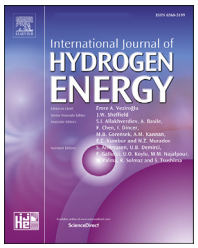
Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/he



Leak localization using distributed sensors and machine learning for hydrogen releases from a fuel cell vehicle in a parking garage

Mingbin Zhao [a], *Teng Huang* [a], *Chenhui Liu* [b], *Mingjia Chen* [a], *Shui Ji* [c], *David M. Christopher* [d], *Xuefang Li* [a,*]

[a] *Institute of Thermal Science and Technology, Shandong University, Jinan, 250061, China*
[b] *Turner-Fairbank Highway Research Center, McLean, VA, 22101, USA*
[c] *State Power Investment Corporation Research Institute, Beijing, 102209, China*
[d] *Key Laboratory of Thermal Science and Power Engineering of Ministry of Education, Tsinghua University, Beijing, 100084, China*

HIGHLIGHTS

- Data acquisition and processing methods for leak localization.
- Scale model study on using machine learning for leak localization.
- Machine learning was used for leak localization in the garage.

ARTICLE INFO

Article history:
Received 6 June 2020
Received in revised form
7 September 2020
Accepted 25 September 2020
Available online 22 October 2020

Keywords:
Parking garage
Hydrogen safety
Hydrogen leak localization
Machine learning

ABSTRACT

The detection and localization of hydrogen releases are of vital importance in large confined spaces to enable safe commercial penetration of hydrogen energy into society. In the present study, a hydrogen leak localization system was developed based on measured concentration data and a machine learning system. A scale parking garage model was used to experimentally model fuel cell vehicle leaks in a confined space with helium used as a surrogate of hydrogen for safety reasons. Twelve gas sensors were placed along the ceiling to measure the helium concentrations. The leak position was changed in each test to gather a range of concentration data. Then, the measured helium concentrations were used as input data to train the machine learning models. Two machine learning algorithms were used to locate the leak position, an artificial neural network (ANN) and the K-DTW algorithm. The helium concentrations recorded by the twelve sensors at a specific time point were used as the input data for the ANN model, while time-series concentration data was used as the input for the K-DTW algorithm. The results show that the ANN model can predict leak locations that were not included in the training datasets and that the K-DTW algorithm can classify the unknown leak location to the given location label. The ANN prediction accuracy was 78.4%. The K-DTW algorithm identified the leak location with 87.5% accuracy. The localization technology developed in the present work can provide safe monitoring of large parking garages and the model accuracy can be improved by feeding more training data to the machine learning models.

* *Corresponding author.*
E-mail addresses: lixf@email.sdu.edu.cn, lixfthu@hotmail.com (X. Li).
https://doi.org/10.1016/j.ijhydene.2020.09.218

Introduction

Hydrogen energy is being deployed around the world as an alternative to traditional petroleum and battery technologies. Hydrogen will play an important role in building a low-carbon economy that will include hydrogen fueled vehicles [1]. Hydrogen will be mainly used in the transportation sector to fuel vehicles through fuel cells. Hydrogen fuel cell vehicles (HFCVs) are expected to be widely used for transportation due to their advantages of very low pollution, quick refueling and long travel ranges. The transition of the transportation sector to HFCVs will be one of the significant steps toward the hydrogen economy [2]. However, HFCVs usually have on-board hydrogen storage tanks at pressures up to 70 MPa to provide comparable ranges as current fossil fuel vehicles, which presents a safety concern since highly compressed hydrogen leakage can possibly result in serious consequences. The hydrogen physical properties differ from hydrocarbon gaseous fuels and the mixing with air and the formation of a flammable region differ from those for other fuels in significant ways. Thus, hydrogen safety is critical for the commercialization of hydrogen energy. There have been numerous hydrogen safety research projects followed by efforts to develop standards and codes appropriate for hydrogen fueled systems. The United States has invested 5–10% of the total funding of hydrogen and fuel cell programs into hydrogen safety research [3]. In Europe, hydrogen safety research has been conducted to analyze various accident scenarios for many years [4]. This hydrogen safety research will affect the commercial penetration of hydrogen energy into future communities.

Hydrogen accidents always begin with unintended hydrogen releases followed by dispersion, deflagration and detonation [5]. Unintended hydrogen releases in a confined space are serious safety issues since the hydrogen will accumulate near the ceiling to form a flammable layer. Therefore, unintended hydrogen leaks in confined structures need to be extensively investigated to understand the gas dispersion behavior. Some researchers have experimentally investigated leaks by injecting hydrogen or other light gases into a confined space to study the gas dispersion characteristics and to provide data for model validation [6–9]. Various release conditions have been studied using hydrogen or helium as the released gas with enclosures including vents and other obstructions. Environmental effects have also been considered, such as wind or temperature. These experimental studies have provided information about the validity of analytical models describing the hydrogen dispersion in an enclosure. Cariteau et al. [10] presented an experimental study of helium dispersion in a 1 m^3 enclosure with the injection Richardson number varied to generate various experimental conditions. They then developed a simple analytical model that gives good estimates of the maximum helium volume fraction. Lacome and Jamois [11] built an 80 m^3 test facility to investigate the dispersion of flammable gases due to hydrogen leaks. Their visual results gave useful information about how the hydrogen/air layer forms, grows and dissipates. Many researchers have been unwilling to use hydrogen for experiments due to safety concerns. Since helium has similar physical properties to hydrogen, helium has been widely used as a surrogate for hydrogen in many studies [12–15]. The experimental results with the two gases are quite similar, so helium can be used to conduct the experiments for hydrogen-powered cars. CFD methods have also been used to study hydrogen releases and deflagrations in confined spaces [16–20].

Hydrogen leaks may result in dangerous consequences in enclosed public parking lots due to the confined spaces. Some papers have described experimental investigations of hydrogen releases from HFCV and dispersion in real or scale enclosures representative of residential garages [21–23]. Simple geometries are normally used that are scaled to preserve the geometric similarity. Pitts et al. [14] investigated helium dispersion in a 1/4 scale two-car garage model. The concentrations were measured at various locations in the garage as functions of the gas flow rates and flow duration. The experimental data has been used to validate numerical models and to improve understanding of the dispersion mechanism. CFD methods have also been used to evaluate methods to quickly disperse the hydrogen, so that first responders can safely approach the vehicle in an accident scenario [24–27]. Choi et al. [28] modeled hydrogen leaks in a parking garage to study the temporal changes of the flammable region and the influences of the leakage flow rate and ventilation on the hydrogen dispersion. Mukai et al. [29] numerically studied hydrogen releases in a multistory parking garage with 18 parking slots and analyzed the dispersion characteristics associated with the safety risk. Their results showed that hydrogen leaks in an enclosed space without sufficient ventilation may lead to dangerous accidents including fires and explosions.

Although some previous studies have investigated hydrogen dispersion in confined spaces, leak detection and localization methods have not been developed for hydrogen applications. Rapid and accurate hydrogen leak detection and localization is essential for alerting response personnel to the formation of potentially explosive gas mixtures and to help prevent explosions. Sensors using acoustics, optics or other technologies have been developed for hydrogen leak detection [30]. Wireless Sensor Networks (WSN) can be used to localize the leak, as have been widely used in the oil and gas industry [31,32]. Data from widely distributed, cheap sensor nodes can be used to estimate the leak positions. However, the WSN can only estimate an approximate leak area and the identification accuracy needs to be improved. Machine learning is a promising tool for fault location and pattern recognition that has been widely used for many industrial problems [33]. Bouthiba [34] detected fault locations in EHV transmission lines using artificial neural networks. Kim et al. [35] used the spatiotemporal distribution of bacterium along a water system to locate pollutants by inversely interpreting the bacterium transport patterns using an artificial neural network (ANN). So et al. [36] proposed a method for monitoring and estimating hazardous gas release rates using optical sensors and neural networks. Hartmann and Link [37] presented an approach for human gesture recognition using inertial sensors and a dynamic time warping (DTW) algorithm. Their gesture recognition

system had a 97.35% accuracy. Thus, a combination of sensors and machine learning algorithms should also be useful for locating hydrogen leaks.

A leak localization system was developed in this study by combining a sensor network with machine learning. A scale parking garage model was constructed with twelve gas sensors mounted along the ceiling to get adequate concentration data for the machine learning. Two machine learning algorithms were used to identify and predict the location of the leaking vehicle from the experimental data for various scenarios. Thus, sensors and machine learning algorithms can be combined for hydrogen leak detection and localization to mitigate the risks of hydrogen accidents.

Methodology

Unintended hydrogen releases in a parking garage are serious safety issues since the hydrogen will accumulate near the ceiling. Dangerous accidents can lead to fires and explosions if the garage does not have sufficient ventilation. A hydrogen leakage scenario in a multilevel parking garage is shown in Fig. 1.

A leak localization system was developed to detect and locate the leak source. The system included five steps, as shown in Fig. 2.

a) Identify many possible leak scenarios before the experiments. The experimental results include both temporal and spatial concentration data.
b) Construct a 1/24 scale model of a real garage according to the China National Standard "Code for design of parking garages (JGJ 100–2015)".
c) Change the leaking vehicle position in each test and record the time-series gas concentration data. Helium was used as a surrogate for hydrogen in the experiments.
d) A data preparation algorithm was developed to preprocess and reorganize the data for the ANN and K-DTW (K-nearest neighbor and dynamic time warping) algorithms. 80% of the experimental data was used for training with the remaining 20% of the experimental data used to evaluate the accuracy of each algorithm by comparing the algorithm predictions with the original target.

Experimental system

Experimental setup and release conditions

A scale parking garage model was constructed to collect concentration data for helium releases from various leak positions for training and testing the machine learning. The 1/24 scale model was selected to ensure the garage model could be built in our lab and that the flow rate could be easily controlled by the present experimental system [38]. The garage model was $24 \times 150 \times 250\ cm^3$. The garage entrance was 24.4 cm wide and 13.3 cm high and was kept open during the release to better simulate the actual gas flow conditions. Interior details like beams and pillars in the parking garage were neglected and sealant was used to seal the crevices between the walls and frames. The garage model was divided into 4 rows (A–D) in the y direction with each being 20 cm wide and 20 columns (1–20) in the x direction with each being 12 cm wide to locate the car models. All the experiments were carried out indoors at the same humidity of 55% and temperature of 24 °C. The temperature of the leaked helium was assumed to be the

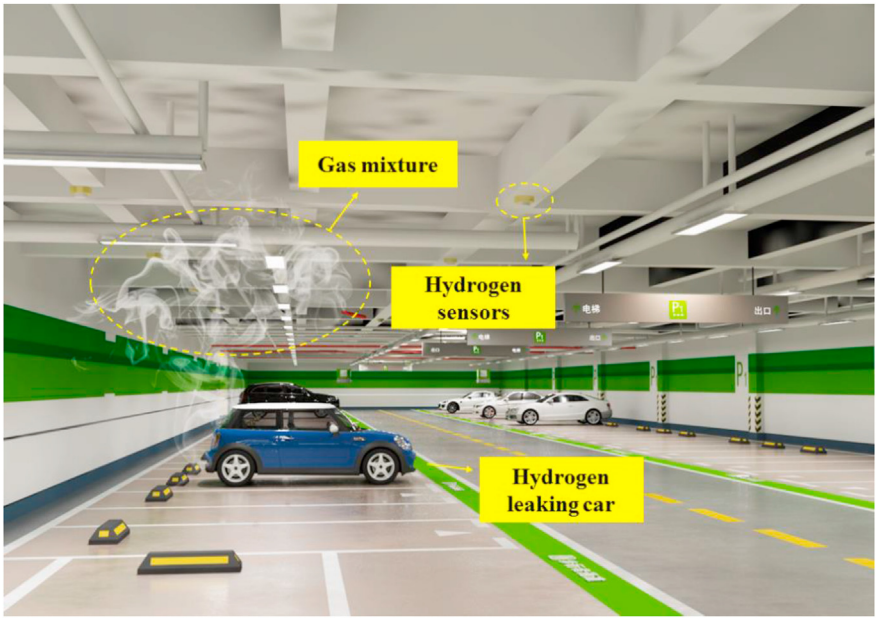


Fig. 1 – Hydrogen leaking from an HFCV in an enclosed parking lot.

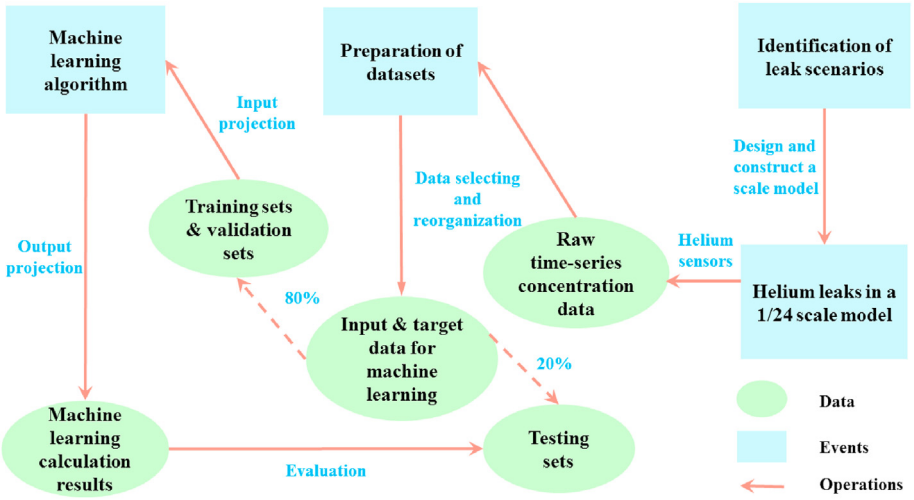


Fig. 2 – Architecture of the leak localization system.

same as room temperature because of the long tube from the helium bottle to the nozzle.

A typical family car model (MINI COOPER, 1/24) was used as the leaking vehicle in the experiment. Helium was used as a hydrogen surrogate for safety reasons. A standard 40 L helium bottle was used as the gas source. The release pressure was controlled by a regulator and measured by a pressure transducer. Then, the flow rates were adjusted by a needle valve and measured by an electronic flowmeter with an accuracy of ±(2.0 + 0.5FS)%. The leak point was set at the bottom of the car to simulate the most common and dangerous scenario. The leak point was circular, since gas releases from circular nozzles results in the longest flammable zones compared to other nozzle shapes [39]. Therefore, the circular nozzle gave conservative leak consequences. The leak point location, shape and size in the scale model setting were selected to be consistent as possible with conditions in a real garage. A 1 m long by 2.5 mm inner diameter PU hose was used at the end of the gas line to connect to the leaking car for the convenience of changing the leak position. A picture of the experimental system is shown in Fig. 3.

Twelve mini-katharometers (XEN-TGC3880Pt) with an accuracy of 2% of the measured value were arranged on the ceiling of the garage model to measure the helium concentrations at various locations. All the sensors were flush with the surface to minimize their influences on the flow. The signals were recorded by an Agilent 34970A data acquisition system every 0.708 s and then saved on the computer. The schematic of the parking spaces and the sensor locations is shown in Fig. 4 with the sensor coordinates listed in Table 1.

Each sensor was calibrated using helium-air mixtures with helium concentrations of 0%, 9.98%, 19.93%, 37.4%, and 99.999%. The calibration process was described in detail by Li et al. [40]. The measured voltage, V, of the sensors was then fit to a quadratic equation for the helium gas mole fraction, X_{He}.

$$V = V_0(1 + aX_{He} + bX_{He}^2) \quad (1)$$

where V_0 is voltage signal for a helium concentration of 0. a and b are coefficients.

The experimental conditions for the helium release in the scale model are listed in Table 2. The flow rate was 3.4 SLPM to simulate momentum dominated leak jets from high storage pressure tanks (up to 35 MPa). The leaking vehicle was located in different parking spaces in each test. The helium release duration was 220 s and the sensors monitored the concentration changes for about 600 s.

The Richardson number, Ri, was used to estimate whether the release was momentum or buoyancy dominated [22]:

$$\mathrm{Ri} = \frac{\pi^2}{32}\frac{\rho_{air} - \rho_{He}}{\rho_{He}}\frac{gD^5}{Q^2} \quad (2)$$

where Q denotes the injection flow rate, D is the exit diameter, g is the gravitational acceleration, ρ_{He} represents the helium density of 0.179 kg/m^3 and ρ_{air} is the ambient air density of 1.293 kg/m^3.

Experimental results

The helium release concentrations at C20 were measured to study the helium dispersion around the leaking vehicle. The data is shown in Fig. 5 where the color blocks represent the sensor ID for each curve. The helium concentrations along the ceiling decrease with the distance from the leak. The helium dispersion can be divided into 3 stages. In Stage 1, the helium concentrations increase rapidly and the time needed for detecting the concentration change increases with the distance from each sensor to the leaking car. Then, the helium concentrations along the ceiling reach a relative steady state after the initial increase in Stage 2, which is suitable for the machine learning training process. In Stage 3, the concentration curve decreases rapidly after the release stops. The maximum helium concentration (sensor VIII) is about 3.5%, which is below the lower flammability limit (LFL) of 4% by volume of hydrogen in air. However, only far-field

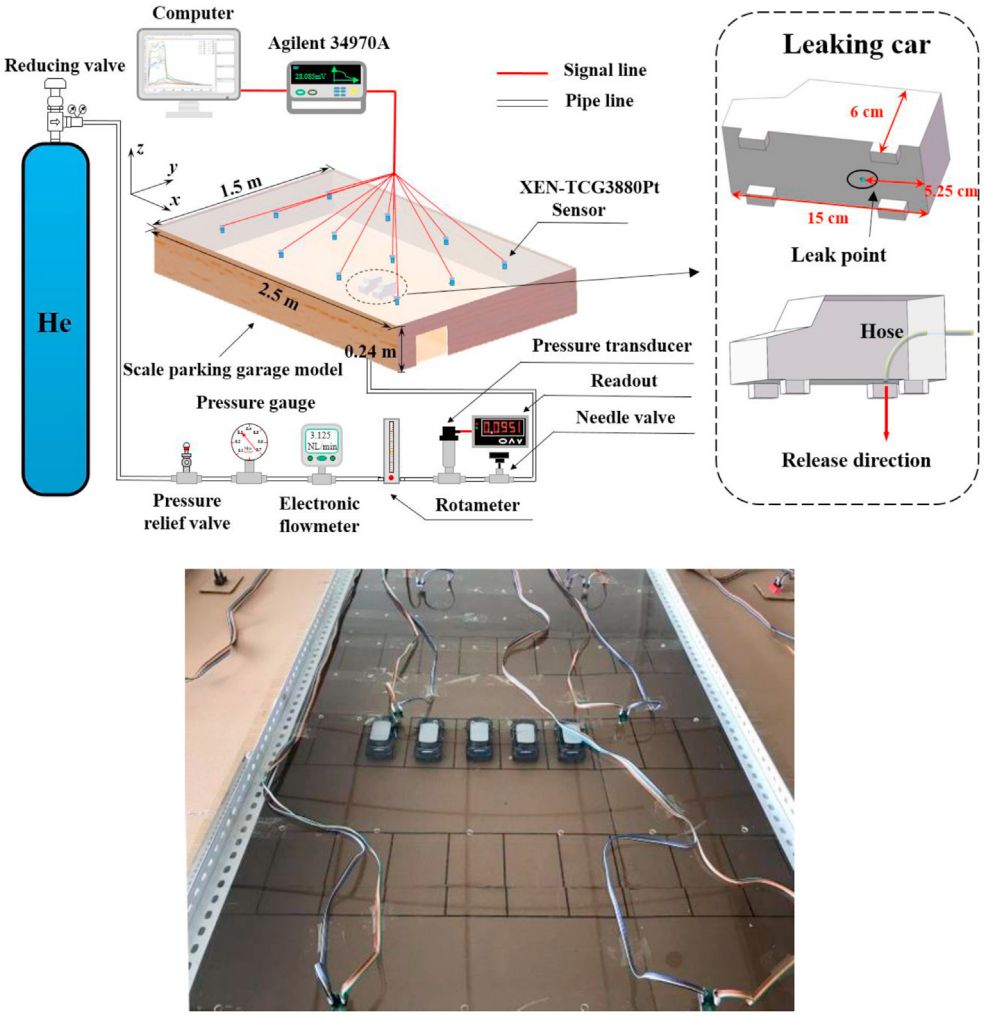


Fig. 3 – Picture of the experimental system.

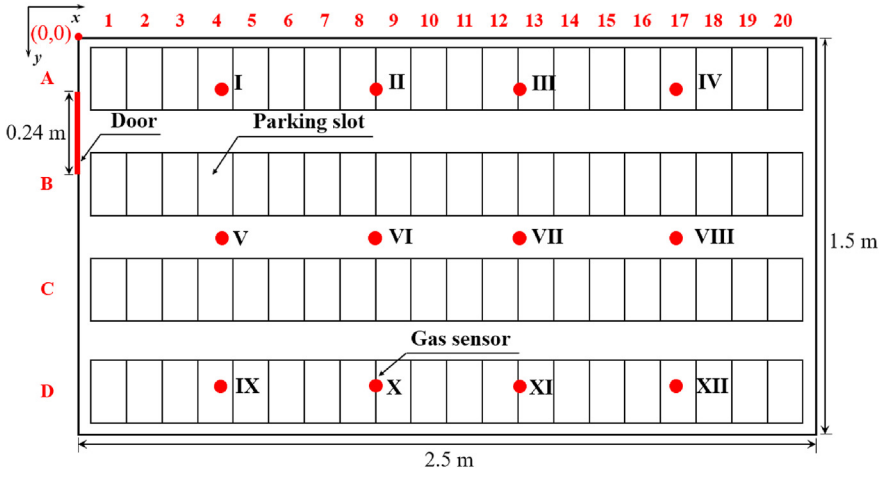


Fig. 4 – Schematic of the parking spaces and sensor locations.

Table 1 – Sensor positions.

Sensor ID	Coordinate (cm)	Sensor ID	Coordinate (cm)
I	(50, 20)	VII	(149, 75)
II	(102, 20)	VIII	(200, 75)
III	(149, 20)	IX	(50, 131)
IV	(201, 20)	X	(102, 130)
V	(51, 75)	XI	(148, 130)
VI	(102, 75)	XII	(201, 131)

*The origin is at the upper left corner of the garage as shown in Fig. 4.

concentrations were measured and an explosive gas mixture might exist near the leak.

The leaked helium rises up to the ceiling and then spreads laterally. Once a sensor detects the concentration change, the measured concentration increases rapidly indicating a rapid step concentration increase, which is called the initial front [6]. The process is shown in Fig. 6. Schlieren visualizations show a significant change in the flow structure of the initial front [6]. The helium dispersion enters Stage 2 when the rapid step concentration increase is detected between adjacent time points. This feature helps the ANN algorithms determine the current leakage stage.

Table 2 – Experimental conditions.

Gas	Flow rate (SLPM)	Exit diameter (mm)	Release direction	Release duration (s)	Monitor duration (s)	Ri (10^{-4})
Helium	3.4	2.5	Downward	220	600	5.83

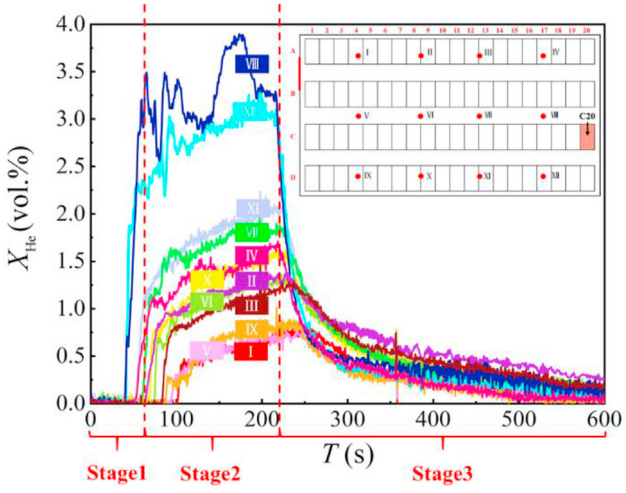


Fig. 5 – Evolution of the helium concentrations with time for the release at C20.

The effects of the neighboring vehicles on the gas dispersion were also studied since the leaking car is most likely surrounded by other cars in the parking garage. The helium concentrations were measured in two tests with one test having only the leaking car in the garage and the other having two cars placed beside the leaking car. The helium concentrations were approximately the same for both tests as shown in Fig. 7. Therefore, the effects of the surrounding cars were ignored and one leaking car without neighboring vehicles was used in subsequent experiments.

In the present experiment, the volumetric flow rate was scaled based on the geometric similarity relative to a release from a 35 MPa leak, but the gas properties were not scaled. Hence, the helium gas distributions in the model garage should be similar to those in a real garage, while the buoyancy effects and dispersion rates were not to scale. This limitation was addressed earlier by introducing a dimensionless time [38]. The differences between the scale model and the real garage will not significantly affect the conclusions, since the

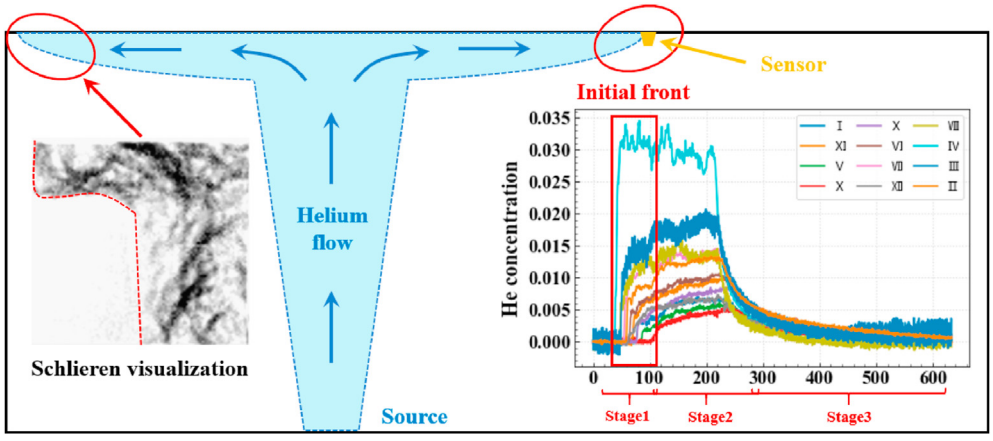


Fig. 6 – Schlieren visualization [6] and schematic of the initial front.

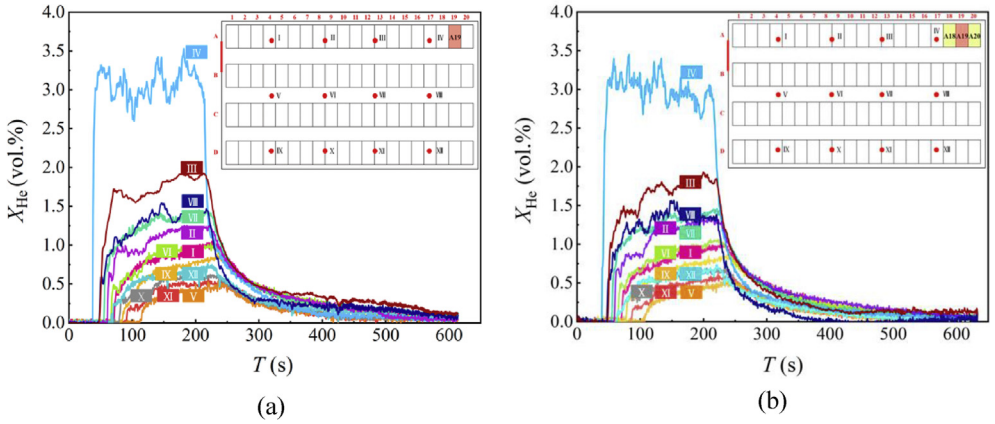


Fig. 7 – Evolution of the helium concentration with time inside the garage for Tests 1 and 2. (a) Test 1: Helium release at A19 without neighboring cars (b) Test 2: Helium release at A19 with 2 neighboring cars at A18 and A20.

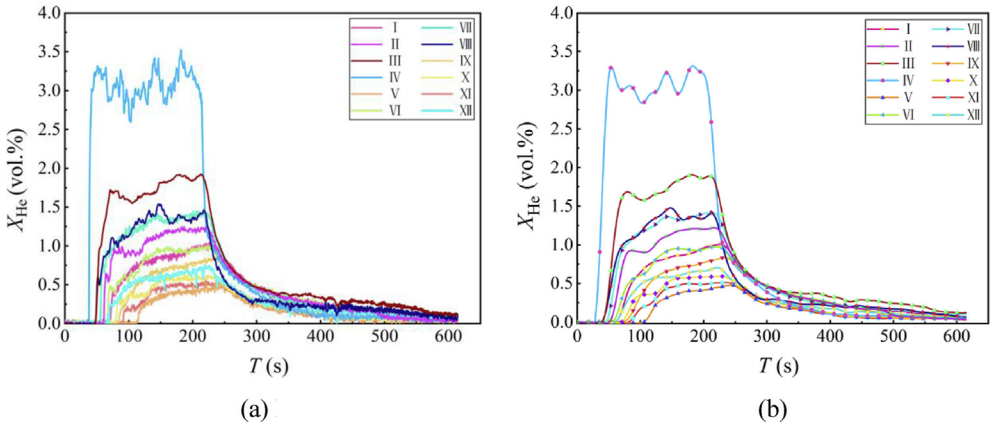


Fig. 8 – Comparisons of the raw and smoothed helium concentration data. (a)Raw helium concentration data. (b) Smoothed helium concentration curves.

primary purpose of the present work was to develop a methodology for accurate leak localization rather than to investigate the gas dispersion in a real garage.

Algorithms

Data preprocessing for the machine learning

Two machine learning algorithms, an artificial neural network and a K-DTW algorithm, were used to locate the leaking car. The concentration data was preprocessed and reorganized before being input to the machine learning algorithms with different input formats for each algorithm.

The neural network is an operational model with a large number of neurons connected to each other. The connections between each neuron pair represent a weight. The network maps each sample input (such as images or data sequences) to its target by adjusting the weights of the neural network. The parameters are difficult to determine because a neural network may contain millions of parameters and changes in each parameter affect all the other parameters. The data fluctuations as seen in Fig. 8(a) reduce the accuracy of the predicted results and prolong the training time. The strong fluctuations in the concentration data are partially due to turbulence in the helium flow along the ceiling and partially due to signal noise. Since these fluctuations will affect the final network training results, the raw concentration data was smoothed before being input into the models. The noisy signals were represented by high-energy signals and low-energy noise with the signals then denoised by frequency domain filtering using Fast Fourier Transforms (FFT). After the low-energy noise was removed from the signals, the remaining signals were converted back into the time domain. Several data points caused by poor sensor contacts were excluded from the curves with the final smoothed curves shown in

Fig. 8(b). The signal oscillations caused by the flow turbulence and signal noise were smoothed while preserving the key trends of the curves.

The smoothed concentration data series were then reorganized for the ANN model and the K-DTW algorithm.

- **ANN format**: Helium concentration data for twelve sensors at selected time points was input into the ANN. The time points were selected from the intermediate stage (stage 2) where the helium concentrations were relatively steady. The trained model can then predict the leak source location using concentrations at adjacent time points because the concentrations vary little over short time periods. The important step is to determine the dispersion stage rather than the time point. The problem was solved by characterizing the sensor concentration readings during the gas dispersion process. The initial front phenomenon described in section Experimental results helped the ANN to identify the current stage. 154 sets of concentration data for various leak locations were then used for network training (some locations may have more than 1 data set). The ANN dataset structure was written as:

$$C_{P,T} = [c_{P,I,T}, c_{P,II,T}, c_{P,III,T}, \cdots, c_{P,X,T}, c_{P,XI,T}, c_{P,XII,T}] \quad (3)$$

where subscript P represents the location of the leaking vehicle such as A1 or B2. The Roman numerals represent the sensor ID, T is the time point and c is the helium concentration.

- **K-DTW format:** Time-series concentration data was input into the K-DTW algorithm. The data points for each sensor were connected end to end with 46 sets of concentration data in the K-DTW dataset. The dataset structure was:

$$C_P = [c_{P,I,1}, \cdots c_{P,I,n}, c_{P,II,1}, \cdots c_{P,II,n}, \cdots\cdots c_{P,XII,1}, \cdots c_{P,XII,n}] \quad (4)$$

where n is the number of time points.

Leak localization using ANN

The helium concentration datasets were used by the artificial neural networks to build a model that could then identify the leak location. The ANN consists of an input layer representing the input data, many hidden layers for the model parameters and an output layer representing the network results. Each layer has several neurons with the neurons connected to each neuron of the previous layer through an adaptable weight, w, and bias, b. The value of each neuron then represents the weighted result of all the neurons in the previous layer. The neural network outputs were calculated from the first layer to the last layer using:

$$y_j^{(k)} = \phi\left(\sum_{i=1}^{n} \omega_{i,j}^{(k)} \times y_i^{(k-1)} + b_j^{(k)}\right) \quad (5)$$

where superscript k represents the layer index, subscripts i, j are the neuron indices, $\omega_{i,j}$ represents the connection weights and b_j represents the bias for the neuron between two adjacent layers. y is the neuron value and ϕ is the activation function (the ReLU squishification function was used here) to normalize the value of each neuron to 0–1.

The loss function, E, is defined as the quality of the calculated result in each epoch which is used to adjust the parameters to improve the training process. In the present work, the cross-entropy was used to define the loss as:

$$E = -\sum_{i=1}^{n} y_{true} \cdot \log\left(y_{pred}\right) \quad (6)$$

where y_{true} is the real value and y_{pred} is the predicted value.

The stochastic gradient descent algorithm and the back-propagation algorithm were used to minimize the loss function.

$$\Delta\omega_{i,j} = -\eta\frac{\partial E}{\partial \omega_{i,j}} \quad \Delta b_{i,j} = -\eta\frac{\partial E}{\partial b_{i,j}}$$
$$\omega'_{i,j} = \omega_{i,j} + \Delta\omega_{i,j} \quad b'_{i,j} = b_{i,j} + \Delta b_{i,j} \quad (7)$$

where the prime represents the updated value. η is a constant called the learning rate which is used to control the parameter updating rate. A suitable value of η enables the network to find a global minimum (training loss is close to zero) and to shorten the training time.

The network training adjusts the weight, ω, and the bias, b, to minimize the loss function. The network weights and biases are randomly initialized in the first step. The network uses the initial parameters to calculate the output, y_{pred}, in the last layer which is compared with the desired output, y_{true}, to calculate the loss. The weights and biases for all the neurons are updated in each iteration to minimize the loss function by the back-propagation algorithm and the stochastic gradient descent algorithm. The iteration repeats until the outputs have an acceptable loss. The network can then do pattern recognition once the appropriate network weights and biases are selected [41].

In the present work, two neural networks were constructed to predict the row and column numbers of the leak location in the garage. The best model accuracy was achieved using the ANN with a 12-30-30-1 architecture (the numbers of neurons in each layer) as shown in Fig. 9. The input layer used the concentration data from twelve sensors at a specific time point as the input data with the one neuron in the output layer representing the predicted result for the row or column number. K-fold-cross-validation, regularization and the dropout algorithm were used to overcome the overfitting problem with a lower generalization error. 80% of the experimental data was used for training with the remaining 20% used to evaluate the accuracy by comparing the predicted results with the target results. The training used a maximum of 500 epochs.

Leak localization using the K-DTW algorithm

Time-series data classification is an essential but difficult problem. In the present work, the helium concentration datasets were typical time-series data. The leak durations and helium peak concentrations varied for the various leakage scenarios. Classifying leaks at unknown locations

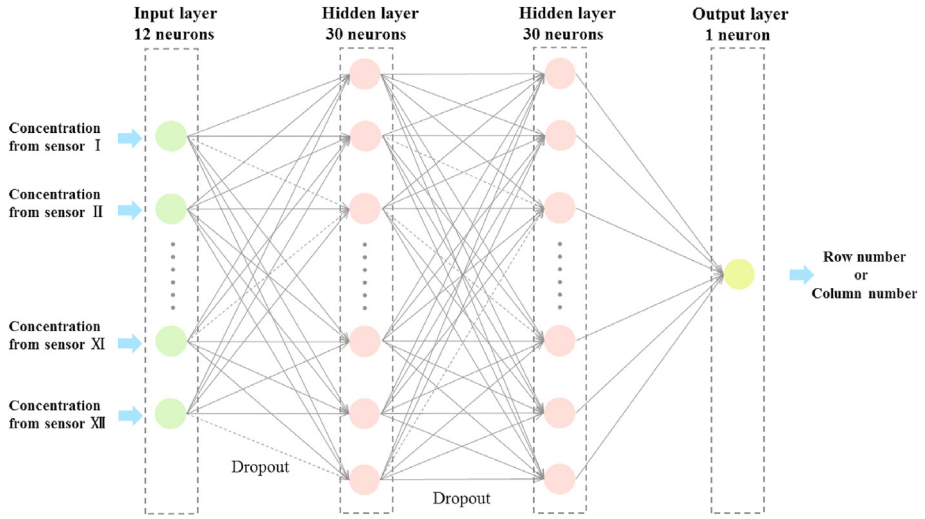


Fig. 9 – Artificial neural network structure.

into known accident scenarios is very difficult. Mitsa [42] suggested that 1 Nearest Neighbor (K = 1) and Dynamic Time warping is useful for analyzing time-series classification problems. Thus, the K-DTW algorithm [43] has been used for fault classification on transmission lines. This algorithm combines the K-nearest neighbor (KNN) algorithm with dynamic time warping (DTW).

KNN is a classical clustering algorithm without any assumptions about the underlying data distribution. In the KNN, K is the number of nearest neighbours, which is the key factor of this algorithm. An unknown case is classified by a majority vote of its neighbours, with the case being assigned to the most common class among its K nearest neighbours measured by a distance function. In this case, the case was simply assigned to the class of its nearest neighbor if K = 1. The distance to the nearest similar point was found using various distance measures such as the Euclidean distance, Hamming distance, Manhattan distance, and Minkowski distance [44].

DTW is a time series alignment algorithm originally developed for tasks related to speech recognition [45]. This algorithm seeks an optimal match between the two sequences by iteratively warping the time axis to align the two sequences of feature vectors. DTW calculates the cumulative distance between each pair of points (called the warp path distance) to measure the similarity between the two time-series datasets. The DTW algorithm principles are as follows. The problem has two time series data, $Q = \{q_1, q_2, \ldots, q_i, \ldots, q_n\}$, and $C = \{c_1, c_2, \ldots, c_j, \ldots, c_m\}$, with different lengths. A $m \times n$ minimum distance matrix, M, is constructed to align the two sequences and the minimum distance matrix is obtained using a dynamic programming algorithm to satisfy the optimization objective:

$$d_{i,j} = |q_i - c_j|, \quad i = 1, 2, \cdots, n \quad j = 1, 2, \cdots m \tag{8}$$

$$M_{i,j} = d_{i,j} + \min(M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}) \tag{9}$$

where $d_{i,j}$ represents the Euclidean distance between the two points q_i and c_j. $M_{i,j}$ are the elements of the minimum distance matrix between Q and C.

A sequence satisfying the following three conditions is defined as a warping path, W, which can be used to evaluate the similarity between the two time-series datasets.

$$W = (w_1, w_2, \cdots, w_k, \cdots, w_K) \quad \max(m, n) \le K \le m + n - 1 \tag{10}$$

- **Boundaries:** The selected path must start at the lower-left corner of the matrix where $w_1 = M_{1,1}$ and end at the upper-right corner where $w_k = M_{m,n}$.
- **Monotonicity:** The warping path has continually increasing indices. If $w_{k-1} = M_{a',b'}$, $w_k = M_{a,b}$ for the next point on the path such that $0 \le a - a' $ and $0 \le b - b'$.
- **Continuity:** The increments between successive elements are limited. If $w_{k-1} = M_{a',b'}$, $w_k = M_{a,b}$ for the next point on the path needs to satisfy $a - a' \le 1$ and $b - b' \le 1$.

There are a very large number of warping paths that satisfy these conditions. The path which minimizes the warping cost was selected as minimum DTW distance to evaluate the similarity between the two time-series.

$$\mathrm{DTW}(Q, C) = \min\left\{\sqrt{\sum_{k=1}^{K} w_k} \Big/ K\right\} \tag{11}$$

The K-DTW algorithm used the steps shown in Fig. 10 to locate the leaking car. An unknown concentration series data in the K-DTW format described in section Data preprocessing for the machine learning without a leak location label was used as the input data. In the first step, the minimum DTW distance was calculated between the inputs and each concentration series using experimental data with known leak locations. In the next step, the KNN compared the minimum DTW distance between the unknown series and other known

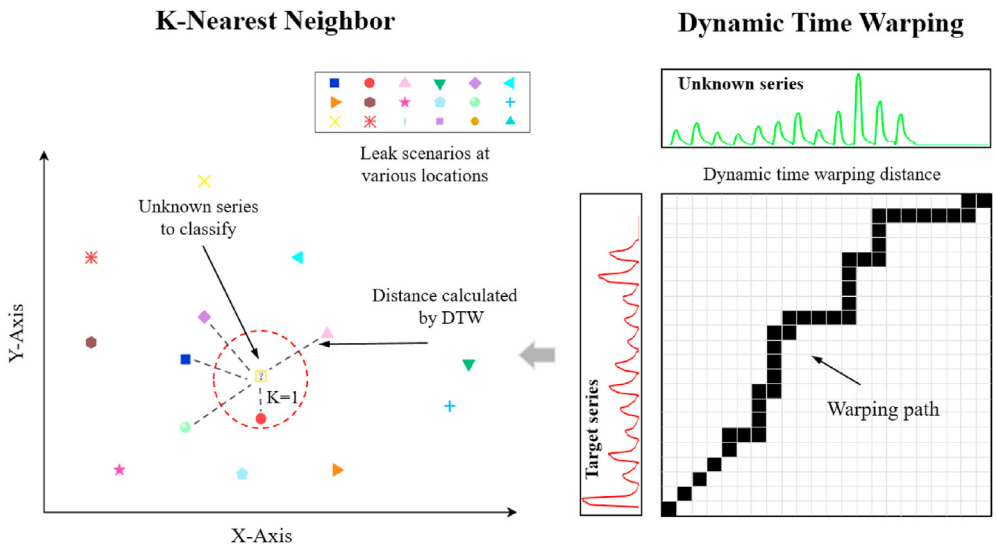


Fig. 10 – K-DTW algorithm schematic.

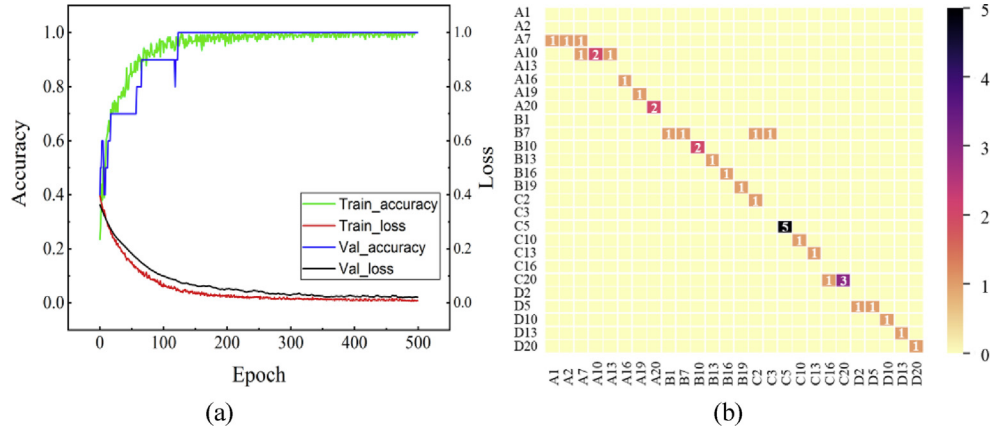


Fig. 11 – ANN model training process and localization accuracy. (a) Convergence of the error and accuracy of the training and validation sets. (b) Confusion matrix of the classification results on the testing sets.

leakage scenarios. The unknown series data was then classified into the closest similar scenario and tagged with its leak location.

Results and discussion

ANN model localization accuracy

The TensorFlow 2.0.0 and Keras 2.3.1 open-source machine learning framework were used to build the ANN model. The ANN model could detect and locate the helium leak position after the training process. The convergence of the training error and the accuracy shown in Fig. 11(a) show that the minimum loss reached 8.8×10^{-3} and the maximum accuracy was 0.996 after 500 epochs. The localization accuracy of the ANN was then evaluated using the test datasets with the localization accuracy represented by a confusion matrix in Fig. 11(b). The abscissa represents the actual leakage locations while the ordinate shows the predicted locations. The diagonal entries show the number of correctly located test samples. As shown in the figure, only a few points deviate from the diagonal. The wrong classification results may be caused by similar helium concentration distributions between cars in adjacent locations. The algorithm has satisfactory localization ability with a test localization accuracy of 78.4%. However, the predict results may be affected by data fluctuations when using only one time point for training. The localization algorithm can be improved by using time-series multivariate data for training which will make the model more robust and

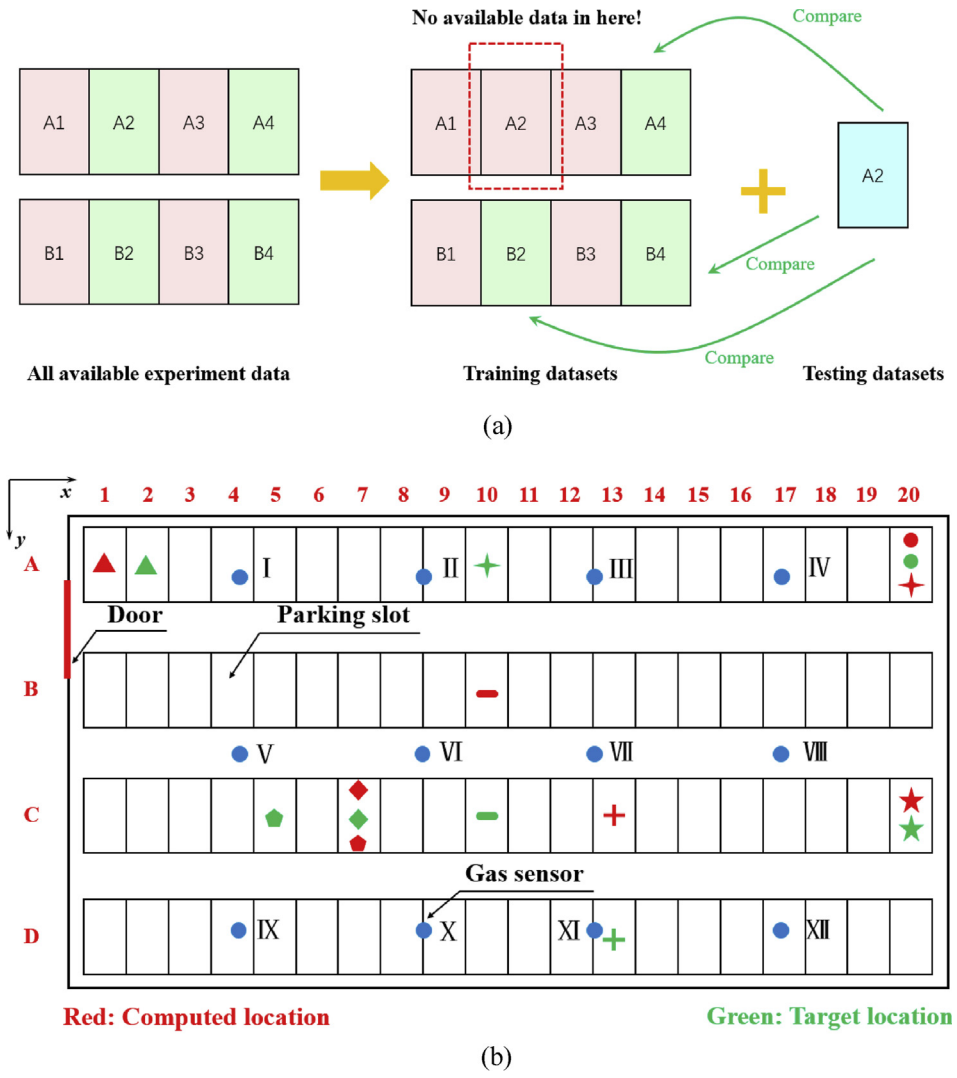


Fig. 12 – Prediction process and localization results of K-DTW algorithm. (a) Prediction process of K-DTW algorithm. (b) Localization results of K-DTW algorithm.

Table 3 – K-DTW algorithm localization results.

Classifier	Testing sets								Accuracy
Target location	A1	C7	D13	C10	A20	A10	C20	C5	87.5%
Computed location	A2	C7	C13	B10	A20	A20	C20	C7	
Result[a]	C	C	C	C	C	IC	C	C	
Error[b]	1	0	1	1	0	10	0	2	

[a] Correctly identified (C), incorrectly identified (IC).
[b] Numbers represent the number of parking spaces from the target leak location.

versatile. The second algorithm, K-DTW, in the present work was used for this purpose.

K-DTW algorithm localization accuracy

The K-DTW algorithm is more suitable for handling the time-series concentration data than the ANN. 8 sets of concentration data were used as test sets with the other experimental data used as the base datasets. The K-DTW calculated the similarity between the unknown inputs from the test sets and various concentration series from the base datasets with known leak locations. Then, the unknown inputs were tagged with the known experimental leak locations. However, the leak locations in the base datasets did not cover

all the parking slots. Thus, the unknown leak scenario may be classified into an adjacent known leak location with some location error. The green blocks in Fig. 12(a) represent parking slots with known experimental data while the red blocks indicate no data. The leak scenario in A2 was used to test the localization accuracy with the data for A2 not included in the training datasets. The K-DTW then compares the test data (A2) and the training data (A4, B2, B4) to classify the leak location as A4, B2 or B4 which were included in the training dataset. Therefore, the classification result was accepted as correct when the error was less than 2 parking slots in the x direction and 1 parking slot in the y direction. The K-DTW algorithm showed the best accuracy with $K = 1$ with the identification results of the K-DTW algorithm shown in Table 3 and Fig. 12(b). The K-DTW algorithm identified the leak locations with an accuracy of 87.5%.

Conclusions and future work

A helium leak localization system was developed using a gas sensor network and machine learning. Helium concentrations in a 1/24 scale parking garage model were measured by the gas sensors to generate datasets for the machine learning models. ANN and K-DTW machine learning algorithms were then used to locate the helium leak positions. The results show that the ANN can predict the leaking car locations from the provided concentration data at the given time points. The K-DTW algorithm can classify the unknown time-series concentrations into known leak locations. The ANN model prediction accuracy was 78.4% while the K-DTW algorithm leak location accuracy was 87.5%. Thus, both algorithms can accurately detect and locate the leak locations.

This localization model can be applied to full size garages but the experimental data collected from the scale model may not be appropriate for a full-size garage. The flow rates in the scale experiments were scaled based on geometric similarity, but the gas properties such as density, viscosity, molecular diffusion rate were not scaled. Hence, the detailed concentration profiles within the scale garage would likely differ from those in a real parking garage, while the experimental data collected in the scale model could be used to validate CFD simulations that could then be applied to full-size garages to generate training data for leak localization in real garages. The garage geometry, the environmental conditions and the leakage locations are the main factors to be considered when making the training sets. For a given garage design, the workload required to produce the training data by CFD models should be acceptable.

The number of training datasets is always an important factor for machine learning. Although both algorithms gave acceptable predictions using the present experimental data, more training data is needed to improve the system accuracy. Therefore, the training database will be extended by including additional experimental data and CFD simulation results and more data for various external factors, such as temperature, humidity, and garage size, for analyses in a future work. The gas dispersion during the release showed self-similar characteristics, so a non-dimensional analysis can be used to describe the dispersion and concentration distributions for various conditions. Similarity laws can be integrated into the localization algorithms to reduce the number of data sets needed to train the model, to enhance the model versatility and robustness, and to expand the model applicability. The leak localization system should be able to quickly identify the locations of dangerous leaks. K-DTW had to iterate over all the datasets to compute their similarities, which is very time-consuming and computationally expensive, so the computational speed can be improved by limiting the comparison range of the data. In addition, the data has obvious spatial correlations, so spatial K-DTW models will provide better accuracy.

The leak localization method developed in this work can be used in real scenarios if adequate data is generated to train the model. The leak location predictions can then help first responders identify a relatively small leak zone to find the leaking car. In the future, hydrogen leak localization programs based on gas sensors and machine learning will play significant roles in developing early warning systems for large confined spaces.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 51706125).

REFERENCES

[1] Molkov V. Fundamentals of hydrogen safety engineering I. BookBoon.com; 2012.
[2] Veziroglu A, MacArio R. Fuel cell vehicles: state of the art with economic and environmental concerns. Int J Hydrogen Energy 2011;36:25–43. https://doi.org/10.1016/j.ijhydene.2010.08.145.
[3] Barilo NF, Weiner SC, James CW. Overview of the DOE hydrogen safety, codes and standards program part 2: hydrogen and fuel cells: emphasizing safety to enable commercialization. Int J Hydrogen Energy 2017;42:7625–32. https://doi.org/10.1016/j.ijhydene.2016.04.070.
[4] Neef HJ. International overview of hydrogen and fuel cell research. Energy 2009;34:327–33. https://doi.org/10.1016/j.energy.2008.08.014.
[5] Safety issues regarding fuel cell vehicles and hydrogen fueled vehicles. 2019. https://dps.mn.gov/divisions/sfm/programs-services/Documents/Responder%20Safety/Alternative%20Fuels/FuelCellHydrogenFuelVehicleSafety.pdf/. [Accessed 19 December 2019].
[6] Cariteau B, Tkatschenko I. Experimental study of the concentration build-up regimes in an enclosure without ventilation. Int J Hydrogen Energy 2012;37:17400–8. https://doi.org/10.1016/j.ijhydene.2012.03.156.
[7] Spijkerboer HP, Beniers JE, Jaspers D, Schouten HJ, Goudriaan J, Rabbinge R. Ability of the Gaussian plume

model to predict and describe spore dispersal over a potato crop. Ecol Model 2002;155:1–18. https://doi.org/10.1016/S0304-3800(01)00475-6.
[8] Pitts WM, Yang JC, Blais M, Joyce A. Dispersion and burning behavior of hydrogen released in a full-scale residential garage in the presence and absence of conventional automobiles. Int J Hydrogen Energy 2012;37:17457–69. https://doi.org/10.1016/j.ijhydene.2012.03.074.
[9] Barley CD, Gawlik K. Buoyancy-driven ventilation of hydrogen from buildings: laboratory test and model validation. Int J Hydrogen Energy 2009;34:5592–603. https://doi.org/10.1016/j.ijhydene.2009.04.078.
[10] Cariteau B, Tkatschenko I. Experimental study of the effects of vent geometry on the dispersion of a buoyant gas in a small enclosure. Int J Hydrogen Energy 2013;38:8030–8. https://doi.org/10.1016/j.ijhydene.2013.03.100.
[11] Lacome JM, Jamois D, Perrette L, Proust CH. Large-scale hydrogen release in an isothermal confined area. Int J Hydrogen Energy 2011;36:2302–12. https://doi.org/10.1016/j.ijhydene.2010.10.080.
[12] He J, Kokgil E, Wang L, Ng HD. Assessment of similarity relations using helium for prediction of hydrogen dispersion and safety in an enclosure. Int J Hydrogen Energy 2016;41:15388–98. https://doi.org/10.1016/j.ijhydene.2016.07.033.
[13] Swain MR, Filoso P, Grilliot ES, Swain MN. Hydrogen leakage into simple geometric enclosures. Int J Hydrogen Energy 2003;28:229–48. https://doi.org/10.1016/S0360-3199(02)00048-4.
[14] Pitts WM, Yang JC, Fernandez MG. Helium dispersion following release in a 1/4-scale two-car residential garage. Int J Hydrogen Energy 2012;37:5286–98. https://doi.org/10.1016/j.ijhydene.2011.12.008.
[15] Prasad K, Pitts W, Yang J. Effect of wind and buoyancy on hydrogen release and dispersion in a compartment with vents at multiple levels. Int J Hydrogen Energy 2010;35:9218–31. https://doi.org/10.1016/j.ijhydene.2010.06.001.
[16] Hourri A, Angers B, Bénard P, Tchouvelev A, Agranat V. Numerical investigation of the flammable extent of semi-confined hydrogen and methane jets. Int J Hydrogen Energy 2011;36:2567–72. https://doi.org/10.1016/j.ijhydene.2010.04.121.
[17] Hourri A, Angers B, Bénard P. Surface effects on flammable extent of hydrogen and methane jets. Int J Hydrogen Energy 2009;34:1569–77. https://doi.org/10.1016/j.ijhydene.2008.11.088.
[18] Middha P, Hansen OR. CFD simulation study to investigate the risk from hydrogen vehicles in tunnels. Int J Hydrogen Energy 2009;34:5875–86. https://doi.org/10.1016/j.ijhydene.2009.02.004.
[19] Malakhov AA, Avdeenkov AV, Du Toit MH, Bessarabov DG. CFD simulation and experimental study of a hydrogen leak in a semi-closed space with the purpose of risk mitigation. Int J Hydrogen Energy 2020;45:9231–40. https://doi.org/10.1016/j.ijhydene.2020.01.035.
[20] Deborah HA, Elena CG. Impact of mechanical ventilation on build-up and concentration distribution inside a 1-m^3 enclosure considering hydrogen energy applications conditions of use experiments and modelling. In: 8th international conference on hydrogen safety; 2019. Australia: Adelaide.
[21] Prasad K, Pitts WM, Yang JC. A numerical study of the release and dispersion of a buoyant gas in partially confined spaces. Int J Hydrogen Energy 2011. https://doi.org/10.1016/j.ijhydene.2011.01.118.
[22] Cariteau B, Brinster J, Tkatschenko I. Experiments on the distribution of concentration due to buoyant gas low flow rate release in an enclosure. Int J Hydrogen Energy 2011;36:2505–12. https://doi.org/10.1016/j.ijhydene.2010.04.054.
[23] Tamura Y, Takeuchi M, Sato K. Effectiveness of a blower in reducing the hazard of hydrogen leaking from a hydrogen-fueled vehicle. Int J Hydrogen Energy 2014;39:20339–49. https://doi.org/10.1016/j.ijhydene.2014.03.231.
[24] Liu W, Christopher DM. Dispersion of hydrogen leaking from a hydrogen fuel cell vehicle. Int J Hydrogen Energy 2015;40:16673–82. https://doi.org/10.1016/j.ijhydene.2015.10.026.
[25] Xie H, Li X, Christopher DM. Emergency blower ventilation to disperse hydrogen leaking from a hydrogen-fueled vehicle. Int J Hydrogen Energy 2015;40:8230–8. https://doi.org/10.1016/j.ijhydene.2015.03.146.
[26] Hussein H, Brennan S, Molkov V. Dispersion of hydrogen release in a naturally ventilated covered car park. Int J Hydrogen Energy 2020;45:23882–97. https://doi.org/10.1016/j.ijhydene.2020.06.194.
[27] Ehrhart BD, Harris SR, Blaylock ML, Muna AB, Quong S. Assessment and ventilation modeling for hydrogen vehicle repair garages. In: 8th international conference on hydrogen safety; 2019. Australia: Adelaide.
[28] Choi J, Hur N, Kang S, Lee ED, Lee KB. A CFD simulation of hydrogen dispersion for the hydrogen leakage from a fuel cell vehicle in an underground parking garage. Int J Hydrogen Energy 2013;38:8084–91. https://doi.org/10.1016/j.ijhydene.2013.02.018.
[29] Mukai JS, Mitsuishi H, Watanabe KOS. CFD simulation of hydrogen leakage caused by fuel cell vehicle accident in tunnel, underground parking lot and multistory parking garage. In: 19th tech. Conf. Enhanced safety of vehicles; 2005. Washington, DC.
[30] Hübert T, Boon-Brett L, Black G, Banach U. Hydrogen sensors - a review. Sensor Actuator B Chem 2011;157:329–52. https://doi.org/10.1016/j.snb.2011.04.070.
[31] Aalsalem MY, Khan WZ, Gharibi W, Khan MK, Arshad Q. Wireless sensor networks in oil and gas industry: recent advances, taxonomy, requirements, and open challenges. J Netw Comput Appl 2018;113:87–97. https://doi.org/10.1016/j.jnca.2018.04.004.
[32] Michaelides MP, Panayiotou CG. Plume source position estimation using sensor networks. In: IEEE international symposium on, mediterrean conference on intelligent control; 2005. Limassol, Cyprus.
[33] Alanis AY, Daniel NA, Franco CL. Artificial neural networks for engineering applications. 1st ed. New York: American academic press; 2019.
[34] Bouthiba T. Fault location in EHV transmission lines using artificial neural networks. Int J Appl Math Comput Sci 2004;14:69–78. https://doi.org/10.1109/TNS2.1962.4316016.
[35] Kim M, Choi CY, Gerba CP. Source tracking of microbial intrusion in water systems using artificial neural networks. Water Res 2008;42:1308–14. https://doi.org/10.1016/j.watres.2007.09.032.
[36] So W, Koo J, Shin D, Yoon ES. Optical sensor and neural networks for real-time monitoring and estimation of hazardous gas release rate. J Chem Eng Jpn 2010;43:682–90. https://doi.org/10.1252/jcej.10we050.
[37] Hartmann B, Link N. Gesture recognition with inertial sensors and optimized DTW prototypes. In: International

conference on systems. Istanbul, Turkey: IEEE; 2010. https://doi.org/10.1109/ICSMC.2010.5641703.

[38] Chen M, Zhao M, Huang T, Ji S, Chen L, Chang H, Li X. Measurements of helium distributions in a scaled-down parking garage model for unintended releases from a fuel cell vehicle. Int J Hydrogen Energy 2020;45:22166–75. https://doi.org/10.1016/j.ijhydene.2020.05.162.

[39] Makarov D, Molkov V. Plane hydrogen jets. Int J Hydrogen Energy 2013;27:8068–83. https://doi.org/10.1016/j.ijhydene.2013.03.017.

[40] Li X, Christopher DM, Hecht ES, Ekoto IW. Comparison of two-layer model for hydrogen and helium jets with notional nozzle model predictions and experimental data for pressures up to 35 MPa. Int J Hydrogen Energy 2017;42:7457–66. https://doi.org/10.1016/j.ijhydene.2016.05.214.

[41] Liu H, Zhou J, Zheng Y, Jiang W, Zhang Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. ISA Trans 2018;77:167–78. https://doi.org/10.1016/j.isatra.2018.04.005.

[42] Mitsa T. Temporal data mining. New York: Chapman and Hall/CRC; 2010. https://doi.org/10.1201/9781420089776.

[43] Costa BG, Freire JCA, Cavalcante HS, Homci M, Castro ARG, Viegas R, Meiguins BS, Morais JM. Fault classification on transmission lines using KNN-DTW. In: international conference on computational science and its applications; 2017. https://doi.org/10.1007/978-3-319-62392-4_13. Trieste, Italy.

[44] KNN classification using Scikit-learn. 2019. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn/. [Accessed 17 December 2019].

[45] Keogh EJ, Pazzani MJ. Derivative dynamic time warping. In: First SIAM international conference on data mining; 2001. https://doi.org/10.1137/1.9781611972719.1. Chicago, USA.