sustainability

MDPI

Article

Digital Twin-Based Hydrogen Refueling Station (HRS) Safety Model: CNN-Based Decision-Making and 3D Simulation

Na Yeon An, Jung Hyun Yang, Eunyong Song, Sung-Ho Hwang, Hyung-Gi Byun and Sanguk Park *

Department of Electronic, Information and Communication Engineering, Kangwon National University, Samcheok 25913, Republic of Korea; ny5290@kangwon.ac.kr (N.Y.A.); 980917@kangwon.ac.kr (J.H.Y.); 5685s@kangwon.ac.kr (E.S.); shhwang@kangwon.ac.kr (S.-H.H.); byun@kangwon.ac.kr (H.-G.B.)
* Correspondence: sanguk@kangwon.ac.kr

Abstract: This study presents a safety management model for hydrogen refueling stations, integrating digital twin technology and artificial intelligence (AI) to enhance operational safety. Given the risks associated with high-pressure gas handling and potential fires from hydrogen leaks, real-time safety monitoring is crucial. The proposed model is based on a digital twin, a virtual replica of the physical system using real-time data, including temperature, pressure, and state of charge, collected from an actual hydrogen refueling station in Samcheok, Gangwon Province. Out of nine tested machine learning and deep learning algorithms, the convolutional neural network (CNN) demonstrated the highest performance (accuracy: 1, F1 score: 0.993) for risk prediction. Using AI libraries like Scikit-Learn and TensorFlow, the model achieved prediction times of 68 milliseconds, enabling decision-making at intervals of 1 s. Developed with the Unity 3D modeling tool, the digital twin visualizes predicted risk situations, allowing users to quickly identify and respond to potential hazards. This approach offers a robust solution for improving the safety of hydrogen refueling stations.

Keywords: digital twin; hydrogen refueling station; prediction model; machine learning; deep learning

check for updates

Citation: An, N.Y.; Yang, J.H.; Song, E.; Hwang, S.-H.; Byun, H.-G.; Park, S. Digital Twin-Based Hydrogen Refueling Station (HRS) Safety Model: CNN-Based Decision-Making and 3D Simulation. *Sustainability* **2024**, *16*, 9482. https://doi.org/10.3390/su16219482

Academic Editor: Eul-Bum Lee

Received: 12 September 2024
Revised: 18 October 2024
Accepted: 28 October 2024
Published: 31 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Hydrogen is an eco-friendly and highly efficient energy source, and its importance has been increasingly emphasized with the emergence of new technologies, such as hydrogen fuel cell vehicles. However, safety issues have become a major concern due to the nature of hydrogen. Hydrogen refueling stations are critical facilities that directly supply hydrogen energy to civilians, and their safety is a key factor in the activation of the hydrogen economy [1].

Hydrogen is the smallest molecule and can easily leak. If the leaked hydrogen reaches explosive concentrations in the air, the risk of explosion is extremely high. Hydrogen is colorless, odorless, and lighter than air, which causes it to quickly disperse upward when leaked, making it difficult to detect the presence of a leak [2]. Hydrogen is typically stored at high pressure (350–700 bar). High-pressure storage poses significant risks if the storage vessel is damaged or ruptured. There are risks associated with pressure changes during the refueling process, making the maintenance and safety management of high-pressure systems essential.

Hydrogen can also be stored in liquid form at extremely low temperatures (−253 °C). In cryogenic environments, materials become more brittle, increasing the risk of accidents. Failure to manage pressure changes due to temperature variations can lead to safety incidents [3,4]. Hydrogen refueling stations contain various electrical installations, and electrical sparks from these installations can combine with hydrogen to increase the risk of explosions.

To address these issues, Internet of Things (IoT) sensors (hydrogen detectors) are installed throughout refueling stations to quickly detect leaks and automatically activate

warning systems. Regular inspections and maintenance ensure the reliability of detection systems [5,6]. Additionally, considering the properties of hydrogen, refueling stations should be designed with explosion-proof structures [7]. Proper ventilation systems should be installed to quickly dissipate leaked hydrogen. Hydrogen storage tanks and refueling equipment are constructed of high-strength materials to enhance their durability. Most importantly, a real-time monitoring system should be implemented to detect leaks and abnormal conditions immediately, enabling a rapid response in emergencies through automated safety devices and shutdown systems. The safety of hydrogen refueling stations is a critical challenge that must be addressed to expand the use of hydrogen energy and activate the hydrogen economy. The comprehensive application of various safety management measures, continuous research, and technological development are necessary to enhance safety [8]. This will ensure the safe operation of hydrogen refueling stations and promote the widespread adoption of hydrogen energy.

This study proposes a digital twin-based smart hydrogen refueling station safety model to establish safe hydrogen refueling stations. By integrating artificial intelligence, the model determines safe or unsafe conditions at hydrogen refueling stations based on previously collected data [9]. In a digital twin-based hydrogen refueling station safety management system, AI-based risk decision-making can be achieved through various methods and technologies [10]. The approach used in this study is as follows.

1.1. Sensor Data Analysis and Real-Time Monitoring

- Various sensors at hydrogen refueling stations monitor the operational conditions, hydrogen, and surrounding environment of the station. AI analyzes these data in real-time to evaluate normal operations and detect abnormal conditions.
- For example, AI systems can analyze data, such as detected hydrogen leaks, high-pressure valve anomalies, and abnormal temperatures, to determine risk situations in real time.

1.2. Data-Driven Risk Prediction and Prevention

- AI models based on historical data can predict the risk factors that may arise during the operation of hydrogen refueling stations. For example, they can identify potential risks that may occur under specific weather conditions or usage patterns and suggest preventive measures in advance.

1.3. Safety Evaluation Through Simulation and Virtual Experimentation

- Simulations using digital twins can be used to test the operation of refueling stations under various scenarios. AI can analyze the results of these simulations to predict potential risks that may occur during actual operations and evaluate safety.

1.4. Emergency Response and Decision Support

- In the event of an accident or emergency, AI systems can quickly analyze data to support an immediate response. For example, if a leak occurs, AI can recommend appropriate response strategies by comprehensively considering the sensor data, environmental conditions, and surrounding situations.

By combining these methods, the digital twin-based hydrogen refueling station safety management model can support faster and more accurate detection and response to risk situations. This study introduces data analysis and prediction through the collection of hydrogen refueling station data (CSV) and digital twin-based simulation technology.

2. Related Studies

The following are current studies related to hydrogen refueling station safety systems:

2.1. Safety Guideline Research

A previous study [11] analyzes the characteristics of hydrogen jets and flames to establish safe distances at hydrogen refueling stations. It simulates hydrogen jet and flame characteristics using the hydrogen-specific analysis program HyRAM, based on the size of the leakage source at refueling stations. Another study [12] predicts jet fires, explosion possibilities, and damage ranges depending on the capacity of high-pressure hydrogen storage tanks in general hydrogen refueling station systems, using the Chamberlain model and vessel rupture. It also proposes guidelines for installing hydrogen refueling stations in other areas of South Korea based on quantitative risk assessments, considering social risks and F-N curves. In another study, leak scenarios were visualized in 3D according to the pressure and hydrogen injection hole size using the commercial computational fluid dynamics tool FLACS.

2.2. Traditional Methods for HRS Safety Systems

A previous study [13] proposes a method for locating hydrogen leakage in fully confined spaces using a deep belief network with pre-training and fine-tuning (DBN-PF). A Gaussian distribution model simulates hydrogen spread, and the DBN is trained using unsupervised and supervised techniques. Experiments with eight sensors across 25 leak locations show that the method achieves an average positioning error of 20.62 mm, reducing error by up to 82.37% compared to traditional neural networks, demonstrating high accuracy in detecting leak sources with minimal data. A previous study [14] explored safety systems such as leak detection, emergency shutdown, and fire suppression systems used at hydrogen refueling stations. Another study [15] mentions the ongoing construction of a system that sends alerts when abnormal sensor signals are detected in real-time safety management at domestic hydrogen refueling stations.

2.3. AI-Based HRS Safety Systems

A previous study [16] analyzed values such as daylight, weekdays, accident types, locations, speed limits, and average speeds using an AI binary classification model to build a traffic safety system, achieving an accuracy of approximately 85%. In another study [17], a smart safety management system for workplaces was proposed. This system suggests real-time monitoring and accident prediction by integrating big data and AI technologies to analyze data collected from sensors connected by IoT technology and CCTV. This study [18] proposes a TD3-based health-conscious EMS for fuel cell hybrid buses, optimizing hydrogen economy, fuel cell durability, and battery health. Validation shows a 28.02% battery life improvement and 8.92% better overall economy compared to existing EMS. Accurately predicting accident consequences in HRS, this study [19] proposes a hybrid surrogate model that combines a Generative Adversarial Network (GAN) and Long Short-Term Memory Network (LSTM), incorporating a Deep Neural Network (DNN) for predicting hydrogen leakage consequences in HRSs based on source parameters. This study [20] presents a Physics-informed Convolutional LSTM (PIConvLSTM) model that predicts hydrogen concentration distribution after leakage at hydrogen refueling stations. Using FLACS-simulated data for training, the model improves accuracy at gas cloud boundaries and enables fast, real-time risk prediction.

2.4. Digital Twin-Based HRS Safety Systems

A previous study [21] introduced a technology that uses AI vision sensors to implement object and screen position information in digital twins, allowing them to determine only object collisions. Another study [22] discusses the prediction of accidents during tunnel construction by displaying values obtained from various sensors using digital twins. This study [20] explores the 3D geometric modeling of HRS to predict hydrogen concentration distribution after leakage at hydrogen refueling stations.

2.5. Current Technology Issues and Solutions

Current research indicates that there is no established dynamic digital twin-based safety system for HRS. It is challenging to respond optimally to hydrogen leak fires and explosions that can cause significant damage with static real-time monitoring alone. Detecting accident precursors before they occur allows quicker and more appropriate actions. Related studies [16–20] have demonstrated the potential of AI-based safety systems to predict accidents with a certain probability based on appropriate data values. However, there are currently no studies addressing AI-based dynamic digital twin safety systems for HRS. Digital twin technology enables more efficient simulations by creating environments similar to reality in 3D. Therefore, this study proposes a dynamic digital twin-based smart hydrogen refueling station safety system that helps to efficiently respond to risk situations in a digital twin-based safety management system by predicting risks using data detected from various sensors. Table 1 shows the analysis of previous research.

Table 1. Analysis of previous research.

Current Research	Safety Guideline	Traditional Method	EMS	AI					Static Digital Twin	Dynamic Digital Twin	Novelty
				GMDH	DNN	CNN	RNN	GAN			
Kang et al., 2019 [11]	√										65%
Kim et al., 2021 [12]	√										50%
Zhou et al., 2024 [13]		√									60%
Genovese et al., 2024 [14]		√									60%
Guido et al., 2020 [16]				√	√						75%
Eom et al., 2020 [17]					√	√					80%
Jia et al., 2024 [18]			√								45%
He et al., 2024 [19]					√		√	√			85%
He et al., 2024 [20]						√					85%
Kim et al., 2023 [21]						√			√		70%
Ye et al., 2023 [22]		√							√		70%
Proposed System					√	√				√	

Novelty was assigned by the author as a percentage based on technological recency and the connectivity of the applied technology.

3. System Overview

3.1. System Architecture

Figures 1 and 2 illustrate the schematics and architecture of the digital twin-based smart hydrogen refueling station safety management model. The entire system is divided into six parts, each with a specific role, as follows. The real world represents the actual hydrogen refueling station infrastructure. The "Virtual World" represents a virtual hydrogen refueling station implemented using a digital twin. The "IoT Infrastructure" is responsible for collecting data from the real world to implement the virtual world. The "Data Hub" collects data from the IoT Infrastructure and facilitates AI training and data analysis. The collected data were analyzed and predicted based on artificial intelligence, and the predicted results were used to determine if the current situation was risky. The following section provides a detailed explanation of each component:

- Real World: The concept of the real world contrasts with the virtual or digital world, referring to the physical reality we experience and perceive in our daily lives. It includes all the elements that exist in the physical environment, such as objects, events, and interactions [23]. The real world possesses characteristics such as complexity and uncertainty, providing various challenges and opportunities for research and application. It is often difficult to conduct real-world experiments in controlled environments, making it a tool to provide practical solutions to real-world problems by training AI models using collected data [24,25].
- Virtual World: The virtual world refers to a digital environment created using computer technology and networks that offer a virtual space where users can interact. The virtual world possesses characteristics distinct from physical reality and is designed to provide an immersive experience for users [26]. It is used for various purposes and is

created through digital technologies such as computer graphics, software algorithms, and databases. It operates according to programmed rules and logic [27]. This allows multiple users to connect and interact simultaneously, providing the functionality for users to create and share content directly. Digital twin technologies fall under this category [28].

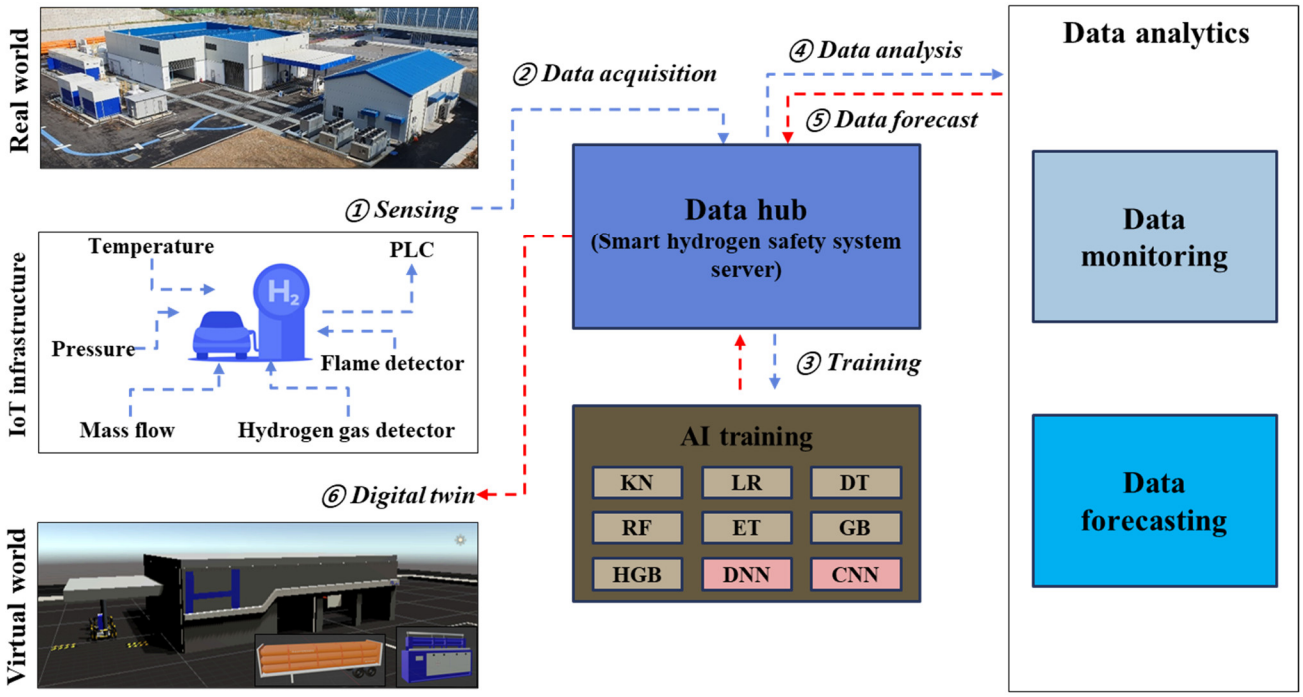


Figure 1. System overview.

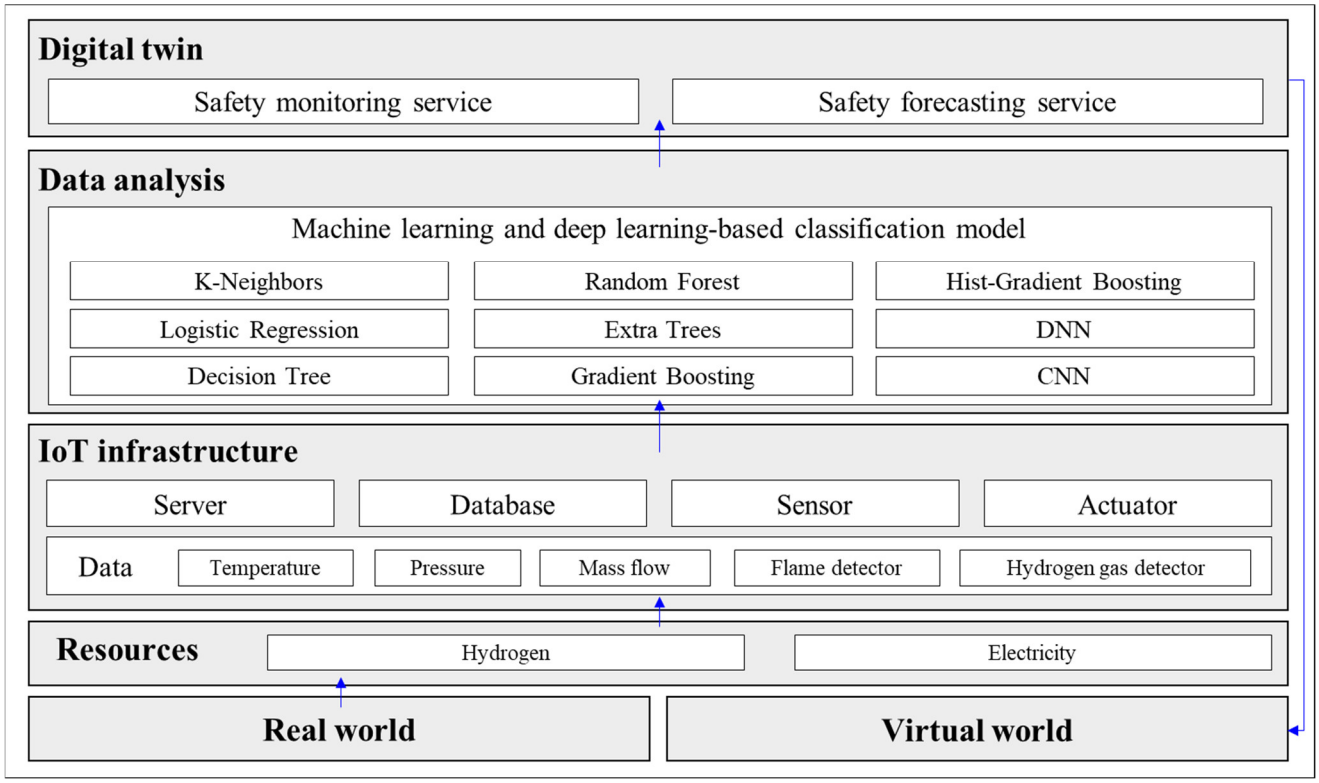


Figure 2. System architecture.

- Cyber-Physical System (CPS): A CPS is a system in which computer-based algorithms closely interact with physical processes by integrating sensors, actuators, networks, and software, enabling seamless interaction between the physical and digital worlds. The main purpose of a CPS is to monitor and control the physical environment in real-time, providing high levels of autonomy and adaptability in various applications. CPSs are implemented by collecting and analyzing data and generating appropriate control commands [29]. They are utilized in fields such as smart grids and smart homes, providing innovative solutions [30] through the convergence of physical environments and digital technologies [31,32].
- Data Hub (Smart Hydrogen Safety System Server): A data hub is a centralized platform that integrates and manages data from various sources, storing, processing, and analyzing data to provide users with necessary information. The Smart Hydrogen Safety System Server is a specialized data hub [33] designed for the safe production, storage, distribution, and use of hydrogen energy sources. This system is a key component of hydrogen safety monitoring and management systems [34]. It ensures the safety of hydrogen-related facilities and infrastructure through real-time data collection and analysis, aiming to improve accident prevention and response capabilities [35].
- AI Training: AI training involves teaching an AI model to perform specific tasks using the given data and algorithms, enabling computers to recognize patterns and make predictions. It utilizes machine learning (ML) and deep learning technologies, acting as a key element in solving complex problems in various industries and research fields. The process involves data collection, preprocessing, model selection, training, evaluation, and deployment, with each stage significantly affecting the performance and accuracy of AI systems. The performance of an AI model relies heavily on the quality and quantity of training data, making it essential to secure high-quality, large-scale datasets, to remove noise through data preprocessing, normalize features, and ensure data diversity [36–38].
- Data Analytics: Data analytics involves organizing and interpreting data collected from various sources to derive meaningful insights. It includes stages such as data collection, preprocessing, analysis, visualization, and interpretation, and is used for multiple purposes, including decision support, problem-solving, and predictive model development. Types of data analytics include descriptive, diagnostic, predictive, and prescriptive analytics, which use methods such as statistical analysis and machine learning [39].

3.2. Status of Hydrogen Refueling Stations

This section provides an overview of hydrogen refueling stations located in Samcheok, Gangwon Province, South Korea. Figure 3 presents a photograph of the hydrogen refueling station in Samcheok, and Figure 4 provides a schematic diagram of the human–machine interface (HMI) of the hydrogen refueling station.



Figure 3. Hydrogen charging station in Samcheok.

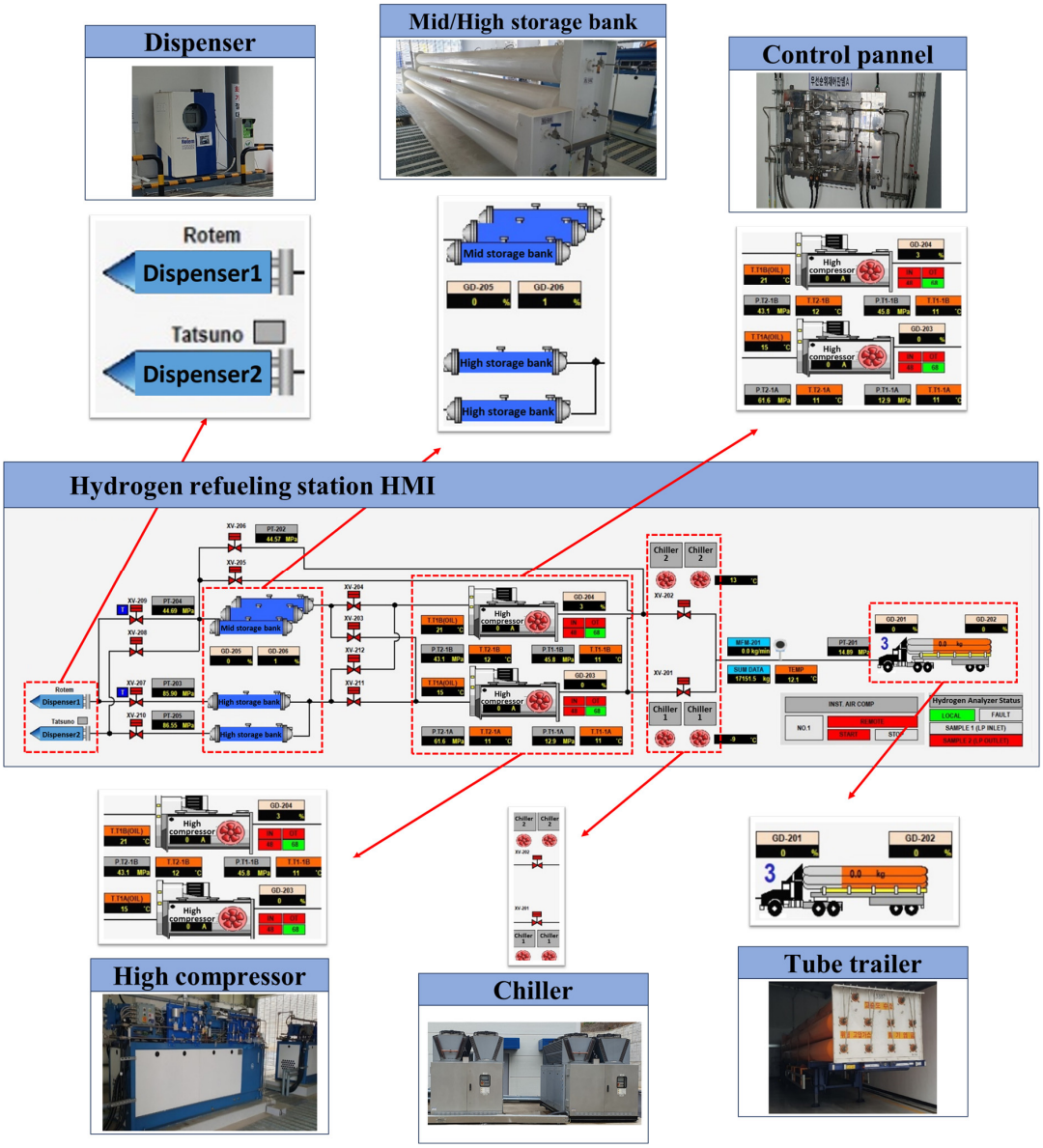


Figure 4. Hydrogen charging station HMI and components.

The descriptions for each component of the hydrogen refueling station are as follows:

- Tube trailer: A mobile trailer for transporting hydrogen equipped with multiple tube-shaped containers that store high-pressure hydrogen. It is used to transport hydrogen from production facilities to refueling stations. Upon arrival, high-pressure hydrogen is transferred to storage tanks within the station. The pressure can typically reach 200–500 bar, and it is one of the primary hydrogen supply sources for refueling stations.
- Chiller: A cooling device used to control the temperature increase during hydrogen refueling. When hydrogen is compressed to high pressure or refueled, heat is generated. If the temperature of hydrogen is excessively high, safety issues can arise. A chiller is used to maintain a constant temperature for hydrogen prevention.
- High compressor: A device that compresses hydrogen to high pressure for storage in station tanks or for supplying hydrogen to vehicles. It typically compresses hydrogen to pressures above 700 bar, generating the high pressure required for rapid refueling of vehicles. This ensures the rapid and safe injection of hydrogen into the fuel tanks of hydrogen vehicles.

- Mid/high storage bank: A tank system for storing hydrogen at medium or high pressures. At the refueling station, hydrogen is stored at various pressure levels to supply vehicles at the appropriate pressure as needed. The mid-storage bank typically maintains a pressure of approximately 350 bar, whereas the high-storage bank maintains pressures above 700 bar, allowing for hydrogen refueling of various types of vehicles.
- Dispenser: A device for directly injecting hydrogen into vehicles. It functions similarly to a gas pump at a fuel station, where the driver connects the dispenser to the fuel tank of the vehicle to inject hydrogen. The dispenser precisely controls the flow rate and pressure of hydrogen, ensuring safety of the refueling process. It also displays necessary data to the user during refueling.
- Hydrogen charging station HMI: An interface for monitoring and controlling the operation of a hydrogen refueling station. The HMI allows the station operator to check the status of the hydrogen refueling station in real time, control the refueling process, and respond immediately to any issues. Through the HMI, key data such as the refueling process, pressure, and temperature can be monitored, and parameters can be adjusted as needed. This is a crucial element that enables interactions between the user and system.

These components are essential for the safe and efficient supply and refueling of hydrogen at refueling stations. Data collection is crucial for implementing a smart hydrogen refueling station safety management model.

Tables 2 and 3 present the types and statuses of data obtained from the hydrogen refueling station. As shown in the tables, 24 types of data were obtained for each component of the hydrogen refueling station. Please refer to the tables for a description of the data.

Table 2. Data description.

Classification	Data Name	Description
Tube trailer	GD-201 (%)	SoC of tube trailer
	GD-202 (%)	SoC of tube trailer
	PT-201 (MPa)	Pressure of tube trailer
	TEMP (°C)	Temperature of input from tube trailer to chiller
Chiller	Chiller No. 1 Dispenser COOLER (°C)	Temperature of chiller No. 1
	Chiller No. 2 Dispenser COOLER (°C)	Temperature of chiller No. 2
High compressor 1	GD-203 (%)	SoC of high compressor 1
	T.T1-1A (°C)	Temperature of input of the high compressor 1
	P.T1-1A (MPa)	Pressure of input of the high compressor 1
	T.T2-1A (°C)	Temperature of output of the high compressor 1
	P.T2-1A (MPa)	Pressure of output of the high compressor 1
	T.T1A (OIL) (°C)	Temperature of T.T1A (OIL)
High compressor 2	GD-204 (%)	SoC of high compressor 2
	T.T1-1B (°C)	Temperature of input of the high compressor 2
	P.T1-1B (MPa)	Pressure of input of the high compressor 2
	T.T2-1B (°C)	Temperature of output of the high compressor 2
	P.T2-1B (MPa)	Pressure of output of the high compressor 2
	T.T1B (OIL)	Temperature of T.T1B (OIL)
Mid/high storage bank	PT-202 (MPa)	Pressure of chiller to mid/high storage bank
	PT-204 (MPa)	Pressure of mid storage bank
	PT-203 (MPa)	Pressure of high storage bank
	PT-205 (MPa)	Pressure of high storage bank
	GD-205 (%)	SoC of storage bank
	GD-206 (%)	SoC of storage bank

Table 3. Statuses of data.

Time	GD-201	GD-202	PT-201	TEMP	Chiller No. 1 Dispenser COOLER	Chiller No. 2 Dispenser COOLER	GD-203	T.T1-1A	P.T1-1A	T.T2-1A	P.T2-1A	T.T1A(OIL)
10:36:18	0	0	11.51	21.0	5	−37	0	22	6.6	22	50.4	19
10:36:19	0	0	11.51	21.0	5	−37	0	22	6.6	22	50.4	19
10:36:20	0	0	11.47	21.0	5	−37	0	22	6.6	21	50.3	19
10:36:21	0	0	11.47	21.0	5	−37	0	22	6.6	21	50.3	19
10:36:22	0	0	11.47	21.0	5	−38	0	22	6.6	21	50.3	19
...		...				...				...		
10:48:56	0	0	11.49	21.1	5	−39	0	22	6.6	22	50.4	19
10:48:57	0	0	11.49	21.1	5	−39	0	22	6.6	22	50.4	19
10:48:58	0	0	11.51	21.1	5	−39	0	22	6.6	22	50.4	19
10:48:59	0	0	11.49	21.1	5	−39	0	22	6.6	22	50.4	19
10:49:00	0	0	11.51	21.1	5	−39	0	22	6.6	22	50.4	19
Time	**GD-204**	**T.T1-1B**	**P.T1-1B**	**T.T2-1B**	**P.T2-1B**	**T.T1B (OIL)**	**PT-202**	**PT-204**	**PT-203**	**PT-205**	**GD-205**	**GD-206**
10:36:18	3	21	8.4	28	52.1	31	44.66	44.75	86.10	86.53	0	0
10:36:19	3	21	8.4	28	52.1	31	44.66	44.75	86.10	86.53	0	0
10:36:20	3	21	8.4	27	52.1	31	44.59	44.69	86.06	86.40	0	0
10:36:21	3	21	8.4	27	52.1	31	44.59	44.69	86.06	86.40	0	0
10:36:22	3	21	8.4	27	52.1	31	44.59	44.69	86.06	86.40	0	0
...				...						...		
10:48:56	3	21	8.4	27	51.9	31	44.67	44.79	86.21	86.52	0	0
10:48:57	3	21	8.4	27	51.9	31	44.67	44.79	86.21	86.52	0	0
10:48:58	3	21	8.4	26	51.8	31	44.69	44.74	86.16	86.45	0	0
10:48:59	3	21	8.4	27	51.9	31	44.67	44.79	86.21	86.52	0	0
10:49:00	3	21	8.4	26	51.8	31	44.69	44.74	86.16	86.45	0	0

3.3. Overall System Operation Flowchart

Figure 5 illustrates the overall operational algorithm of the DT-based smart hydrogen refueling station safety management model. The hydrogen refueling data (CSV) are received from the hydrogen refueling station and stored in a database (this study does not cover the IoT infrastructure for real-time data collection and database storage). The stored data are then transmitted to the digital twin server and displayed as graphs to the users and administrators. Conversely, the accumulated data in the database are also used for data analysis and referred to as historical data. Historical data were used for AI-based training and model building, where attributes were extracted, and the data were cleansed and normalized to create the final AI model. If real-time data collected in the future are inputted into the created AI model, it will be possible to predict the potential for future hydrogen leaks and explosions. The AI model, trained based on past historical data, predicts whether current incoming data indicate a dangerous or safe situation. Based on the predicted values, the current risk situation is simulated using a digital twin screen. This simulation provides critical information for deciding whether to control hydrogen panels to cut off the hydrogen supply in the future.

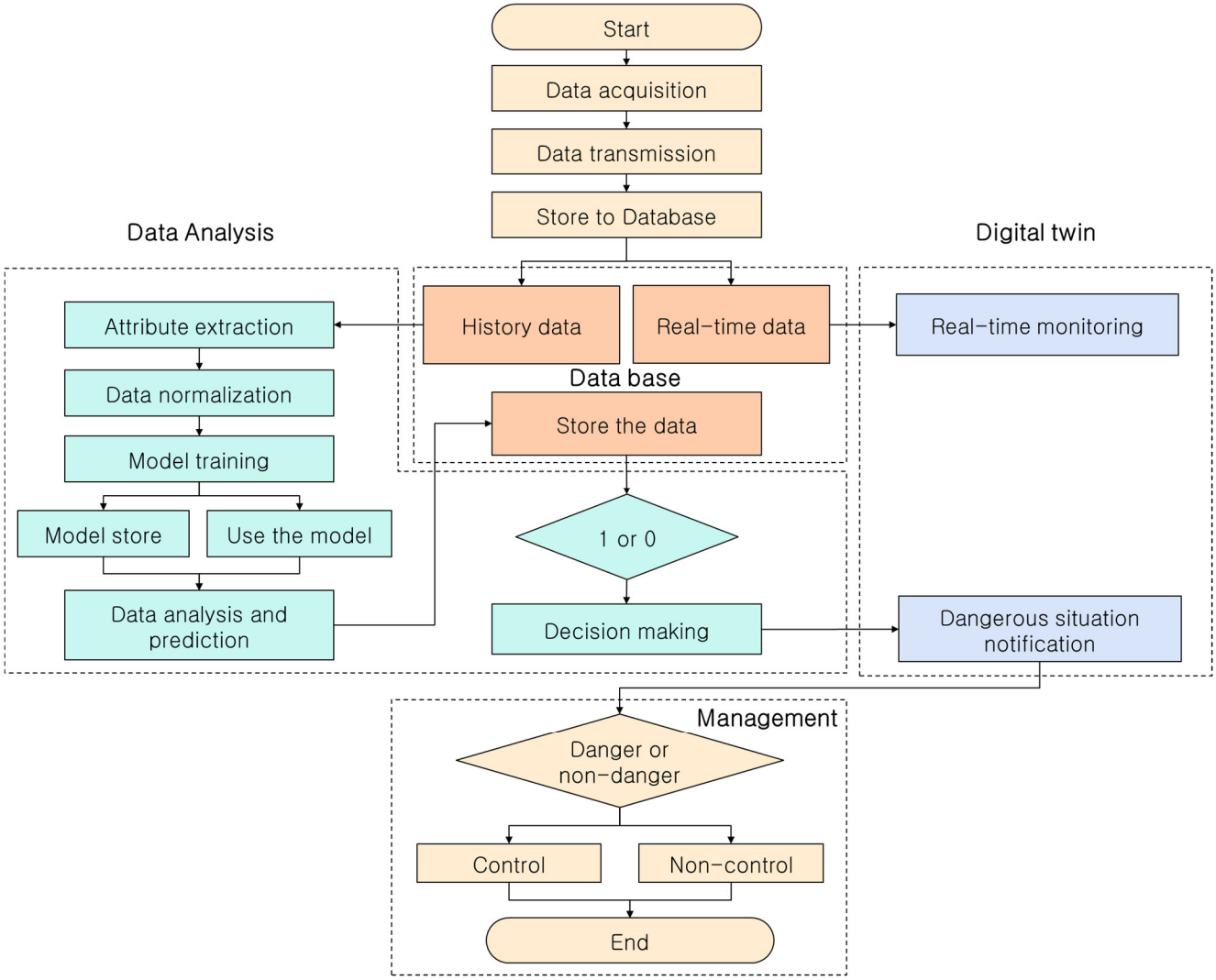


Figure 5. Overall operation algorithm of the digital twin-based smart hydrogen refueling station safety management model.

This study did not address IoT infrastructure for real-time data collection or an automated control system for managing panels in dangerous situations. The focus of this study

is to store the previously collected hydrogen refueling station data (24 attribute data) in CSV files, analyze these data to create a predictive model for future inference, and reflect the current situation in a digital twin model similar to the real environment.

4. Data Analysis Model

This section describes the data analysis process from data collection to training, model generation, and prediction. The data were implemented as a binary classification model using the 24 attributes presented above as input data with normal and abnormal states (hydrogen leaks) used as target data. The 24-attribute data are stored in a CSV file.

4.1. Data Collection

Figure 6 presents the data for each part of the hydrogen refueling station. Each data point was normalized using a standard scaler to make it visually identifiable from the original data. Upon examining the data, a periodic flow of data related to the overall processes occurring within the hydrogen refueling station was observed.

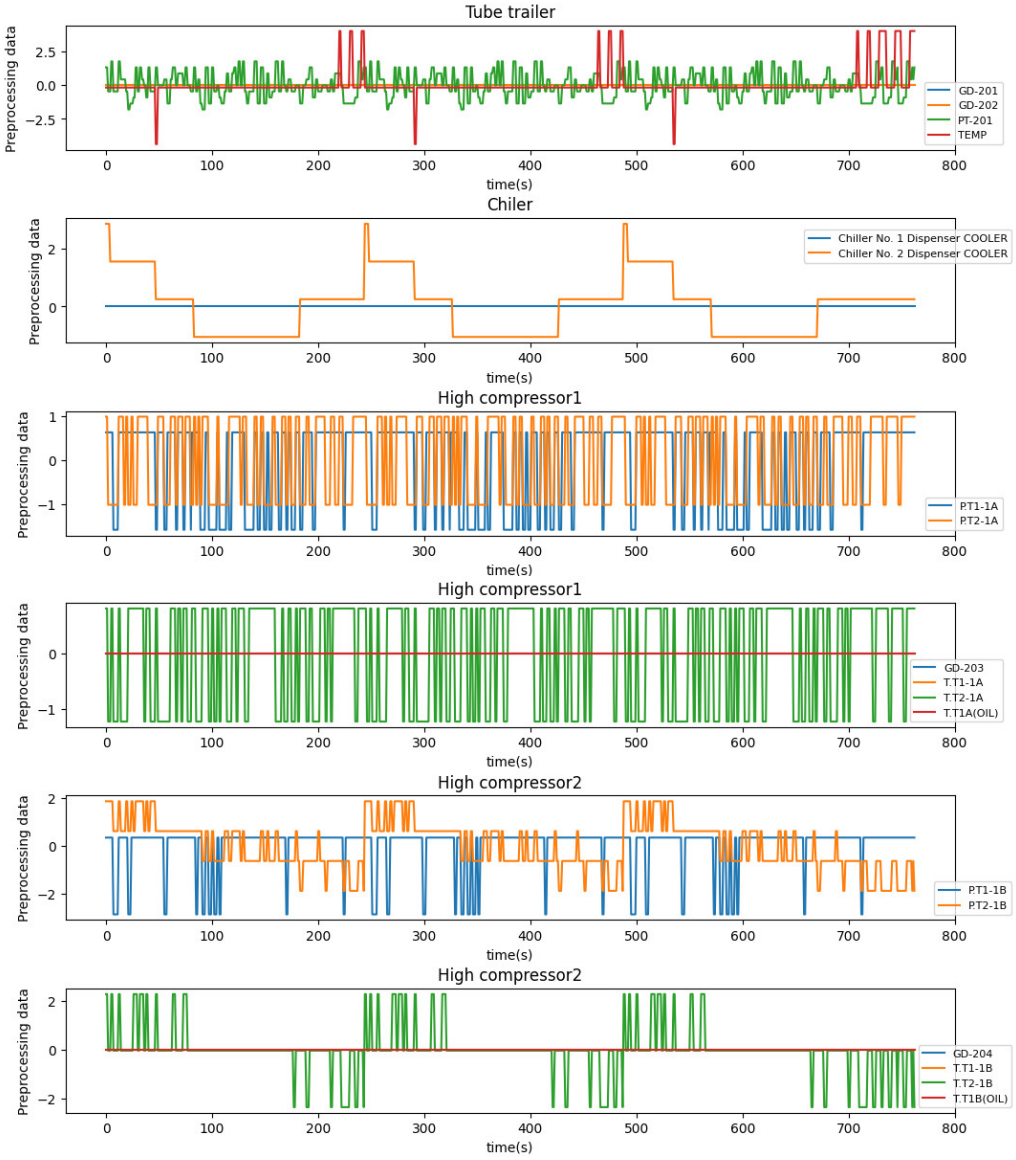


Figure 6. *Cont.*

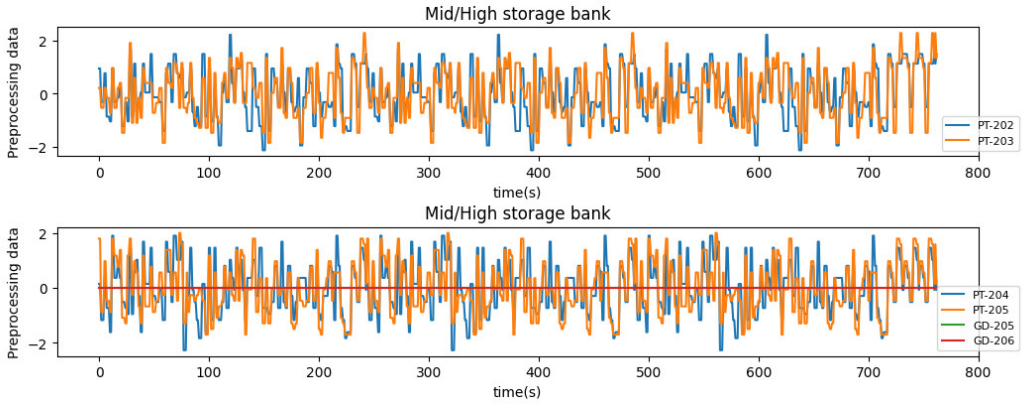


Figure 6. Data status for each component of the hydrogen refueling station.

The input data utilized the 24 types of hydrogen refueling station data mentioned previously, while the target data used both normal and abnormal states (hydrogen leaks). In other words, the target was set to 1 for normal conditions and 0 for abnormal conditions, and the status values of the data were transmitted to the digital twin server, triggering a 3D graphic-based alarm that was easy for administrators to identify visually.

However, data representing abnormal conditions for components such as compressors, chillers, and dispensers within hydrogen refueling stations have not been presented in the literature. Therefore, in this study, the abnormal condition data were arbitrarily created by referencing the range of hydrogen sensor values in abnormal or dangerous states as described in previous hydrogen sensor-related studies. The collection of data from hydrogen refueling stations under abnormal conditions must be addressed in the future. Various studies have proposed threshold values for detecting abnormal conditions in hydrogen sensors, and we simulated using virtual data based on the pressure range of compressed gas facilities presented in existing papers [40,41].

4.2. Data Preprocessing

The following concerns address missing data. The most basic strategy for handling missing data is to either delete the data or fill them with other data. In this study, missing data were handled by replacing them with immediately preceding data. Out of a total of 885 data points, one time point was missing. This was considered negligible, so instead of using interpolation, we filled the missing data with the immediately preceding value (1 s earlier).

Data normalization was performed using a standard scaler, a normalization method in Scikit-Learn, a popular machine learning library. The data normalization methods defined in Scikit-Learn are described below:

- MinMaxScaler: A normalization method that assigns values uniformly within the 0–1 range based on the minimum and maximum values of each feature.
- StandardScaler: A method that normalizes the data by setting the mean of each feature column to zero and a standard deviation of 1.

4.3. Data Analysis Algorithms

The following describes the artificial intelligence algorithms used for data analysis. In this study, nine AI algorithms (seven machine learning algorithms and two deep learning algorithms) were used.

4.3.1. Machine Learning Models

- **KNN (K-nearest neighbor):**

This is a flexible learning algorithm that does not learn any explicit model based on training data. It performs all computations at the time of prediction, making it usable without prior distributions or assumptions for specific data [42]. Depending on the nature

of the data, Euclidean, Manhattan, and Minkowski distances were used to identify the nearest K-neighbors [43]. Once the neighbor data are selected, predictions for a new data point are made based on the labels or values of the neighbors. The performance of KNN is heavily influenced by the value of K, the distance measurement method, and the data preprocessing method. A smaller K value may cause the model to overfit the training data, whereas a larger K value may overgeneralize the model and reduce its performance [44,45].

- **Logistic regression:**

A supervised learning algorithm primarily used for binary classification problems that focuses on modeling the relationship between independent and dependent variables and predicting the probability that a given input belongs to a specific class [46]. Teaching a logistic regression model involves determining the optimal regression coefficients for the given training data. As the model output is expressed as a probability, it can also provide confidence in the prediction results. However, as it is inherently a linear classifier, its performance may degrade if the classes are not linearly separable [47].

- **Decision tree classifier:**

This algorithm focuses on learning clear decision rules by splitting data. It is based on a tree structure in which each internal node represents a decision on a specific feature, each branch represents the result of that decision [48], and each leaf node represents a final class label [49]. The learning process involves recursively splitting the given data to build a clear tree, typically following a binary tree structure. At each node, a condition is set for a specific feature, and the data are divided into two subsets according to that condition. In a decision tree, the selection criteria for splitting are crucial concepts that determine how the data are divided, with the aim of growing the tree in a manner that minimizes impurities at each split [50,51].

- **Random forest classifier:**

As an ensemble learning technique in machine learning, random forest combines several decision trees to create a stronger and more generalized classification model. It is based on the bagging (bootstrap aggregating) technique [52,53], which effectively prevents the overfitting of individual decision trees and improves prediction performance. Random forest considers only a randomly selected subset of features to identify the optimal split at each node. Since each tree learns on different data samples and feature sets, it demonstrates robust performance on high-dimensional or complex data patterns. It is also relatively robust in the context of data noise and outliers and can naturally extend to multiclass classification problems [54].

- **Extra trees classifier:**

Similar to random forest, extra trees is an ensemble learning technique that combines multiple decision trees to create a more accurate and robust classification model. Unlike random forest, extra trees does not search for the optimal split criteria but selects the split points randomly [55]. The features used at each node and the split values for these features are selected randomly, resulting in a higher level of randomness [56], which reduces the correlation between individual trees and lowers the variance of the model. Additionally, extra trees does not use bootstrap sampling but instead uses the entire training dataset to train each tree, helping to reduce model bias and shorten computation time [57].

- **Gradient boosting classifier:**

This ensemble learning method combines multiple weak learners to create a strong predictive model. It is a boosting technique that, unlike random forest, sequentially learns to gradually reduce prediction errors using an optimization algorithm called gradient descent to minimize the loss function [58,59]. Although gradient boosting models can evaluate the importance of the features used during the learning process, they can be time-consuming to train, as trees are learned sequentially, and the computational cost can be very high for large datasets or high-dimensional data.

- **Hist-gradient boosting classifier:**

A variation of gradient boosting that maximizes the efficiency of data preprocessing and learning while maintaining high-performance characteristics. It is suitable for large-scale or high-dimensional data and is an ensemble learning technique that sequentially trains multiple weak learners, typically decision trees, to complement the prediction errors of previous trees [60]. The goal of gradient boosting is to gradually improve model performance by adding new trees at each stage to minimize the loss function. Hist-gradient boosting enhances performance by partitioning the data into histograms and training trees based on each histogram bin [61].

4.3.2. Deep Learning Models

- **DNN (Deep neural network):**

DNN is an artificial neural network (ANN) used to model complex nonlinear relationships through deep structures with multiple hidden layers. It demonstrates strong performance on large datasets and complex learning problems and is widely used in various AI and machine learning applications such as image recognition and natural language processing [62,63]. A DNN consists of a structure with multiple layers of neurons or nodes, where each layer receives output values from the previous layer, applies weights and biases, and passes them to the next layer. The learning process is based on backpropagation and gradient descent optimization algorithms [36] that adjust the network weights and biases to minimize the difference between the predicted and actual values [64,65]. Due to its high computational cost and risk of overfitting, it is a widely used deep learning model that can leverage large datasets and has a flexible structure.

- **CNN (Convolutional neural network):**

CNN is an artificial neural network that excels at processing 2D data such as images and videos. It is inspired by the structure of the biological visual cortex and effectively learns the spatial information and patterns of image data. A CNN consists of input, convolutional, activation function, pooling, and fully connected layers. It utilizes the spatial hierarchy of image data to extract important features and perform tasks such as classification. Similar to a DNN, the learning process involves optimizing weights using backpropagation and gradient descent algorithms. Due to its local pattern learning based on convolution and pooling operations, it possess strong features against small translations or distortions; however, owing to the high computational cost of handling convolution operations and large maps, there is a risk of overfitting the training data if sufficient data are not provided.

4.3.3. Model Parameters

The following Table 4 shows the model hyperparameters.

Table 4. Hyperparameters of models.

Hyperparameters	**Model**	**Description**
	K-Neighbors	
n_neighbors	5	Number of neighbors to consider
metric	minkowski	The method for measuring the distance between neighbors
Hyperparameters	**Model**	**Description**
	Logistic Regression	
Penalty	L2	L2 regularization (Ridge) is used when applying regularization

Table 4. *Cont.*

c		1.0		Parameter to determine the strength of regularization
solver		lbfgs		Optimization algorithm
max_iter		100		Maximum number of iterations for the optimization algorithm
Hyperparameters	**Model**			**Description**
	Decision Tree	**Random Forest**	**Extra Trees**	
max_depth	None	None	None	Maximum depth of the tree
min_samples_split	2	2	2	Minimum number of samples required to split a node
min_samples_leaf	1	1	1	Minimum number of samples required to be at a leaf node
Hyperparameters	**Model**			**Description**
	Gradient Boosting	**Hist Gradient Boosting**		
loss	log_loss	log_loss		Defines the loss function to use
learning_rate	0.1	0.1		Rate at which each tree's contribution is reduced
max_iter	100	100		Maximum number of trees to learn
max_leaf_nodes	1.0	31		Maximum number of leaf nodes that can be created for each tree
Hyperparameters	**Model**			**Description**
	DNN	**CNN**		
Total params	144,385	1,070,501		Total number of parameters
Activation	sigmoid	sigmoid		Defines the activation function
Optimizer	adam	adam		Defines the optimization algorithm
loss	binary crossentropy	binary crossentropy		Defines the loss function for the binary classification model
epochs	50	30		Defines the maximum number of epochs
batch_size	24	32		Defines the batch size for training

4.4. Model Training and Prediction

Figure 7 presents the structure of the input and target data.

The data were divided into input and target data (Table 5). The input data consist of attribute data, whereas the target data are binary (1, 0) data indicating whether hydrogen has leaked. The dataset consisted of 885 data samples with a data collection interval of 1 s. There were 24 data attributes, including temperature, pressure, and SoC. The data were split into training and test datasets (80% and 20%, respectively). The nine algorithms presented above were used to train the model on the training dataset and to evaluate the model on the test dataset. In this study, the evaluation was conducted using a training dataset, test dataset evaluation, and 5-fold cross-validation. The next section explains the evaluation methods used in the proposed model.

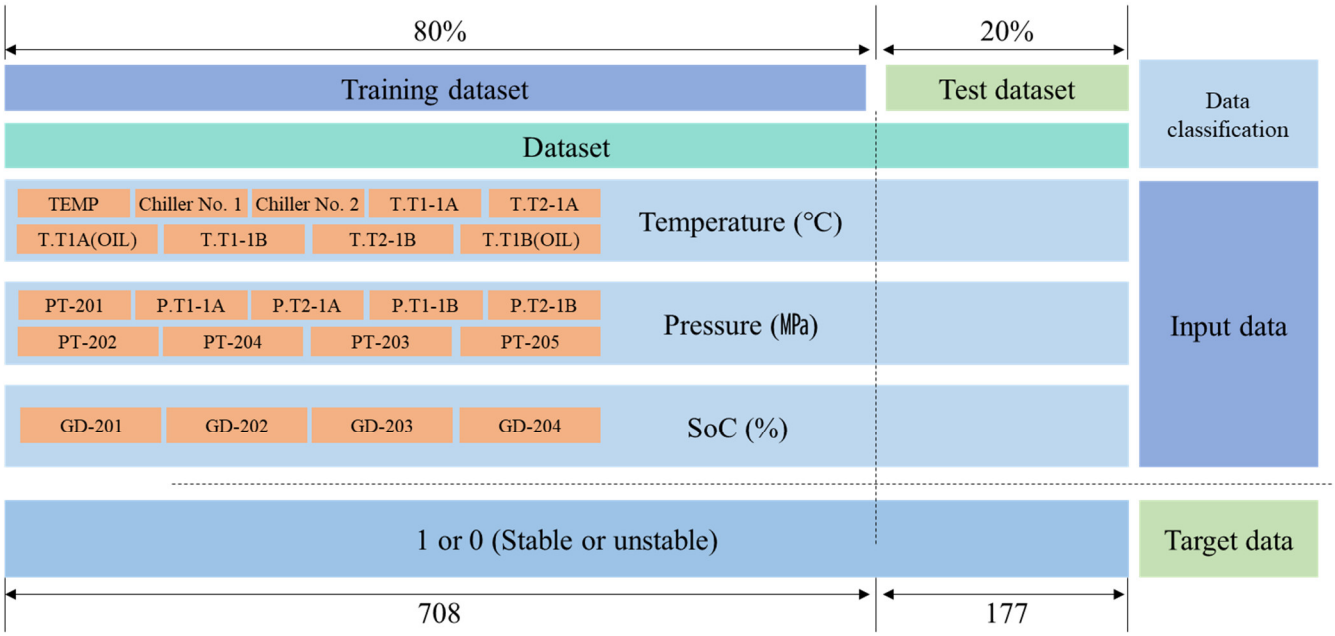


Figure 7. Structure of the input and target data.

Table 5. Data status.

Classification	Number
Number of data features	24
Number of data target	1
Number of samples	885
Training dataset	708
Test dataset	177

4.5. Evaluation

List of evaluation methods:

The machine learning models were validated based on a 5-fold cross-validation, and a 5-fold cross-validation was used to evaluate the performance of the model. The data were divided into five distinct parts, with one part used as the validation data and the remaining four parts used for training. This process was repeated five times so that each part was used once as validation data, and the final result was obtained by averaging the five evaluation results. This method allows a more accurate assessment of a model's generalization performance, prevents overfitting, and increases the reliability of the model.

Deep learning models employ techniques such as model checkpoints and early stopping. A model checkpoint is a technique by which the state of a machine learning/deep learning model is saved when certain conditions are met during the training process. This method records the best-performing model (e.g., with minimum validation loss) during training, thereby allowing easy restoration of the optimal model after training ends. This prevents the use of an overfitted model and ensures that the training progress is not lost in the case of an interruption or restart. Early stopping stops training when there is no performance improvement in a certain number of epochs. Typically, it monitors the performance of the validation data (e.g., validation loss or accuracy), and if the performance does not improve or begins to degrade, the training is stopped. This technique helps to prevent overfitting and avoids unnecessary training while maintaining an efficient training process.

The following Figure 8 and Table 6 present the performance evaluation results for each model. Each item represents the results for the training set, test set, and cross-validation. 'Accuracy (1)' is the ratio of the number of correct predictions to the total number of predictions. 'Precision (2)' measures how many of the predicted positive instances are

actually positive. 'Recall (3)' measures how many of the actual positive instances were correctly identified. 'F1-Score (4)' is the harmonic mean of Precision and Recall, providing a single metric that balances both.

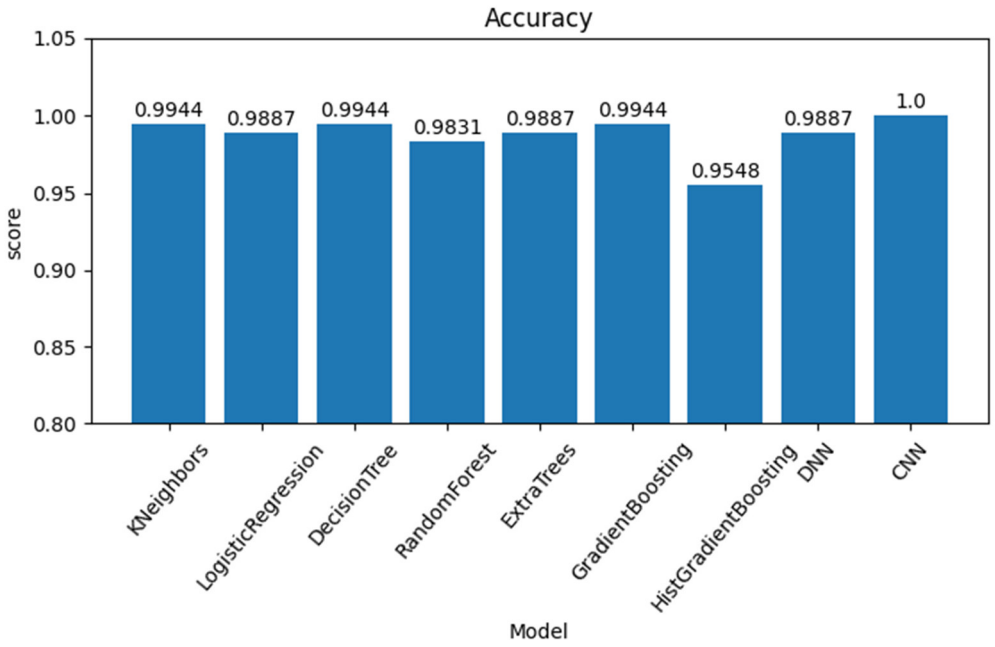


Figure 8. Accuracy for each model.

Table 6. Performance evaluation results for each model.

Model	5-Fold Cross-Validation	Accuracy	Precision	Recall	F1-Score
K-Neighbors	0.996	0.994	0.993	1.000	0.997
Logistic Regression	0.996	0.989	0.987	1.000	0.993
Decision Tree	0.996	0.994	0.993	1.000	0.997
Random Forest	0.984	0.983	0.980	1.000	0.990
Extra Trees	0.996	0.989	0.987	1.000	0.993
Gradient Boosting	0.999	0.994	0.993	1.000	0.997
Hist Gradient Boosting	0.954	0.955	0.949	1.000	0.974
DNN	-	0.989	0.987	1.000	0.993
CNN	-	1.000	0.987	1.000	0.993

It is calculated as:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Prediction} \quad (1)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$F1 = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

The evaluation results showed that the CNN model had the highest performance. Due to the lack of data, it was observed that the accuracy of most binary classification models was generally high, and the model that produced the best results was the CNN model.

The following Figure 9 and Table 7 show the hyperparameters of the final selected CNN model. Since most of the data collected from the hydrogen refueling station is

1-dimensional time-series data, a convolutional neural network based on 1D was used. The model consists of two convolutional layers and two max-pooling layers. After flattening the data into a 1D format through a flatten layer, it is fed into the input layer of the feedforward network. To address the vanishing gradient problem, the relu function was used as the activation function for each convolutional layer, and the sigmoid function was used in the output layer of the feedforward network for binary classification results.

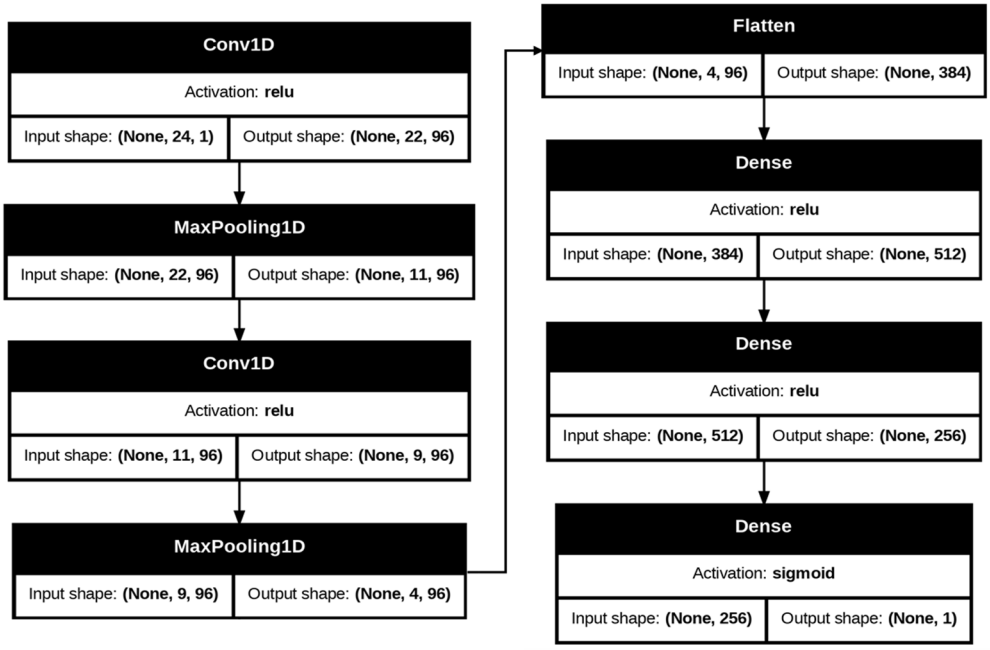


Figure 9. CNN model layers.

Table 7. CNN model layers and parameters.

Layer	Output Shape	Activation	Number of Parameters
Conv1D	(22, 96)	relu	384
Max pooling1D	(11, 96)	-	0
Conv1D	(9, 96)	relu	27,744
Max pooling1D	(4, 96)	-	0
Flatten	384	-	0
Dense1	512	relu	197,120
Dense2	256	relu	131,328
Dense3	1	sigmoid	257

5. Implementation: Decision-Making-Based Digital Twin Smart Hydrogen Refueling Station Safety Management Model Implementation

5.1. IoT Infrastructure

The system is configured as presented in Figure 9. The "Real World" represents an actual hydrogen refueling station, where the IoT infrastructure is installed to collect and control data. It consists of sensors and actuators; sensors are responsible for data collection, and actuators control the equipment by turning it on or off. This study does not cover IoT infrastructure such as sensors and actuators; this will be supplemented in the future. Instead, it analyzes data using previously collected CSV files.

5.2. Communication Protocol

The data hub stores and manages the collected data. The database (MySQL) stores the hydrogen refueling station data (CSV). The DBMS (Database Management System) manages the data to be input and output from the database. The data stored in the database

are transmitted to digital twins and machine learning/deep learning servers via socket communication (TCP).

5.3. Machine Learning/Deep Learning Model Training, Storage, and Prediction

The machine learning/deep learning server extracts and normalizes the data and then trains it to create a model. The model was created using nine machine learning/deep learning algorithms as presented earlier, and the model with the highest performance was saved after validation. In this study, the CNN model with the best performance was used. Based on the saved model, incoming real-time data are predicted using binary classification to determine whether the current situation is dangerous, and the result is sent to a digital twin server and stored in a database.

5.4. Unity-Based Digital Twin 3D Modeling

Digital twin technology is a digital replica of a physical system that enables efficient safety management through real-time monitoring and simulations. The digital twin server was configured based on the Unity 3D program, displaying the hydrogen refueling station data transmitted in real time on a virtual 3D model identical to the actual hydrogen refueling station. It also displays the prediction data (1 or 0 binary data, danger/nondanger data) sent from the machine learning/deep learning server. Users or administrators visually inspect the digital twin 3D model to intuitively monitor dangerous situations.

5.5. Digital Twin Simulation

Figures 10–14 present the flowchart and implementation photos of the entire system. The 24 sensor data stored in the CSV file are transmitted from the client to the server based on socket communication (TCP) and stored in the database (MySQL). The stored data can be monitored in real time, as presented in Figure 6, and the data in the database are sent to the ML/DL server for model training. The ML/DL server trains the model based on the data and saves the optimal model. Model training is conducted periodically in a loop, and the model is retrained based on the latest data collected so far (training takes approximately 35 s). The saved model makes predictions using incoming real-time data (there is no incoming real-time data in this study as IoT infrastructure is not implemented; this will be supplemented in the future). The prediction time is approximately 50–60 ms, and predictions are made every 0.5–1 s in a loop. A shorter cycle allows for a faster alert for dangerous situations. With a prediction time of approximately 55 ms, predictions can be made at a minimum cycle of 100 ms. The predicted data are binary (0 or 1), and depending on the result, it is immediately sent to the digital twin (Unity) server to display the situation.

In Unity, the particle system is turned on/off based on the predicted data received from the ML/DL server. A particle system is a 3D graph that effectively represents fire and flame situations. By turning this particle system on or off, the digital twin screen can indicate whether the current situation is dangerous. Table 8 shows the training time, prediction time, and data transmission intervals.

Table 8. Training/prediction time, and data transmission intervals.

Classification	Time
Training time (s)	35
Forecasting time (ms)	68
Data transmission interval (s)	1

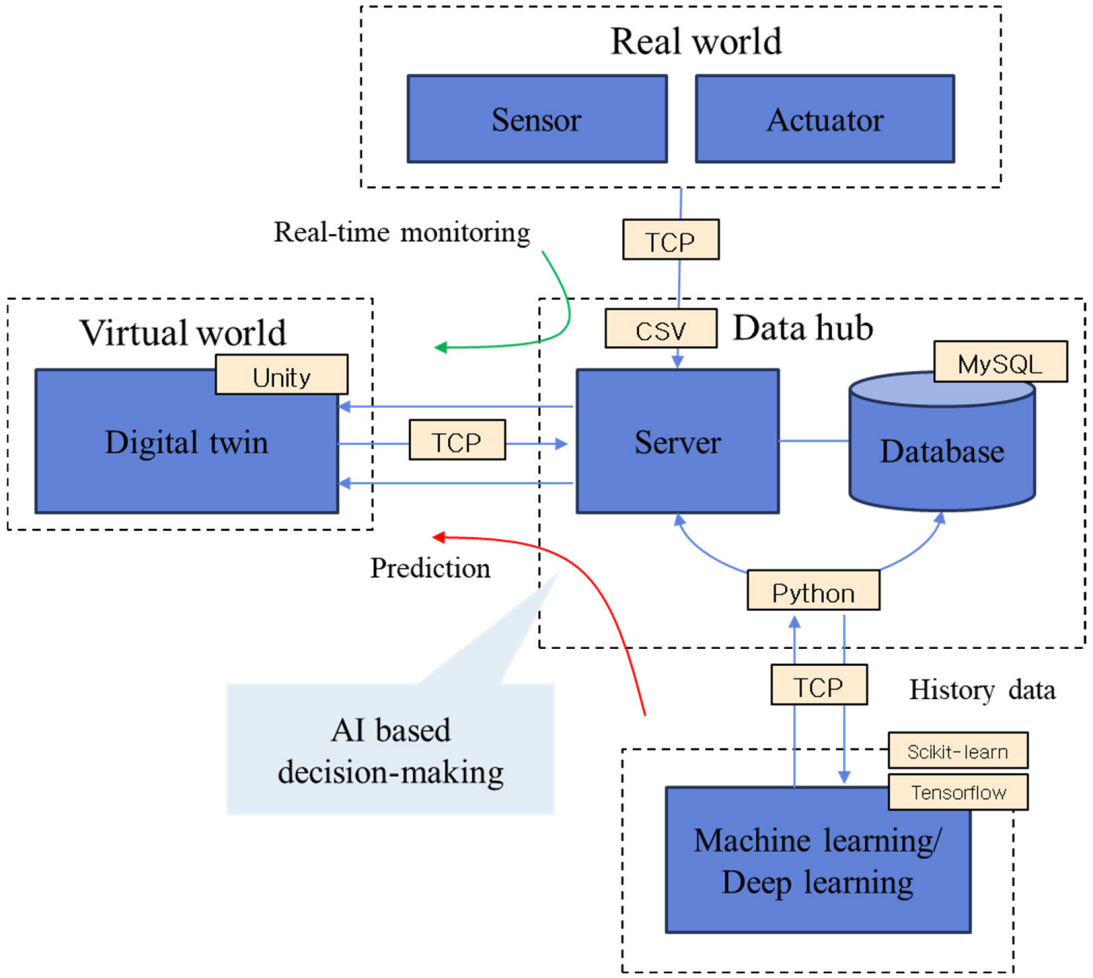


Figure 10. System configuration.

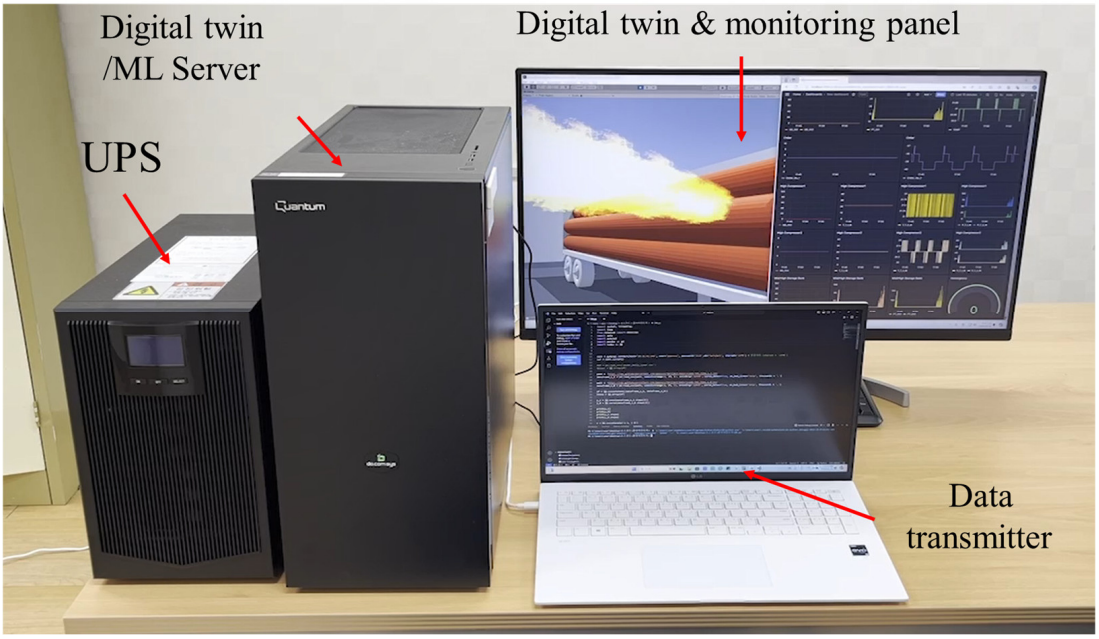


Figure 11. Prototype of risk prediction of hydrogen refueling station.

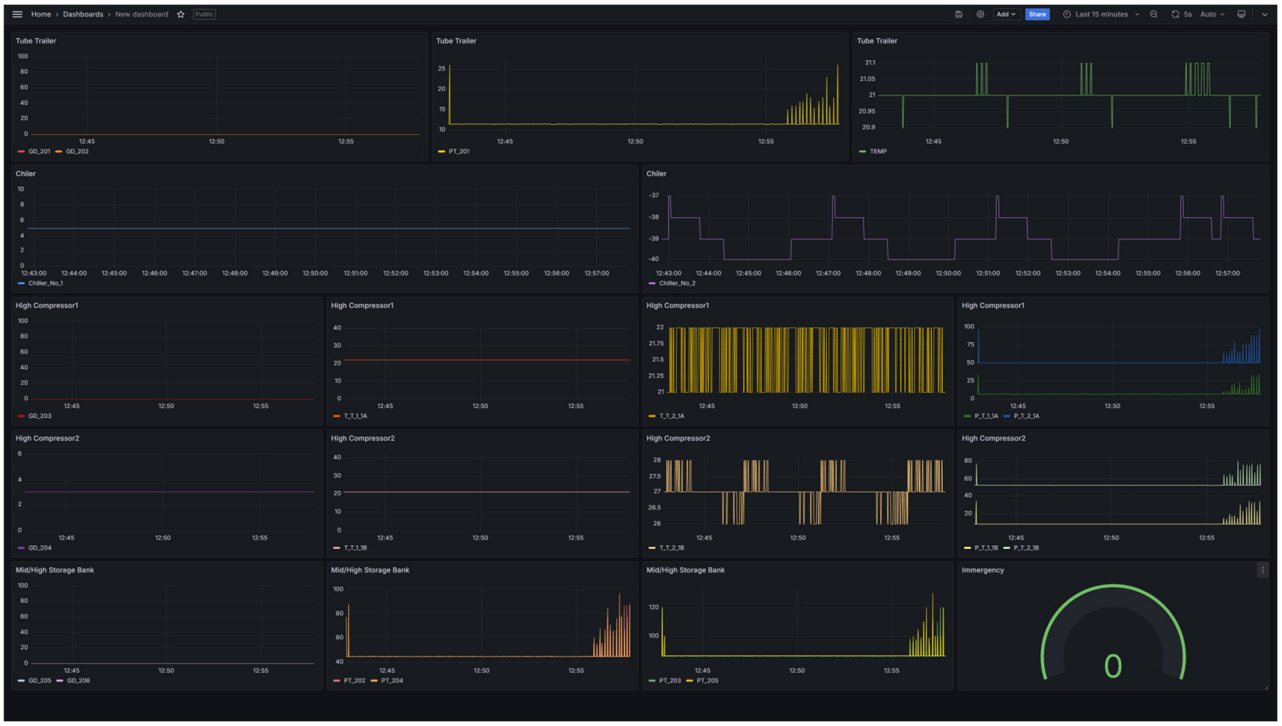
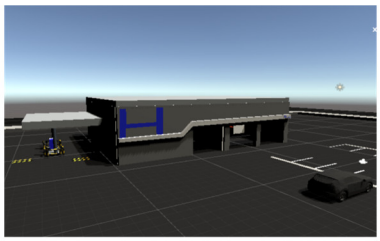


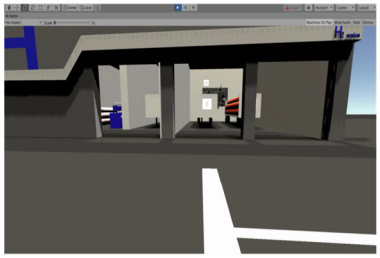
Figure 12. Prototype of monitoring panel.



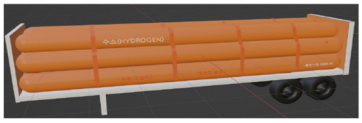
(a) Hydrogen charging station



(b) Hydrogen charging station digital twin



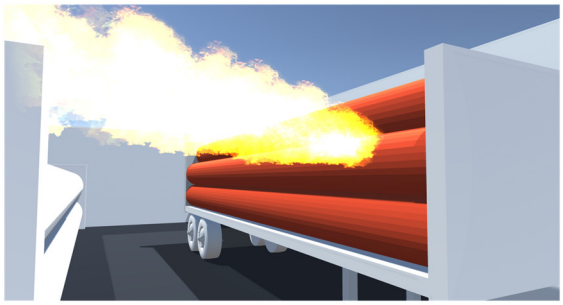
(c) Front of the hydrogen charging station digital-twin



Tube trailer digital-twin



(d) High compressor and panel digital-twin



(e) Risk prediction of hydrogen charging station (hazardous state)

Figure 13. System implementation and prototype.

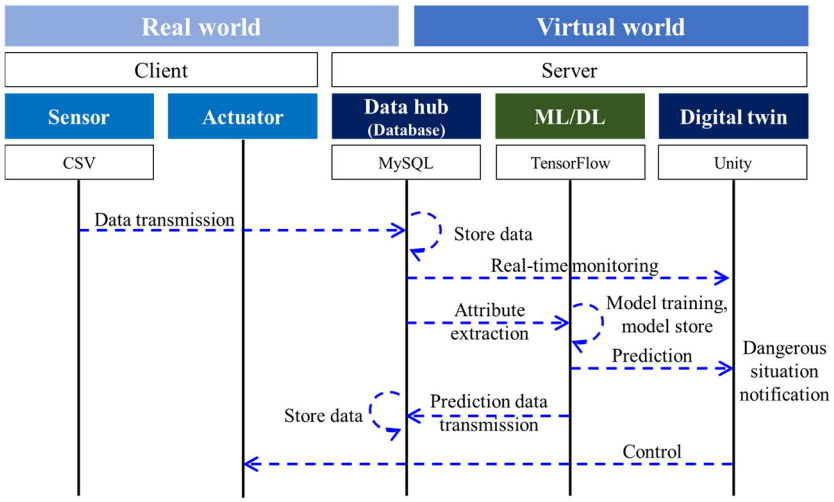


Figure 14. Flowchart of the system.

6. Conclusions

In this study, we proposed a safety model for hydrogen refueling stations utilizing digital twin technology and presented a more intuitive safety management approach by integrating AI-based decision-making and 3D simulations. By building a digital twin platform capable of monitoring and analyzing various potential hazards during the operation of hydrogen refueling stations, we expect to significantly enhance the speed and accuracy of decision-making for accident prevention and emergency response.

In particular, the AI-based decision-making model plays a crucial role in preemptively identifying potential risks through predictive analysis and suggesting appropriate countermeasures. Additionally, the 3D situational awareness technology allows for real-time identification of hazardous situations in a visual and intuitive manner, enabling more efficient and effective safety management.

The results of this study demonstrate the potential application of this model not only to hydrogen refueling stations but also to the safety management systems of various high-risk industrial facilities. Digital twin-based risk prediction and automatic control will become a significant topic in the field of 'safety'. Beyond the 'hydrogen' sector, it appears that applying this system to areas such as building fires, wildfires, earthquakes, floods, and landslides could have a major impact [66–68].

This research suggests the future direction for the development of safety management systems that combine digital twin and AI-based technologies. Future studies should focus on enhancing the model's accuracy by incorporating various environmental conditions and real-world operational data, as well as validating its effectiveness through field-based case studies. In particular, this study did not involve the installation of IoT infrastructure within the hydrogen refueling station, so real-time data could not be collected. Instead, data analysis was performed using previously stored CSV files. Additionally, actual data for abnormal situations or hazardous events within the hydrogen refueling station were not collected. Addressing the acquisition of such data is a task that must be undertaken in future research. Furthermore, in the implementation of the binary classification-based situational assessment model, future research should focus on classification methods that adapt to various environments and sensors, rather than relying solely on binary classification.

Author Contributions: Conceptualization, J.H.Y.; Investigation, N.Y.A. and E.S.; Supervision, S.P.; Project administration, S.-H.H. and H.-G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2022RIS-005) and by the 2023 Research Grant from Kangwon National University.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dawood, F.; Anda, M.; Shafiullah, G.M. Hydrogen production for energy: An overview. *Int. J. Hydrogen Energy* **2020**, *45*, 3847–3869. [CrossRef]
2. Züttel, A.J.N. Hydrogen storage methods. *Sci. Nat.* **2004**, *91*, 157–172. [CrossRef] [PubMed]
3. Graetz, J. New approaches to hydrogen storage. *Chem. Soc. Rev.* **2009**, *38*, 73–82. [CrossRef]
4. Usman, M.R.; Reviews, S.E. Hydrogen storage methods: Review and current status. *Renew. Sustain. Energy Rev.* **2022**, *167*, 112743. [CrossRef]
5. Brown, T.; Stephens-Romero, S.; Samuelsen, G.S. Quantitative analysis of a successful public hydrogen station. *Int. J. Hydrogen Energy* **2012**, *37*, 12731–12740. [CrossRef]
6. Lin, R.-H.; Ye, Z.-Z.; Wu, B.-D. A review of hydrogen station location models. *Int. J. Hydrogen Energy* **2020**, *45*, 20176–20183. [CrossRef]
7. An, S.; Oh, S.; Kim, E.; Lee, J. Optimization of Designing Barrier to Mitigate Hazardous Area in Hydrogen Refueling Stations. *J. Hydrogen New Energy* **2023**, *34*, 734–740. [CrossRef]
8. Kim, H.; Kang, S. Analysis of damage range and impact of on-site hydrogen fueling station using quantitative risk assessment program (Hy-KoRAM). *Trans. Korean Hydrogen New Energy Soc.* **2020**, *31*, 459–466. [CrossRef]
9. Gerard, B.; Carrera, E.; Bernard, O.; Lun, D. Smart design of green hydrogen facilities: A digital twin-driven approach. In Proceedings of the E3S Web of Conferences, Hyderabad, India, 15–17 December 2021; EDP Sciences: Les Ulis, France, 2022; p. 02001.
10. Jaribion, A.; Khajavi, S.H.; Öhman, M.; Knapen, A.; Holmström, J. A digital twin for safety and risk management: A prototype for a hydrogen high-pressure vessel. In Proceedings of the 15th International Conference on Design Science Research in Information Systems and Technology, Kristiansand, Norway, 2–4 December 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 369–375.
11. Kang, S.-K. A study of jet dispersion and jet-fire characteristics for safety distance of the hydrogen refueling station. *J. Korean Inst. Gas* **2019**, *23*, 74–80. [CrossRef]
12. Kim, H.J.; Jang, K.M.; Kim, S.H.; Kim, G.B.; Jung, E.S. A Study on Safety Guidelines for Hydrogen Refueling Stations at Expressway Service Area using Quantitative Risk Assessment. *Trans. Korean Hydrogen New Energy Soc.* **2021**, *32*, 551–564. [CrossRef]
13. Zhou, J.; Zhang, J.; Zhang, J.; Yi, F.; Wang, X.; Sun, Y.; Zhang, C.; Hu, D.; Wu, G. Hydrogen leakage source positioning method in deep belief network based on fully confined space Gaussian distribution model. *Int. J. Hydrogen Energy* **2024**, *63*, 435–445. [CrossRef]
14. Genovese, M.; Blekhman, D.; Fragiacomo, P.J.H. An exploration of safety measures in hydrogen refueling stations: Delving into hydrogen equipment and technical performance. *Hydrogen* **2024**, *5*, 102–122. [CrossRef]
15. Hydrogen Charging Station, Real-Time Safety Management 24 Hours a Day. Available online: https://www.motie.go.kr/kor/article/ATCL3f49a5a8c/165645/view (accessed on 27 August 2024).
16. Guido, G.; Haghshenas, S.S.; Haghshenas, S.S.; Vitale, A.; Gallelli, V.; Astarita, V.J.S. Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm. *Sustainability* **2020**, *12*, 6735. [CrossRef]
17. Eom, J.J. An Architecture of a Smart Safety Management System to prevent safety Accidents in Workplace. *J. Digit. Contents Soc.* **2020**, *21*, 817–823. [CrossRef]
18. Jia, C.; Zhou, J.; He, H.; Li, J.; Wei, Z.; Li, K. Health-conscious deep reinforcement learning energy management for fuel cell buses integrating environmental and look-ahead road information. *Energy* **2024**, *290*, 130146. [CrossRef]
19. He, X.; Kong, D.; Yang, G.; Yu, X.; Wang, G.; Peng, R.; Zhang, Y.; Dai, X. Hybrid neural network-based surrogate model for fast prediction of hydrogen leak consequences in hydrogen refueling station. *Int. J. Hydrogen Energy* **2024**, *59*, 187–198. [CrossRef]
20. He, X.; Kong, D.; Yu, X.; Ping, P.; Wang, G.; Peng, R.; Zhang, Y.; Dai, X. Prediction model for the evolution of hydrogen concentration under leakage in hydrogen refueling station using deep neural networks. *Int. J. Hydrogen Energy* **2024**, *51*, 702–712. [CrossRef]
21. Kim Tony, S. The Development of an Intelligent Risk Recognition System for Construction Safety by Combining Artificial Intelligence and Digital Twin Technology. *J. Korea Inst. Build. Constr.* **2023**, *23*, 405–406.
22. Ye, Z.; Ye, Y.; Zhang, C.; Zhang, Z.; Li, W.; Wang, X.; Wang, L.; Wang, L. digital twin approach for tunnel construction safety early warning and management. *Comput. Ind.* **2023**, *144*, 103783. [CrossRef]
23. Liu, F.; Panagiotakos, D. Real-world data: A brief review of the methods, applications, challenges and opportunities. *BMC Med. Res. Methodol.* **2022**, *22*, 287. [CrossRef]
24. Kusters, R.; Misevic, D.; Berry, H.; Cully, A.; Le Cunff, Y.; Dandoy, L.; Díaz-Rodríguez, N.; Ficher, M.; Grizou, J.; Othmani, A.; et al. Interdisciplinary research in artificial intelligence: Challenges and opportunities. *Front. Big Data* **2020**, *3*, 577974. [CrossRef] [PubMed]
25. Parmar, J.; Chouhan, S.; Raychoudhury, V.; Rathore, S. Open-world machine learning: Applications, challenges, and opportunities. *ACM Comput. Surv.* **2023**, *55*, 1–37. [CrossRef]

26. Lastowka, F.G.; Hunter, D. The laws of the virtual worlds. In *Popular Culture and Law*; Routledge: London, UK, 2017; pp. 363–435. [CrossRef]
27. Gaidon, A.; Wang, Q.; Cabon, Y.; Vig, E. Virtual worlds as proxy for multi-object tracking analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June 2016; pp. 4340–4349.
28. Liaw, S.Y.; Carpio, G.A.C.; Lau, Y.; Tan, S.C.; Lim, W.S.; Goh, P.S. Multiuser virtual worlds in healthcare education: A systematic review. *Nurse Educ. Today* **2018**, *65*, 136–149. [CrossRef] [PubMed]
29. Alguliyev, R.; Imamverdiyev, Y.; Sukhostat, L. Cyber-physical systems and their security issues. *Comput. Ind.* **2018**, *100*, 212–223. [CrossRef]
30. Shim, S.; Kim, J.Y.; Hwang, S.W.; Oh, J.M.; Kim, B.K.; Park, J.H.; Hyun, D.J.; Lee, H. Computing. A Comprehensive Review of Cyber-physical System (CPS)-based Approaches to Robot Services. *IEIE Trans. Smart Process. Comput.* **2024**, *13*, 69–80. [CrossRef]
31. Tao, F.; Qi, Q.; Wang, L.; Nee, A.J.E. Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison. *Engineering* **2019**, *5*, 653–661. [CrossRef]
32. Yao, X.; Zhou, J.; Lin, Y.; Li, Y.; Yu, H.; Liu, Y. Smart manufacturing based on cyber-physical systems and beyond. *J. Intell. Manuf.* **2019**, *30*, 2805–2817. [CrossRef]
33. Javid, S.S.M.; Derakhshan, G.; mehdi Hakimi, S. Energy scheduling in a smart energy hub system with hydrogen storage systems and electrical demand management. *J. Build. Eng.* **2023**, *80*, 108129. [CrossRef]
34. Abohamzeh, E.; Salehi, F.; Sheikholeslami, M.; Abbassi, R.; Khan, F. Review of hydrogen safety during storage, transmission, and applications processes. *J. Loss Prev. Process Ind.* **2021**, *72*, 104569. [CrossRef]
35. Calabrese, M.; Portarapillo, M.; Di Nardo, A.; Venezia, V.; Turco, M.; Luciani, G.; Di Benedetto, A.J.E. Hydrogen safety challenges: A comprehensive review on production, storage, transport, utilization, and CFD-based consequence and risk assessment. *Energies* **2024**, *17*, 1350. [CrossRef]
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June 2016; pp. 770–778.
37. Park, K. Towards intelligent agents to assist in modular construction: Evaluation of datasets generated in virtual environments for AI training. In Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC), Dubai, United Arab Emirates, 2–4 November 2021. [CrossRef]
38. Ignjatović, D.; Bailey, D.W.; Bajić, L. The wormhole ai training processor. In Proceedings of the 2022 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 20–26 February 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 356–358.
39. Runkler, T.A. *Data Analytics*; Springer: Berlin/Heidelberg, Germany, 2020.
40. Bong, S.; Oh, S.C.P.; Joong, S.; Jang, J. A Study on Inference Methods for Functional Safety of Hydrogen Refueling Stations Through Analysis of Similar Device Health Data. *J. Appl. Reliab.* **2021**, *21*, 367–380.
41. Bong, S.; Oh, D.S.S.; Yong, H.; Lee, Y.; Joong, J.S. Development of Technology for CBM+ Data Acquisition of Hydrogen Refueling Station Compressor. *J. Appl. Reliab.* **2023**, *23*, 51–72.
42. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1774–1785. [CrossRef] [PubMed]
43. Saadatfar, H.; Khosravi, S.; Joloudari, J.H.; Mosavi, A.; Shamshirband, S.J.M. A new K-nearest neighbors classifier for big data based on efficient data pruning. *Mathematics* **2020**, *8*, 286. [CrossRef]
44. Syaliman, K.; Nababan, E.; Sitompul, O. Improving the accuracy of k-nearest neighbor using local mean based and distance weight. In *Journal of Physics: Conference Series, Proceedings of the 2nd International Conference on Computing and Applied Informatics, Medan, Indonesia, 28–30 November 2017*; IOP Publishing: Bristol, UK, 2018; p. 012047.
45. Yu, Z.; Chen, H.; Liu, J.; You, J.; Leung, H.; Han, G. Hybrid $ k $-nearest neighbor classifier. *IEEE Trans. Cybern.* **2015**, *46*, 1263–1275. [CrossRef]
46. Harrell, J.; Frank, E.; Harrell, F.E. Ordinal logistic regression. In *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 311–325. [CrossRef]
47. Zabor, E.C.; Reddy, C.A.; Tendulkar, R.D.; Patil, S. Logistic regression in clinical studies. *Int. J. Radiat. Oncol. Biol. Phys.* **2022**, *112*, 271–277. [CrossRef]
48. Parvin, H.; MirnabiBaboli, M.; Alinejad-Rokny, H. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Eng. Appl. Artif. Intell.* **2015**, *37*, 34–42. [CrossRef]
49. Yoo, S.H.; Geng, H.; Chiu, T.L.; Yu, S.K.; Cho, D.C.; Heo, J.; Choi, M.S.; Choi, I.H.; Cung Van, C.; Nhung, N.V. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front. Med.* **2020**, *7*, 427. [CrossRef]
50. Charbuty, B.; Abdulazeez, A.; Trends, T. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
51. Li, X.; Zhao, H.; Zhu, W.J.K.-B.S. A cost sensitive decision tree algorithm with two adaptive mechanisms. *Knowl. Based Syst.* **2015**, *88*, 24–33. [CrossRef]
52. Parmar, A.; Katariya, R.; Patel, V. A review on random forest: An ensemble classifier. In Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI), Coimbatore, India, 7–8 August 2018; Springer: Heidelberg, Germany, 2019; pp. 758–763.
53. Paul, A.; Mukherjee, D.P.; Das, P.; Gangopadhyay, A.; Chintha, A.R.; Kundu, S. Improved random forest for classification. *IEEE Trans. Image Process.* **2018**, *27*, 4012–4024. [CrossRef] [PubMed]

54. Masetic, Z.; Subasi, A. Congestive heart failure detection using random forest classifier. *Comput. Methods Programs Biomed.* **2016**, *130*, 54–64. [CrossRef] [PubMed]
55. Sharaff, A.; Gupta, H. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences, Proceedings of the IC4S, Bangkok, Thailand, 20–21 October 2018*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 189–197.
56. Zdravevski, E.; Lameski, P.; Kulakov, A.; Trajkovikj, V. *Performance Comparison of Random Forests and Extremely Randomized Trees*; Faculty of Computer Science and Engineering, Ss. Cyril and Methodius: Skopje, North Macedonia, 2016.
57. Kocev, D.; Ceci, M. Ensembles of extremely randomized trees for multi-target regression. In *Discovery Science, Proceedings of the 18th International Conference, DS 2015, Banff, AB, Canada, 4–6 October 2015*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 86–100.
58. Velthoen, J.; Dombry, C.; Cai, J.-J.; Engelke, S.J.E. Gradient boosting for extreme quantile regression. *Extremes* **2023**, *26*, 639–667. [CrossRef]
59. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
60. Tamim Kashifi, M.; Ahmad, I. Efficient histogram-based gradient boosting approach for accident severity prediction with multisource data. *Transp. Res. Rec.* **2022**, *2676*, 236–258. [CrossRef]
61. Cui, J.; Hang, H.; Wang, Y.; Lin, Z. GBHT: Gradient boosting histogram transform for density estimation. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021; pp. 2233–2243, PMLR: 2021.
62. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
63. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
64. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef]
65. Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X. DNN-based prediction model for spatio-temporal data. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016; pp. 1–4.
66. Almatared, M.; Liu, H.; Abudayyeh, O.; Hakim, O.; Sulaiman, M. Digital-Twin-Based Fire Safety Management Framework for Smart Buildings. *Buildings* **2024**, *14*, 4. [CrossRef]
67. Ariyachandra, M.R.M.F.; Wedawatta, G. Digital Twin Smart Cities for Disaster Risk Management: A Review of Evolving Concepts. *Sustainability* **2023**, *15*, 11910. [CrossRef]
68. Park, S.; Park, S.H.; Park, L.W.; Park, S.; Lee, S.; Lee, T.; Lee, S.H.; Jang, H.; Kim, S.M.; Chang, H.; et al. Design and Implementation of a Smart IoT Based Building and Town Disaster Management System in Smart City Infrastructure. *Appl. Sci.* **2018**, *8*, 2239. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.