The Fourteenth International Conference on Smart Cities, Systems, Devices and Technologies, SMART 2025.

Deep Learning-based Failure Detection for Safety Diagnostics of Hydrogen Storage Vessels

Dong-Ju Kim

Institute of Artificial Intelligence, POSTECH, South Korea

Contact email : kkb0320@postech.ac.kr

POSTECH

IARIA

Contents

1 Introduction

2 Related Works

3 Background

4 Failure Detection Methodology

5 Experiment Result

6 Conclusion

Why Hydrogen ?

- Hydrogen
  - Clean energy source, zero emissions of greenhouse gas
  - Key to climate change & energy security
  - Gas state at room temperature → **bulky, flammable**
  - Therefore, requires advanced storage and transportation technologies

- Hydrogen storage vessels
  - Classify into types 1 to 4 according to their materials

  - Type 1 : entirely metals such as aluminum or steel
  - Type 2 : metal (inner layer) + Glass Fiber Reinforced Plastic (outer layer)
  - Type 3 : metal (inner layer) + Carbon Fiber Reinforced Plastic (outer layer)
  - Type 4 : resin liner (inner layer) + CFRP(outer layer)


(a) Type 1 (b) Type 2 (c) Type 3 (d) Type 4

[Types of hydrogen storage vessels]

Type 1 Safety Challenges

- Metal vessels (Type 1)
  - Be exposed to fatigue, corrosion, cracking
  - Increase risk of leakage or explosion
  - So, needs the periodic inspection and failure detection
- Traditional inspection method
  - Often require disassembly of the vessel, but it is not feasible during operation
  - So, **Non-Destructive Testing (NDT) technologies**, such as ultrasonic, radiographic and Acoustic Emission Testing (AET) **are essential** → provide real-time in-service safety diagnostics.
- AE-Based failure detection method
  - **Effective method** because AET analyses acoustic signals generated during failure
  - Previous works : mostly focused on failure detection of Type 2 and Type 3 vessels.

- So, this study proposes a deep learning-based failure detection method of Type 1 vessels.

FEM-Based vs. AE-Based Research

- Research on hydrogen storage failure detection
  - Divide into Finite Element Model (FEM)-based approaches and AE signal analysis methods.

- FEM simulates stress & fatigue under operational condition → but limited for real-time diagnostics
- **AET enables real-time failure detection** → No need for complex numerical models

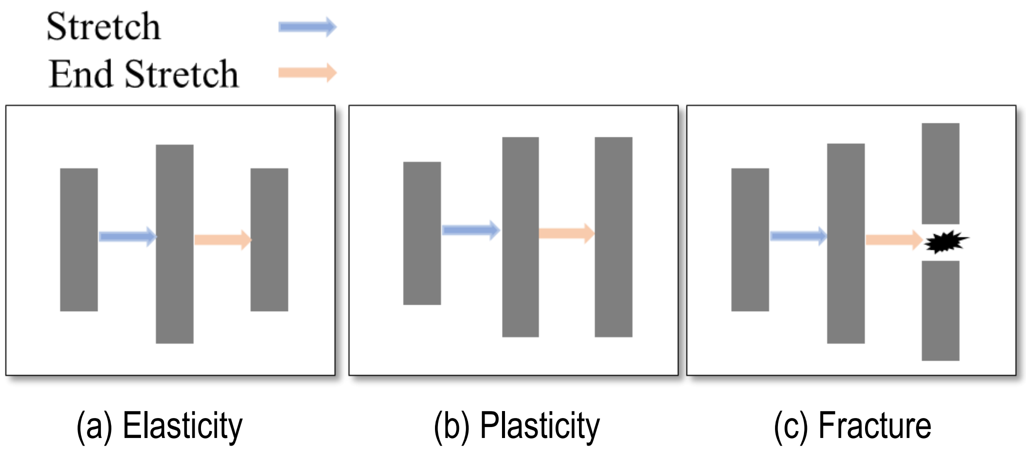
Current Limitations in AE-based Research

- Mostly focused on composite vessels (Type 2, 3, 4 with CFRP)
- However, **Research on Type 1 vessels is lacking**, despite their widespread use

Metal Failure Modes

- **Elasticity** : Stress exceeds yield strength → Reversible deformation (returns to original shape)
- **Plasticity** : Permanent deformation, Material does not return to original shape after stress removal
- **Fracture** : Cracks or full breakage due to excessive stress, Irreversible structural failure

- Metals have **sequential failure behavior** according to fatigue level



[Failure modes of Type 1 vessel]

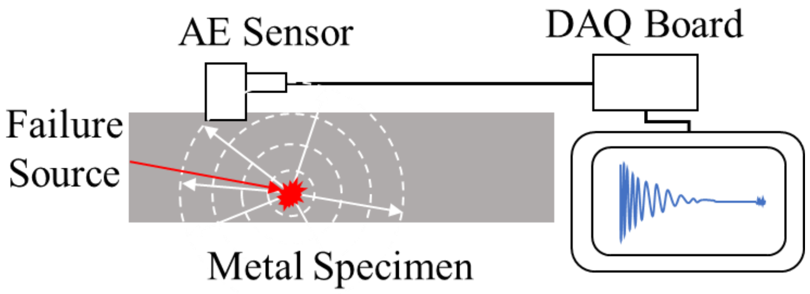
POSTECH PIAI POSTECH Institute of Artificial Intelligence

IARIA

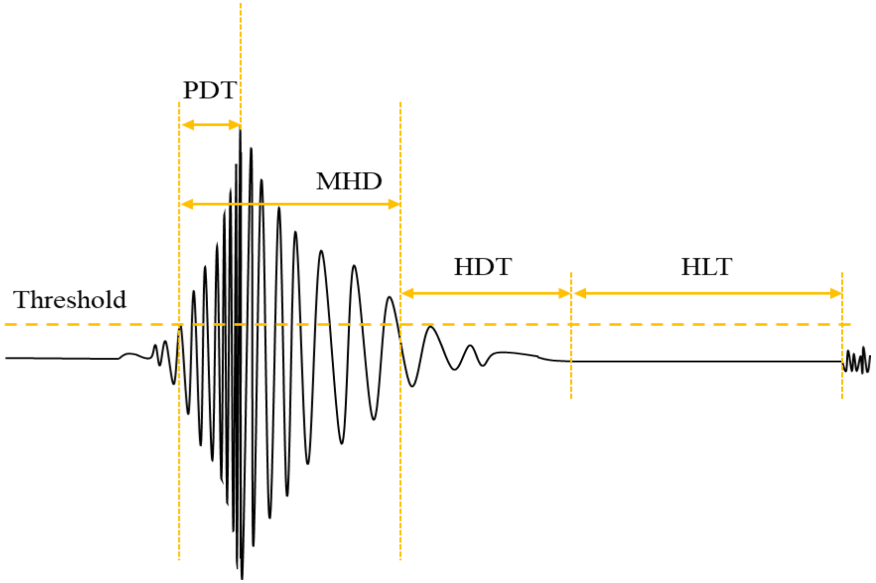
Acoustic Emission Testing (AET) System

- NDT : Can be inspect without damaging the object
  → Enables real-time failure detection
- AET overview
  - **Detects elastic waves** from material deformation
  - System components:
    → AE Sensors (signal detection)
    → DAQ System (digital conversion)
    → Signal Analysis (interpretation)

- Data acquisition from AE sensors
  - Hit definition : use parameters such as threshold, Peak Definition Time (PDT), Hit Definition Time (HDT), Maximum Hit Duration (MHD) and Hit Lockout Time (HLT)
  - Features extracted:
    ✓ Time domain: Max amplitude, Rise time
    ✓ Frequency domain: Peak/Avg frequency



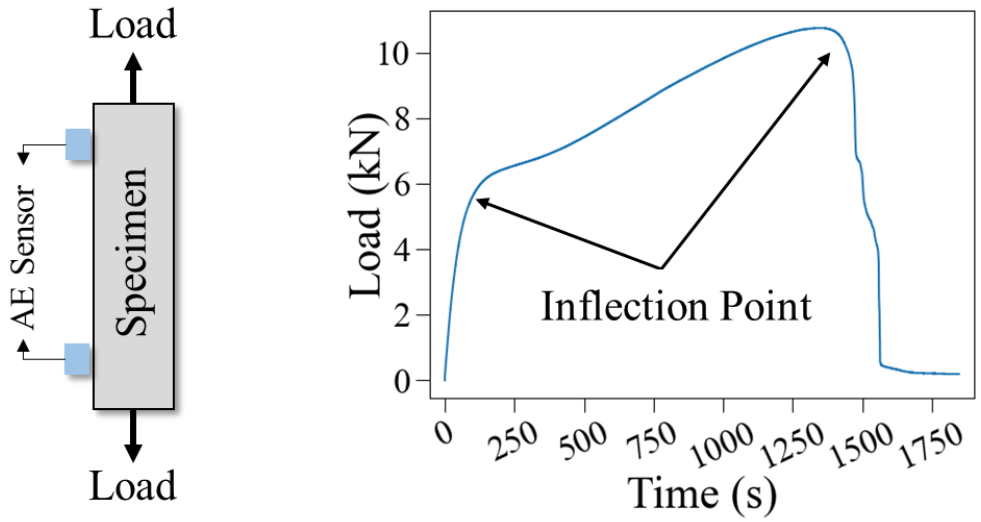
[An example of an AET system]



[Example of DAQ parameters for defining hits]

Tensile Testing with AET

- Tensile testing:
  - Process that pulls customized specimens using Universal Testing Machine (UTM)
  - Here, **stress applied until failure occurs**
- AET-integrated test
  - Monitors acoustic signals during loading
- Failure detection:
  - **Failures appear as inflection points on load graph**
  - By analyzing these inflection points, changes in the specimen properties can be understood.



(a) Tensile testing (b) Tensile-load graph and inflection point

[AET-based tensile test setup and load graph]

4. Failure Detection Methodology

Data Collection – Specimens

- Specimen materials : Stainless Steel (SUS304), Steel (SS400), Aluminum (AL6106-T6)

  → Widely used in hydrogen storage vessels

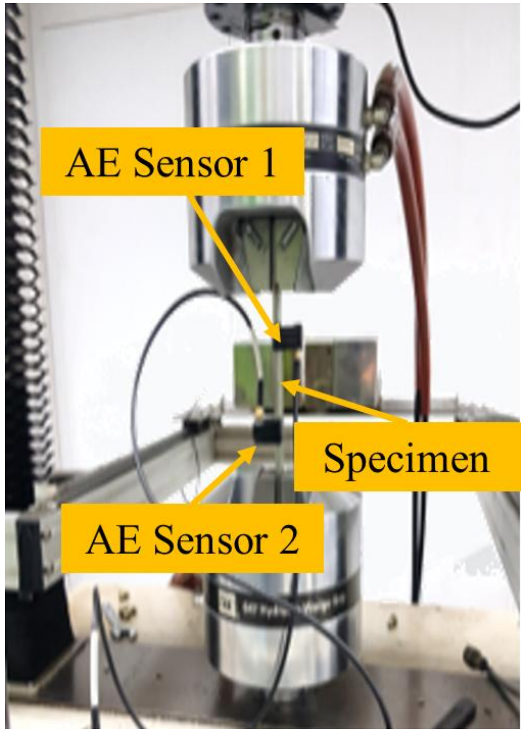
  → These specimens were fabricated in accordance with Korean standard KS B 0801 No. 5

[Example of specimens]

Material	Standard	Example Images
Stainless steel	SUS304	
Steel	SS400	
Aluminum	AL6106-T6	

Data Collection – Tensile testing

- Actuality, perform the tensile testing to induce material specific failures
- Testing setup
  - Equipment: Sintech MTS System
  - AE Sensor: IDK-AES-H150 Resonant Sensor (1 MHz)
- Frequency filtering applied
  - Failures-related AE Signals occur mostly < 500 kHz
  - Noise < 10 kHz removed



[Tensile testing environment]

Data Collection – Data acquisition & labeling

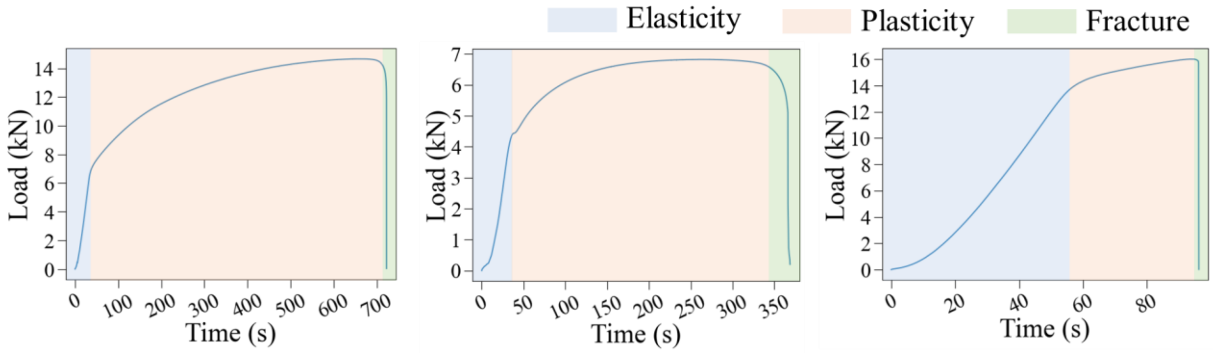
- Data acquisition
  - Collect waveform data during tensile failure events
  - Normal AE signals collected from Type 1 vessels under safe pressure

[Number of acquisition data]

Specimen	Number of data
Stainless steel	333
Aluminum	2,056
Steel	44,792
Type 1 Storage (Normal)	69,243

- Failure region labeling
  - Failure occurs when tensile load exceeds yield strength
  - **Inflection points** on the load-time curve **help segment failure regions**

  - Elastic region: **Stress < Yield Point** (Minimal deformation)
  - Plastic region: **Stress > Yield Point** (Large deformation)
  - Fracture: **Material fails to support load**, Load drops sharply to zero



(a) Stainless steel (b) Steel (c) Aluminum

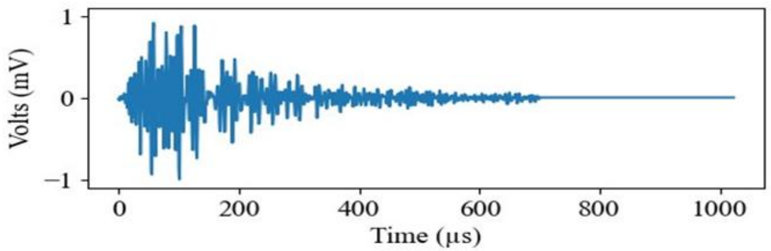
[Tensile-load graph and failure region of each material specimens]

Data Collection – Construction of frequency spectrum dataset

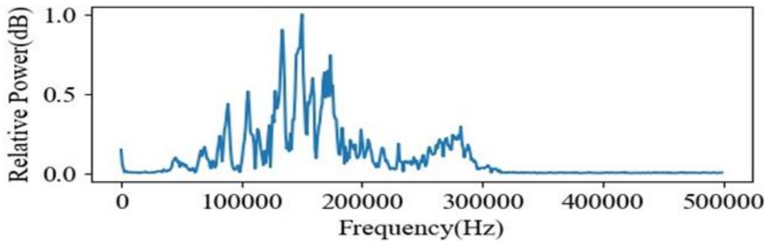
- Frequency features help reduce sensor-related variability
  → Enables more generalizable classification

- So, Discrete Fourier Transform (DFT) applied using FFT algorithm

- FFT :
  - Perform every 1024 samples of waveform
  - Due to symmetry, only positive frequencies (512 samples) are used
  - Scaled to range [0, 1] for normalization



(a) Waveform of AE signal



(b) Frequency Spectrum

[Example of waveform and frequency spectrum]

Augmentation

- Class imbalance can negatively affect model training and classification performance
- So, applied Synthetic Minority Over-sampling Technique(SMOTE) augmentation to the training set

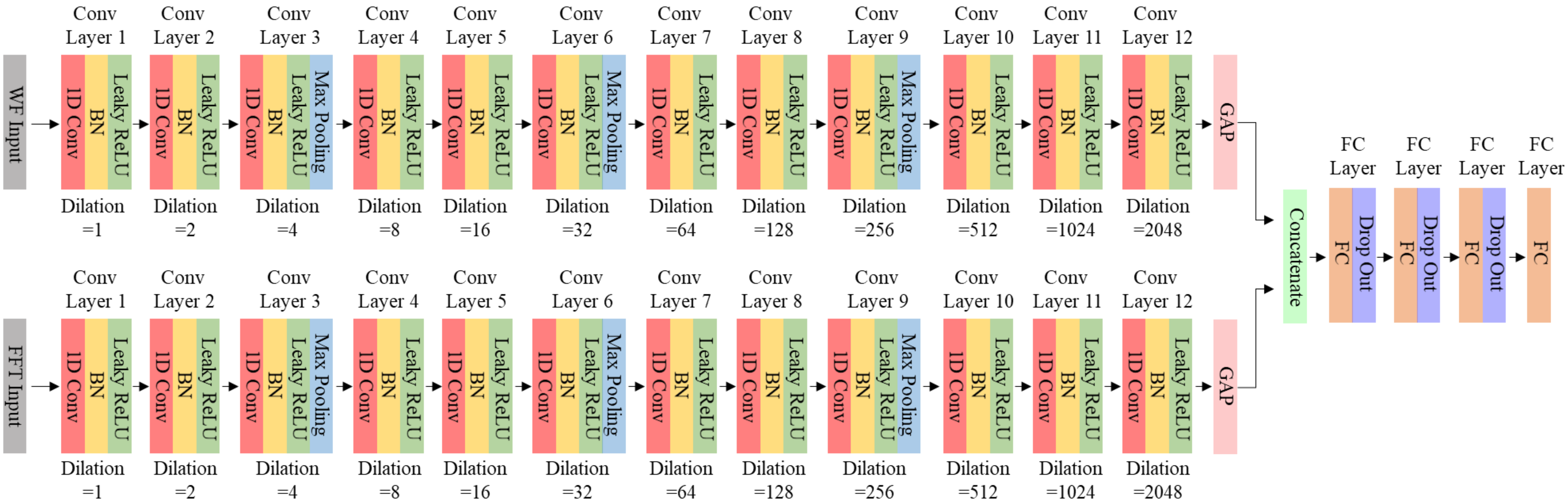
- The table below shows the number of samples in each dataset after SMOTE augmentation

[Number of train / validation / test dataset]

Failure Mode	**Train**		**Validation**	**Test**
	Before Augmentation	After Augmentation		
Elasticity	3,516	**41,545**	1,172	1,172
Plasticity	24,717	**41,545**	8,239	8,239
Normal	41,545	**41,545**	13,849	13,849
Total	69,778	**124,635**	23,260	23,260

Model Architecture

- To classify failure states, Design one-dimensional convolutional neural network (1D-CNN)
- Finally, **12-Layer 1D-CNN** and **ResNet-50 1D-CNN network** are built
- Also, Extracts features from **both waveform and frequency spectrum** for **multimodal learning**



[Proposed 1D-CNN multimodal model architecture]

Experimental Setup

- Deep Learning Model
  - Use batch normalization and max-pooling (size: 2)
  - Softmax activation for probability output
  - Nadam optimizer for model training

- Performance evaluation scenarios
  1. Dataset split : 60% training (SMOTE applied) / 20% validation / 20% test
  2. Classes: Elasticity, Plasticity, Normal
  3. Input : ① Waveform data only, ② Frequency spectrum data only, ③ Multimodal (combined waveform and frequency spectrum)

- Performance comparison with different model

  → 12-Layer 1D-CNN and ResNet-50 1D-CNN used as feature extractor to compare performance

  → Assessed benefits of shallower vs. deeper architectures

Experimental Results

- Table below : classification performance for models using different inputs and architectures.

[Classification Performance]

Input Data Type	Model Structure	Accuracy	Precision	Recall	F1-Score
Waveform Only	12-Layer 1D-CNN	98.87%	0.9603	0.9622	0.9613
	ResNet-50 1D-CNN	98.83%	0.9602	0.9655	0.9628
Frequency Spectrum Only	12-Layer 1D-CNN	98.94%	0.9599	0.9712	0.9654
	ResNet-50 1D-CNN	98.93%	0.9604	0.9624	0.9614
Multimodal	**12-Layer 1D-CNN**	**99.19%**	**0.9723**	**0.9743**	**0.9733**
	ResNet 1D-CNN	98.89%	0.9590	0.9686	0.9637

Experimental Results & Best Model

- Best result
  → **12-layer 1D CNN (Multimodal)**
  → **Accuracy: 99.19%, F1 Score: 0.9733**

- Insights
  - Multimodal model outperformed single-input models
    → Waveform-only and Spectrum-only performed lower
    → Confirms that combining waveform and spectrum is complementary and effective

  - Shallower model (12-layer) outperformed deeper model (ResNet-50)
    → More complex architecture not required for this task

Conclusion

- This study proposed a multimodal deep learning model for failure detection in Type 1 hydrogen storage vessels
  - **Collected AE signals** via tensile tests → Dataset built for elastic, plastic, and normal regions
  - Employ **multimodal model** used both waveform and frequency spectrum data
  - Consequently, Achieved **99.19% accuracy and 0.9733 F1 score**
  - **Demonstrated excellent performance** in failure detection

Thank you

kkb0320@postech.ac.kr