

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

Dong-Ju Kim, Da-Hyun Kim, Hyo-Jin Kim, and Young-Joo Suh  
Institute of Artificial Intelligence, POSTECH, Pohang, South of Korea  
e-mail: {kkb0320, kdhyun8011, hyojinkim, yjsuh}@postech.ac.kr

**Abstract** - Hydrogen is a clean energy source that is essential for responding to climate change and ensuring energy security. Typically, hydrogen storage vessels are exposed to high pressure environments, which can pose an immediate risk of explosion in the event of failure. Therefore, technologies are needed to detect and resolve failures early through diagnostics of hydrogen storage vessels. In this paper, we propose a deep learning-based multimodal failure detection technique to ensure the safety of hydrogen storage vessels. To develop the failure detection technique, we first performed tensile tests on the storage vessel material to collect Acoustic Emission (AE) signals, and also collected failure and normal data based on tensile load graphs. The Synthetic Minority Over-sampling Technique (SMOTE) method was applied to solve the data imbalance. Finally, we developed a multimodal deep learning model using time-domain waveforms and frequency spectra for failure detection, and the proposed method achieved an accuracy of 99.19% and an F1 score of 0.9733, demonstrating excellent failure detection performance. Furthermore, we confirmed that the proposed method shows better performance than using only time-domain waveforms or frequency spectra, and we expect that this research will contribute to the safety diagnosis and maintenance of hydrogen storage vessels.

**Keywords** - Hydrogen Storage Vessels; Acoustic Emission; Multimodal; Deep-learning.

## I. INTRODUCTION

Hydrogen is a clean energy source that emits no greenhouse gases when burned, and is essential for responding to climate change and ensuring energy security. However, hydrogen exists as a gas at room temperature and is highly flammable and bulky, requiring advanced storage and transport technologies. Therefore, high-pressure vessels for compressed hydrogen storage are essential for hydrogen mobility and infrastructure development [1].

In general, hydrogen storage vessels are classified into types 1 to 4 according to their materials and structures, as shown in Figure 1. Type 1 vessels, made entirely of metals such as aluminum or steel, are cost-effective and ideal for transporting hydrogen at low pressures of 200 bar [2]. Type 2 vessels consist of a metal liner reinforced with an outer layer of Glass Fiber Reinforced Plastic (GFRP), allowing a maximum pressure limit of 300 bar. Type 3 vessels enhance this design by using Carbon Fiber Reinforced Plastic (CFRP) as the outer layer, significantly increasing the pressure limit to 700 bar. Unlike Type 2, which does not fully wrap the liner with fibers, Type 3 uses a fully wrapped

structure, providing superior reinforcement. Type 4 vessels, on the other hand, use a resin liner as the inner layer and CFRP as the outer layer, achieving a lightweight design while maintaining the same pressure rating as Type 3.

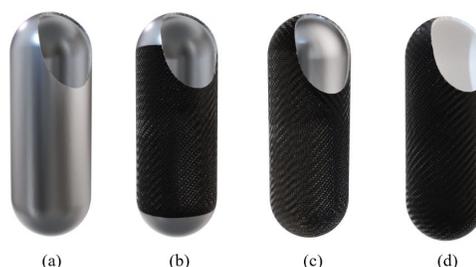


Figure 1. Types of hydrogen gas storages: (a) Type 1, (b) Type 2, (c) Type 3, (d) Type 4.

The metallic composition of Type 1 hydrogen storage vessels makes them susceptible to fatigue, corrosion and cracking, increasing the risk of hydrogen leakage or explosion. Periodic inspection and failure detection are therefore essential for safety. Traditional diagnostics often require disassembly of the vessel, which is not feasible during operation. Failure of a high-pressure vessel poses an immediate risk of explosion, emphasizing the need for in-service diagnostics to detect and resolve failures early, while maintaining reliability and efficiency. Non-Destructive Testing (NDT) technologies are therefore essential. NDT technologies, such as ultrasonic, radiographic and Acoustic Emission Testing (AET) provide real-time in-service safety diagnostics. AET is particularly effective because it analyses acoustic signals generated during failure, making it ideal for high-pressure vessel diagnostics. These methods can ensure safety while preventing catastrophic events such as explosions. However, previous research has focused on specific materials or single failure modes, and in particular on Type 2 and Type 3 vessels.

Therefore, this paper proposes a deep learning based multimodal failure detection technique to ensure the safety of Type 1 hydrogen storage vessels. To do so, we first perform tensile tests on specimens made of aluminum, stainless steel and steel to collect AE signals during failures, and construct a dataset of AE signals using tensile load plots. A multimodal deep learning model using time-domain waveform and frequency spectrum data is developed to improve detection accuracy and reliability. As a result, the

multimodal model achieved an accuracy of 99.19% and an F1 score of 0.9733, demonstrating excellent performance.

## II. BACKGROUNDS

### A. Failure Modes of Hydrogen Storage Vessels

Failures in hydrogen storage vessels refer to structural deformations caused by external impacts, exceeding allowable pressure limits or material durability issues. The high temperature and high pressure conditions resulting from repeated loading and unloading cycles lead to fatigue-related failures, and Type 1 vessels are particularly susceptible due to their lower allowable pressure limits.

As Type 1 vessels are made entirely of metal, any potential failures are also limited to metal failures. Metals have a single molecule structure and failures occur sequentially depending on the fatigue level of the material. Failures in metals are classified into elasticity, plasticity and fracture. Elasticity occurs when the stress exceeds the yield strength, i.e., a deformation that is reversible when the stress is removed. Plasticity, on the other hand, refers to the permanent deformation that occurs even after the stress is removed [3]. Fracture refers to cracks and ultimate fracture caused by excessive stress. Figure 2 shows examples of the three failure modes in Type 1 hydrogen storage vessels.

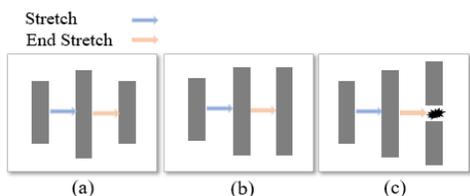


Figure 2. Failure modes of Type 1 vessel: (a) Elasticity, (b) Plasticity, (c) Fracture.

### B. AET-Based Non-Destructive Testing

NDT inspects objects without damage, enabling real-time failure detection in hydrogen storage tanks. AET evaluates material failure by analyzing the elastic waves generated during deformation. AET systems consist of AE sensors for signal detection, Data Acquisition (DAQ) systems for digital signal conversion, and analysis for interpretation. Figure 3 shows an example of an AET system.

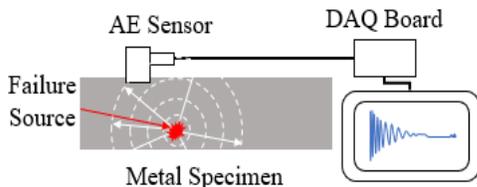


Figure 3. An Example of an AET system.

Accurate AE data acquisition requires appropriate sampling rates. Hits are defined using parameters such as preamplifier gain, threshold and Hit Definition Time (HDT).

Noise is filtered with High Pass Filters (HPF) and Low Pass Filters (LPF), and signal features are extracted in the time domain (e.g. maximum amplitude, rise time) and frequency domain (e.g. peak frequency, average frequency). Figure 4 illustrates the AE waveform and the DAQ parameters used to define hits. To collect accurate data from AE sensors, an appropriate sampling rate must be set. Event occurrences, or hits, are defined using parameters such as preamplifier gain, threshold, Peak Definition Time (PDT), Hit Definition Time (HDT), Maximum Hit Duration (MHD) and Hit Lockout Time (HLT).

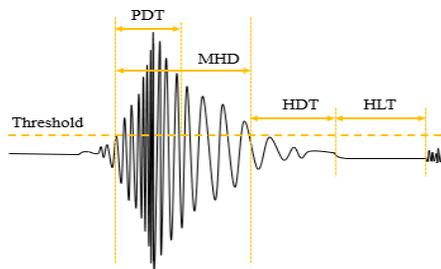


Figure 4. Example of DAQ parameters for defining hits.

### C. Tensile Testing

Tensile testing involves pulling customized specimens using a Universal Testing Machine (UTM) to apply stress until failure occurs. Figure 5 shows an example of an AET-based tensile test and a tensile stress graph. When a specimen reaches failure, its properties change, resulting in variations in the applied stress. Generally, failures appear as inflection points on the tensile load graph. By analyzing these inflection points, changes in the specimen properties can be identified.

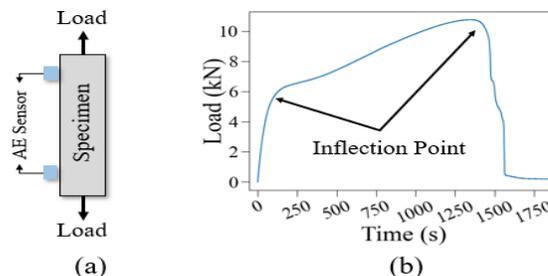


Figure 5. (a) Tensile testing, (b) Tensile-load graph and inflection point.

## III. RELATED WORKS

Research on hydrogen storage failure is divided into Finite Element Model (FEM)-based approaches and AE signal analysis methods. FEM simulates stress and fatigue under operational conditions and analyses potential failures experimentally [4]. However, FEM studies primarily focus on correlations between failure phenomena and fatigue levels, limiting their application for real-time detection during operation. In contrast, AET-based research is simpler as it avoids detailed numerical analysis and relies on acoustic signals generated during failures. This allows for real-time

failure detection. However, AET studies have mainly focused on composite materials, such as CFRP in Type 2, 3 and 4 vessels, and there is a lack of research on Type 1 vessels, despite their widespread use and advantages.

Type 1 vessels are made from a variety of metals such as steel, stainless steel and aluminum and require extensive failure analysis. Recent AET and deep learning studies have analyzed failures in metallic vessels, but have been limited to 4130X steel [5]. Therefore, this study addresses this gap by collecting AE signals from steel, stainless steel, and aluminum through tensile testing and constructing an accurate dataset. A multimodal classification model was also developed using time domain waveforms and frequency spectrum data.

IV. DATA PROCESSING

For safety reasons, it is impractical to directly charge and discharge Type 1 hydrogen storage vessels or to apply destructive pressure. Instead, AE failure signals have been obtained by performing tensile tests on specimens of container materials under predetermined parameters.

A. Data Acquisition

In this study, the specimens included stainless steel (SUS304), steel (SS400) and aluminum (AL6106-T6), all of which are widely used in hydrogen storage vessels [6][7]. These specimens were fabricated in accordance with Korean standard KS B 0801 No. 5, and Table I shows example images of each specimen.

TABLE I. EXAMPLE IMAGES OF SPECIMENS

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

The specimens were subjected to tensile testing to induce material specific failures. Tensile loads were applied using the Sintech 30/G model (MTS system) and AE signals were recorded using the IDK-AES-H150 resonant sensor at 1 MHz. Failures typically occur below 500 kHz, while signals below 10 kHz are often noise or equipment vibration, so a digital filter was applied to remove noise. Hit detection parameters were set to accurately capture peak values and event intervals. Table II summarizes the DAQ settings used in the experiment.

Tensile tests were performed on three specimen types to collect AE waveform data. Specimens were loaded to failure and only event waveform data was collected based on the sensor settings. Normal data was also collected by attaching AE sensors to Type 1 vessels operated within allowable pressure limits. Figure 6 shows the test environment and failed specimens, while Table III lists the number of samples collected.

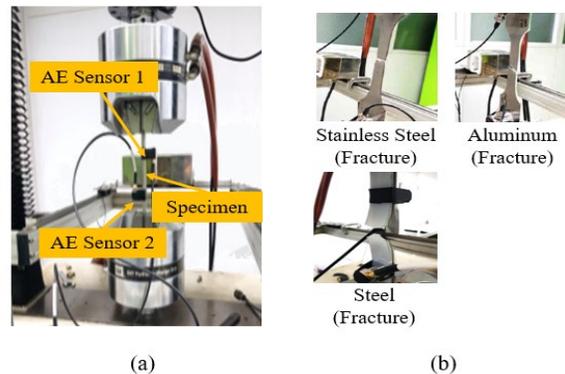


Figure 6. (a) Tensile testing environment, (b) Specimens after testing.

TABLE II. DAQ PARAMETERS CONFIGURED TO DEFINE HITS

Parameter Type	Parameters	Value	Unit
Sensor	Sampling Rate	1	MHz
	Pre-Amp Gain	40	dB <sub>ae</sub>
	Threshold	30	dB
Digital Filter	High Pass Filter (HPF)	10	kHz
	Low Pass Filter (LPF)	500	kHz
Hit Detection	Peak Definition Time (PDT)	200	μs
	Hit Definition Time (HDT)	400	μs
	Maximum Hit Duration (MHD)	1	ms
	Hit Lock-out Time (HLT)	10	ms

TABLE III. NUMBER OF COLLECTED DATA

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

B. Data Labeling

Metal failure occurs when the tensile load exceeds the yield strength, changing atomic arrangements and material properties. These changes vary with the rate of load increase, allowing failure regions to be identified using a time dependent load curve. The elastic region occurs when stresses remain below the yield point, causing minimal deformation and a continuous increase in load. Plasticity begins when the stress exceeds the yield point, resulting in significant deformation and slower load increase. An inflection point marks the transition from elasticity to plasticity. Fracture occurs when the material can no longer support the stress, causing the load to drop rapidly to zero,

creating another inflection point. These points divide failure regions and matching their times to the event waveforms allows failure labelling. Figure 7 illustrates tensile load graphs and failure region subdivisions, summarized in Table IV.

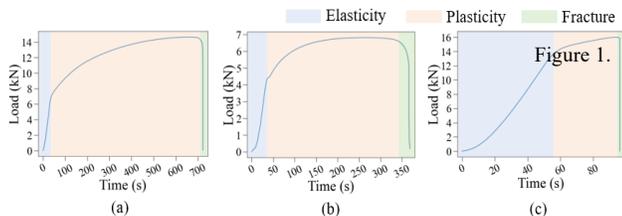


Figure 7. Tensile-load graph and failure region of each material specimens: (a) Stainless steel, (b) Steel, (c) Aluminum.

### C. Data Preprocessing

The dataset was pre-processed for deep learning training, incorporating frequency domain information to improve performance.

#### 1) Unify Waveform Lengths and Min-Max Scaling

All waveform lengths were unified to 1024 samples. AE waveforms vary in length depending on event duration, but consistent input sizes are required for deep learning. This study determined the optimal length to minimize information loss and computational load. Waveform lengths were statistically analyzed and outlier information segments were identified to set the unified length to 1024, as shown in the histogram in Figure 8.

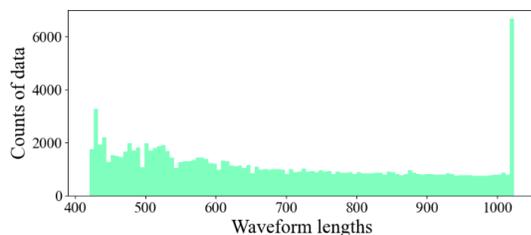


Figure 8. Histogram of waveform lengths.

Second, the amplitude of the waveform was scaled to [-1, 1] using min-max scaling. Without scaling, data values could

vary widely, causing instability and inefficiency during the weight update process. Scaling creates a uniform distribution, reducing variability and stabilizing training. Since waveform data includes negative values, the scaling range was set to [-1, 1], as expressed below:

$$x' = 2 \left( \frac{x_i - \min(x)}{\max(x) - \min(x)} \right) - 1 \quad (1)$$

where  $\max(x)$  and  $\min(x)$  is:

$$\max(x) = \begin{cases} |\max(x)|, & |\max(x)| \geq |\min(x)| \\ |\min(x)|, & |\max(x)| < |\min(x)| \end{cases} \quad (2)$$

$$\min(x) = \begin{cases} -|\max(x)|, & |\max(x)| \geq |\min(x)| \\ -|\min(x)|, & |\max(x)| < |\min(x)| \end{cases} \quad (3)$$

For  $x$ , the baseline is 0, but  $x'$  maps the maximum and minimum of  $x$  to 1 and -1 respectively. If their absolute values differ, the centers of the waveforms can vary, increasing the variance of the data. To overcome this, the larger absolute value is mapped to 1 and the smaller to -1, centering the waveform at 0. Figure 9 shows the original waveforms after min-max scaling and alignment to 1024 length, comparing standard scaling and the adapted method.

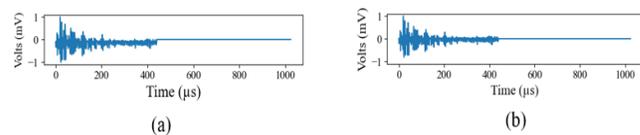


Figure 9. Plot a waveform after preprocessing: (a) General min-max, scaling (b) Ours.

#### 2) Construction of Frequency Spectrum Dataset

Understanding both intrinsic and frequency characteristics is essential in waveform analysis. Frequency characteristics minimize the effect of sensor type and placement, facilitating generalized classification methods. To incorporate this, the frequency domain data was constructed using the Fourier Transform (FT). As the signals were digital and discrete, the Discrete Fourier Transform (DFT) was applied using the Fast Fourier Transform (FFT) algorithm for computational efficiency.

TABLE IV. TIME AND NUMBER OF ACQUIRED WAVEFORMS ACCORDING TO FAILURE REGIONS IN EACH SPECIMEN

Specimen	Time by Region (Seconds)				Number of Waveforms		
	Total Experiment	Elasticity	Plasticity	Fracture	Elasticity	Plasticity	Fracture
Aluminum	100.54	0 ~ 57.21	57.2 ~ 100.4	100.4 ~ 100.5	1,583	467	6
Steel	379.85	0 ~ 37.47	37.4 ~ 365.9	366.7 ~ 379.8	4,093	40,621	78
Stainless steel	728.94	0 ~ 33.94	34.6 ~ 706.7	706.5 ~ 728.9	184	107	42
<b>Total</b>					<b>5,860</b>	<b>41,195</b>	<b>126</b>

The FFT transformation produced discrete frequency spectrum graphs as line plots of the frequency distribution. Unlike the waveform data, the frequency spectrum lacks negative values and was scaled to [0, 1]. Each FFT result contained 1024 samples, corresponding to the length of the waveform. Due to y-axis symmetry, only positive frequencies were retained, reducing the length to 512. Table V shows examples of the original waveform, the pre-processed waveform and the scaled spectrum data after FFT transformation.

TABLE V. EXAMPLE OF PREPROCESSED WAVEFORM AND FREQUENCY SPECTRUM

Data Type	Specimen	Example
Waveform	Stainless steel	
	Aluminum	
	Steel	
Frequency Spectrum	Stainless steel	
	Aluminum	
	Steel	

### V. CLASSIFICATION MODEL FOR FAILURE DETECTION

A deep learning model was developed using the constructed dataset to classify type 1 storage failures. Fracture, which indicates material rupture and explosion, was excluded as it is irrelevant to safety diagnostics. Only elasticity, plasticity and normal data were used to distinguish these states. The data set was divided into training, validation and test sets in a ratio of 60:20:20. However, there is a significant imbalance between elasticity and plasticity data. To address this, the SMOTE method was used to balance the training data [8]. Table VI shows the number of training, validation and test data augmented by SMOTE.

A binary failure and normal classification model was trained and tested using the collected dataset. The model, designed as a one-dimensional convolutional neural

network (1D-CNN), extracted features from both waveforms and frequency spectra for classification. To better capture temporal characteristics, an extended causal 1D-CNN architecture was used. Figure 10 shows the structure of the diluted causal 1D CNN.

TABLE VI. NUMBER OF TRAIN/VALID/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

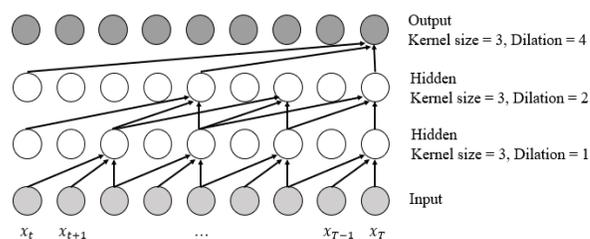


Figure 10. Dilated Causal 1D-CNN

Convolution, batch normalization and max-pooling (size: 2) layers were used to process the input data to extract features, which were then transformed into a 1-dimensional vector with 3 outputs using global average pooling, bypassing the need for a fully connected layer. The outputs were passed through the softmax activation function to compute the final probabilities for each failure type. The Nadam optimizer was used to train the model. This study evaluated the classification performance under three input scenarios: waveform data only, frequency spectrum data only, and combined features of both. In addition, the ResNet-50 architecture was used as the feature extractor to evaluate the performance improvements from a deeper network design.

Table VII shows the experimental results for models using different inputs and architectures. The result shows that the 12-layer multimodal model achieved the best performance with 99.19% and an F1 score of 0.9733. The superior results of the flatter model compared to ResNet suggest that a more complex architecture is not necessary for this classification task. Furthermore, the higher performance of the multimodal model compared to single input models (waveform or frequency spectrum) indicates that the two types of data are complementary for this classification problem. Figure 11 shows the structure of the best performing model, i.e. the 12-layer 1D CNN multimodal model.

TABLE VII. 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	<b>99.19%</b>	<b>0.9723</b>	<b>0.9743</b>	<b>0.9733</b>
	ResNet 1D-CNN	98.89%	0.9590	0.9686	0.9637

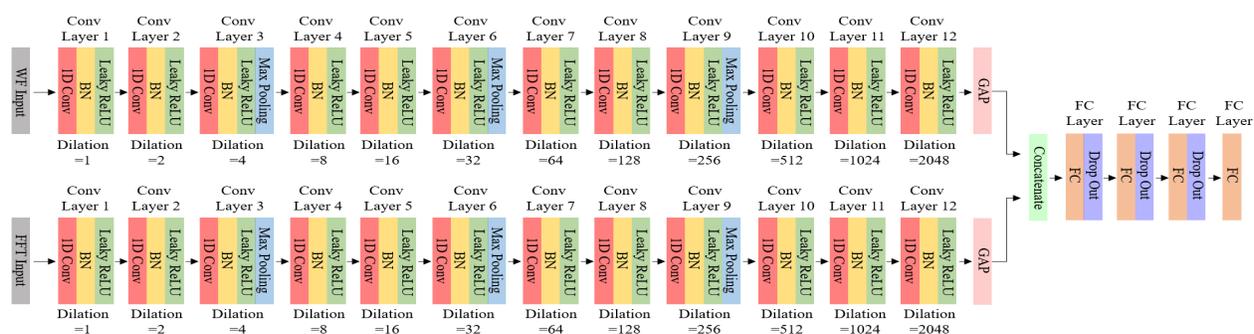


Figure 11. Proposed 1D-CNN multimodal model.

VI. CONCLUSION

This paper proposes a deep learning-based multimodal fault detection technique to ensure the safety of hydrogen storage vessels. To this end, we first collected AE signals from Type 1 hydrogen storage vessels by tensile testing and constructed data sets for elastic, plastic and normal regions. We then developed a multimodal deep learning fault detection model using waveform and frequency spectral data. From the experimental results, we confirmed that the proposed multimodal model achieved an accuracy of 99.19% and an F1 score of 0.9733, demonstrating excellent failure detection performance. In the future, the proposed method is expected to enable real-time fault detection of Type 1 vessels, contributing to efficient and reliable safety diagnostics.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.2022R1A6A1A03052954)

REFERENCES

- [1] D. Cecere, E. Giacomazzi, and A. Ingenito, "A review on hydrogen industrial aerospace applications," *Int. J. Hydrogen Energy*, vol. 39, no. 20, pp. 10731–10747, Jul. 2014.
- [2] H. Barthélémy, "Hydrogen storage – Industrial perspectives," *Int. J. Hydrogen Energy*, vol. 37, no. 22, pp. 17364–17372, 2012.
- [3] R. Hill, *The Mathematical Theory of Plasticity*. Oxford, U.K.: Oxford University Press, 1950.
- [4] A. El-Abbassi, A. Mekki, and A. Medles, "A comparative study of hydrogen storage in solid-state materials," *Int. J. Hydrogen Energy*, vol. 45, no. 9, pp. 5367–5375, 2020.
- [5] R. M. Gaikwad, A. Kulkarni, and S. V. Joshi, "Experimental investigation of hydrogen storage in composite vessels," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 810, no. 1, p. 012016, 2020.
- [6] S. Kumar and S. K. Dwivedi, "Composite Pressure Vessels: Design, Analysis, and Manufacturing," *J. Pressure Vessel Technol.*, vol. 140, no. 6, 2018.
- [7] O. Comond, D. Perreux, F. Thiebaud, and M. Weber, "Methodology to Improve the Lifetime of Type III HP Tank with a Steel Liner," *Int. J. Hydrogen Energy*, vol. 34, no. 2, pp. 576–586, 2009.
- [8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.