 A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing 7

1.2 RESEARCH OUTLINES 8

2. **DEFECT CLASSIFICATION OF COMPOSITE MATERIALS USING TRANSFER LEARNING METHODS 10**

2.1 MOTIVATION 10

2.2 RELATED WORK 12

2.3 MATERIALS AND METHODS 14

2.3.1 Ultrasonic inspection system 14

2.3.2 Datasets 16

2.3.3 Performance metrics 17

2.4 EXPERIMENTS 18

2.5 RESULTS AND DISCUSSION 21

3. **ENSEMBLE FEATURE SELECTION FOR CLUSTERING DAMAGE MODES IN CARBON FIBER-REINFORCED POLYMER SANDWICH COMPOSITES USING ACOUSTIC EMISSION 25**

3.1 MOTIVATION 25

3.2 RELATED WORK 28

3.3 EXPERIMENTS 29

3.3.1 Materials 29

3.3.2 Testing Methods 29

3.4 AE ANALYSIS 30

3.4.1 Ensemble-Supervised Feature Selection 33

3.4.2 Clustering Models 34

3.4.2.1 K-Means Clustering 34

3.4.2.2 Agglomerative Clustering 35

3.4.2.3 Gaussian Mixture Clustering 36

3.4.2.4 Birch 37

3.4.2.5 MiniBatch K-Means 38

3.4.3 Performance Metrics 39

3.5 RESULTS AND DISCUSSION 40

3.5.1 Ensemble Supervised Feature Selection 40

3.5.2 Clustering 41

4. **A FEDERATED LEARNING FRAMEWORK FOR CLASSIFYING THE IMAGES IN ULTRASONIC NONDESTRUCTIVE TESTING 45**

4.1 INTRODUCTION 45

4.2 Related Work 46
4.3 Materials and Methods 47
4.3.1 Preliminaries *48*
4.3.2 Model Architecture *49*
4.4 Experiments 50
4.4.1 Dataset preparation *50*
4.4.2 Model settings and evaluation metric *51*
4.5 Results 52
5. CONCLUSIONS AND FUTURE PROSPECTS 55
5.1 Conclusions 55
5.2 Societal Impact and Contribution to Global Sustainability 56
5.3 Future Prospects 57

LIST OF FIGURES

Figure 2.1 The schematic view of the experimental setup of the ultrasonic NDT system for composite materials classification. 14
Figure 2.2 Ultrasonic inspection images of various composite materials without defects. 16
Figure 2.3 Ultrasonic inspection images of various composite materials with defects. . 17
Figure 2.4 A workflow diagram summarizing the transfer learning process across two tasks. 19
Figure 2.5 The flow diagram of the composite materials classification process. 21
Figure 3.1 Setup of the compression test system and AE equipment, including sensors, preamplifier, and the compression test system with the specimen 30
Figure 3.2 Position of the AE sensors and the deformed specimen after the compression test. 31
Figure 3.3 Illustration of key signals: amplitude, rise time, duration, energy, and counts, with a marked threshold. 32
Figure 3.4 Schematic representation of the proposed methodology, including compression test, AE signal acquisition, feature extraction, feature selection and ranking, and clustering to identify damage characterization. 33
Figure 3.5 Results of four individual and one ensemble supervised feature selection methods labeled with amplitude. 41
Figure 3.6 Results of four individual and one ensemble supervised feature selection methods labeled with peak frequency. 41
Figure 3.7 Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the peak frequency target. 42
Figure 3.8 Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the amplitude. 43
Figure 3.9 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. 44

Figure 4.1 Federated Learning approach for defect classification of ultrasonic inspection images. 49
Figure 4.2 Comparative visualization of defective and non-defective samples in the UsimgAIST dataset. 51
Figure 4.3 Comparison of test accuracy between FL and traditional methods using an ultrasonic inspection image dataset. The number of clients are denoted as C2, C3, C4, and C5, corresponding to 2, 3, 4, and 5 clients, respectively. TA represents the traditional approach. The horizontal axis indicates training progress: each point for traditional methods represents an epoch - a complete pass over the dataset. For FL, each point denotes a round in which every client trains simultaneously on their local data. 53

LIST OF TABLES

Table 2.1 Sample size of the UNDT dataset. 16
Table 2.2 Sample size of the USimgAIST dataset. 17
Table 2.3 Parameters and depths of TL models. 19
Table 2.4 Grid search hyper-parameter values for the transfer learning methods. 20
Table 2.5 Performance results of TL models on UNDT dataset. 22
Table 2.6 Performance results of TL models on USimgAIST dataset. 23
Table 3.1 Sandwich specimen geometry and characteristics. 29
Table 3.2 Descriptions of the features extracted from the AE signals. 31
Table 4.1 Layer configurations of the CNN model. 50
Table 4.2 Distribution of Classes in the UsimgAIST Dataset 51
Table 4.3 Accuracy, Precision, Recall and F1 scores of both FL and traditional learning approach. 54

LIST OF ABBREVIATIONS

NDT	Non-destructive Testing
UT	Ultrasonic Testing
AE	Acoustic Emission
DL	Deep Learning
ML	Machine Learning
DNN	Deep Neural Networks
CNN	Convolutional Neural Networks
TL	Transfer Learning
FL	Federated Learning
CFRP	Carbon Fiber-Reinforced Polymer
AFT	ANOVA F-test
CS	Chi-Squared
MI	Mutual Information
RD	Ridge
GMM	Gaussian Mixture Model
IID	Independently and Identically Distributed
TT	Through Transmission
PE	Pulse-Echo
WOD	Without Defects
WD	With Defects
TA	Traditional Approach

To my family

Chapter 1

Introduction

Composites are advanced materials composed of multiple base materials, each contributing unique structures and properties [1]. The main objective of composite materials is to combine two or more materials to produce unique characteristics. Building up 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 [2]. 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 [3]. These advanced properties come from their complex hierarchical and/or hybrid microstructures. In 2019, the global demand for composites reached up to 90 billion USD, and it is expected to reach up to 160 billion USD by 2027, representing an annual growth rate of 7.6% [4]. However, this complexity also makes them harder to use safely. Composite materials often serve as critical structural components in several technologies. Notably, the failure of these composites is frequently due to microcracks or manufacturing defects, which can lead to catastrophic damage [5,6]. 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 [7].

Non-destructive testing (NDT) encompasses various analysis techniques that evaluate whether a material meets its requirements without causing damage. In recent years, NDT has become an indispensable tool in the industry, allowing quick, safe, reliable, and accurate testing of materials, structures, and components without compromising their integrity or future utility. NDT methods such as ultrasonic testing (UT), and acoustic emission (AE) are available for monitoring the health of composite materials [8]. UT stands out as one of the most widely utilized techniques in the field of NDT [1]. UT is particularly favored for detecting, localizing, and evaluating defects in

composite materials due to its superior accuracy, sensitivity and speed [9-11]. Additionally, its straightforward setup and safety features make it especially preferred compared to other NDT methods. This method entails the transmission of ultrasonic waves through the material and the analysis of the energy reflected from internal discontinuities, such as defects [12]. AE is known for its capability to detect damage at the exact moment it happens, including on very small scales [13]. This method monitors acoustic waves that result from microstructural changes, offering important insights into the sources of damage [14]. This thesis presents machine learning (ML) and deep learning (DL) techniques to detect and classify defects in composite materials using UT and AE data. ML involves developing computational models that generate predictions by inferring patterns from data using statistical and mathematical methods. These models examine input data, produce outcomes, and continually refine themselves to enhance predictive accuracy. Rather than adhering to predetermined rules, ML algorithms learn directly from datasets, allowing them to adapt and improve over time [15]. When applied to UT and AE, these algorithms extract meaningful features from raw data, thereby facilitating accurate defect classification and detection. Deep learning, a specialized subset of ML, relies on artificial neural networks with multiple layers—often referred to as deep neural networks—to learn hierarchical representations of data by capturing complex patterns directly from raw input. In this framework, lower-level layers detect simple patterns, while deeper layers capture increasingly complex abstractions, reducing the need for manual feature engineering [16]. ML and DL approaches not only improve the accuracy, robustness, and efficiency of defect detection and classification in composite materials but also simplify the workflow by minimizing the need for manual feature extraction. Ultimately, this integrated methodology offers a scalable and reliable solution for real-world industrial applications, enhancing quality assurance and structural integrity assessments of advanced composite materials.

1.1 Research Objectives and Solutions

Despite the advantages of UT, such as high inspection speed and high detection rates [17,18], accurately classifying defects in ultrasonic images is an open research challenge. Evaluating NDT data is a time-consuming and error-prone process that largely depends on the inspector's experience, which is acquired over many years of training. An analytical tool that alerts inspectors to potential anomalies would be beneficial and

timesaving. Accurate defect classification minimizes measurement errors, enhances test reliability, and consequently increases the safety of the materials [19]. In recent years, significant advancements in computer vision and ML have enabled computers to achieve or even surpass human performance in various visual tasks, such as object detection/recognition and tracking [20]. With the goal of reducing human errors and effort in ultrasonic imaging inspection, a new trend is emerging that applies computer vision technology to data interpretation challenges [21]. These developments have enhanced the ability to handle more complex UT data, including B-scans, C-scans, and their image representations. Particularly, the advancement of ML algorithms and increased computational power have enabled the development of sophisticated DL models such as deep neural networks (DNN) and convolutional neural networks (CNN) [22], which are a specialized version of artificial neural networks that are extensively used for the classifying of images, audio, and text, opening possibilities for classifying more complex ultrasonic data. Still, the challenge of acquiring a substantial and diverse dataset that captures various defect types across different composite structures persists. Additionally, the research and application of neural networks can be constrained by the scarcity of appropriate datasets and the limitations of computing resources. Under these circumstances, transfer learning (TL) offers an effective alternative for training CNN models, even when data is limited [23]. TL is particularly valuable because it can efficiently learn fundamental features such as edges, textures, patterns, and objects from images, making it a reliable approach for enhancing model performance. Recent advances in TL have led to more sophisticated and robust models for image classification tasks. By leveraging TL models, this approach facilitates the adaptation of these models to new, albeit similar tasks, using relatively limited data. TL enables simpler DL models to be trained with available data and integrated into a larger model on demand by learning the relationships between them [24]. Such strategies significantly diminish the need for extensive data collection for each new type of composite material or structure. Furthermore, TL accelerates the adaptation of models to different environments and materials, enhancing their practical utility in real-world applications. It also maintains high accuracy and performance under varied conditions, thereby playing a crucial role in deploying intelligent diagnostic systems within the field of composite material engineering [25]. The scarcity of publicly available, realistic ultrasonic datasets hampers the development of an effective model for defect classification and also impacts the repeatability and reliability of research findings in this field. Most available datasets are

limited and often rely on data augmentation techniques that fail to accurately mimic real-world conditions [26]. To address challenges scarcity of datasets, Federated Learning (FL), introduced by Google in 2017 [27], emerges as a promising solution. By allowing data to remain distributed and enabling collaborative model training without data exchange, FL addresses privacy concerns, bypasses the limitations of data silos, and harnesses decentralized data sources effectively. This study explores the application of FL in developing advanced ultrasonic inspection models, highlighting how this approach can overcome the traditional barriers posed by data privacy and availability while enhancing the accuracy and applicability of ultrasonic inspection technologies in industrial settings.

In AE analysis, interpreting AE 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 [28]. ML models are rapidly evolving in the field of damage characterization in composite structures and can be broadly categorized into supervised and unsupervised learning approaches [14]. 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 [14]. Adding more in situ observations, like computer tomography [29], microscopy [30], digital image correlation [31] or simulating AE patterns[32] 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 respect, supervised ensemble feature selection methods and unsupervised clustering models are utilized to examine damage mechanisms.

Overall, the thesis begins with an in-depth evaluation of defect classification and characterization. The research primarily delves into three specific studies:

1. Defect classification of composite materials using transfer learning methods [33].
2. Ensemble Feature Selection for Clustering Damage Modes in Carbon Fiber-Reinforced Polymer Sandwich Composites Using Acoustic Emission [34].

3. A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing [35]

1.1.1 Defect classification of composite materials using transfer learning methods

This study aims to improve defect classification in composite materials by applying advanced TL models on ultrasonic inspection images. The research introduces a novel dataset, UNDT, which contains 1,150 C-scan images derived from scans of 60 different composite materials. These materials include artificially induced defects created through projectile impact, simulating real-world defect scenarios. The dataset comprises 850 defective images and 300 non-defective images, with each image having a resolution of 75 × 75 pixels in grayscale format. The utilized UT system comprises an immersion-type ultrasonic bench, pulser-receiver and digitiser unit for acquiring and digitizing ultrasonic signals, transmitter and receiver transducers, and a PC-based software for data analysis, visualization, and control of ultrasonic bench motion.

Two datasets are evaluated in this study: the newly created UNDT dataset and a publicly available USimgAIST dataset, which contains 7,004 ultrasonic inspection images derived from scans of 18 stainless steel materials. The TL approach involves applying state-of-the-art models, including VGG 16, VGG 19, Xception, MobileNet, MobileNet V2, DenseNet 121, DenseNet 169, DenseNet 201, ResNet 50, ResNet 50 V2, ResNet 101, ResNet 101 V2, ResNet 152, ResNet 152 V2, InceptionV3, Inception ResNet V2, and EfficientNetB0. The models are implemented in Python using Keras, and they are optimized using the Adam optimizer with binary cross-entropy loss. The datasets are split into training (80%) and testing (20%) subsets, with a rigorous 5-fold cross-validation process ensuring robust model performance.

Two strategies are explored: firstly, freezing the convolutional base layers and retraining the upper layers which is TL; and secondly, retraining the entire DL model from scratch. The convolutional base layers of the models are weighted using the ImageNet dataset. To optimize performance, grid search hyperparameter tuning is applied, adjusting parameters like the number of neurons, dropout rates, L2 regularization, learning rates, and batch sizes. This fine-tuning helps mitigate overfitting and enhances the models' ability to generalize. The performance metrics used include accuracy, precision, recall, and F1-measure, which are crucial for handling class imbalances in the datasets.

Overall, the findings show that TL enhances defect classification. On the UNDT dataset, the DenseNet121, with frozen convolutional layers, achieved 97.8% accuracy, which improved to 98.8% after grid search and 5-fold cross-validation. Similarly, VGG19, retrained from scratch on the USimgAIST dataset, reached 96.9% accuracy, further refined to 98.6% with hyperparameter tuning. These results highlight the importance of balancing generalization and specialization in TL. Freezing layers prevent overfitting, while retraining adapts models to dataset-specific features. This approach enhances performance, reduces computational costs, and ensures versatility, making TL efficient for real-world applications.

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

This study aims to improve damage characterization in carbon fiber-reinforced polymer (CFRP) sandwich composites by integrating innovative ensemble feature selection techniques and advanced unsupervised clustering models. The research focuses on analyzing AE signals collected during edgewise compression tests on honeycomb-core CFRP composites. The methodology addresses the challenge of identifying damage modes by leveraging supervised feature selection and clustering models to achieve more accurate and interpretable results.

The AE system captures stress waves generated during damage formation and propagation. The experimental setup includes sensors positioned on the specimen during compression tests, following the ASTM C364 standard for edgewise compressive strength. 19 features such as counts, energy, rise time, and partial powers are derived from the AE waveforms for analysis. AE signals are analyzed based on two primary features: amplitude and peak frequency. These features are used to label the AE data due to their relevance to defect detection. Based on the literature, amplitude intervals are defined as 40–60 dB (matrix cracking), 60–80 dB (delamination), and 80–100 dB (fiber breakage). Peak frequency intervals are set at 60–120 kHz (matrix cracking), 120–210 kHz (delamination), and 200–350 kHz (fiber breakage). The dataset is labeled according to these intervals.

The study employs an ensemble-supervised feature selection approach that combines four techniques: ANOVA F-test (AFT), Chi-Squared (CS), Mutual Information (MI), and Ridge (RD). This method ranks AE features based on their relevance to

amplitude and peak frequency. The ensemble method aggregates the results from the four techniques to provide a robust feature ranking, ensuring that the most relevant features are selected for clustering.

Five clustering models are applied to the features ranked by the ensemble-supervised feature selection method. These models include K-Means, Minibatch K-Means, Gaussian Mixture Model (GMM), Agglomerative Clustering, and BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies). These models are chosen for their strengths in handling different types of data. The clustering performance is evaluated using the V-measure, a metric that quantifies the quality of clustering results by measuring homogeneity (the extent to which each cluster contains only data points from a single class) and completeness (the degree to which all data points of a given class are assigned to the same cluster).

The results show that ensemble feature selection enhances clustering performance by identifying key features for damage characterization in CFRP composites. For amplitude-based labeling, counts, energy, duration, and partial power 2 are most relevant, while for peak frequency-based labeling, partial powers 1 and 2, reverberation frequency, and frequency centroid are most significant. GMM and Minibatch K-Means achieve the highest V-measure scores for amplitude 48.75% and 40.68%, and BIRCH performs best for peak frequency 39.7%. These findings highlight the need for concise feature selection to avoid noise and improve clustering efficiency. This combined approach of ensemble feature selection and clustering provides robust damage characterization insights for real-time structural health monitoring in composite materials.

1.1.3 A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing

This paper introduces a FL framework for defect classification using ultrasonic inspection images. Unlike traditional centralized methods, FL enables decentralized training on private datasets, addressing privacy concerns and dataset scarcity issues. The primary aims of this study are to enhance defect classification accuracy, ensure robust model performance across distributed datasets, and facilitate secure collaboration among organizations by preserving sensitive data. Additionally, the study aims to demonstrate the scalability of FL for increasing numbers of clients and highlight its potential to overcome limitations caused by the lack of publicly available, realistic ultrasonic datasets.

The proposed FL framework consists of multiple clients that train models locally on distributed datasets and share model updates with a central server for aggregation using the Federated Averaging algorithm. Experiments are conducted on the publicly available UsimgAIST dataset containing 7,004 ultrasonic inspection images, categorized into defective and non-defective samples. The dataset is split into 80% for training and 20% for testing. The FL approach is tested with 2 to 5 clients, and each client's dataset is balanced and IID (Independently and Identically Distributed). A traditional centralized learning method is also implemented for comparison. In the traditional approach, the entire dataset is used to train the CNN model in a single, centralized location.

A basic CNN architecture is implemented, consisting of convolutional layers, max pooling, and dense layers, with a sigmoid activation function for binary classification. The training process involves 50 rounds for FL, 50 epoch for traditional learning, with performance evaluated on test set using accuracy, precision, recall, and F1-score metrics.

Results show that the FL framework achieves performance comparable to traditional centralized methods. Specifically, the FL model achieves an accuracy of up to 97.1%, while the traditional approach reaches 97.3%. The framework demonstrates robustness and scalability, maintaining accuracy even as the number of clients increases. The findings underscore the potential of FL in preserving data privacy and enhancing defect classification models in NDT applications. Future work suggests exploring advanced CNN architectures and developing a federated platform for broader collaboration in NDT scenarios.

1.2 Research Outlines

The thesis is meticulously organized to ensure thoroughness and clarity. Chapter 2 begins with a review of the existing literature on defect classification in composite materials using TL models. This exploration highlights the challenges and recent advancements in NDT and ML techniques. A comprehensive explanation of the UT system follows, describing the datasets used, including the newly created UNDT dataset and the publicly available USimgAIST dataset. The chapter details the application of TL models such as DenseNet121 and VGG19, covering the classification process, data preprocessing, model training and evaluation metrics. The chapter concludes by presenting the experimental results, demonstrating the effectiveness of TL in defect classification and emphasizing the robustness of the proposed methods.

Chapter 3 expands on the defect classification approach introduced in Chapter 2 by exploring defect characterization through ensemble feature selection for clustering damage modes in CFRP composites using AE. This chapter examines the classification process from a different perspective, focusing on identifying and understanding specific defects within composite materials. The chapter discusses the importance of AE techniques for real-time structural health monitoring and damage characterization, emphasizing key features like amplitude and peak frequency. The experimental setup, including edgewise compression tests and the composite structures analyzed, is explained in detail. An ensemble-supervised feature selection methodology is introduced, utilizing methods like AFT, CS, MI, and RD. Clustering models such as K-Means, Minibatch K-Means, GMM, Agglomerative, and BIRCH are applied to identify damage modes. The chapter concludes with a comparative analysis of the clustering results with V-measure scores, highlighting correlations between selected AE features and specific damage mechanisms such as matrix cracking, delamination, and fiber breakage.

Chapter 4 introduces a FL framework for classifying ultrasonic inspection images. The chapter reviews the significance of FL in addressing data privacy concerns and the challenges posed by the scarcity of publicly available ultrasonic datasets. A detailed exposition of the proposed FL framework follows, explaining the architecture of the CNN and the Federated Averaging algorithm used for model aggregation. The experimental setup involves distributing the USimgAIST dataset across multiple clients to simulate a decentralized training environment. The performance of FL is compared with traditional centralized methods using metrics such as accuracy, precision, recall, and F1-score. The chapter concludes by demonstrating the robustness and scalability of the FL approach, emphasizing its potential for privacy-preserving defect classification.

Finally, Chapter 5 synthesizes the research findings from the previous chapters, highlighting the contributions of TL, ensemble feature selection, and FL to the field of defect classification and characterization in composite materials. The societal impact of these advancements is discussed, particularly their role in enhancing the safety, reliability, and longevity of critical infrastructure. The chapter also explores the contributions to global sustainability through improved defect classificaiton efficiency and reduced material waste. Several potential areas for future investigation are outlined, including the integration of advanced neural network architectures, the exploration of hybrid learning approaches, and the expansion of datasets to improve model generalization and performance.

Chapter 2

Defect classification of composite materials using transfer learning methods

This chapter focuses on the classification of defects in composite materials using UT data and TL models. It presents a novel approach that leverages TL techniques to improve the accuracy of defect classification in ultrasonic inspection images. The study utilizes the newly introduced UNDT dataset—comprising 1,150 images collected from 60 composite materials—as well as the publicly available USimgAIST dataset. Through a comprehensive evaluation of various TL models, the approach demonstrates its effectiveness in enhancing defect classification performance. Among the tested models, DenseNet121 and VGG19 achieve the highest accuracy, underscoring the significant advantages that TL methods offer for defect classification tasks.

2.1 Motivation

Composite materials are extensively employed in various high-technology industries—such as aerospace, automotive, sporting goods, and aviation—due to their exceptional strength, durability, and manufacturability [3]. Despite their advantageous properties, the presence of defects within these materials remains a critical concern, as minor flaws can propagate into fractures and compromise structural integrity. To ensure their suitability for intended applications, comprehensive testing is essential.

In this context, NDT has emerged as an indispensable method for assessing material quality without causing damage. NDT techniques enable reliable and accurate inspection, minimizing both the time and expertise required to ensure the safety and performance of critical components. Among the array of available NDT methods, such as visual testing [36,37], thermography [38-40], electroluminescence [41], and UT [42-44], ultrasonic imaging techniques are frequently employed to detect and characterize defects. Ultrasonic energy captured during inspections is typically visualized in A-scan, B-scan, or C-scan formats [45]. An A-scan displays the amplitude of received ultrasonic energy over time,

generated by the probe's transmission and reception of a single pulse. As the probe moves, it emits and captures ultrasound waves, producing multiple A-scans. These A-scans are then compiled into a B-scan, which presents a coherent image-based representation that simplifies data analysis compared to individual A-scans. Furthermore, a C-scan is constructed by compiling several B-scans into a top-view projection and is used to preliminary assess defect locations. This study employs the C-scan format, which is highly effective in precisely locating defects and assessing the structure of alloys with varying acoustic properties, making it a preferred choice for evaluating composite materials.

Despite recent advancements, accurately classifying defects within ultrasonic images remains an open research challenge. Manual interpretation of NDT data is both time-consuming and susceptible to human error, as it relies heavily on the expertise developed through extensive training and experience. The scarcity of comprehensive datasets and the limited availability of computational resources further constrain the research and application of neural network models for automated defect classification. Under these circumstances, transfer learning has emerged as a promising approach, enabling effective training of CNN models even with limited data, thereby offering a viable pathway toward more reliable and efficient NDT processes [23].

This study introduces a novel dataset composed of ultrasonic inspection images of composite materials, obtained with ultrasonic testing. Transfer learning methods are employed to evaluate state-of-the-art DL models. Our main contributions can be summarized as follows:

- We have created a new ultrasonic inspection dataset called the UNDT dataset, comprising of 1150 C-scan images from the scans of 60 different composite materials. We have made this dataset publicly available so that other researchers can develop different methods and test on this dataset. This dataset is the first of its kind in terms of size and diversity for ultrasonic inspection images of composite materials. By releasing the UNDT dataset, we aim to address the noticeable scarcity of publicly accessible datasets in this domain.
- We have implemented a wide range of state-of-the-art DL models – including VGG16, VGG19, Xception, MobileNet, MobileNetV2, DenseNet variants, ResNet variants, InceptionV3, InceptionResNetV2, and EfficientNetB0 – which vary in depth and parameters. The convolutional

base layers of these models were initialized with pretrained weights from the ImageNet dataset, leveraging learned features from a large-scale image classification task. We have systematically evaluated the performance of these developed models. We analyzed the models' performance on both the newly generated UNDT dataset and on the existing USimgAIST dataset. Across these models, we compared two transfer learning strategies: (a) retraining only the upper layers while freezing the ImageNet-weighted convolutional base layers, and (b) retraining the entire model from scratch.

- We optimized model performance using grid search hyper-parameter tuning and 5-fold cross-validation, essential for adapting models to specific datasets. Our study offers insights into balancing generalization (freezing layers) and specialization (retraining) in transfer learning. By enhancing defect classification, we improve inspection processes and reduce human errors in industries using composite materials.

This chapter is organized as follows: Section 2.2 reviews the existing literature on defect classification. Section 2.3 provides a comprehensive explanation of the UT system and describes the UNDT and USimgAIST ultrasonic inspection image datasets. Section 2.4 details the transfer learning methods and classification steps used in this study. The final section demonstrates the performance of the TL methods.

2.2 Related Work

Earlier studies primarily utilized ML approaches for defect detection and classification in ultrasonic testing, analyzing ultrasonic A-scans. More recently, advanced ML algorithms such as Support Vector Machines (SVMs) [46-50], basic Artificial Neural Networks (ANNs) [51-52], CNNs [53], Long Short-Term Memory networks (LSTM) [54-55], Fully Connected Network-Gated Recurrent Units (FCN-GRU) [56], and DNNs [57-58] have been evaluated for their efficacy in detecting and classifying defects from A-scans [59-60]. Although many researchers have reported promising results, the datasets employed for these evaluations fail to encompass the entire spectrum of potential variations in signal appearance due to their limited sample sizes and the small proportion of scans included. Since these evaluations are limited to small-scale datasets, generalizing their performance to unseen cases and datasets has proven challenging [61]. Furthermore,

a significant limitation of A-scan representation is the lack of contextual information from the surrounding area.

With the emergence of ML techniques and the growth of computational capabilities, sophisticated DL architectures [62] have enabled the classification of more complex UT data, such as B-scans and C-scans, which incorporate spatial information from the surrounding area. DL models are primarily used for image classification and offer the advantage of not requiring manual feature engineering and selection. This benefit arises since these networks, with their convolutional architecture, can autonomously learn the optimal features necessary for accurate classification. In [42], the authors conducted a comprehensive evaluation of transfer learning models, such as ResNet, DenseNet, and VGGNet, in the context of ultrasonic image classification. Similarly, in [43], the authors created various DL models. Their investigation revealed the DNN's superior performance in discriminating defective samples from non-defective samples, surpassing conventional ANN techniques. The author of [44] introduced a specialized CNN architecture, which is designed to classify individual rows of C-scans, achieving remarkable success in defect classification.

Acquiring samples from real or synthetic defects in UT is costly, often resulting in inadequate datasets. This scarcity can undermine model performance due to biased sampling across the entire pattern space of ultrasonic images. Conversely, employing highly complex models risks overfitting during training, thereby diminishing their effectiveness on unseen datasets in practical applications. To mitigate these limitations, strategies such as data augmentation and the use of simulation techniques have been implemented to generate expansive datasets [26,63,64].

In the existing literature, there is a noticeable scarcity of datasets, with only a limited number of studies using publicly accessible datasets. Typically, researchers create their own datasets for experimental analysis. While these efforts provide valuable insights, the absence of standardized datasets limits the ability to conduct fair comparisons between various ML and DL methodologies. This study introduces a new dataset of real-world ultrasonic inspection images, enabling researchers to further explore ML techniques in this domain and to assess the classification performance of their models. Furthermore, the study highlights the effectiveness of utilizing the transfer learning approach.

2.3 Materials and Methods

2.3.1 Ultrasonic inspection system

In this study, the Ultrasonar Deep Structure Analyzer [65] ultrasonic NDT system is employed to capture images from composite materials. The system scans 60 different composite materials. Defects in composite materials are manufactured artificially by applying a projectile's impact load. The projectile's kinetic energy is transferred to the fibres of composites and results in defects. 60 inspection images are collected and annotated as defective or not-defective.

The UT system comprises an immersion-type ultrasonic bench, a pulser-receiver and digitizer unit for acquiring and digitizing ultrasonic signals, transmitter and receiver transducers, and a PC-based software for data analysis, visualization, and control of ultrasonic bench motion. Figure 2.1 provides a schematic view of the experimental setup. The bench operates in an immersion setup and is equipped with servo motors that enable movement along two axes. It features two equal-length long arms, with ultrasonic transducers mounted at their ends to facilitate vertical movement. The movement of the transducers in the water is ensured by filling the glass tank on the workbench with water. The bench can move manually or automatically. Meanwhile, there are two encoders connected to the axes of the machine. The position is determined according to the step signals coming from the encoders. Figure 2.1 shows the setup of the experimental system.

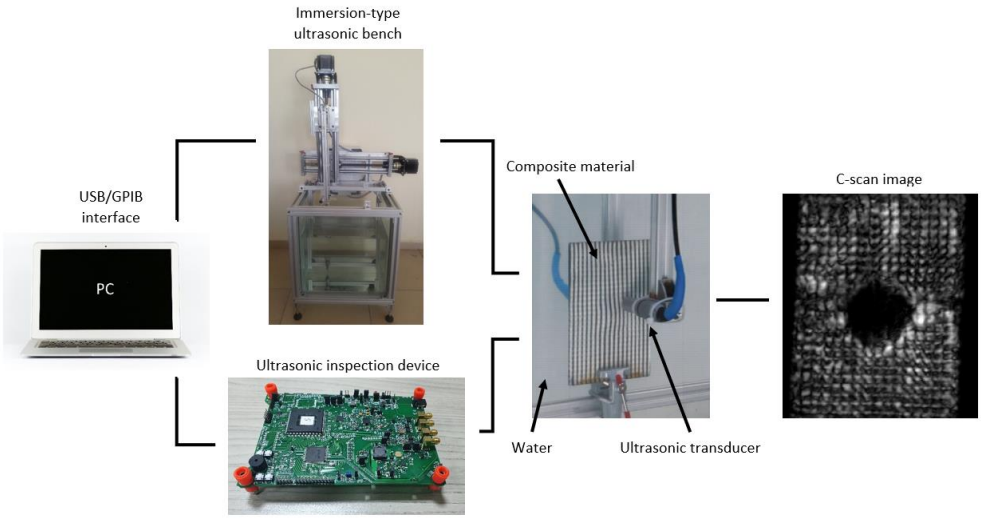

Figure 2.1 The schematic view of the experimental setup of the ultrasonic NDT system for composite materials classification.

The US1000 device from Ultrasonar Defense and Aeronautics Technologies Inc. [65] was utilized to perform UT. This device generates electrical pulses ranging from 500 kHz to 15 MHz, necessary for producing ultrasound waves from piezoelectric transducers. It also captures returning electrical signals at a sampling rate of 100 MSPS. The device connects to a PC via a USB 2.0 interface. C-scan images are produced by integrating encoder position data with ultrasonic signal amplitude values. The voltage of the transducer excitation signal can be adjusted between 45 and 200 Vpp. A voltage-controlled amplifier in the system allows amplification of low-amplitude received signals from 0 dB to 65 dB. To facilitate the inspection of thick materials, the device can operate in tone burst mode. Additionally, conducted and radiated sounds received by the US 1000 ADC can be filtered out using minimum, median, and maximum filters. All images were collected using a 3 MHz inspection frequency, selected based on the composite material thickness and type. The experiments employed immersion-type ultrasonic transducers housed in casings measuring 17 mm and 45 mm in height. The UT system was implemented Through Transmission (TT) and pulse-echo (PE) modes. In the PE mode, one transducer operates as both the transmitter and the receiver, while in TT mode, one transducer operates as the sender and the other as the receiver. The control card is used for transmitting high-amplitude electrical impulses to ultrasonic piezoelectric probes, detecting reflected ultrasonic signals, and completing data acquisition. During the inspection, the amplitude of impulses is adjusted to 90 Vpp with a gain of 40 dB in TT mode and 20 dB in PE mode. These parameters were determined experimentally and were not altered throughout tests. The PE mode of UT often requires greater amplification than the TT mode due to the ultrasonic waves' longer flight through the material in the PE mode. The most and least attenuated portions of the material were taken into consideration while determining the pulse count and pulse voltage. As cut-off and saturation both result in the loss of information about the condition of the material, we set these parameters for signal reception via all material surfaces without passing into any of those zones. We chose 0.25 mm scanning resolution since c-scan image quality improves with higher resolution. Ultrasonic analysis was performed, and C-scan images have been created using raw ultrasonic data. These C-scan images have been utilized in the greyscale color scheme. Images include 8-bit data.

2.3.2 Datasets

This study introduces an ultrasonic inspection image dataset UNDT and utilizes the publicly available USimgAIST ultrasonic image dataset [42]. The UNDT dataset comprises 60 composites of various sizes, each embedded with artificially created defects in their internal and surface structures. To increase the dataset's sample size, operations such as flipping, rotating, and cropping are performed. As a result, the dataset consists of 1150 images; each image is 75 × 75 pixels in greyscale, stored in '.bmp' format, and categorized into two classes: 300 images without defects (WOD) and 850 images with defects (WD), as illustrated in Figures **Figure 2.2** and **Figure 2.3**. The details of the sample sizes in UNDT dataset are shown in

Table 2.1. The UsimgAIST dataset consists of 7004 ultrasonic inspection images, of which 3615 are classified as WD and 3389 as WOD, derived from 18 different stainless-steel materials. The details of the sample sizes in UsimgAIST dataset are presented in **Table 2.2**.

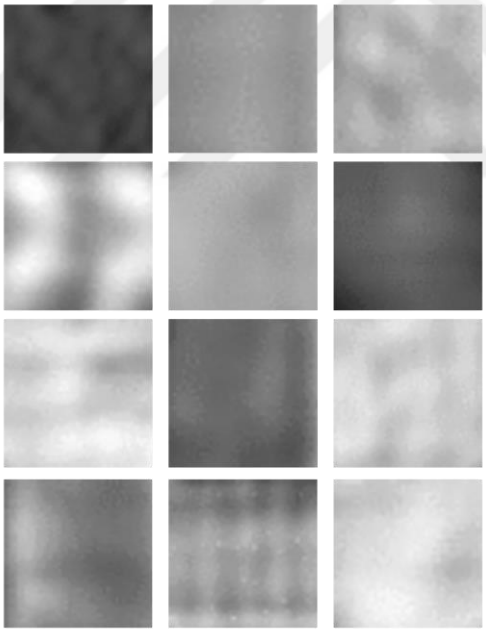

Figure 2.2 Ultrasonic inspection images of various composite materials without defects.

Table 2.1 Sample size of the UNDT dataset.

Class	WD	WOD	Total
Train	544	192	736
Validation	136	48	184
Test	170	60	230

Total	850	300	1150

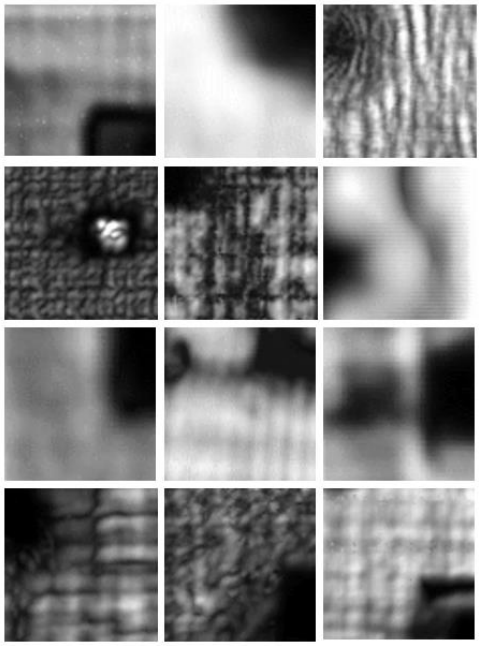

Figure 2.3 Ultrasonic inspection images of various composite materials with defects.

Table 2.2 Sample size of the USimgAIST dataset.

Class	**WD**	**WOD**	**Total**
Train	2314	2169	4483
Validation	578	542	1120
Test	723	678	1401
Total	3615	3389	7004

2.3.3 Performance metrics

While accuracy of defect classification on composite materials is an essential metric, its importance diminishes in the presence of asymmetric datasets. In this study, accuracy alone is inadequate for assessing the performance of a classifier due to the imbalance in dataset. Therefore, to overcome the limitations of using accuracy, this research also considers additional performance metrics such as precision, recall, and f-measure. A true positive (TP) indicates that it predicts defective materials as defective, and true negative (TN) indicates that it predicts a non-defective material as non-defective. Conversely, a false positive (FP) indicates that non-defective materials are predicted to be defective, and a false negative (FN) indicates that defective materials are predicted to

be non-defective materials. Accuracy is calculated as the ratio of correctly predicted labels to the total number of predictions. Precision is the ratio of correctly predicted defective data points to the total predicted as defective. Recall, or sensitivity, reflects the ratio of actual defective materials that are correctly identified. The F-measure is a metric that combines precision and sensitivity under a single parameter. The equations for these metrics are presented in Equation (4.2), (2.2), (2.3) and (4.2), respectively.

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{2.1}$$

$$Precision = \frac{TP}{\mathrm{TP} + \mathrm{FP}} \tag{2.2}$$

$$Recall = \frac{TP}{\mathrm{TP} + \mathrm{FN}} \tag{2.3}$$

$$F - measure = \frac{TP + TN}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{2.4}$$

2.4 Experiments

In this study, the objectives include evaluating the effectiveness of various transfer learning models in classifying ultrasonic inspection images as either with or without having defects, using the proposed UNDT and existing USimgAIST datasets. Two approaches are evaluated. The first approach involves freezing the convolutional base layers while retraining the top layers. The second approach entails retraining the entire transfer learning model from scratch, as illustrated in **Figure 2.4**. The transfer learning models, including VGG 16, VGG 19, Xception, MobileNet, MobileNet V2, DenseNet 121, DenseNet 169, DenseNet 201, ResNet 50, ResNet 50 V2, ResNet 101, ResNet 101 V2, ResNet 152, ResNet 152 V2, InceptionV3, Inception ResNet V2, and EfficientNetB0 [66], which vary in depth and number of parameters, as detailed in **Table 2.3**, have been implemented and trained on the ImageNet dataset. In both approaches, after convolutional base layers, which vary depending on the chosen transfer learning model, a flatten layer is used to convert multi-dimensional tensor output into one dimensional vector.

Subsequently, a single hidden layer is then added, followed by the final layer, which utilizes a sigmoid activation function for binary classification. The Adam optimizer, with binary cross-entropy as the loss function, is employed alongside a specified minimum batch size and a set number of 100 epochs. Model Checkpoint is used to save the model weights at each epoch. The validation loss determines the epoch chosen for test data.

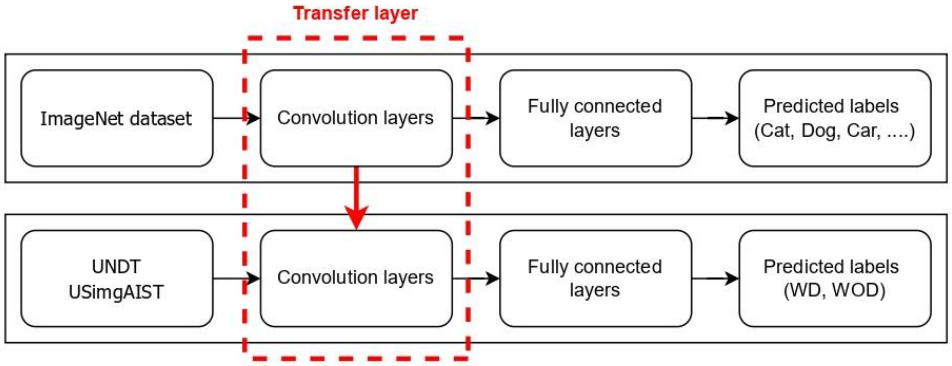

Figure 2.4 A workflow diagram summarizing the transfer learning process across two tasks.

Table 2.3 Parameters and depths of TL models.

Model	Parameters (Million)	Depth
Xception	22.9	81
VGG16	138.4	16
VGG19	143.7	19
ResNet50	25.6	107
ResNet50V2	25.6	103
ResNet101	44.7	209
ResNet101V2	44.7	205
ResNet152	60.4	311
ResNet152V2	60.4	307
InceptionV3	23.9	189
InceptionResNetV2	55.9	449
MobileNet	4.3	55
MobileNetV2	3.5	105
DenseNet121	8.1	242
DenseNet169	14.3	338
DenseNet201	20.2	402
EfficientNetB0	5.3	132

During the training phase of the model, the dataset is divided into training and test sets using stratified random sampling: 80% of the data is selected for the training and the remaining data is selected for the testing. Initially, the models are evaluated using the

training set, then validated on the validation set, and finally tested on the test set. For each best resulting model on both datasets, parameters such as neurons, dropout rate, L2 regularization, learning rate, and batch size are fine-tuned using grid search optimization combined with 5-fold cross-validation on training and validation set to determine the best hyperparameter values. These values are subsequently applied to the test set. The hyperparameter values are presented in **Table 2.4**. The codes for the transfer learning models are developed in Python using the Keras libraries [67]. The process of classifying composite materials is illustrated in the block diagram in **Figure 2.5**.

Table 2.4 Grid search hyper-parameter values for the transfer learning methods.

Parameters	Values
Neurons	32, 64, 128
Dropout Rate	0.1, 0.2, 0.3, 0.4, 0.5
L2 Regularization	0.01, 0.001, 0.0001
Learning Rate	0.01, 0.001, 0.0001
Batch Size	16, 32, 64, 128
Number of Epochs	100

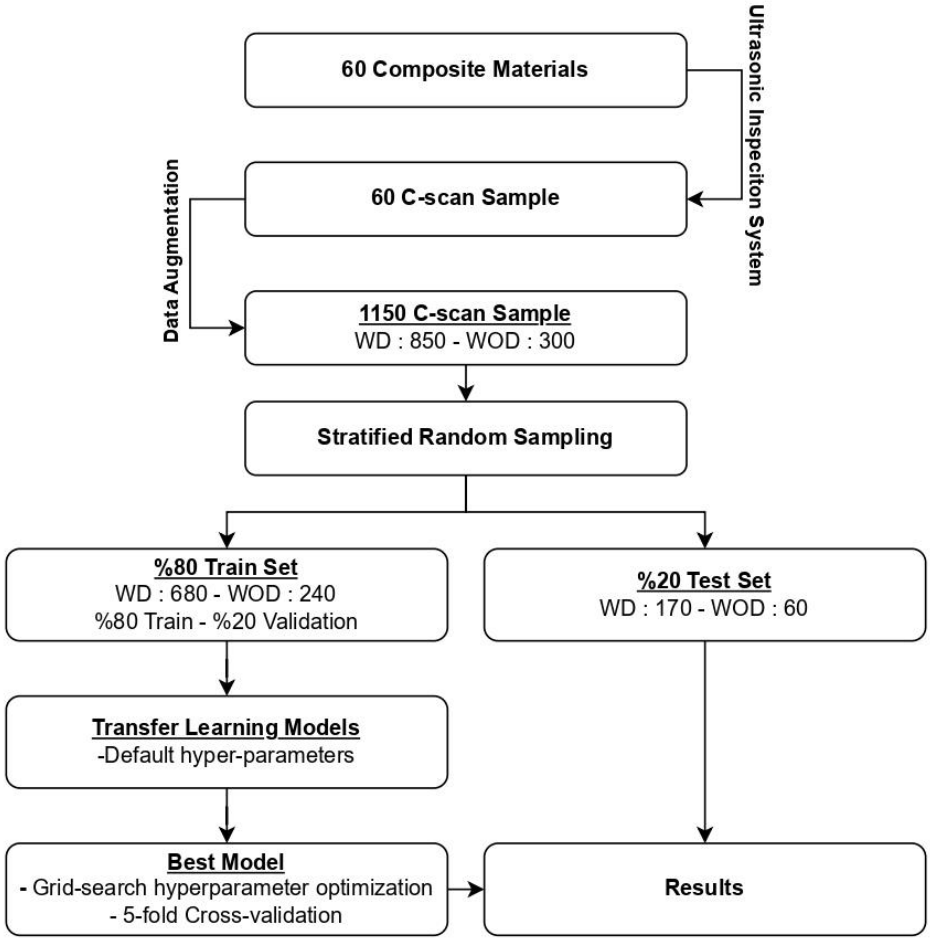

Figure 2.5 The flow diagram of the composite materials classification process.

2.5 Results and Discussion

The aim of this study is to classify ultrasonic inspection images of composite materials using transfer learning models, enhanced by grid search hyper-parameter optimization and 5-fold cross-validation techniques. In this study, 60 different composite materials are scanned using an ultrasonic inspection system, resulting in a total of 1150 C-scan images through data augmentation. Additionally, the publicly available ultrasonic image dataset, USimgAIST, is used for comparative analysis. Various kinds of transfer learning models, each with different depths and parameters, have been employed. In implementing these transfer learning models, two approaches are used: firstly, freezing the convolutional base layers and retraining the upper layers; and secondly, retraining the entire transfer learning model from scratch. The convolutional base layers of the models are weighted using the ImageNet dataset. For both approaches, the accuracies of all models are determined using the following initial hyperparameters: learning rate = 0.01, batch size = 32, dropout = 0.1, L2 regularization = 0.01, with Adam as the optimizer. The DenseNet121 model achieves the highest scores when its convolutional base layers are frozen, recording a 97.8% accuracy. Subsequently, grid search hyper-parameter optimization, and 5-fold cross-validation are applied to the best-performing model DenseNet121. The best results are obtained with optimal hyper-parameters, which are a learning rate of 0.0001, batch size of 16, dropout of 0.1, L2 regularization of 0.0001, and Adam as the optimizer. The accuracy and F1-score for the DenseNet121 model reach up to 98.8%. The performance results of the proposed UNDT dataset are displayed in **Table 2.5**.

For the USimgAIST dataset, using both approaches, the accuracies of all models are determined using the initial hyper-parameters: learning rate = 0.0001, batch size = 128, dropout = 0.1, L2 regularization = 0.01, and Adam as the optimizer. The best result is obtained with the VGG19 model, retraining from scratch, achieving 96.9% accuracy. Subsequently, grid search hyper-parameter optimization, and 5-fold cross-validation are applied to the best-performing model, VGG19. The accuracy and F1-score of the VGG19 model are 98.6% using the following hyperparameters: learning rate = 0.001, batch size = 64, dropout = 0.1, L2 regularization= 0.001, and Adam as the optimizer. **Table 2.6** shows the performance results obtained for the USimgAIST dataset. The results

demonstrate that the transfer learning models utilized in this approach are successful in defect classification.

Table 2.5 Performance results of TL models on UNDT dataset.

Weights	Model	Accuracy (%)	F1-measure (%)	Precision (%)	Recall (%)
Unfreezed	VGG16	73.9	62.8	54.6	50.0
	VGG19	73.9	62.8	54.6	50.0
	EfficientNetB0	73.9	62.8	54.6	50.0
	InceptionV3	74.4	62.8	61.1	57.0
	DenseNet169	77.8	79.1	85.7	82.8
	MobileNet	83.2	62.7	74.8	54.1
	ResNet50V2	84.3	65.6	76.9	57.2
	MobileNetV2	86.2	84.1	74.5	71.8
	Xception	86.9	76.5	72.0	81.7
	DenseNet121	89.4	77.2	87.5	69.2
	ResNet152	90.9	84.5	75.7	95.8
	ResNet152V2	93.0	87.6	81.9	94.2
	ResNet50	93.9	94.0	94.6	94.8
	ResNet101V2	94.7	94.5	95.1	96.0
	DenseNet201	94.7	94.9	95.6	96.4
	InceptionResNetV2	94.8	90.1	88.8	91.6
	ResNet101	96.0	96.1	96.1	95.7
Freezed	VGG19	67.3	63.0	60.7	49.3
	EfficientNetB0	73.9	62.8	54.6	50.0
	VGG16	80.8	81.8	86.0	83.8
	ResNet152	84.7	85.6	90.3	89.7
	ResNet152V2	84.7	85.4	87.5	85.9
	InceptionResNetV2	87.8	88.2	89.2	87.9
	ResNet50V2	89.1	89.4	90.1	88.8
	ResNet50	90.8	91.0	91.5	90.5
	ResNet101	90.8	90.7	90.7	86.8
	InceptionV3	91.3	91.5	92.1	91.4
	DenseNet201	91.3	91.5	92.3	91.9
	Xception	91.7	84.5	82.5	86.6
	MobileNetV2	91.7	91.2	92.2	84.7
	ResNet101V2	92.1	92.3	92.9	92.5
	DenseNet169	94.3	94.3	94.4	93.4
	MobileNet	96.0	96.1	96.2	96.2
	DenseNet121	98.8	98.8	98.9	97.8

The findings from this study highlight the robust capabilities of transfer learning models in classifying ultrasonic inspection images of composite materials. The DenseNet121 model, when its convolutional base layers are frozen, achieved an impressive accuracy of 97.8%. The application of grid search hyper-parameter

optimisation and 5-fold cross-validation further refined the model's performance, boosting the accuracy of DenseNet121 to 98.8%. This underscores the importance of fine-tuning hyper-parameters in achieving optimal model performance. Conversely, the VGG19 model, when retrained from scratch for the USimgAIST dataset, demonstrated a high accuracy of 96.9%, with further enhancements through hyperparameter tuning leading to an accuracy of 98.6%.

Table 2.6 Performance results of TL models on USimgAIST dataset.

Weights	Model	Accuracy (%)	F1-measure (%)	Precision (%)	Recall (%)
Unfreezed	EfficientNetB0	88.4	88.3	90.5	88.5
	ResNet50	90.6	90.5	92.1	90.6
	ResNet101	93.6	93.6	93.7	93.6
	DenseNet201	94.4	94.4	94.5	94.4
	MobileNet	95.0	95.0	95.4	95.0
	MobileNetV2	95.0	95.0	95.4	95.0
	ResNet152	95.0	95.0	95.5	95.0
	ResNet152V2	95.8	95.8	96.0	95.8
	ResNet50V2	95.9	95.9	96.2	95.9
	DenseNet121	96.1	96.1	96.4	96.1
	DenseNet169	96.2	96.2	96.4	96.2
	ResNet101V2	96.2	96.2	96.4	96.2
	VGG16	96.3	96.3	96.5	96.3
	InceptionV3	96.5	96.5	96.7	96.5
	InceptionResNetV2	96.8	96.8	96.8	96.8
	Xception	97.0	97.1	94.4	94.4
	VGG19	98.6	98.6	98.6	98.6
Freezed	EfficientNetB0	49.9	33.2	24.9	50.0
	ResNet152	86.6	88.3	86.4	86.6
	ResNet50	86.7	86.6	87.8	86.7
	ResNet101	87.7	87.7	88.2	87.7
	VGG16	89.0	89.0	90.1	89.1
	Xception	89.0	89.9	83.4	97.4
	InceptionResNetV2	91.2	91.1	91.6	91.2
	InceptionV3	91.2	91.1	91.7	91.2
	MobileNetV2	91.5	91.5	92.3	91.5
	DenseNet169	92.1	92.1	92.2	92.1
	DenseNet121	92.3	92.3	92.5	92.3
	ResNet50V2	92.4	92.3	92.9	92.4
	ResNet101V2	92.9	92.9	92.9	92.9
	VGG19	93.0	93.0	93.0	93.0
	ResNet152V2	93.6	93.6	94.1	93.6
	DenseNet201	94.4	94.4	94.5	94.4
	MobileNet	94.9	94.9	95.0	94.9

These results reveal two critical aspects of transfer learning: the value of leveraging transfer learning methods for common features and the advantage of adapting models to specific datasets through extensive training. Both strategies, depending on the nature of the dataset and its similarity to pre-trained data, have proven effective in enhancing the accuracy and robustness of the model. These observations underscore the critical balance between generalization and specialization. Freezing layers helps to maintain generalization and prevents overfitting, while unfreezing layers allows for greater specialization, optimizing performance by adapting more specifically to the unique features of each dataset. This nuanced approach ensures that transfer learning not only reduces the need for extensive computational resources and training time but also enhances model versatility and effectiveness across different operational conditions. These benefits of efficiency and adaptability make transfer learning an invaluable strategy in the rapid deployment of models in practical applications, where both time and resource efficiency are paramount.

Chapter 3

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

This study focuses on the characterization of damage modes in CFRP composites under lateral loading through AE analysis. It presents a novel methodology that integrates AE measurements, ensemble-supervised feature selection, and clustering models to precisely identify key signal features correlated with damage modes. The work utilizes both time and frequency AE data extracted during compression testing, incorporating reference intervals for amplitude and peak frequency from existing literature. Through a comprehensive evaluation of multiple feature selection and clustering methods, the proposed approach enhances the accuracy of damage characterization and offers new insights into complex damage mechanisms. Among the examined models, partial power features emerge as critical indicators, underscoring their significance alongside commonly used features. The findings highlight the effectiveness and generalizability of the adopted techniques for improving AE-based damage characterization in CFRP composites.

3.1 Motivation

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. 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 [68]. 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 [69]. These complex damage dynamics underscore the critical need for rigorous health monitoring to ensure the structural integrity and longevity of composites.

AE is known for its capability to detect damage at the exact moment it happens, including on very small scales [13]. This method monitors acoustic waves that result from microstructural changes, offering important insights into the sources of damage [14]. 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 [28].

ML methods are rapidly evolving in the field of damage characterization in composite structures and can be broadly categorized into supervised and unsupervised learning approaches [14]. 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 [14]. Adding more in situ observations, like computer tomography [29], microscopy [30], digital image correlation [31] or simulating AE patterns [32] 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 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 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 ML methods to improve damage characterization and description. The novelty of this study lies in three primary areas:

- The examination of crack formation and propagation in carbon fiber composite sandwiches under edgewise loading through AE measurements.
- The application of an ensemble supervised feature selection method that utilizes reference intervals for amplitude and peak frequency features as established in existing literature.
- A comprehensive analysis of the most effective features through the evaluation of clustering model outcomes.

The article's main contributions are summarized as follows:

- 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.
- 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.
- 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 chapter is organized as follows: Section 3.2 reviews the literature on AE techniques for classifying damage characterization. Section 3.3 describes the experimental setup, and the composite structures used. Section 3.4 outlines ensemble-supervised feature selection and clustering methods. Finally, Section 3.5 offers a comparative analysis of the results and discusses the findings.

3.2 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 [70]. Numerous studies have investigated the various damage modes associated with AE waves generation [71]. 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 [72]. In composites, particularly CFRP, common AE features for damage assessment include amplitude, peak frequency, counts, rise time, energy, and average frequency [70,73]. Most studies utilize unsupervised learning models such as K-Means [74,75], fuzzy C-means [75,76], K-Means++ [77,78], and Gaussian mixture models [79,80] 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 [81,82], correlation coefficient [78,81], Davies–Bouldin index [83,84], and principal component analysis [85,86] are employed to identify the essential features. In [82], 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. The study in ref. [75] utilized AE features, which include count, signal strength, 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 values. In ref. [81], 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 peak frequency are the main AE features used to group AE signals [87,88]. There is a consensus in the literature regarding the correlation between these features and specific damage mechanisms in composite materials [70], [75,80,89-92]. This agreement reinforces the reliability of using these AE features to effectively identify and characterize various types of damage.

3.3 Experiments

3.3.1 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 3.1**. Composite face sheets were produced by the hot-pressing method by stacking prepregs and placing them in the mold. The production was completed under appropriate pressure and temperature conditions and cut into sandwich structure geometry. After the honeycomb material was cut to the 100 x 150 mm dimensions, the composite plates were ready for sandwich-making. The composite plates were cleaned with ethanol. After the surface of the composite plates 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.

Table 3.1 Sandwich specimen geometry and characteristics.

Length [mm]	Width [mm]	Thickness [mm]	Number of plies
Face-sheets-CFRP-150	100	2.1	10 [0/90°]
Core-polypropylene honeycomb-150	100	15	–

3.3.2 Testing Methods

The mechanical test was carried out following ASTM C364– Standard Test Method for Edgewise Compressive Strength of Sandwich Constructions [93]. The compression

test system and AE equipment are shown in **Figure 3.1**. The compression load was applied under displacement-controlled mode at a constant crosshead displacement of 1 mm min1. **Figure 3.2** 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 are 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.

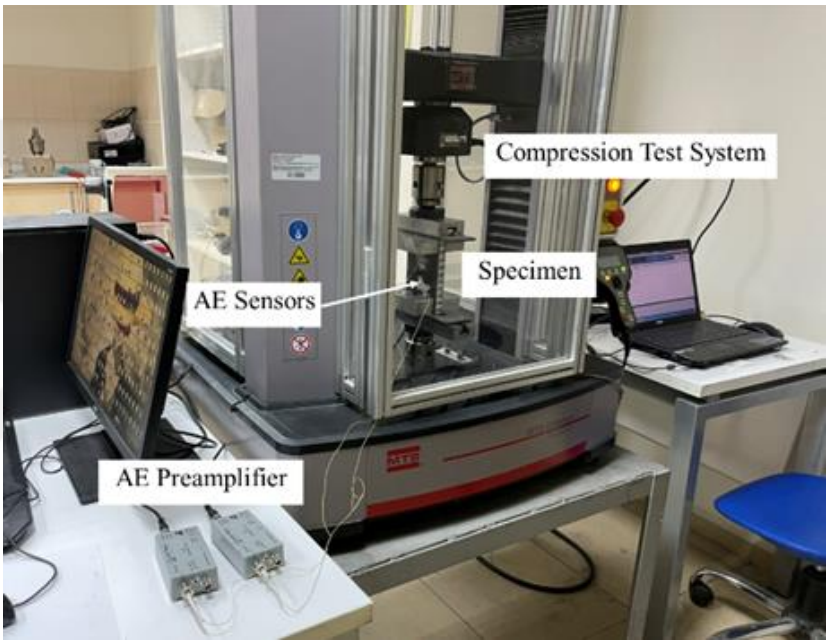

Figure 3.1 Setup of the compression test system and AE equipment, including sensors, preamplifier, and the compression test system with the specimen.

3.4 AE Analysis

In this research, 19 features are derived from the AE waveforms for analysis, with each feature detailed in **Table 3.2**. **Figure 3.3** illustrates a typical AE signal with commonly used features, which have been extensively discussed in the literature [70,71,94]. **Figure 3.4** provides a visual representation of the methodology.

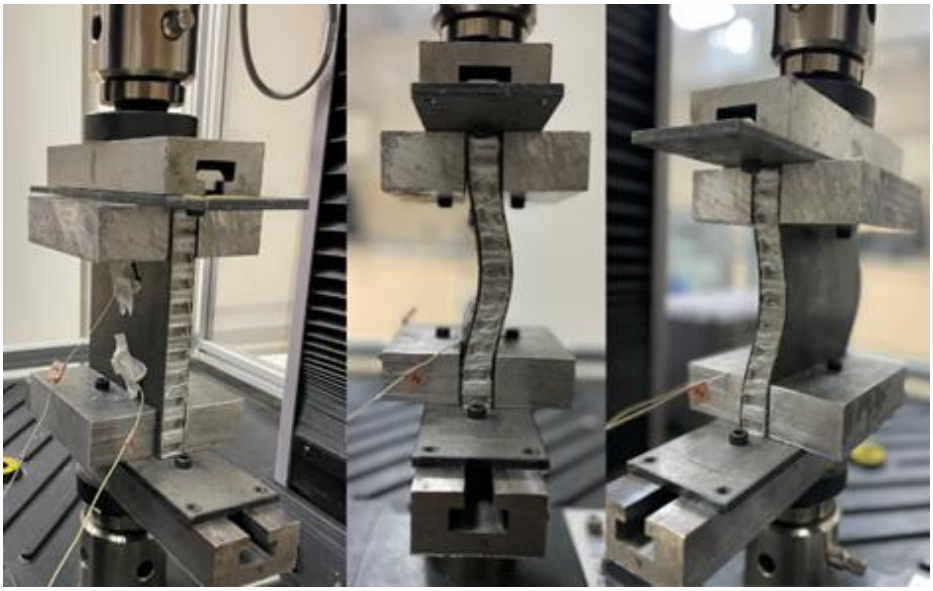

Figure 3.2 Position of the AE sensors and the deformed specimen after the compression test.

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 peak frequency. 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.

Table 3.2 Descriptions of the features extracted from the AE signals.

Feature	Description
Counts	The aggregate number of times the AE signal exceeds the predefined threshold throughout the measurement period [count].
Energy	The total energy of the AE signal from the time it exceeds the detection threshold until the signal ends [aJ].
Duration	The measure is the time from when the signal first crosses the detection threshold to when it last [μs].
Counts to peak	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	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.
Reverberation frequency	(Counts – Counts to peak) / (Duration – Rise time) [kHz].
Initiation frequency	Counts to peak / Rise time (kHz).
Signal strength	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].
Average frequency	The proportion of the total counts and duration of the waveform [kHz].

Frequency centroid	Weighted mean of frequency components, determined by conducting fast Fourier transform (FFT) and executing calculations on each FFT element [kHz].
Peak frequency	The frequency at which the AE signal's amplitude is at its maximum, measured from the FFT data [kHz].
Amplitude	The maximum voltage level of the AE signal, often expressed in decibels relative to a reference voltage [dB].
Partial power 1	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 0 and 200 [kHz].
Partial power 2	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 200 and 400 [kHz].
Partial power 3	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 400 and 600 [kHz].
Partial power 4	Indicates the portion of the signal power that exists in the frequency spectrum of the signal between 600 and 800 [kHz].

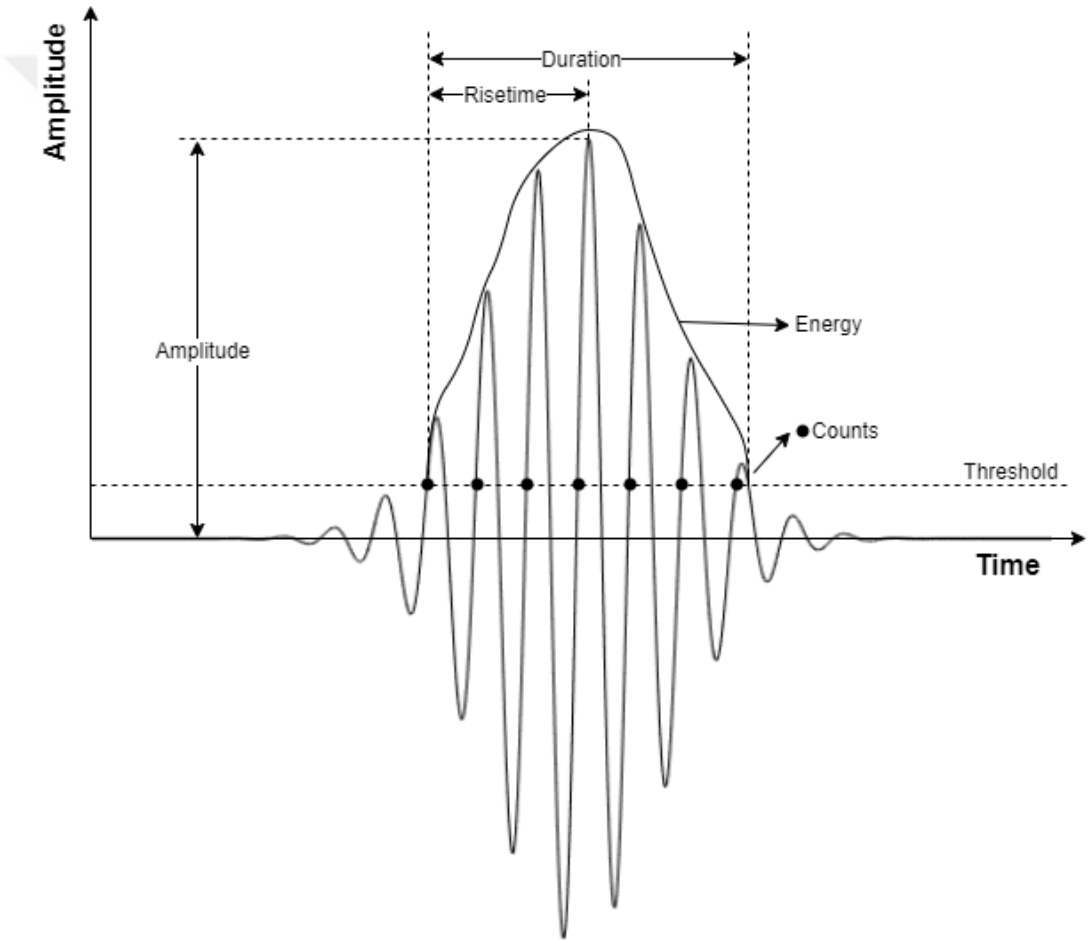

Figure 3.3 Illustration of key signals: amplitude, rise time, duration, energy, and counts, with a marked threshold.

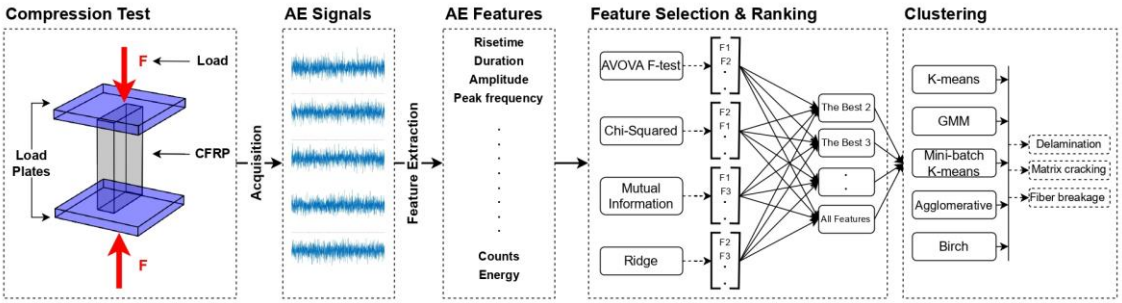

Figure 3.4 Schematic representation of the proposed methodology, including compression test, AE signal acquisition, feature extraction, feature selection and ranking, and clustering to identify damage characterization.

3.4.1 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: AFT, CS, MI, and RD [95] 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, peak frequency 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 [96]. The reference intervals of peak frequency 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 [97]. 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 peak frequency during training. Second, peak frequency 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 peak frequency 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.

3.4.2 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 peak frequency as the target variable.

3.4.2.1 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 [98].

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 $X = \{x_1, x_2, \ldots, x_n\}$, 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 ((3.1).

$$J(C, \mu) = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \tag{3.1}$$

In Equation ((3.1), 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 ((3.2).

$$\mathrm{C}_k = \{x_i : ||x_i - \mu_k|| \leq ||x_i - \mu_j|| \forall j, 1 \leq j \leq K\} \tag{3.2}$$

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.

3.4.2.2 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 [99].

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.

3.4.2.3 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 [100].

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 $X = \{x_1, x_2, \ldots, x_n\}$, where each x_i is a *d*-dimensional vector, the GMM aims to model the data using the likelihood function in Equation ((3.3).

$$p(x_i|\Theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \tag{3.3}$$

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 ((3.4). In the M-step, these probabilities are used to update the parameters Θ, thereby maximizing the likelihood of the observed data.

$$w_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)} \tag{3.4}$$

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

3.4.2.4 Balanced Iterative Reducing and Clustering Using Hierarchies (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 [101]. Additionally, BIRCH is used for anomaly detection in network traffic [102]. 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 (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 $X = \{x_1, x_2, \ldots, x_n\}$, where each x_i is a *d*-dimensional vector, the *CF* for a cluster *C* containing *N* data points is defined as in Equation ((3.**5**).

$$\mathrm{CF} = \left(N, \sum_{i=1}^{N} x_i, \sum_{i=1}^{N} x_i^2 \right) \tag{3.5}$$

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 xi 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.

3.4.2.5 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 [103]. 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.

3.4.3 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 ((3.6), 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.

$$h = 1 - \sum_{k=1}^{K} \sum_{c=1}^{C} \frac{n_{k,c}}{N} \log \left(\frac{n_{k,c}}{n_k} \right) \tag{3.6}$$

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 ((3.7). 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.

$$c = 1 - \sum_{c=1}^{C} \sum_{k=1}^{K} \frac{n_{k,c}}{N} \log\left(\frac{n_{k,c}}{n_c}\right) \tag{3.7}$$

The V-measure is defined as in Equation ((3.**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.

$$V = 2 \times \frac{h \times c}{h + c} \tag{3.8}$$

3.5 Results and Discussion

3.5.1 Ensemble Supervised Feature Selection

In this study, the dataset is labeled separately using two features: amplitude and peak frequency. 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 3.5**. 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, RMS, and partial power 2 are found to have a strong correlation with amplitude, while reverberation frequency, average frequency, partial power 1, initiation frequency, and signal stregth exhibited a weak correlation with amplitude. The same process is applied by selecting peak frequency as the target. The results of the feature selection methods and the ensemble feature selection for peak frequency are shown in **Figure 3.6**. According to the ensemble feature selection, partial power 1, partial power 2, reverberation frequency, frequency centroid, and partial power 3 exhibited a strong correlation with peak frequency, whereas counts to peak, duration, average signal level, counts, and enery showed a weak correlation with peak frequency. This method offers a more thorough and robust set of features, effectively capturing a broad range of feature interactions and dependencies.

Anova F-Test	
Rank	Feature
17	Counts
16	ASL
15	Energy
14	Duration
13	RMS
12	Partial Power 4
11	Partial Power 2
10	FrequencyCentroid
9	CountstoPeak
8	PartialPower3
7	Risetime
6	PartialPower1
5	InitiationFrequency
4	AverageFrequency
3	AbsoluteEnergy
2	SignalStrength
1	ReverberationFrequency

Chi-squared	
Rank	Feature
17	PartialPower4
16	ASL
15	Energy
14	RMS
13	Counts
12	PartialPower2
11	Duration
10	PartialPower3
9	CountstoPeak
8	FrequencyCentroid
7	Risetime
6	InitiationFrequency
5	AbsoluteEnergy
4	SignalStrength
3	AverageFrequency
2	PartialPower1
1	ReverberationFrequency

Mutual-info	
Rank	Feature
17	Energy
16	SignalStrength
15	Counts
14	AbsoluteEnergy
13	CountstoPeak
12	Duration
11	RMS
10	Risetime
9	ReverberationFrequency
8	FrequencyCentroid
7	InitiationFrequency
6	PartialPower2
5	ASL
4	AverageFrequency
3	PartialPower4
2	PartialPower1
1	PartialPower3

Ridge	
Rank	Feature
17	Counts
16	Energy
15	PartialPower2
14	Duration
13	AbsoluteEnergy
12	Risetime
11	SignalStrength
10	FrequencyCentroid
9	PartialPower3
8	RMS
7	PartialPower4
6	PartialPower1
5	ASL
4	AverageFrequency
3	CountstoPeak
2	ReverberationFrequency
1	InitiationFrequency

Ensemble Feature Selection	
Median of Ranks	Feature
16	Counts
15,5	Energy
13	Duration
12	RMS
11,5	PartialPower2
10,5	ASL
9,5	PartialPower4
9	AbsoluteEnergy
9	FrequencyCentroid
9	CountstoPeak
8,5	Risetime
8,5	PartialPower3
7,5	SignalStrength
5,5	InitiationFrequency
4	PartialPower1
4	AverageFrequency
1,5	ReverberationFrequency

Figure 3.5 Results of four individual and one ensemble supervised feature selection methods labeled with amplitude.

Anova F-Test	
Rank	Feature
17	PartialPower1
16	PartialPower2
15	ReverberationFrequency
14	PartialPower3
13	FrequencyCentroid
12	PartialPower4
11	AverageFrequency
10	InitiationFrequency
9	Counts
8	RMS(16)
7	SignalStrength
6	AbsoluteEnergy
5	ASL
4	Energy
3	CountstoPeak
2	Duration
1	Risetime

Chi-squared	
Rank	Feature
17	PartialPower2
16	PartialPower3
15	PartialPower4
14	PartialPower1
13	ReverberationFrequency
12	AverageFrequency
11	SignalStrength
10	AbsoluteEnergy
9	FrequencyCentroid
8	InitiationFrequency
7	RMS(16)
6	ASL
5	Counts
4	Energy
3	CountstoPeak
2	Duration
1	Risetime

Mutual-info	
Rank	Feature
17	PartialPower1
16	PartialPower2
15	AverageFrequency
14	ReverberationFrequency
13	FrequencyCentroid
12	Risetime
11	PartialPower3
10	PartialPower4
9	InitiationFrequency
8	RMS(16)
7	Duration
6	SignalStrength
5	AbsoluteEnergy
4	Energy
3	ASL
2	CountstoPeak
1	Counts

Ridge	
Rank	Feature
17	PartialPower1
16	PartialPower2
15	FrequencyCentroid
14	ReverberationFrequency
13	SignalStrength
12	AbsoluteEnergy
11	PartialPower3
10	Energy
9	AverageFrequency
8	Risetime
7	PartialPower4
6	CountstoPeak
5	Duration
4	InitiationFrequency
3	Counts
2	RMS(16)
1	ASL

Ensemble Feature Selection	
Median of Ranks	Feature
17	PartialPower1
16	PartialPower2
14	ReverberationFrequency
13	FrequencyCentroid
12,5	PartialPower3
11,5	AverageFrequency
11	PartialPower4
9	SignalStrength
8,5	InitiationFrequency
8	AbsoluteEnergy
7,5	RMS
4,5	Risetime
4	Energy
4	Counts
4	ASL
3,5	Duration
3	CountstoPeak

Figure 3.6 Results of four individual and one ensemble supervised feature selection methods labeled with peak frequency.

3.5.2 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 peak frequency. The performance of each model is evaluated using the V-measure. In the clustering performance results for the peak frequency 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 3.7**. This notable difference in V-measure scores indicates that the preset labeling intervals for peak frequency 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 peak frequency. The inclusion of the third (reverberation frequency), fourth (frequency centroid), 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 complexity rather than clarity, potentially introducing noise or irrelevant variables.

Figure 3.7 Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the peak frequency target.

The clustering performance results for the amplitude target show varying trends across different numbers of features, as illustrated in **Figure 3.8**. 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.

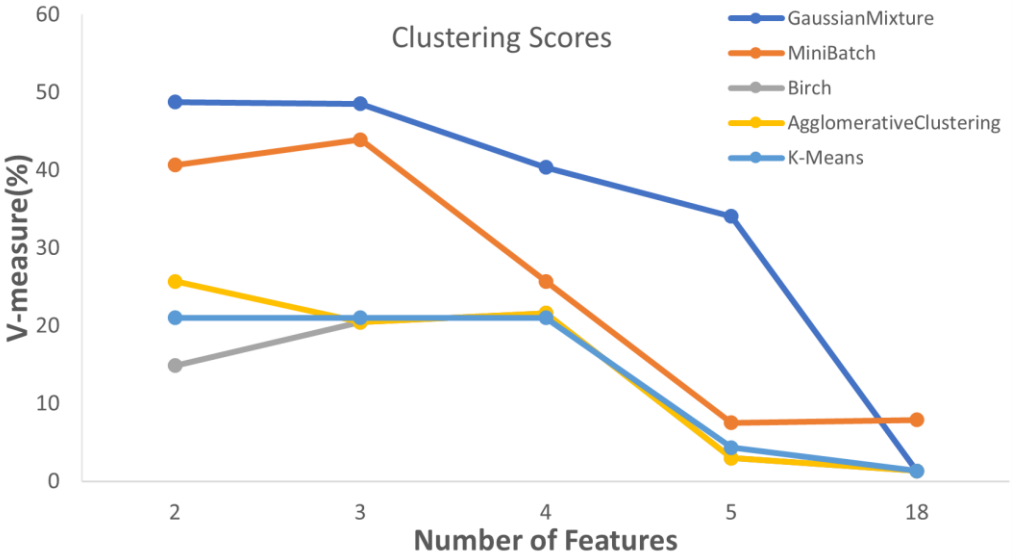

Figure 3.8 Comparison of clustering scores (V-measure) for various algorithms across different numbers of features under the amplitude.

The GMM consistently achieves high V-measure scores in both targets, indicating its effectiveness as an initial clustering model in AE studies. **Figure 3.9** 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, d) all features. In **Figure 3.9**a, 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 ≈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 3.9**b,c, the overlap between clusters 1 and 2 contributes to the decrease in V-measure scores. In **Figure 3.9**d, cluster 1 shows precise detection, whereas clusters 2 and 3 display notable misclassification.

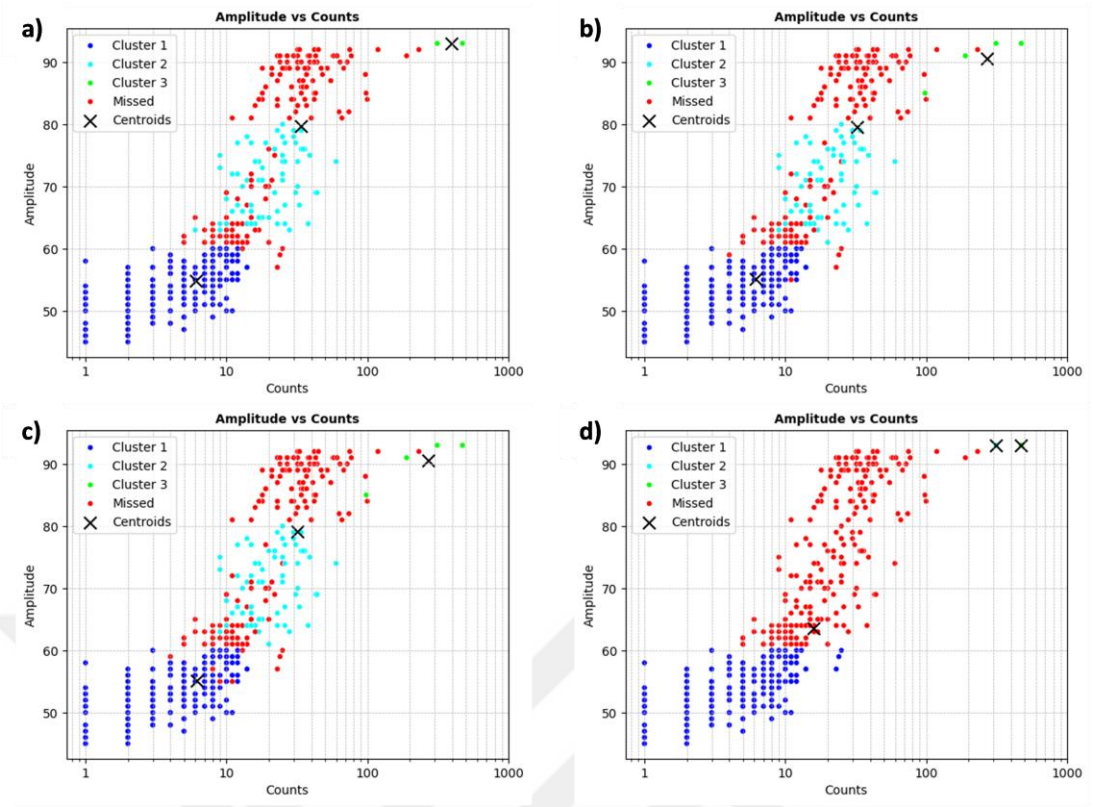

Figure 3.9 GMM clustering results with different number of features for amplitude target. Missed indicates incorrectly clustered points. The x-axis is scaled logarithmically. a) Amplitude versus counts with counts and energy features, b) amplitude versus counts with counts, energy, duration, and RMS features, c) amplitude versus counts with counts, energy, duration, RMS, and partial power 2 features, and d) amplitude versus counts with all features.

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.

Chapter 4

A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing

This chapter explores the classification of defects in ultrasonic inspection images using FL to address challenges related to data privacy and availability. The study introduces a centralized FL framework that enables collaborative model training while preserving data privacy, overcoming traditional data-sharing limitations. By applying FL to UT images, the study represents the first known implementation of FL for defect classification in this context. The findings demonstrate that FL achieves performance comparable to traditional centralized methods, highlighting its potential to enhance the defect classification accuracy in industry.

4.1 Introduction

UT faces considerable challenges, primarily due to the scarcity of publicly available, realistic ultrasonic datasets. Most available datasets are limited and often rely on data augmentation techniques that fail to accurately mimic real-world conditions [26]. This not only hampers the development of an effective model for defect classification but also impacts the repeatability and reliability of research findings in this field. The lack of authentic datasets stymies the progression of ultrasonic testing technologies, which is crucial for improving the accuracy and reliability of safety inspections across various industries.

To address these multifaceted challenges, FL, introduced by Google in 2017 [27], emerges as a promising solution. By allowing data to remain distributed and enabling

collaborative model training without data exchange, FL addresses privacy concerns, bypasses the limitations of data silos, and harnesses decentralized data sources effectively. This study explores the application of FL in developing advanced ultrasonic inspection models, highlighting how this approach can overcome the traditional barriers posed by data privacy and availability while enhancing the accuracy and applicability of ultrasonic inspection technologies in industrial settings.

In this study, we have dedicated our efforts to the development and validation of a system based on FL for the classification of defects from ultrasonic inspection images. This initiative represents the core innovation of our research. To the best of our knowledge, this study is the inaugural exploration of applying FL to ultrasonic inspection images for defect classification. The principal contributions of our paper are as follows:

- We introduce a centralized FL framework that enables researchers to access and leverage private ultrasonic inspection data, while simultaneously preserving data privacy.
- Our findings demonstrate that the FL approach achieves comparable performance to traditional methods.
- This study represents the first known application of FL techniques to ultrasonic inspection images.

The remainder of this chapter is organized as follows: Section 4.2 reviews the relevant literature. Section 4.3 provides a detailed exposition of our proposed framework. Section 4.4 elaborates on the experimental setup, and Section 4.5 presents the results obtained, comparing the performance of both federated and traditional learning approaches using an ultrasonic inspection image dataset.

4.2 Related Work

The integration of DL in image classification for ultrasonic inspection has emerged as a focal area of research, corroborated by a series of pivotal studies as highlighted in [53]. In [43], researchers conducted a comprehensive evaluation of various methodologies for classifying UT images, pinpointing DL models as the most effective. In [57], the authors examined a DL network augmented with dropout regularization. This model demonstrated superior performance over traditional defect classification methods, which rely on prior manual feature extraction.

Despite these advancements, the field faces a significant challenge: the scarcity of datasets containing defective examples. This limitation restricts the development of automated ultrasonic analysis, a problem partially mitigated by generating synthetic examples, as discussed in [63,64]. The impact of limited datasets is twofold. On one hand, small-scale datasets hinder the ability of deep neural networks to develop robust feature representations, leading to inadequate model performance due to biased sampling across the entire pattern space of ultrasonic inspection images. On the other hand, employing an overly complex model, which is characterized by numerous layers predisposes the system to overfit on the limited training data, thereby impairing its performance on new, unseen data in practical settings [104]. To address these challenges, recent initiatives have sought to expand the dataset size through simulation and data augmentation techniques, as discussed in [26,105]. While these methods are useful, unfortunately, they do not fully resolve the issues related to data privacy and the unavailability of large, diverse datasets.

Recognizing these persistent gaps, this paper proposes the development of a collaborative framework designed to facilitate defect classification research on large datasets without compromising privacy. This framework employs FL, a centralized approach that allows for the collective use of distributed data sources. By processing data locally and only sharing model updates, FL ensures data privacy and security while enabling the use of extensive, diverse datasets. Through FL, we demonstrate that it is possible to meet the stringent requirements for data privacy without sacrificing the model's performance. This approach not only addresses the immediate needs of the ultrasonic inspection field but also sets a precedent for future research in similar areas requiring data confidentiality.

4.3 Materials and Methods

In this section, we provide a comprehensive overview of our proposed methodology for classifying defects in ultrasonic inspection images. Initially, we outline the foundational concepts within the context of FL, followed by a detailed description of the model architecture used during the training phase.

4.3.1 Preliminaries

FL is a distributed machine learning framework where multiple clients collaboratively train a global model using their private datasets $\{D_i\}_{i=1}^{N}$ without data exchange, enhancing privacy and reducing communication costs. The architecture of the proposed FL approach is illustrated in Figure 4.1. The process involves initializing a global model on a central server with parameters, distributing it to selected clients for local training, and then aggregating the locally updated models using the Federated Averaging algorithm [27]. Federated Averaging centralizes the aggregation phase by computing a weighted average of the model parameters returned by the clients. The weights are given by $\frac{n_k}{n}$, indicating the proportion of total data contributed by each client *k*, where n_k is the number of data samples from client *k* and *n* is the aggregate number of samples across all clients. The aggregation formula, which is expressed in Equation ((4.1).

$$\theta_{t+1} = \theta_t + \eta \sum_{k=1}^{K} \frac{n_k}{n} (\theta_{t+1}^k - \theta_t) \tag{4.1}$$

Equation ((4.1) incorporates θ_t, the global model parameters at iteration *t*, and computes θ_{t+1}^k, the parameters after local training on client *k*. The parameter η serves as the learning rate, adjusting the impact of each update on the global model, while *K* represents the number of clients participating in that round of training. The overall learning objective aims to minimize the global loss function defined in Equation ((4.2).

$$f(\theta) = \sum_{i=1}^{N} p_i F_i(\theta) \tag{4.2}$$

In Equation ((4.2), $F_i(\theta)$ is the loss function associated with the dataset of client *i*, and p_i, the proportion of the total dataset represented by client *i*, emphasizes the significance of each client's data in the global model. This strategic framework paves the way for further developments in enhancing model performance within diverse and dynamic federated environments.

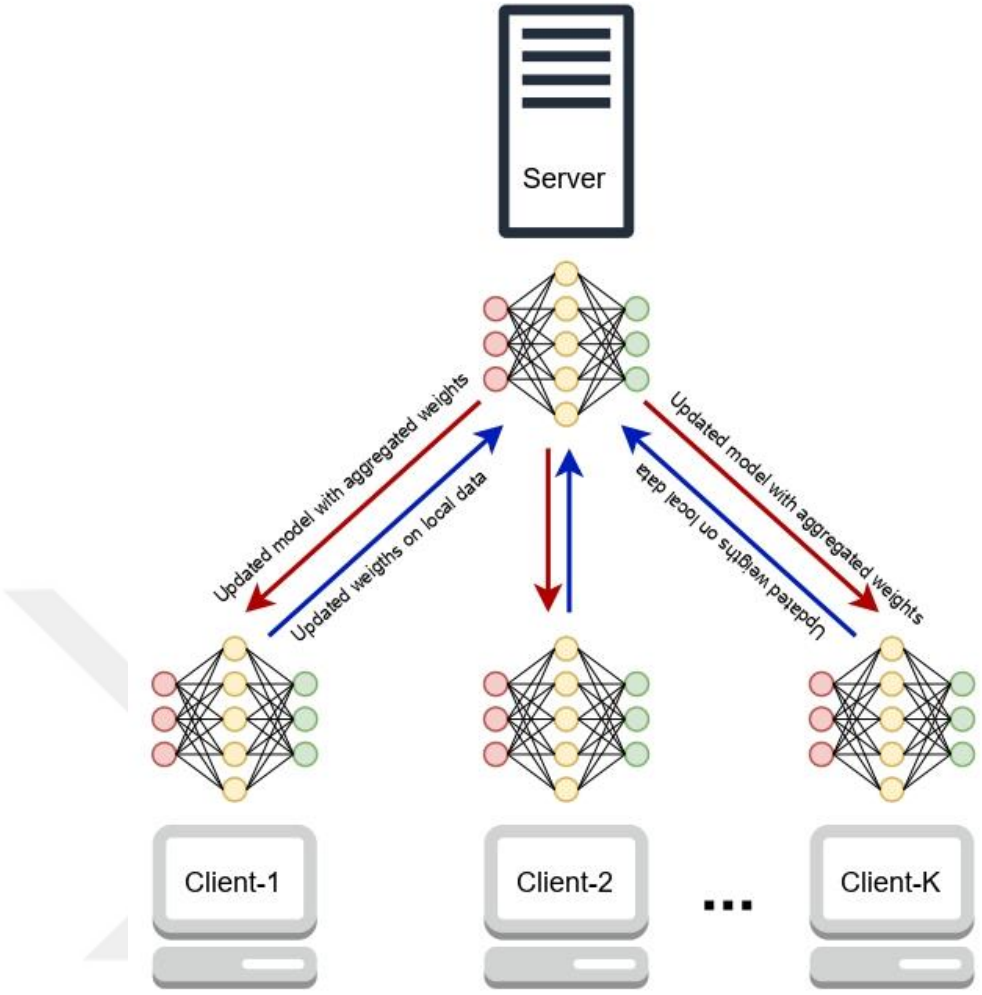

Figure 4.1 Federated Learning approach for defect classification of ultrasonic inspection images.

4.3.2 Model Architecture

In this research, we propose a collaborative framework for classifying defects in ultrasonic testing images through a FL methodology. The principal aim of this study is to illustrate that FL applied to a CNN model enables the utilization of shared private data while maintaining data privacy. Consequently, the selection of a specific CNN architecture is not our focal point, as various architectural configurations can influence performance metrics. For ease of implementation, we opted for a basic CNN architecture. The structure of the CNN model is detailed in Table 4.1. The architecture detailed in our model includes an initial convolution layer with 32 filters of size 3x3 followed by a max pooling layer, which helps in reducing the dimensionality of the feature maps while retaining essential features. This is followed by a second convolution layer with 64 filters of size 3x3 and another max pooling layer. After flattening the output from the

convolutional layers, the data passes through a dense layer of 128 units with ReLU activation to introduce non-linearity and aid in learning complex patterns in the data. The final output layer consists of a single unit with a sigmoid activation function, making it suitable for binary classification tasks. Our CNN takes as input an ultrasonic inspection image of size 224 × 224 pixels, outputting a probability value indicating the likelihood of each of the two classes. This design choice reflects our aim to classify ultrasonic images into two categories, which correspond to the presence or absence of defects.

Table 4.1 Layer configurations of the CNN model.

Layer Number	Layer Type	Configuration
L1	Input	Shape: (224, 224)
L2	Conv2D	32 filters, 3x3, activation='relu'
L3	MaxPooling2D	Pool size: (2, 2)
L4	Conv2D	64 filters, 3x3, activation='relu'
L5	MaxPooling2D	Pool size: (2, 2)
L6	Flatten	
L7	Dense	128 units, activation='relu'
L8	Dense	1 unit, activation='sigmoid'

4.4 Experiments

In this study, we simulate experiments involving 2 to 5 clients to investigate the effects of varying client numbers on the system's performance. Each client processes their entire local dataset in each round of the experiment. This setup allows us to focus on the impact of different client counts on system performance. By altering the number of clients, we aim to understand how scalability and performance are affected in a distributed learning environment.

4.4.1 Dataset preparation

In this study, we employed the UsimgAIST dataset [42], a publicly accessible collection of ultrasonic inspection images. The dataset comprises 7,004 images, each with a resolution of 224x224 pixels, sourced from 18 different stainless-steel materials, categorized into 3,615 defective and 3,389 non-defective samples. The specifics of the dataset composition are detailed in Table 4.2, while Figure 4.2 displays representative images from both categories.

Table 4.3 Distribution of Classes in the UsimgAIST Dataset

Class	Defective	Non-defective	Total
Train	2,892	2,711	5,603
Test	723	678	1,401
Total	3,615	3,389	7,004

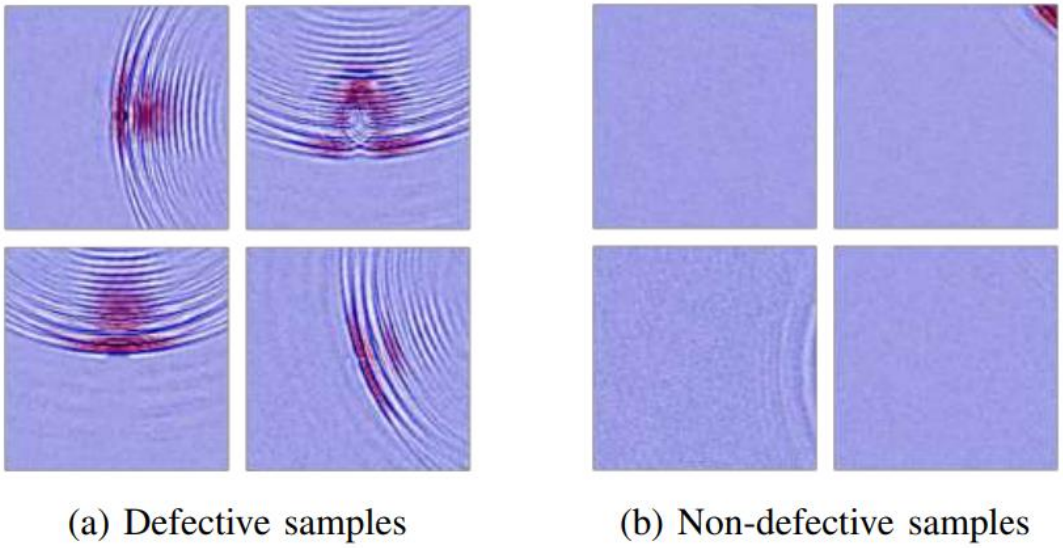

Figure 4.2 Comparative visualization of defective and non-defective samples in the UsimgAIST dataset.

4.4.2 Model settings and evaluation metric

In our research, we utilized the same CNN model for both federated and traditional learning approaches to analyze ultrasonic inspection images for defect classification. Our CNN architecture is implemented using TensorFlow [106] and TensorFlow Federated [107]. In both approaches, we normalize the ultrasonic images to a range of [0, 1] to maintain consistent input quality. The model training is conducted with a batch size of 32, chosen to balance computational efficiency and model stability.

In the FL approach, the model operates across different numbers of clients, ranging from 2 to 5, each managing a segment of the dataset to mimic real-world distributed data scenarios. The learning process involves utilizing the FedAvg method to update the model weights on a central server. The client-side utilizes stochastic gradient descent optimizer to minimize the loss function with a learning rate of 0.02. The training process consists of 50 rounds, and following each round, the central server model undergoes evaluation on the testset to maintain consistent progress, adaptation, and continual model improvement. In the traditional learning approach, the optimizer is Adam with a learning

rate of 0.0001 with a binary cross-entropy loss function. The maximum number of epochs is set to 50. After each epoch, the model is evaluated on the testset. We evaluate the model's performance using a comprehensive set of metrics, which include accuracy, precision, recall, and F1 score. These metrics are particularly chosen to provide a detailed measure of the model's classification performance. This structured evaluation facilitates a comprehensive comparison of the advantages between federated and traditional learning approaches, underscoring the potential of FL for ultrasonic inspection image analysis.

4.5 Results

Our study investigates the effectiveness of FL for defect classification in contexts where publicly available datasets are scarce. We aim to compare centralized FL with traditional machine learning approaches that directly utilize the entire dataset in a single model. Additionally, we explore how changing the number of clients in the federated network affects model performance. Performance metrics are evaluated after each training round, and the dataset in the FL setup is distributed to be IID across different clients. This approach highlights the potential of centralized and collaborative learning models in enhancing data privacy and utility in contexts with limited public data resources.

In our study, we adopted a balanced and IID setting for data distribution. Based on this IID setting, we distributed the training data among a varying number of clients, ranging from 2 to 5. This distribution ensures that each client receives an equal portion of the dataset: 50% for two clients, approximately 33.3% for three clients, 25% for four clients, and 20% for five clients. For each client configuration, we applied a CNN model with 50 rounds. The performance of these models was evaluated using accuracy scores on a separate, held-out test dataset. The results, illustrated in Figure 4.3, provide a comparison between FL and traditional learning approaches. Specifically, for the FL approach, the figure plots the model's accuracy on the test set against the number of rounds. For the traditional learning approach, the figure plots the model's accuracy on the test set against the number of training epochs. This comparative analysis aims to highlight the efficacy and efficiency of FL across different client configurations.

In Figure 4.3, the FL approach exhibits remarkable consistency after only two rounds. Despite an increase in the number of clients, the performance of the FL model remains comparably effective to that of traditional learning methods. Additionally, while

the accuracy of the traditional learning model fluctuates between 20 and 30 epochs, the FL model demonstrates stability across varying numbers of clients throughout all rounds. This stability indicates that FL efficiently maintains accuracy and ensures consistency irrespective of the number of clients involved.

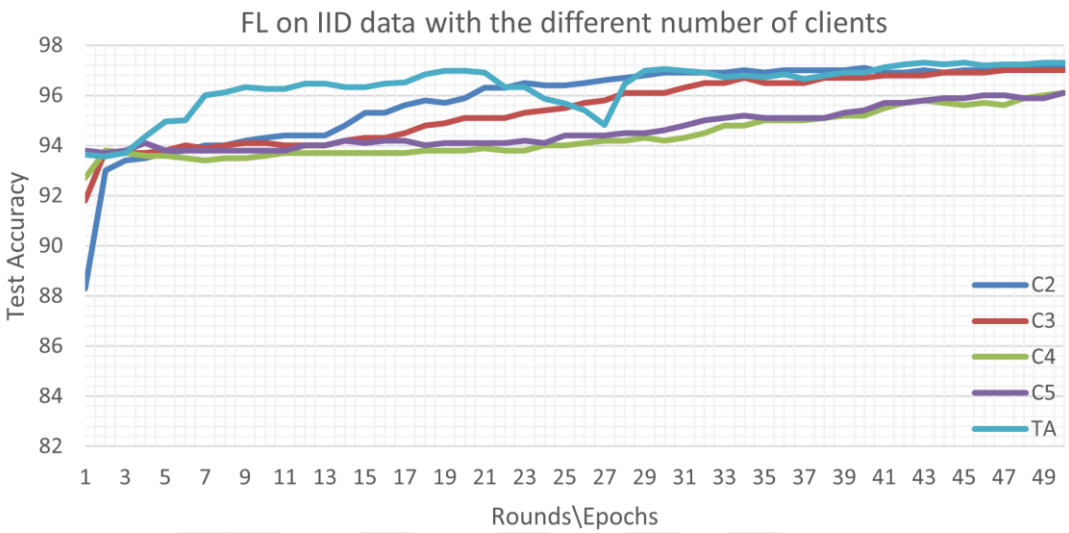

Figure 4.3 Comparison of test accuracy between FL and traditional methods using an ultrasonic inspection image dataset. The number of clients are denoted as C2, C3, C4, and C5, corresponding to 2, 3, 4, and 5 clients, respectively. TA represents the traditional approach. The horizontal axis indicates training progress: each point for traditional methods represents an epoch - a complete pass over the dataset. For FL, each point denotes a round in which every client trains simultaneously on their local data.

Another important result that can be underlined from Figure 4.3 that after 35 rounds, the accuracy for FL models with 2 and 3 clients stabilizes in subsequent rounds. Conversely, for clients numbered 4 and 5, there is a continuous increase in accuracy up to the last round. This pattern suggests that conducting additional rounds can enhance model performance, even with a higher number of clients. This underlines the pivotal role of the number of rounds in performance of FL. Furthermore, the continuous increase in accuracy for a higher number of clients underscores the scalability and adaptability of FL to accommodate varying degrees of client participation.

The results presented in Table 4.4 provide a comparative analysis of the performance metrics—including accuracy, precision, recall, and F1 score—of various methods tested in the final round for FL approaches and the final epoch for traditional methods. Notably, after 50 epochs, our CNN model demonstrates the highest accuracy among traditional methods, reaching up to 97.3%. This model also yields a matching recall rate of 97.3%. However, in the context of FL, when evaluating performance with

two and three clients, the accuracy metrics are 97.1% and 97%, respectively. These findings underscore the efficacy of FL-based methods, which offer performance comparable to that of traditional approaches, even with limited client participation. Moreover, the FL-based method excels in precision, achieving 97.8% with five clients. This result suggests that the collaborative nature of FL fosters more consistent and reliable outcomes across diverse client settings. Overall, this comparison illuminates the potential of FL to enhance both the performance and generalization capabilities of models in distributed environments.

Table 4.4 Accuracy, Precision, Recall and F1 scores of both FL and traditional learning approach.

Method	Accuracy(%)	Precision(%)	Recall(%)	F1(%)
FL (2 Clients)	97.1	97.7	96.4	97.0
FL (3 Clients)	97.0	97.3	96.5	96.9
FL (4 Clients)	96.1	96.8	95.2	96.0
FL (5 Clients)	96.1	97.8	94.1	95.9
TA	97.3	97.3	97.3	97.3

Chapter 5

Conclusions and Future Prospects

5.1 Conclusions

This thesis advances the fields of DL and ML for defect classification and characterization in composite materials, contributing to the development of modern inspection systems. It addresses the challenges associated with accurately identifying defects and characterizing damage mechanisms in composite materials, as well as the increasing demand for privacy-preserving collaborative learning approaches. Effective defect classification and damage characterization are vital to ensuring the structural integrity, performance, and safety of components across industries such as automotive and aerospace. By reducing inspection costs, improving reliability, and enhancing customer satisfaction, these advancements form the foundation of next-generation quality assurance processes.

This research has introduced key innovations in ultrasonic inspection image analysis, demonstrating that DL models, especially those employing transfer learning, can achieve remarkable accuracies of up to 98.8% and 98.6% on newly presented datasets. Such achievements underscore the potential of these models to support and augment human inspectors. Equally significant, the integration of ensemble supervised feature selection methods and unsupervised clustering has yielded deeper insights into damage mechanisms within CFRP composites. This approach not only refines the understanding of complex damage behaviors but also streamlines the identification of essential features for robust clustering performance, ultimately paving the way toward more informed and reliable characterization strategies.

In addition, the introduction of a FL framework provides a groundbreaking methodology for leveraging diverse, distributed, and privacy-sensitive datasets without centralized data aggregation. By coordinating locally trained models through a central server and sharing only model parameters, the FL approach attains performance levels

comparable to traditional methods. This not only safeguards proprietary data but also expands collaborative opportunities, enabling stakeholders to continually improve defect classification accuracy as more participants join the global network.

In conclusion, this thesis has made substantial contributions to the improvement of defect classification and characterization for composite materials. The utilization of advanced DL techniques, ensemble-based feature selection, unsupervised clustering, and FL methodologies exemplifies how converging models and approaches can yield a highly accurate, efficient, and adaptable defect inspection ecosystem. This synergy is critical for advancing inspection technologies, fostering safer products, and encouraging sustainable innovation.

5.2 Societal Impact and Contribution to Global Sustainability

In the pursuit of sustainable development, ensuring the reliability, integrity, and efficiency of materials and infrastructures is paramount. This thesis integrates advanced data-driven methodologies to enhance NDT and facilitate better defect classification, damage characterization, and collaborative learning. By refining the ability to detect and understand material defects, these efforts contribute to more durable and resilient industrial components, aligning closely with Sustainable Development Goals (SDG) 9 and 11. Improved defect characterization not only fosters the development of robust composite materials and more sustainable manufacturing processes, but also curbs material waste and prolongs the operational life of critical components. In doing so, it mitigates the environmental footprint of industrial activities, aligning closely with SDG 13’s focus on climate action.

The introduction of transfer learning methods to ultrasonic inspection illustrates how adapting pre-trained models to new datasets allows for efficient and accurate defect classification in composite materials. Such advancements not only improve product quality and safety but also foster more sustainable manufacturing processes. By reducing the need for extensive trial-and-error testing and material wastage, these innovations pave the way for more responsible consumption and production patterns. The adaptability of these models ensures that industrial components—from aerospace parts to structural

elements—can be monitored reliably, thus contributing to both SDG 9 (Industry, Innovation, and Infrastructure) and SDG 12 (Responsible Consumption and Production).

Similarly, the integration of ensemble feature selection techniques and unsupervised clustering models for characterizing damage modes in CFRP composites aligns with the global push toward more efficient and sustainable use of materials. By leveraging AE data, these methods refine our understanding of damage mechanisms, enabling proactive maintenance and reducing the risk of catastrophic failures. This approach not only enhances the safety and longevity of high-performance composite structures but also reduces unnecessary replacements and material discard, contributing to sustainable resource management (SDG 12).

The FL framework elevates this pursuit of sustainability to a new dimension by ensuring data privacy, promoting collaborative research, and overcoming dataset scarcity. FL enables decentralized training on private data, thus encouraging partnerships among various organizations and institutions. These collaborations exemplify SDG 17 (Partnerships for the Goals), showcasing how knowledge exchange across boundaries can enhance the collective ability to develop advanced detection methods. By ensuring robust performance even across distributed datasets and diverse operational settings, FL supports scalable solutions that respect data security (aligned with SDG 16's focus on peace, justice, and strong institutions) while reinforcing trust in emerging technologies.

Overall, these collective studies align with multiple Sustainable Development Goals, including SDG 9 (Industry, Innovation, and Infrastructure), SDG 11 (Sustainable Cities and Communities), SDG 13 (Climate Action), SDG 12 (Responsible Consumption and Production), SDG 17 (Partnerships for the Goals), and SDG 16 (Peace, Justice, and Strong Institutions). This thesis contributes to a sustainable future by ensuring that industrial components are intelligently monitored, infrastructures are more resilient, data-driven methods are responsibly deployed, and global partnerships flourish.

5.3 Future Prospects

For each of the three studies carried out as part of this thesis work, we can briefly describe our prospects for the future as follows:

- **Future prospects of study 1:** The defect classification models developed using ultrasonic inspection data could be further improved by significantly expanding and diversifying the dataset. Larger, more heterogeneous datasets—encompassing

various composite materials, defect types, and real-world inspection conditions—can enhance model reliability and enable more accurate defect detection in complex industrial environments. Additionally, incorporating Generative Adversarial Networks for data augmentation will help create more realistic synthetic inspection images, reducing the need for extensive data collection efforts. In the future, these techniques may support the classification of an even wider range of defect types, ultimately improving the robustness and accuracy of inspection systems.

- **Future prospects of study 2:** To gain a more comprehensive understanding of material degradation, future efforts could focus on analyzing hybrid composite structures through integrated AE-based damage characterization and ultrasonic inspection data. By combining both supervised and unsupervised feature selection models, it will be possible to identify the most informative features more effectively, improving both the reliability and interpretability of the results. Such multimodal, feature-optimized approaches can facilitate proactive maintenance decisions, minimize downtime, and offer deeper insights into the underlying damage mechanisms.
- **Future prospects of study 3:** Building upon the FL framework for defect classification, future research could explore the use of more sophisticated convolutional neural network architectures. Additionally, expanding the FL framework to incorporate multiple NDT methods and larger, more distributed datasets can help realize its full potential. Ultimately, this could foster global collaboration, streamline data sharing practices, and further improve the accuracy and efficiency of automated defect classification in various industrial applications.

BIBLIOGRAPHY

[1] S. Gholizadeh, "A review of non-destructive testing methods of composite materials," in *Procedia Structural Integrity*, 2016. doi: 10.1016/j.prostr.2016.02.008.

[2] A. Kelly and C. Zweben, "Comprehensive composite materials," *Materials Today*, vol. 2, no. 1, 1999, doi: 10.1016/s1369-7021(99)80033-9.

[3] D. Palumbo, P. Cavallo, and U. Galietti, "An investigation of the stepped thermography technique for defects evaluation in GFRP materials," *NDT & E International*, vol. 102, pp. 254–263, Mar. 2019, doi: 10.1016/J.NDTEINT.2018.12.011.

[4] "Composites & Sustainability | Sustainable Plastics." Accessed: Dec. 03, 2024. [Online]. Available: https://www.sustainableplastics.com/news/composites-sustainability

[5] H. T. Bang, S. Park, and H. Jeon, "Defect identification in composite materials via thermography and deep learning techniques," *Compos Struct*, vol. 246, p. 112405, Aug. 2020, doi: 10.1016/J.COMPSTRUCT.2020.112405.

[6] U. Caliskan, F. Yildiz, S. Teke, and A. T. Ozdemir, "Impact-Delamination Detection in Repaired-Composite Laminates Using Numerical and Ultrasonic Method," *J Nondestr Eval*, vol. 41, no. 2, 2022, doi: 10.1007/s10921-022-00878-x.

[7] R. S. M. Almeida, M. D. Magalhães, M. N. Karim, K. Tushtev, and K. Rezwan, "Identifying damage mechanisms of composites by acoustic emission and supervised machine learning," *Mater Des*, vol. 227, 2023, doi: 10.1016/j.matdes.2023.111745.

[8] Z. Quiney, E. Weston, P. Ian Nicholson, S. Pattison, and M. R. Bache, "Volumetric assessment of fatigue damage in a SiCf/SiC ceramic matrix composite via in situ X-ray computed tomography," *J Eur Ceram Soc*, vol. 40, no. 11, 2020, doi: 10.1016/j.jeurceramsoc.2020.04.037.

[9] M. Gupta, M. A. Khan, R. Butola, and R. M. Singari, "Advances in applications of Non-Destructive Testing (NDT): A review," *Advances in Materials and Processing Technologies*, vol. 8, no. 2, 2022, doi: 10.1080/2374068X.2021.1909332.

[10] J. Jodhani, A. Handa, A. Gautam, Ashwni, and R. Rana, "Ultrasonic non-destructive evaluation of composites: A review," in *Materials Today: Proceedings*, 2023. doi: 10.1016/j.matpr.2022.12.055.

[11] B. Wang, S. Zhong, T. L. Lee, K. S. Fancey, and J. Mi, "Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review," 2020. doi: 10.1177/1687814020913761.

[12] B. Wang, P. He, Y. Kang, J. Jia, X. Liu, and N. Li, "Ultrasonic Testing of Carbon Fiber-Reinforced Polymer Composites," *J Sens*, vol. 2022, no. 1, p. 5462237, Jan. 2022, doi: 10.1155/2022/5462237.

[13] M. Roth, E. Mojaev, E. Dul'kin, P. Gemeiner, and B. Dkhil, "Phase transition at a nanometer scale detected by acoustic emission within the cubic phase Pb(Zn1/3Nb2/3)O3-xPbTiO3 relaxor ferroelectrics," *Phys Rev Lett*, vol. 98, no. 26, 2007, doi: 10.1103/PhysRevLett.98.265701.

[14] C. Muir *et al.*, "Damage mechanism identification in composites via machine learning and acoustic emission," 2021. doi: 10.1038/s41524-021-00565-x.

[15] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/S12525-021-00475-2/TABLES/2.
[16] Y. LeCun, Y. Bengio, G. H.- nature, and undefined 2015, “Deep learning,” *nature.comY LeCun, Y Bengio, G Hintonnature, 2015•nature.com*, 2015, doi: 10.1038/nature14539.
[17] A. A. Carvalho, L. V. S. Sagrilo, I. C. Silva, and J. M. A. Rebello, “The PoD curve for the detection of planar defects using a multi-channel ultrasonic system,” *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 44, no. 11, 2002.
[18] J. A. Ogilvy, “Model for predicting ultrasonic pulse-echo probability of detection,” *NDT & E International*, vol. 26, no. 1, pp. 19–29, Feb. 1993, doi: 10.1016/0963-8695(93)90160-V.
[19] J. L. B. C. Veiga, A. A. De Carvalho, I. C. Da Silva, and J. M. A. Rebello, “The use of artificial neural network in the classification of pulse-echo and TOFD ultrasonic signals,” *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 27, no. 4, 2005, doi: 10.1590/S1678-58782005000400007.
[20] G. LeCun, Y., Bengio, Y., Hinton, “Deep learning. nature 521 (7553): 436,” *Nature*, vol. 521, 2015.
[21] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. 2012. doi: 10.1002/9781118443118.
[22] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *Pattern Recognit*, vol. 77, 2018, doi: 10.1016/j.patcog.2017.10.013.
[23] S. J. Pan, “Transfer learning,” *Learning*, vol. 21, pp. 1–2, 2020.
[24] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, “A survey of transfer learning,” *J Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/S40537-016-0043-6.
[25] J. Zhao *et al.*, “Deep transfer learning approach for localization of damage area in composite laminates using acoustic emission signal,” *mdpi.com*, Accessed: Dec. 03, 2024. [Online]. Available: https://www.mdpi.com/2073-4360/15/6/1520
[26] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, and J. Rinta-aho, “Augmented Ultrasonic Data for Machine Learning,” *J Nondestr Eval*, vol. 40, no. 1, 2021, doi: 10.1007/s10921-020-00739-5.
[27] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*, 2017.
[28] N. Godin, P. Reynaud, and G. Fantozzi, “Challenges and limitations in the identification of acoustic emission signature of damage mechanisms in composites materials,” *Applied Sciences (Switzerland)*, vol. 8, no. 8, 2018, doi: 10.3390/app8081267.
[29] J. P. Forna-Kreutzer, J. Ell, H. Barnard, T. J. Pirzada, R. O. Ritchie, and D. Liu, “Full-field characterisation of oxide-oxide ceramic-matrix composites using X-ray computed micro-tomography and digital volume correlation under load at high temperatures,” *Mater Des*, vol. 208, 2021, doi: 10.1016/j.matdes.2021.109899.
[30] W. Zhou, R. Qin, K. ning Han, Z. yuan Wei, and L. H. Ma, “Progressive damage visualization and tensile failure analysis of three-dimensional braided composites by acoustic emission and micro-CT,” *Polym Test*, vol. 93, 2021, doi: 10.1016/j.polymertesting.2020.106881.
[31] E. Özaslan, A. Yetgin, B. Acar, and M. A. Güler, “Damage mode identification of open hole composite laminates based on acoustic emission and digital image

correlation methods," *Compos Struct*, vol. 274, 2021, doi: 10.1016/j.compstruct.2021.114299.
[32] M. G. R. Sause and S. Horn, "Simulation of acoustic emission in planar carbon fiber reinforced plastic specimens," *J Nondestr Eval*, vol. 29, no. 2, 2010, doi: 10.1007/s10921-010-0071-7.
[33] A. Gulsen, B. Kolukisa, A. T. Ozdemir, B. Bakir-Gungor, and V. C. Gungor, "Defect classification of composite materials using transfer learning methods," *Nondestructive Testing and Evaluation*, pp. 1–17, 2024.
[34] A. Gulsen, B. Kolukisa, U. Caliskan, B. Bakir-Gungor, and V. C. Gungor, "Ensemble Feature Selection for Clustering Damage Modes in Carbon Fiber-Reinforced Polymer Sandwich Composites Using Acoustic Emission," *Adv Eng Mater*, vol. 26, no. 22, p. 2400317, 2024.
[35] A. Gulsen, H. Hacilar, B. Kolukisa, and B. Bakir-Gungor, "A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing," *2024 9th International Conference on Computer Science and Engineering (UBMK)*, pp. 360–364, Oct. 2024, doi: 10.1109/UBMK63289.2024.10773541.
[36] J. K. Park, B. K. Kwon, J. H. Park, and D. J. Kang, "Machine learning-based imaging system for surface defect inspection," *International Journal of Precision Engineering and Manufacturing - Green Technology*, vol. 3, no. 3, pp. 303–310, Jul. 2016, doi: 10.1007/S40684-016-0039-X/METRICS.
[37] Y. Yu, H. Cao, X. Yan, T. Wang, and S. S. Ge, "Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor," *Neurocomputing*, vol. 376, pp. 1–9, Feb. 2020, doi: 10.1016/J.NEUCOM.2019.09.071.
[38] Y. He *et al.*, "Intelligent Detection Algorithm Based on 2D/3D-UNet for Internal Defects of Carbon Fiber Composites," *Nondestructive Testing and Evaluation*, vol. 39, no. 4, 2024, doi: 10.1080/10589759.2023.2234548.
[39] Q. Luo, B. Gao, W. L. Woo, and Y. Yang, "Temporal and spatial deep learning network for infrared thermal defect detection," *NDT & E International*, vol. 108, p. 102164, Dec. 2019, doi: 10.1016/J.NDTEINT.2019.102164.
[40] L. Ruan, B. Gao, S. Wu, and W. L. Woo, "DeftectNet: Joint loss structured deep adversarial network for thermography defect detecting system," *Neurocomputing*, vol. 417, pp. 441–457, Dec. 2020, doi: 10.1016/J.NEUCOM.2020.07.093.
[41] M. W. Akram *et al.*, "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," *Energy*, vol. 189, p. 116319, Dec. 2019, doi: 10.1016/J.ENERGY.2019.116319.
[42] J. Ye and N. Toyama, "Benchmarking Deep Learning Models for Automatic Ultrasonic Imaging Inspection," *IEEE Access*, vol. 9, pp. 36986–36994, 2021, doi: 10.1109/ACCESS.2021.3062860.
[43] J. Ye, S. Ito, and N. Toyama, "Computerized Ultrasonic Imaging Inspection: From Shallow to Deep Learning," *Sensors 2018, Vol. 18, Page 3820*, vol. 18, no. 11, p. 3820, Nov. 2018, doi: 10.3390/S18113820.
[44] B. Filipovic, F. Milkovic, M. Subasic, S. Loncaric, T. Petkovic, and M. Budimir, "Automated ultrasonic testing of materials based on C-scan flaw classification," *International Symposium on Image and Signal Processing and Analysis, ISPA*, vol. 2021-September, pp. 230–234, Sep. 2021, doi: 10.1109/ISPA52656.2021.9552056.
[45] J. Krautkrämer and H. Krautkrämer, "Ultrasonic Testing of Materials," *Ultrasonic Testing of Materials*, 1983, doi: 10.1007/978-3-662-02357-0.

[46] C. Fei, Z. Han, and J. Dong, “An ultrasonic flaw-classification system with wavelet-packet decomposition, a mutative scale chaotic genetic algorithm, and a support vector machine and its application to petroleum-transporting pipelines,” *Russian Journal of Nondestructive Testing*, vol. 42, no. 3, 2006, doi: 10.1134/S1061830906030077.

[47] F. C. Cruz, E. F. Simas Filho, M. C. S. Albuquerque, I. C. Silva, C. T. T. Farias, and L. L. Gouvêa, “Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing,” *Ultrasonics*, vol. 73, 2017, doi: 10.1016/j.ultras.2016.08.017.

[48] Y. Chen, H. W. Ma, and G. M. Zhang, “A support vector machine approach for classification of welding defects from ultrasonic signals,” *Nondestructive Testing and Evaluation*, vol. 29, no. 3, 2014, doi: 10.1080/10589759.2014.914210.

[49] H. Hu, G. Peng, X. Wang, and Z. Zhou, “Weld defect classification using 1-D LBP feature extraction of ultrasonic signals,” *Nondestructive Testing and Evaluation*, vol. 33, no. 1, 2018, doi: 10.1080/10589759.2017.1299732.

[50] K. Virupakshappa and E. Oruklu, “Ultrasonic flaw detection using Support Vector Machine classification,” in *2015 IEEE International Ultrasonics Symposium, IUS 2015*, 2015. doi: 10.1109/ULTSYM.2015.0128.

[51] W. Yi, “The defect detection and non-destructive evaluation in weld zone of austenitic stainless steel 304 using neural network-ultrasonic wave,” *KSME International Journal*, vol. 12, no. 6, pp. 1150–1161, 1998, doi: 10.1007/BF02942589/METRICS.

[52] A. Masnata and M. Sunseri, “Neural network classification of flaws detected by ultrasonic means,” *NDT and E International*, vol. 29, no. 2, 1996, doi: 10.1016/0963-8695(95)00053-4.

[53] N. Munir, H. J. Kim, J. Park, S. J. Song, and S. S. Kang, “Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions,” *Ultrasonics*, vol. 94, 2019, doi: 10.1016/j.ultras.2018.12.001.

[54] L. Huang, X. Hong, Z. Yang, Y. Liu, and B. Zhang, “CNN-LSTM network-based damage detection approach for copper pipeline using laser ultrasonic scanning,” *Ultrasonics*, vol. 121, 2022, doi: 10.1016/j.ultras.2022.106685.

[55] F. Zhang, L. Wang, W. Ye, Y. Li, and F. Yang, “Ultrasonic lamination defects detection of carbon fiber composite plates based on multilevel LSTM,” *Compos Struct*, vol. 327, 2024, doi: 10.1016/j.compstruct.2023.117714.

[56] Y. Guo *et al.*, “Fully Convolutional Neural Network with GRU for 3D Braided Composite Material Flaw Detection,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2946447.

[57] N. Munir, H. J. Kim, S. J. Song, and S. S. Kang, “Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments,” *Journal of Mechanical Science and Technology*, vol. 32, no. 7, 2018, doi: 10.1007/s12206-018-0610-1.

[58] M. Meng, Y. J. Chua, E. Wouterson, and C. P. K. Ong, “Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks,” *Neurocomputing*, vol. 257, 2017, doi: 10.1016/j.neucom.2016.11.066.

[59] H. Sun, P. Ramuhalli, and R. E. Jacob, “Machine learning for ultrasonic nondestructive examination of welding defects: A systematic review,” 2023. doi: 10.1016/j.ultras.2022.106854.

[60] J. B. Harley and D. Sparkman, “Machine learning and NDE: Past, present, and future,” in *AIP Conference Proceedings*, 2019. doi: 10.1063/1.5099819.

[61] T. Latête, B. Gauthier, and P. Belanger, “Towards using convolutional neural network to locate, identify and size defects in phased array ultrasonic testing,” *Ultrasonics*, vol. 115, 2021, doi: 10.1016/j.ultras.2021.106436.
[62] G. Marcus, “Deep Learning: A Critical Appraisal,” *arXiv preprint arXiv:1801.00631*, 2018.
[63] L. Posilović, D. Medak, M. Subašić, M. Budimir, and S. Lončarić, “Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans,” *Neurocomputing*, vol. 459, 2021, doi: 10.1016/j.neucom.2021.06.094.
[64] A. Boikov, V. Payor, R. Savelev, and A. Kolesnikov, “Synthetic data generation for steel defect detection and classification using deep learning,” *Symmetry (Basel)*, vol. 13, no. 7, 2021, doi: 10.3390/sym13071176.
[65] Ultrasonar Defense and A. T. Inc., “US DSA Deep Structure Analyzer: Automated Immersion Type Ultrasonic Scanning System,” 2020.
[66] Francois. Chollet, “Deep learning with Python Francois Chollet,” 2022, Accessed: Dec. 04, 2024. [Online]. Available: https://books.google.com/books/about/Deep_Learning_with_Python_Second_Edition.html?id=XHpKEAAAQBAJ
[67] G. Van Rossum and F. L. Drake, “Python 3 Reference Manual CreateSpace Independent Publishing Platform,” 2009.
[68] J. H. Hwang, C. S. Lee, and W. Hwang, “Effect of crack propagation directions on the interlaminar fracture toughness of carbon/epoxy composite materials,” *Applied Composite Materials*, vol. 8, no. 6, 2001, doi: 10.1023/A:1012663722334.
[69] S. M. Lee, “A Comparison of Fracture Toughness of Matrix Controlled Failure Modes: Delamination and Transverse Cracking,” *J Compos Mater*, vol. 20, no. 2, 1986, doi: 10.1177/002199838602000205.
[70] M. Saeedifar and D. Zarouchas, “Damage characterization of laminated composites using acoustic emission: A review,” 2020. doi: 10.1016/j.compositesb.2020.108039.
[71] C. Barile, C. Casavola, G. Pappalettera, and V. P. Kannan, “Application of different acoustic emission descriptors in damage assessment of fiber reinforced plastics: A comprehensive review,” 2020. doi: 10.1016/j.engfracmech.2020.107083.
[72] W. Harizi, S. Chaki, G. Bourse, and M. Ourak, “Damage mechanisms assessment of Glass Fiber-Reinforced Polymer (GFRP) composites using multivariable analysis methods applied to acoustic emission data,” *Compos Struct*, vol. 289, 2022, doi: 10.1016/j.compstruct.2022.115470.
[73] C. U. Grosse and M. Ohtsu, *Acoustic emission testing: Basics for Research-Applications in Civil Engineering*. 2008. doi: 10.1007/978-3-540-69972-9.
[74] F. Pashmforoush, M. Fotouhi, and M. Ahmadi, “Acoustic emission-based damage classification of glass/polyester composites using harmony search k-means algorithm,” *Journal of Reinforced Plastics and Composites*, vol. 31, no. 10, 2012, doi: 10.1177/0731684412442257.
[75] S. Gholizadeh, Z. Leman, and B. T. H. T. Baharudin, “State-of-the-art ensemble learning and unsupervised learning in fatigue crack recognition of glass fiber reinforced polyester composite (GFRP) using acoustic emission,” *Ultrasonics*, vol. 132, 2023, doi: 10.1016/j.ultras.2023.106998.
[76] A. Refahi Oskouei, H. Heidary, M. Ahmadi, and M. Farajpur, “Unsupervised acoustic emission data clustering for the analysis of damage mechanisms in

glass/polyester composites," *Mater Des*, vol. 37, 2012, doi: 10.1016/j.matdes.2012.01.018.

[77] L. Li, Y. Swolfs, I. Straumit, X. Yan, and S. V. Lomov, "Cluster analysis of acoustic emission signals for 2D and 3D woven carbon fiber/epoxy composites," *J Compos Mater*, vol. 50, no. 14, 2016, doi: 10.1177/0021998315597742.

[78] F. E. Oz, N. Ersoy, M. Mehdikhani, and S. V. Lomov, "Multi-instrument in-situ damage monitoring in quasi-isotropic CFRP laminates under tension," *Compos Struct*, vol. 196, 2018, doi: 10.1016/j.compstruct.2018.05.006.

[79] H. A. Sawan, M. E. Walter, and B. Marquette, "Unsupervised learning for classification of acoustic emission events from tensile and bending experiments with open-hole carbon fiber composite samples," *Compos Sci Technol*, vol. 107, 2015, doi: 10.1016/j.compscitech.2014.12.003.

[80] M. Saeedifar, M. A. Najafabadi, D. Zarouchas, H. H. Toudeshky, and M. Jalalvand, "Clustering of interlaminar and intralaminar damages in laminated composites under indentation loading using Acoustic Emission," *Compos B Eng*, vol. 144, 2018, doi: 10.1016/j.compositesb.2018.02.028.

[81] L. Li, S. V. Lomov, X. Yan, and V. Carvelli, "Cluster analysis of acoustic emission signals for 2D and 3D woven glass/epoxy composites," *Compos Struct*, vol. 116, no. 1, 2014, doi: 10.1016/j.compstruct.2014.05.023.

[82] C. Barile, C. Casavola, G. Pappalettera, and V. Paramsamy Kannan, "Laplacian score and K-means data clustering for damage characterization of adhesively bonded CFRP composites by means of acoustic emission technique," *Applied Acoustics*, vol. 185, 2022, doi: 10.1016/j.apacoust.2021.108425.

[83] D. D. Doan, E. Ramasso, V. Placet, S. Zhang, L. Boubakar, and N. Zerhouni, "An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer-composite materials," *Mech Syst Signal Process*, vol. 64–65, 2015, doi: 10.1016/j.ymssp.2015.04.011.

[84] J. Tang, S. Soua, C. Mares, and T. H. Gan, "A pattern recognition approach to acoustic emission data originating from fatigue of wind turbine blades," *Sensors (Switzerland)*, vol. 17, no. 11, 2017, doi: 10.3390/s17112507.

[85] M. Assarar, M. Bentahar, A. El Mahi, and R. El Guerjouma, "Monitoring of damage mechanisms in sandwich composite materials using acoustic emission," *International Journal of Damage Mechanics*, vol. 24, no. 6, 2015, doi: 10.1177/1056789514553134.

[86] H. Heidary, N. Z. Karimi, M. Ahmadi, A. Rahimi, and A. Zucchelli, "Clustering of acoustic emission signals collected during drilling process of composite materials using unsupervised classifiers," *J Compos Mater*, vol. 49, no. 5, 2015, doi: 10.1177/0021998314521258.

[87] J. Yousefi, M. A. Najfabadi, H. H. Toudeshky, and M. Akhlaghi, "Damage evaluation of laminated composite material using a new acoustic emission Lamb-based and finite element techniques," *Applied Composite Materials*, vol. 25, no. 5, 2018, doi: 10.1007/s10443-017-9649-x.

[88] M. Fotouhi, P. Suwarta, M. Jalalvand, G. Czel, and M. R. Wisnom, "Detection of fibre fracture and ply fragmentation in thin-ply UD carbon/glass hybrid laminates using acoustic emission," *Compos Part A Appl Sci Manuf*, vol. 86, 2016, doi: 10.1016/j.compositesa.2016.04.003.

[89] R. Gutkin, C. J. Green, S. Vangrattanachai, S. T. Pinho, P. Robinson, and P. T. Curtis, "On acoustic emission for failure investigation in CFRP: Pattern recognition and peak frequency analyses," *Mech Syst Signal Process*, vol. 25, no. 4, 2011, doi: 10.1016/j.ymssp.2010.11.014.

[90] S. F. Karimian and M. Modarres, “Acoustic emission signal clustering in CFRP laminates using a new feature set based on waveform analysis and information entropy analysis,” *Compos Struct*, vol. 268, 2021, doi: 10.1016/j.compstruct.2021.113987.
[91] W. Zhou, W. zheng Zhao, Y. nan Zhang, and Z. jun Ding, “Cluster analysis of acoustic emission signals and deformation measurement for delaminated glass fiber epoxy composites,” *Compos Struct*, vol. 195, 2018, doi: 10.1016/j.compstruct.2018.04.081.
[92] M. Haggui, A. El Mahi, Z. Jendli, A. Akrout, and M. Haddar, “Static and fatigue characterization of flax fiber reinforced thermoplastic composites by acoustic emission,” *Applied Acoustics*, vol. 147, 2019, doi: 10.1016/j.apacoust.2018.03.011.
[93] A. International, *Standard test method for edgewise compressive strength of sandwich constructions*. ASTM International, 2016.
[94] M. A. Hamstad, “Thirty years of advances and some remaining challenges in the application of acoustic emission to composite materials,” in *Acoustic Emission - Beyond the Millennium*, 2000. doi: 10.1016/b978-008043851-1/50005-4.
[95] J. Li *et al.*, “Feature selection: A data perspective,” 2017. doi: 10.1145/3136625.
[96] P. F. Liu, J. K. Chu, Y. L. Liu, and J. Y. Zheng, “A study on the failure mechanisms of carbon fiber/epoxy composite laminates using acoustic emission,” *Mater Des*, vol. 37, 2012, doi: 10.1016/j.matdes.2011.12.015.
[97] R. Boominathan, V. Arumugam, C. Santulli, A. Adhithya Plato Sidharth, R. Anand Sankar, and B. T. N. Sridhar, “Acoustic emission characterization of the temperature effect on falling weight impact damage in carbon/epoxy laminates,” *Compos B Eng*, vol. 56, 2014, doi: 10.1016/j.compositesb.2013.09.002.
[98] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognit Lett*, vol. 31, no. 8, 2010, doi: 10.1016/j.patrec.2009.09.011.
[99] F. Murtagh and P. Contreras, “Algorithms for hierarchical clustering: An overview,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 2, no. 1, 2012, doi: 10.1002/widm.53.
[100] N. Bouguila and W. Fan, *Mixture models and applications*. Springer, 2020.
[101] K. Konhäuser and T. Werner, “Uncovering the financial impact of energy-efficient building characteristics with eXplainable artificial intelligence,” *Appl Energy*, vol. 374, p. 123960, 2024.
[102] A. G. Roselin, P. Nanda, S. Nepal, and X. He, “Intelligent Anomaly Detection for Large Network Traffic with Optimized Deep Clustering (ODC) Algorithm,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3068172.
[103] Z. Wang, Y. Zhou, and G. Li, “Anomaly Detection by Using Streaming K-Means and Batch K-Means,” in *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, 2020. doi: 10.1109/ICBDA49040.2020.9101212.
[104] L. Posilović, D. Medak, F. Milković, M. Subašić, M. Budimir, and S. Lončarić, “Deep learning-based anomaly detection from ultrasonic images,” *Ultrasonics*, vol. 124, 2022, doi: 10.1016/j.ultras.2022.106737.
[105] S. Uhlig, I. Alkhasli, F. Schubert, C. Tschöpe, and M. Wolff, “A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation,” 2023. doi: 10.1016/j.ultras.2023.107041.
[106] G. TensorFlow, “Large-scale machine learning on heterogeneous systems,” *Google Research*, vol. 10, p. s15326985ep4001, 2015.
[107] “TensorFlow.” Accessed: Dec. 06, 2024. [Online]. Available: https://www.tensorflow.org/

CURRICULUM VITAE

2013 – 2018	B.Sc., Electrical-Electronics Engineering, Computer Engineering (Double Major), Erciyes University, Kayseri, TÜRKİYE
2018 – 2020	M.Sc., Electrical-Electronics Engineering, Erciyes University, Kayseri, TÜRKİYE
2020 – 2024	Ph.D., Electrical and Computer Engineering, Abdullah Gül University, Kayseri, TÜRKİYE
2020 – Present	Research Assistant, Computer Engineering, Abdullah Gül University, Kayseri, TÜRKİYE

SELECTED PUBLICATIONS AND PRESENTATIONS

J1) A. Gulsen, B. Kolukisa, A. T. Ozdemir, B. Bakir-Gungor, V. C. Gungor, Defect classification of composite materials using transfer learning methods, published in Taylor & Francis Journal of Nondestructive Testing and Evaluation (2024).

J2) A. Gulsen, B. Kolukisa, U. Caliskan, B. Bakir-Gungor, V. C. Gungor, Ensemble feature selection for clustering damage modes in carbon fiber-reinforced polymer sandwich composites using acoustic emission, published in Wiley Journal of Advanced Engineering Materials (2024).

C1) A. Gulsen, B. Kolukisa, H. Hacilar B. Bakir-Gungor, A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing. In 9th IEEE International Conference on Computer Science and Engineering (2024).

C2) B. Kolukisa, B. K. Dedeturk, B. A. Dedeturk, A. Gulsen, G. Bakal, A Comparative Analysis on Medical Article Classification Using Text Mining & Machine Learning Algorithms. In 6th IEEE International Conference on Computer Science and Engineering (2021).