A hybrid deep belief network-based label distribution learning system for seismic damage estimation of liquid storage tanks

Men, Jinkun; Chen, Guohua; Reniers, Genserik; Rao, Xiaohui; Zeng, Tao

DOI
10.1016/j.psep.2023.02.079

Publication date
2023

Document Version
Final published version

Published in
Process Safety and Environmental Protection

Process Safety and Environmental Protection 172 (2023) 908–922

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

A hybrid deep belief network-based label distribution learning system for seismic damage estimation of liquid storage tanks

Jinkun Men [a,b], Guohua Chen [a,b,*], Genserik Reniers [c,d,e], Xiaohui Rao [a,b], Tao Zeng [a,b,d]

[a] Institute of Safety Science & Engineering, South China University of Technology, Guangzhou 510640, China
[b] Guangdong Provincial Science and Technology Collaborative Innovation Center for Work Safety, Guangzhou 510640, China
[c] Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), TU Delft, 2628 BX, Delft, the Netherlands
[d] CEDON, KULeuven, Campus Brussels, Brussels 1000, Belgium
[e] Faculty of Applied Economics, Antwerp Research Group on Safety and Security (ARGoSS), University Antwerp, Antwerp 2000, Belgium

ARTICLE INFO

Keywords:
Seismic damage estimation
Label ambiguity
Damage data imbalance
Liquid storage tanks
Label distribution learning

ABSTRACT

Liquid storage tanks play a vital role in the modern chemical process industry (CPI). The strong ground motion caused by large-scale earthquakes may easily impose severe structural damage on liquid storage tanks, leading to a series of catastrophic cascaded events. The seismic damage estimation of liquid storage tanks is a challenging problem, as the fluid-structure interaction exhibits extremely complicated and non-stationary response behavior. This study develops a novel data-driven methodology to estimate the seismic damage state probability distribution of liquid storage tanks in the contexts of label ambiguity and data imbalance. With the support of the advanced deep learning framework, synthetic oversampling methods, and label enhancement techniques, a hybrid deep belief network-based label distribution learning system (HDBN-LDLS) is proposed for probability distribution learning. The proposed HDBN-LDLS is evaluated on the widely used ALA database. Simulation results indicate that HDBN-LDLS can achieve a balanced estimation for all damage states while maintaining sufficient robustness to cope with label ambiguity. The reliability of the obtained data-driven model is validated by a damaged tank in the 2006 Silakhor earthquake. For practical applications, a more natural way to estimate a seismic damaged tank is to assign a membership degree to each possible damage state. The proposed methodology can quickly obtain the seismic damage state probability curves of a specific liquid storage tank, which can be used to support quantitative risk assessment and seismic design.

1. Introduction

Liquid storage tanks are one of the common storage facilities in the modern chemical process industry (CPI), which can be used to contain various products such as diesel, gasoline, liquefied natural gas, and other hazardous chemicals (Huang et al., 2022a; Men et al., 2022a; Amin et al., 2019). The functionality of liquid storage tanks is essential for the safe operation of a wide range of petrochemical and processing operations (Berahman and Behnamfar, 2009; Wang et al., 2022). Earthquakes may rapidly lead to a series of loss of containment (LOC) events in chemical tank farms, causing massive fires, explosions, or toxic cloud emissions (Antonioni et al., 2007; Yang et al., 2020; Men et al., 2022b; Meng et al., 2015). For example, a large-scale earthquake of magnitude 7.4 occurred off the Kocaeli, Turkey, on 17 August 2011 (Girgin, 2011a). As shown in Fig. 1, many storage tanks at the TUPRAS Izmit refinery were damaged and triggered massive fires (Scawthorn and Johnson, 2000). Damages in liquid storage tanks due to seismic events have been widely reported in the existing literature (Chakraborty et al., 2018; D'Amico and Buratti, 2019; Krausmann and Cruz, 2013; Ricci et al., 2021; Sezen and Whittaker Andrew, 2006), which highlighted the importance of liquid storage tanks in performance-based earthquake engineering (PBEE). Accordingly, many scholars (Men et al., 2022b; Salzano et al., 2003; Fabbrocino et al., 2005; Bakalis et al., 2018) have pointed out seismic hazards should be integrated into the quantitative risk assessment of industrial facilities.

As one of the core tasks of in process safety and risk management, the seismic damage estimation of liquid storage tanks has received considerable attention in recent years (Antonioni et al., 2007; Zuluaga Mayorga et al., 2019). As mentioned by Marta et al (D'Amico and Buratti, 2019)., most of the related studies focus on simulating and analyzing the

* Correspondence to: South China University of Technology, No.381, Wushan Rd., Tianhe District, Guangzhou 510640, China.
E-mail address: mmghchen@scut.edu.cn (G. Chen).

https://doi.org/10.1016/j.psep.2023.02.079
Received 30 August 2022; Received in revised form 14 January 2023; Accepted 26 February 2023
Available online 1 March 2023

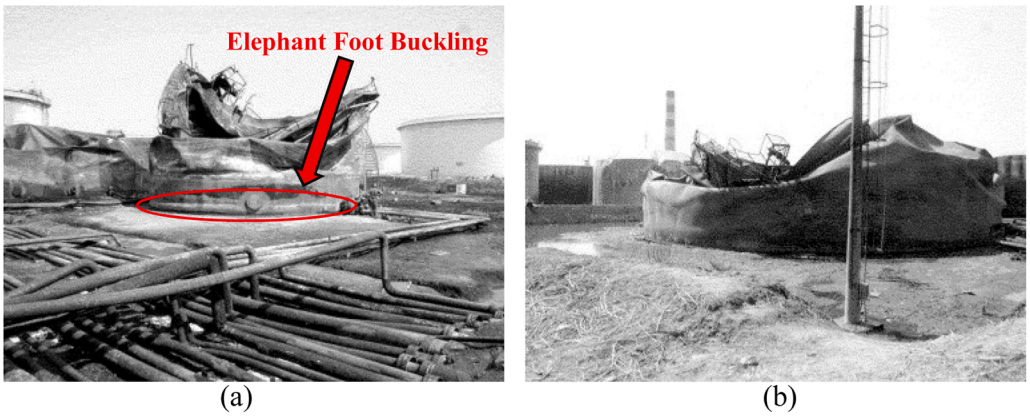

Fig. 1. Tank damage at TUPRAS Izmit refinery (Scawthorn and Johnson, 2000). (a) The elephant foot buckling caused by the earthquake; (b) Tanks collapsed due to fire.

Table 1
Description of five seismic damage states (Bezir et al., 2022; D'Amico and Buratti, 2019; O'Rourke and So, 2000; Alliance, 2001b; Alliance, 2001a).

Notations	Damage State	Physical Damage Description
DS_1	No damage	No damage to tank structure and accessories
DS_2	Minor	Damage to roof, minor loss of content, minor shell damage, minor piping damage, cracked foundation of the tank, no buckling
DS_3	Moderate	Buckling with no leak or minor loss of contents
DS_4	Severe	Elephant foot buckling with major loss of contents, severe damage
DS_5	Collapsed	Total failure, tank structure collapse

fluid-structure interaction system (Zhou and Zhao, 2021; Bakalis and Karamanos, 2021; Ozdemir et al., 2010) or assessing the seismic fragility of liquid storage tanks (D'Amico and Buratti, 2019; Lee et al., 2019; Salzano et al., 2003; Zuluaga Mayorga et al., 2019). Given the stochastic nature of accident evolution, probabilistic fragility is essential to understand the response behavior of liquid storage tanks under extreme circumstances (D'Amico and Buratti, 2019; Men et al., 2022c; Yang et al., 2020; Zuluaga Mayorga et al., 2019). Research on fragility analysis is abundant and most probabilistic models were driven by the post-earthquake damage data (D'Amico and Buratti, 2019; Salzano et al., 2003; Vílchez et al., 2001; Alliance, 2001b; Alliance, 2001a) or the analytical and numerical approximations (Berahman and Behnamfar, 2009; Fabbrocino et al., 2005; Zuluaga Mayorga et al., 2019; Lee et al., 2019; Saha et al., 2016; Gabbianelli et al., 2022). An overview of representative studies for the seismic fragility of liquid storage tanks is shown in the Appendix. Table A1. Many useful statistical methods such as probit regression (Salzano et al., 2003; Fabbrocino et al., 2005), logistic regression (Yang et al., 2020; Gabbianelli et al., 2022; O'Rourke and So, 2000), Monte-Carlo simulation (Zuluaga Mayorga et al., 2019; Saha et al., 2016; Huang et al., 2022b), and Bayesian analysis (Berahman and Behnamfar, 2009; D'Amico and Buratti, 2019) were widely used to obtain parametric fragility curves. Despite the increasing attention devoted to this research field, related studies are still limited by the imperfections of post-earthquake damage data (Bezir et al., 2022; D'Amico and Buratti, 2019; Men et al., 2022c).

Following the performance-based earthquake engineering (PBEE) framework (Krawinkler, 2000; Ghosh et al., 2017), the damage states are usually divided into five levels (DS_1*: No damage,* DS_2*: Minor damage,* DS_3*: Moderate damage,* DS_4*: Severe damage,* DS_5*: Collapsed*), and the damage description shown in Table 1 has been widely used to capture the damage states of liquid storage tanks (Bezir et al., 2022; D'Amico and Buratti, 2019; O'Rourke and So, 2000; Alliance, 2001b; Alliance, 2001a). In real applications, the corresponding damage states were assigned to damaged tanks according to the experts' opinions on the physical damage during past earthquake (Saha et al., 2016; Gao et al., 2017; Geng, 2016). Such manual labeling method is difficult to collect sufficient samples with precise damage states. Given the fuzzy nature of natural language variables, qualitative terms such as *minor*, *moderate*, *severe*, etc., lead to the significant label ambiguity. Uncertainties associated with tank properties, hazard characteristics, measurement errors, incomplete information, and modeling errors also impose significant limitations on the above crisp models (Berahman and Behnamfar, 2009; D'Amico and Buratti, 2019; Saha et al., 2016).

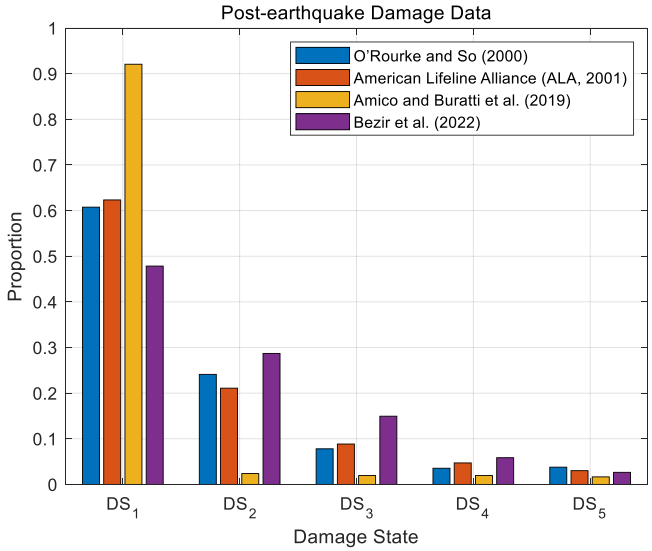

Fig. 2. Proportion of each damage state in common databases (1). O'Rourke and So (O'Rourke and So, 2000); (2). (Alliance, 2001a, 2001b); (3). (D'Amico and Buratti, 2019).; (4). (Bezir et al., 2022).).

Moreover, from the perspective of data analysis, the size and distribution of sample space significantly affect the reliability of fragility analysis (Bezir et al., 2022; D'Amico and Buratti, 2019; Hu and Jiang, 2019). As shown in Fig. 2, most of the existing databases (Bezir et al., 2022; D'Amico and Buratti, 2019; O'Rourke and So, 2000; Alliance, 2001b; Alliance, 2001a) show the characteristics of imbalance, which is rarely noticed by related studies. It leads to the probabilistic models failing to assign the same attention to minority damage states as the majority and further leads to the lack of generalization ability, although the overall performance is considerable (Hu and Jiang, 2019; Debowski

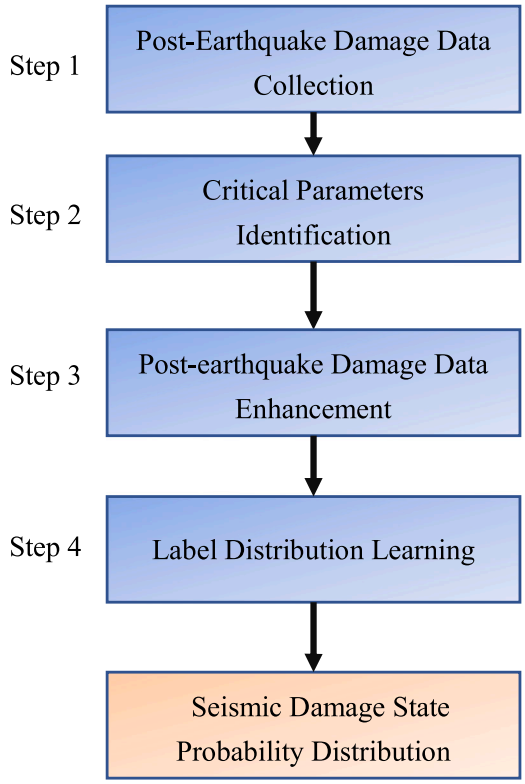

Fig. 3. The flowchart of the proposed methodology.

et al., 2012; Gu et al., 2021). Due to the low frequency, samples with high damage states are often difficult to estimate. Meanwhile, these low-probability high-consequence events may lead to catastrophic interlocking chemical accidents (Men et al., 2022b).

The central theme of this work is an expansion of fragility analysis, which aims to provide the probabilistic mapping relationship between structural physical damage and seismic intensity, theoretical support required for QRA. Two tough challenges caused by the imperfect post-earthquake damage data are especially considered: label ambiguity and data imbalance. To solve the research gaps mentioned above, a hybrid deep neural network-based label distribution learning system (HDBN-LDLS) is proposed, of which the core is the enhancement of post-earthquake damage data. To be specific, a broad artificial database is established to modify the imbalanced post-earthquake damage data and fill in the missing crucial information. A Niche-based synthetic over-sampling method is proposed to modify the imbalanced data. The label propagation algorithm is adopted to convert logical labels to label distributions. Compared with the traditional crisp data-driven models, the proposed methodology deals with label ambiguity by allowing the answer to consist of more than one damage state. Finally, the advanced deep learning approach (Low et al., 2020; Arunthavanathan et al., 2021) is adopted to estimate the unknown probability distribution over the complex multivariate factors.

The novelty of this study mainly has three aspects.

Firstly, a novel data-driven methodology is proposed to estimate the seismic damage state probability distribution of liquid storage tanks, which can achieve a balanced estimation for all damage states while maintaining sufficient robustness to cope with label ambiguity. Compared with traditional crisp models, the proposed methodology can provide a more natural mapping relation of physical damage to probability space.

Secondly, the obtained seismic damage state probability curves provide a new and comprehensive perspective for seismic fragility analysis. Five compatible damage states are covered by the seismic damage state probability curves. The traditional crisp models focused on solving the issue "*which damage state can represent the physical damage*?", while the proposed methodology deals with damage scale ambiguity by allowing the answer to consist of more than one damage state. Once the seismic damage state probability distribution is quantified, consequence analysis can be performed for those subsequent technological hazards such as fires, explosions and toxic gas emissions.

Finally, the intersection of the probability curve corresponding to each seismic damage state is defined as the critical damage point, which can characterize the seismic performance of liquid storage tanks more conveniently. The proposed methodology has the potential to be a promising tool for the seismic design of liquid storage tanks.

The rest of this paper is organized as follows. The proposed methodology is developed in Section 2. The performance of the proposed methodology is evaluated in Section 3 as well as some discussions about the methodology applications. This work is concluded in Section 4. This work includes an additional Supplemental material. Some preliminaries about seismic response characteristics, seismic damage characterization, and material properties are available in the Supplemental material.

Table 2
Critical parameters affecting seismic performance of liquid storage tanks.

Notations	Definitions
D	Tank Diameter, m
H	Tank Height, m
t_r	Roof Thicknesses, mm
t_s	Shell Thicknesses[a], mm
t_b	Base Thicknesses, mm
t_a	Annular Ring Thicknesses, mm
R_t	Roof Type, 1. Floating Roof Tank; 2. Fixed Roof Tank
S_t	Steel Type
A_S	Anchorage System, 1. Self-anchored; 2. Mechanically-anchored
h	Height of Liquid Level, m
ρ_L	Density of Stored Liquid, kg/m^3
S_p	Peak Ground Acceleration, %g[b]
SC	Soil Classification[c]

[a] In this work, the shell thickness of the bottom plate is of particular concern. the thicknesses of other courses are assumed to the same as the bottom plate.
[b] Acceleration due to gravity in consistent units, m/s^2
[c] API 650 differentiates soil types in six different "site classes" which range from site class A to F.

2. Methodology

The proposed methodology consists of four main steps.

Step 1 Post-Earthquake Damage Data Collection: The post-earthquake damage data of liquid storage tanks is first collected, which mainly includes tank properties and seismic intensity. Each sample is assigned to a specific damage state according to the experts'

Table 3
Seismic damage parameters.

Notations	Definitions
F_u	Uplift Force of Tank Bottom Perimeter, N/m
F_w	Force Resisting Uplift in Tank Bottom, N/m
F_L	Force Resisting Uplift in Annular Region, N/m
J	Anchorage Ratio[a]
σ_C	Maximum Longitudinal Shell Compression Stress, MPa
σ_T	Hoop Tensile Stress, MPa
F_C	Allowable Longitudinal Shell-membrane Compression Stress, MPa
Sl_h	Sloshing Wave Height, m

[a] According to the API 650, $J < 0.785$ indicates the tank is self-anchored; $0.785 < J \leq 1.54$ indicates the tank is uplifting but stable; $J > 1.54$ indicates the tank is not stable and cannot be self-anchored.

Table 4
Distributions of random parameters.

Parameters	Distribution	Reference
D, H, t_s, t_r, t_b	Discrete	Information of 653 tanks are collected from previous studies (Yang et al., 2020; Spritzer and Guzey, 2017; Miladi and Razzaghi, 2019)
S_t	Discrete	Seven steel types listed in Chinese Standard GB 50341–2014 (P. Ministry of Housing and Urban-Rural Development, 2014). Material properties are stated in theSupplemental Material. B.
h	Uniform (0, 0.95H)	(Spritzer and Guzey, 2017; Miladi and Razzaghi, 2019)
ρ_L	Uniform (750, 1000)	(Yang et al., 2020)
t_a	Uniform ($t_a^{\min}$, t_s)	The value of $t_a^{\min}$ can be obtained according to theSupplemental Material. D. Table. D. 1.
S_p	Uniform (0, 1.5)	(Zuluaga Mayorga et al., 2019; Alliance, 2001a, b)
SC	Discrete (A, B, C, D, E, F)	(Spritzer and Guzey, 2017)

opinions on the physical damage during past earthquake (Saha et al., 2016; Gao et al., 2017; Geng, 2016).

Step 2 Critical Parameters Identification: Thirteen critical parameters affecting the seismic performance of liquid storage tanks and seven seismic damage parameters are identified and summarized.

Step 3 Post-earthquake Damage Data Enhancement: The enhancement of post-earthquake damage data is the core of the proposed methodology. To be specific, a broad artificial database is established to modify the imbalanced post-earthquake damage data and fill in the missing crucial information. The label propagation algorithm is adopted to convert the original logical labels to the robust label distribution.

Step 4 Label Distribution Learning: The obtained label distribution database is regarded as the input layer of the proposed hybrid deep neural network-based label distribution learning system. Through the training process, the seismic damage state probability distribution of liquid storage tanks can be obtained. (Fig. 3).

2.1. Post-earthquake damage data collection

Liquid storage tanks are often located in highly seismic regions, such as the case of oil storage facilities placed along the coasts of countries like Japan, California, Peru, Alaska, and Turkey (D'Amico and Buratti, 2019). Several studies have constructed post-earthquake damage databases from different sources (Bezir et al., 2022; D'Amico and Buratti, 2019; O'Rourke and So, 2000; Alliance, 2001b; Alliance, 2001a), which provide strong support for evaluating the seismic fragility of liquid storage tanks. However, most databases are not publicly available. In this work, the open-source ALA post-earthquake damage database (Alliance, 2001a,b) is adopted to demonstrate the proposed methodology, as it is publicly available and the reliability of the data is acceptable. This post-earthquake damage database consists of 531 liquid storage tanks in 21 earthquakes, and the observational information includes: (1). *Tank Diameter*, (2). *Tank Height*, (3). *Maximum Design Product Level*, (4). *Aspect Ratio* (ratio of tank height to tank radius), (5). *Filling Ratio*, (6). *Peak Ground Acceleration*. Each tank is labeled with a specific damage state based on the definition shown in Table 1. It is worth mentioning that the proposed methodology is not limited to the ALA database (Alliance, 2001a,b), as many other databases also show the characteristics of label ambiguity and data imbalance.

2.2. Identification of crucial parameters

In this work, the cylindrical welded steel tanks in CPI are especially concerned, which can be regarded as a typical fluid-structure interaction system (Ozdemir et al., 2010; Hernandez-Hernandez et al., 2021). As mentioned by Ozdemir et al. (2010), the seismic damage estimation of liquid storage tanks is extremely complex since the fluid-structure interaction system possesses many different nonlinear behavior mechanisms which may be triggered simultaneously or independently depending on the several factors, such as, characteristics of earthquake, contained liquid properties and its depth, dimensions of the tank, material properties and supporting conditions and stiffness of underlying soil medium.

Some preliminaries about the seismic response characteristics of cylindrical steel liquid storage tanks under the excitation of ground motion are stated in the Supplemental Material. A. Through summarizing, aggregating, organizing, and comparing the knowledge extracted from the existing studies (Scawthorn and Johnson, 2000; Bakalis and Karamanos, 2021; Housner, 1963, 1957; Spritzer and Guzey, 2017; API, 2020; Phan et al., 2019), thirteen critical parameters affecting the seismic performance of liquid storage tanks are extracted and stated in Table 2. We cannot guarantee that the extracted parameters can represent all the factors affecting the seismic performance of the storage tank. However, from the widely available literature, the extracted parameters

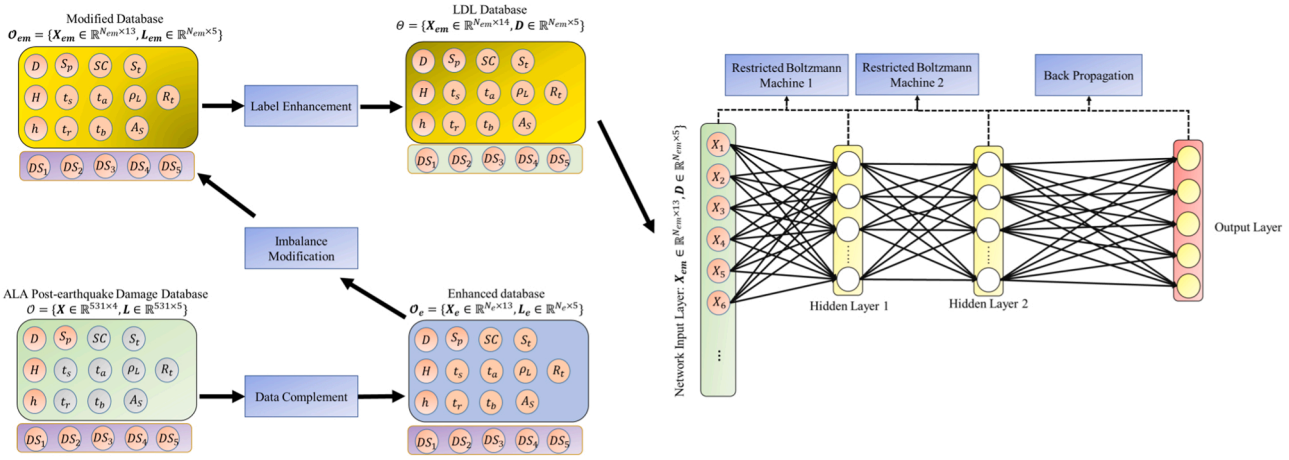

Fig. 4. Structure of the hybrid deep neural network-based label distribution learning system for estimating seismic damage state probability distribution of liquid storage tanks.

are sufficiently representative.

Based on the seismic response analysis, various design codes are developed to guide the seismic design of storage tanks, such as the Chinese Standard GB 50341–2014 Annex D (P. Ministry of Housing and Urban-Rural Development, 2014), the American Standard API 650 Annex E (API, 2020), NZSEE Recommendation (NZSEE, 2009), etc. The following eight parameters are usually adopted to characterize the seismic damage. (Table 3)

2.3. Post-earthquake damage data enhancement

Although a lot of efforts have been made in the research on the seismic fragility of liquid storage tanks (D'Amico and Buratti, 2019; Lee et al., 2019; O'Rourke and So, 2000; Saha et al., 2016), existing studies are still limited by the imperfections of past databases. Thus, in this work, the enhancement of post-earthquake damage data is the core of the proposed methodology. In light of the above-mentioned limits, the enhancement of post-earthquake damage data is mainly carried out from the following three aspects.

2.3.1. Data complement

$\mathscr{O} = \{\boldsymbol{X} \in \mathbb{R}^{N\times Q}, \boldsymbol{L} \in \mathbb{R}^{N\times P}\}$ ($N = 531$ is the number of instances; $Q = 4$ is the dimension of the feature vector; $P = 5$ is the number of potential damage states) is the ALA post-earthquake damage database (only *Tank Diameter D*, *Tank Height H*, *Liquid level h*, *Peak Ground Acceleration* S_p are available.). The enhanced database is denoted as $\mathscr{O}_{\boldsymbol{e}} = \{\boldsymbol{X_e} \in \mathbb{R}^{N_e\times Q_e}, \boldsymbol{L_e} \in \mathbb{R}^{N_e\times P}\}$ (N_e is the number of instances; $Q_e = 13$ is the dimension of the feature vector). In the initialization phase, the enhanced database $\mathscr{O}_{\boldsymbol{e}}$ is assumed to be empty, and a broad artificial database $\mathscr{O}_{\boldsymbol{h}}$ is constructed. The feature vector $\boldsymbol{X}_h \in \mathbb{R}^{N_h\times Q_e}$ of $\mathscr{O}_{\boldsymbol{h}}$ is generated by random sampling (Yang et al., 2020; Zuluaga Mayorga et al., 2019), the number of artificial instances N_h is set to be ten times N (Zhang and Qiu, 2022; Gan et al., 2022; Sawant and Prabukumar, 2020). The missing information of the ALA post-earthquake damage database can be obtained by random sampling (Yang et al., 2020; Zuluaga Mayorga et al., 2019), and the generated database $\mathscr{O}_{\boldsymbol{m}} = \{\boldsymbol{X}_m \in \mathbb{R}^{N\times Q_e}, \boldsymbol{L} \in \mathbb{R}^{N\times P}\}$ is obtained. For each generated instance, the *label sharing assumption* is made. Thus, the damage state of each generated instance is assumed to be the same as the original ALA instance.

Algorithm 1. . Data Complement.

Input: The ALA post-earthquake damage database $\boldsymbol{\mathcal{O}} = \{\boldsymbol{X} \in \mathbb{R}^{N\times Q}, \boldsymbol{L} \in \mathbb{R}^{N\times P}\}$
Output: The enhanced database $\boldsymbol{\mathcal{O}_e} = \{\boldsymbol{X_e} \in \mathbb{R}^{N_e\times Q_e}, \boldsymbol{L_e} \in \mathbb{R}^{N_e\times P}\}$
Initialization:
Construct an empty database $\boldsymbol{\mathcal{O}_e}, N_e = 0$
Construct a broad artificial database $\boldsymbol{\mathcal{O}_h} = \{\boldsymbol{X}_h \in \mathbb{R}^{N_h\times Q_e}, \boldsymbol{S_h} \in \mathbb{R}^{N_h\times S}\}$
While $N_e \le N_{max}$ ***Do***
  Data Generation: $\boldsymbol{\mathcal{O}_m} \leftarrow RSam(\mathcal{O}), \boldsymbol{\mathcal{O}_m} = \{\boldsymbol{X}_m \in \mathbb{R}^{N\times Q_e}, \boldsymbol{L} \in \mathbb{R}^{N\times P}\}$
  Seismic Response Analysis: $\boldsymbol{S_m} \in \mathbb{R}^{N\times S} \leftarrow SRA(\boldsymbol{X}_m \in \mathbb{R}^{N\times Q_e})$
  Damage Feature Space: $\boldsymbol{\mathcal{S}} = \boldsymbol{S_m} \cup \boldsymbol{S_h}$
  Clustering Analysis: $\boldsymbol{C} \leftarrow Kmeans(\boldsymbol{\mathcal{S}}, P)$
  For all $x \in \boldsymbol{X}_m \in \mathbb{R}^{N\times Q_e}$ ***Do***
    If $\boldsymbol{C_x} = \boldsymbol{L_x}$ ***Do***
      Update $\boldsymbol{\mathcal{O}_e}$ *with* $\{x, \boldsymbol{L_x}\}$
      $N_e = N_e + 1$
    End If
  End For
End While

The original database (Alliance, 2001a,b) may only contains limited information. Many critical parameters affecting the seismic performance of liquid storage tanks are not available. The reliability of fragility models driven by such incomplete databases cannot be guaranteed (Bezir et al., 2022; D'Amico and Buratti, 2019). To facilitate the representation, the database obtained by the data complement process is denoted as the enhanced database. The database is required to include all critical parameters shown in Table 2.

In this section, a data complement algorithm is proposed to fill in the missing information in the original database, while the information contained in the original database should be retained as far as possible. The missing information is regarded as random parameters and the random distributions of these parameters are stated in Table 4.

The data complement process is stated in Algorithm 1. Suppose that

For each instance x, the corresponding seismic damage feature $s_x = \{F_u^x, F_w^x, F_L^x, J^x, \sigma_C^x, \sigma_T^x, F_C^x, S_h^x\}$ can be obtained. Detailed calculations of these seismic damage parameters can be found in the Supplemental Material B. Thus, we have the seismic damage vector $\boldsymbol{S_h} \in \mathbb{R}^{N_h\times S}$ for the artificial database $\mathscr{O}_{\boldsymbol{h}}$, the $\boldsymbol{S_m} \in \mathbb{R}^{N\times S}$ for the original database $\mathscr{O}$. The union of S_m and S_h constitutes an extensive feature space damage feature space $\mathcal{S} = \boldsymbol{S_m} \cup \boldsymbol{S_h}$. The seismic damage parameters can be used to assist to evaluate the deviation degree between the randomly generated information and the actual situation. To be specific, the widely used K-means algorithm (Brus et al., 2006) is adopted for clustering analysis of $\boldsymbol{\mathcal{S}}$. The predefined number of cluster centers is set to five, and the clustering labels $\boldsymbol{C}$ can be obtained. For each generated instance $x \in \boldsymbol{X}_m \in \mathbb{R}^{N\times Q_e}$, when its clustering label $\boldsymbol{C}_x$ is consistent with the

original label $\boldsymbol{L_x}$, the label sharing assumption is considered to be valid. Instances that satisfy the label sharing assumption largely retain information of the original data. Thus,$\{x, \boldsymbol{L_x}\}$ is introduced into the enhanced database $\mathscr{O}_{\boldsymbol{m}}$.$N_{\max}$ is a hyper-parameter, which is used to limit the number of the enhanced instances. The above process is repeated until a sufficient enhanced database is obtained.

2.3.2. Imbalance modification

As mentioned above, imbalanced data is one of the major challenges in estimating seismic damage state probability distribution of liquid storage tanks. It leads to the data-driven model failing to assign the same attention to minority damage states as the majority and further leads to the lack of generalization ability, although the overall performance is considerable (Hu and Jiang, 2019).

Algorithm 2. . Imbalance Modification.

Input:
The imbalanced database $\boldsymbol{\mathcal{O}_e} = \{\boldsymbol{X_e} \in \mathbb{R}^{N_e \times Q_e}, \boldsymbol{L_e} \in \mathbb{R}^{N_e \times P}\}$
Predefined rarity threshold μ
Output: The modified database $\boldsymbol{\mathcal{O}_{em}} = \{\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em} \times Q_e}, \boldsymbol{L_{em}} \in \mathbb{R}^{N_{em} \times P}\}$
Initialization: $\boldsymbol{\mathcal{O}_{em}} \coloneqq \boldsymbol{\mathcal{O}_e}$
For all $\{x, \boldsymbol{L_x}\} \in \boldsymbol{\mathcal{O}_e}$ ***Do***
  Rarity Calculation: $r(x) = \sum_{x' \in N_k(x), x' \neq x} sf(x, x')$
  While $r(x) \leq \mu$ ***Do***
    Synthetic Oversampling: $x_{new} = x + \alpha(x_{neo} - x), x_{neo} \in N_k(x)$
    $\boldsymbol{L}_{x_{new}} \leftarrow KNN(x_{new}, \boldsymbol{\mathcal{O}_{em}})$
    If $\boldsymbol{L}_{x_{new}} = \boldsymbol{L_x}$ ***Do***
      Update $\boldsymbol{\mathcal{O}_{em}}$ ***with*** $\{x_{new}, \boldsymbol{L}_{x_{new}}\}$
      Update $r(x)$
    End If
  End While
End For

However, in CPI, rare events often cause catastrophic consequences (Scawthorn and Johnson, 2000; Krausmann and Cruz, 2013; Zhou and Zhao, 2021; Girgin, 2011b). The synthetic oversampling methods (Hu and Jiang, 2019; Chawla et al., 2002; Han et al., 2005) are widely used to deal with imbalanced data, but are often limited by information loss and self-adaption. Thus, in this work, a Niche-based synthetic oversampling method is proposed to modify the imbalanced data. The imbalance modification process is stated in Algorithm 2. To facilitate the representation, the database obtained by the imbalance modification process is denoted as the modified database.

The Niche technique (Goldberg and Richardson, 1987) is adopted to identify the rarity of instances. For $\forall\{x, \boldsymbol{L_x}\} \in \mathscr{O}_{\boldsymbol{e}}$, its rarity $r(x)$ can be calculated as follows:

$$r(x) = \sum_{x' \in N_k(x), x' \neq x} sf(x, x') \tag{1}$$

where $N_k(x)$ is the neighborhood of x; the rarity of x is defined as the sum of the similarity between x and other samples in the neighborhood $\forall x' \in N_k(x) \wedge x' \neq x$. The sharing function $sf(x, x')$ is used to describe the similarity between x and x' Goldberg and Richardson (1987). Suppose that μ is the predefined rarity threshold, for all the instances $x \in \boldsymbol{X_e}$, $r(x) \leq \mu$, the synthetic oversampling method (Hu and Jiang, 2019; Chawla et al., 2002; Han et al., 2005) is adopted to generate artificial instances.

$$x_{new} = x + \alpha(x_{neo} - x), x_{neo} \in N_k(x) \tag{2}$$

where x_{neo} is a random sample from the neighborhood of x;$\alpha \in (0, 1]$ is a random variable. The damage state of the generated artificial instance x_{new} can be obtained by the k-nearest neighborhood (KNN) algorithm (Pourbahrami et al., 2020). If x_{new} and x have the same fault pattern, then x_{new} is included in enhanced database $\mathscr{O}_{\boldsymbol{e}}$. This process is repeated until the rarities of all instanses are greater than the threshold. Finally, the modified database is denoted as $\mathscr{O}_{\boldsymbol{em}} = \{\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em} \times Q_e}, \boldsymbol{L_{em}} \in \mathbb{R}^{N_{em} \times P}\}$.

2.3.3. Label enhancement

Label distribution learning (LDL) is a novel learning paradigm, which labels an instance with a label distribution and learns a mapping from instance to label distribution straightly (Xu et al., 2021). In most of the post-earthquake damage data, an instance x is assigned with $l_x^{DS_i} \in \{0, 1\}$ to each potential damage state DS_i, representing whether DS_i describes x. To be specific, $l_x^{DS_i}$ is denoted as the logical label, which reflects the logical relationship between the damage state and the instance. Such crisp classification answers the essential question "*which damage state can describe the instance*", but does not involve the explicit relative importance of each damage state (Gao et al., 2017; Geng, 2016; Xu et al., 2021).

Following the concept of LDL learning paradigm, in this work, an instance x is to assign a real number $d_x^{DS_i} \in [0, 1]$ to each potential damage state DS_i. As shown in Table 1, the definitions of five damage states are derived from qualitative language variables such as minor, moderate, severe, etc. The fuzzy nature of these qualitative language variables leads to the significant label ambiguity. Thus, the term "*membership degree*" in fuzzy theory is adopted. Thus, $d_x^{DS_i}$ is denoted as the membership degree of DS_i to x. The higher the value of $d_x^{DS_i}$, the higher the degree to which x belongs to the damage state DS_i. As

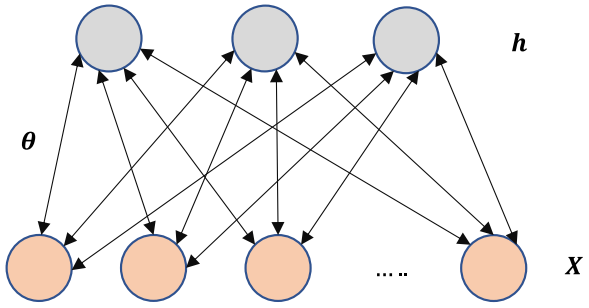

Fig. 5. Structure of the restricted Boltzmann machine.

Table 5
The composition of enhanced databases obtained from ten independent runs.

No.	N_{DS_1}	N_{DS_2}	N_{DS_3}	N_{DS_4}	N_{DS_5}	N_e	L_r
1	522	83	31	15	14	665	3.201e-2
2	532	99	29	16	12	688	3.951e-2
3	490	91	31	14	17	643	4.122e-2
4	512	82	27	15	13	649	2.823e-2
5	533	89	29	14	12	677	3.012e-2
6	484	94	34	13	13	638	3.213e-2
7	512	87	33	12	11	655	3.581e-2
8	498	86	36	13	16	649	2.811e-2
9	514	82	31	14	13	654	2.931e-2
10	524	87	32	12	15	670	3.012e-2

mentioned above, the set of potential damage state can be defined as $DS = \{DS_1, DS_2, DS_3, DS_4, DS_5\}$. To satisfy the form of seismic damage state probability distribution, we have $\sum_{DS} d_x^{DS_i} = 1$. Given the database $\mathscr{O}_{em} = \{\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em}\times Q}, \boldsymbol{L_{em}} \in \mathbb{R}^{N_{em}\times P}\}$, label enhancement recovers the damage state distribution d_x of $\boldsymbol{x} \in \boldsymbol{X_{em}}$ from the logical label vector $l_x \in \{0,1\}^{\boldsymbol{P}}$, and thus transforms $\mathscr{O}_{em}$ into a LDL data set $\Theta = \{\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em}\times Q_e}, \boldsymbol{D} \in \mathbb{R}^{N_{em}\times P}\}$. There are several algorithms designed for label enhancement (Xu et al., 2021).

In this work, the label propagation (LP) algorithm is adopted. The basic idea of LP is to iteratively update the original logical labels by the following similarity matrix $\boldsymbol{S} = (s_{ij})_{N\times N}$.

$$s_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right) if\, i \neq j \\ 0, if\, i = j \end{cases} \tag{3}$$

Following the structure of *Laplacian matrix*, the label propagation matrix $\mathcal{P}$ is constructed as follows:

$$\mathcal{P} = \widehat{\boldsymbol{S}}^{-\frac{1}{2}} \boldsymbol{S} \widehat{\boldsymbol{S}}^{-\frac{1}{2}} \tag{4}$$

where $\widehat{\boldsymbol{S}}$ is a diagonal matrix with the elements $\widehat{s}_{ij} = \sum_{j=1}^{N} s_{ij}$. Suppose that $\boldsymbol{D}$ is the label distribution matrix, at iteration g, $\boldsymbol{D}$ is updated as follows:

$$\boldsymbol{D}^{(g)} = \boldsymbol{\alpha}\mathcal{P}\boldsymbol{D}^{(g-1)} + (1-\alpha)\boldsymbol{L} \tag{5}$$

where the initial $D^{(0)}$ is set as the logical label matrix $\boldsymbol{L}$; $\alpha \in (0,1)$ is the balancing parameter. Finally, $\boldsymbol{D}^{(g)}$ will converge to $\boldsymbol{D}^*$, and the softmax normalization is adopted to normalize the label distribution matrix.

2.4. Label distribution learning

A hybrid deep neural network-based label distribution learning system (HDBN-LDLS) is developed to estimate the seismic damage state probability distribution of liquid storage tanks. The structure of the proposed HDBN-LDLS is shown in Fig. 4. The obtained label distribution database is regarded as the input layer of HDBN-LDLS. The training process is driven by the deep belief network (DBN) (Cao et al., 2022). As one of the advanced deep learning technologies, the DBN can overcome the defects of the traditional shallow neural network due to its powerful linear approximation (Cao et al., 2022; Li et al., 2022). The DBN is a probabilistic generative model that contrasts with the discriminative nature of traditional neural networks (Low et al., 2020; Cao et al., 2022).

As shown in Fig. 4, DBN is a deep network obtained by stacking multiple restricted Boltzmann machine (RBM) and adding a Back Propagation neural network (BP-NN) into the last layer to accept the output vector of the topmost RBM (Tian et al., 2022). Following the basic framework of DBN, the training process of the proposed HDBN-LDLS is divided into two sub-processes: pretraining process and fine-tuning process.

2.4.1. Pretraining process

The pretraining process mainly adopts the RBM to map the feature

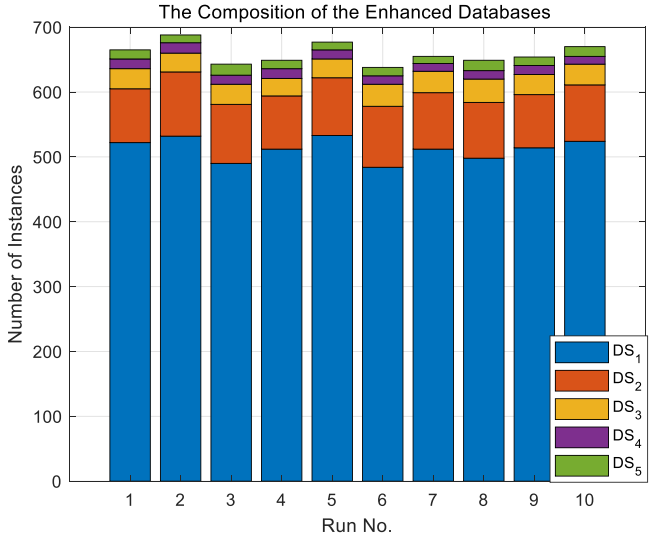

Fig. 6. The composition of the enhanced databases obtained from ten independent runs.

Table 6
The composition of the modified databases obtained from ten independent runs.

No.	N_{Ds_1}	N_{Ds_2}	N_{Ds_3}	N_{Ds_4}	N_{Ds_5}	N_{em}
1	511	334	221	131	112	1309
2	513	299	218	133	117	1280
3	519	311	232	127	96	1285
4	498	321	239	119	100	1277
5	502	331	232	122	105	1292
6	511	312	266	121	99	1309
7	531	311	222	113	110	1287
8	528	306	235	116	96	1281
9	528	321	232	112	106	1299
10	533	312	235	121	99	1300

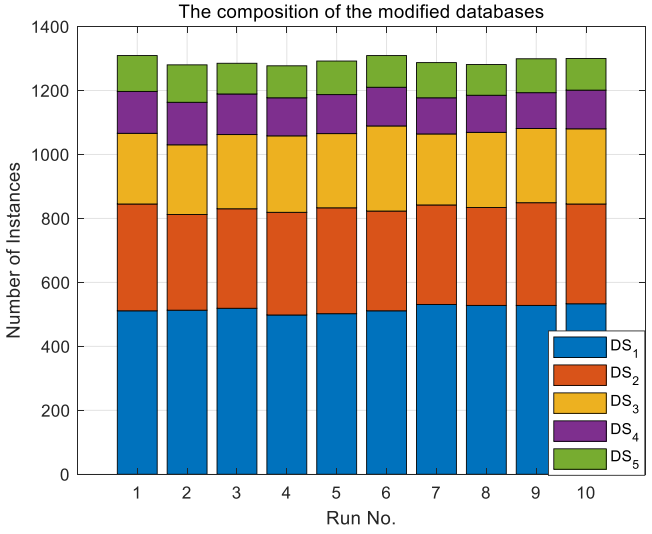

Fig. 7. The composition of the modified databases.

vector $\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em}\times 13}$ to different feature spaces, so as to retain as much feature information as possible. The RBM is an energy-based probabilistic model.

As shown in Fig. 5, an RBM usually consists of discrete visible units $\boldsymbol{x} \in \boldsymbol{X_{em}}$ and discrete hidden units $\boldsymbol{h}$ (Mohan et al., 2021). The joint probability distribution $(\boldsymbol{x}, \boldsymbol{h}) \sim P(\boldsymbol{x}, \boldsymbol{h}; \theta)$ is defined by the energy function $E(\boldsymbol{x}, \boldsymbol{h}; \theta)$ Mohan et al. (2021).

$$P(\boldsymbol{x}, \boldsymbol{h}; \theta) = \frac{e^{-E(\boldsymbol{x},\boldsymbol{h};\theta)}}{Z} \tag{6}$$

$$E(\boldsymbol{x}, \boldsymbol{h}; \theta) = -\boldsymbol{b}^T\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{h} - \boldsymbol{h}^T\boldsymbol{W}\boldsymbol{x} \tag{7}$$

where $\theta = \{\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{W}\}$ is the model parameter; $Z = Z = \sum_{\boldsymbol{x}}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{x},\boldsymbol{h};\theta)}$ is the partition function. During the pretraining process, the initial visible layer vector of RBM is the input training sample vector $\boldsymbol{X_{em}} \in \mathbb{R}^{N_{em}\times 13}$. Then, the hidden layer output vector calculated by $\{\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{W}\}$ can be input to the next RBM. The Contrastive Divergence (CD) algorithm can be adopted to train RBM layer by layer quickly (Ning et al., 2018). The whole pretraining process can be regarded as an unsupervised learning process, which make the DBN have superior performance in feature extraction and automatic data dimension reduction (Cao et al., 2022).

2.4.2. *Fine-tuning process*

The BP-NN is adopted to fine-tune the DBN model. The hidden layer output vector output vector of the final RBM is regarded as the input layer of BP-NN. The output of the BP-NN is compared with the label distribution matrix $\boldsymbol{D} \in \mathbb{R}^{N_{em}\times 5}$, thus the model error can be obtained. The back-propagation algorithm (Cao et al., 2022; Li et al., 2022) can be used to adjust the network parameter values, so as to minimize the model error. The obtained data-driven model can be used to estimate seismic damage state probability distribution of liquid storage tanks.

3. Methodology validation

The proposed methodology is verified in this section. Methodology validation is performed on Matlab R2021a under a computer that is equipped with Intel(R) Core(TM) i7–8750H CPU @2.20 GHz and 8.00 GB RAM.

3.1. *Training database*

Following the flowchart of the proposed methodology, the original ALA database $\mathcal{O} = \{\boldsymbol{X} \in \mathbb{R}^{531\times 4}, \boldsymbol{L} \in \mathbb{R}^{531\times 5}\}$ is first converted to the enhanced database $\mathscr{O}_e = \{\boldsymbol{X_e} \in \mathbb{R}^{N_e\times 13}, \boldsymbol{L_e} \in \mathbb{R}^{N_e\times 5}\}$. To preserve as much information as possible about the original instances, the size of $\mathscr{O}_e$ should be similar to that of $\mathcal{O}$ (Zhang and Qiu, 2022; Gan et al., 2022). Thus, $N_{\max}$ is set to be 531. Due to the stochastic property of the proposed data complement process, ten independent data complement processes are performed to obtain different enhanced databases $\mathscr{O}_e$. The composition of the obtained enhanced databases is shown in Table 5. Suppose that N_r is the number of the instances $x \in \mathcal{O} \bigwedge x \notin \mathscr{O}_e$, the loss rate $L_r = \frac{N_r}{N}$ is adopted to represent the information loss of the data complement process. As shown in Table 5, the loss rate is very low. Simulation results indicate that the proposed data complement process can obtain sufficient instances that satisfy the *label sharing assumption*, which not only fills in the missing information, but also preserves the information contained in the original ALA database. However, as shown in Fig. 6, the enhanced databases still present an imbalanced distribution. In addition, we found that instances with minority damage states

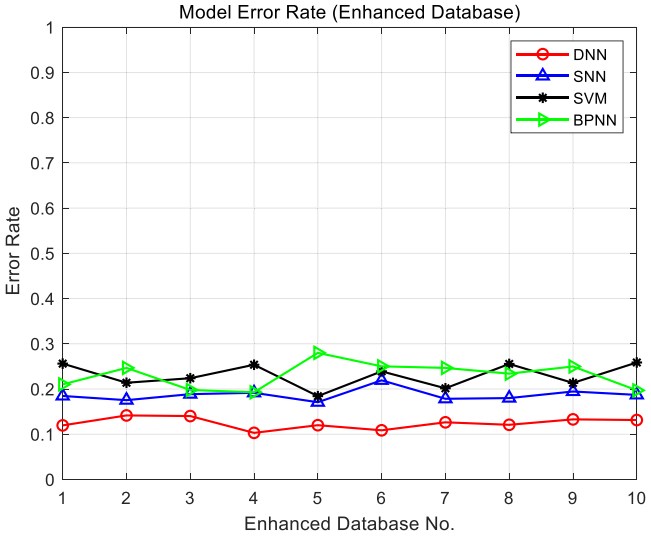

Fig. 8. Model error rate (enhanced databases).

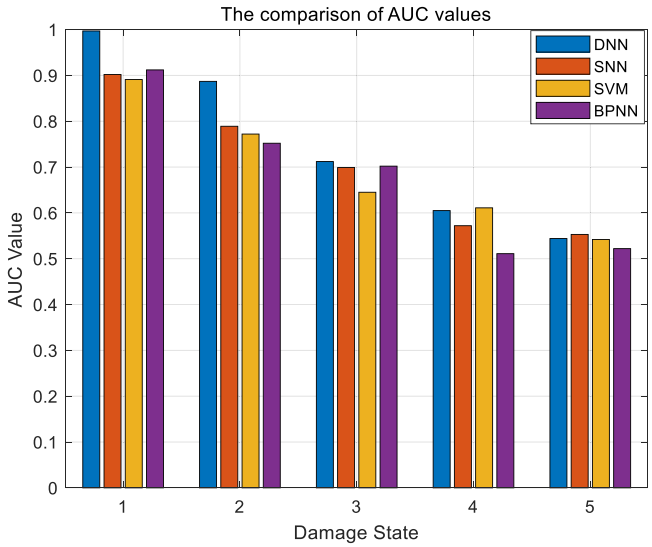

Fig. 9. The comparison of AUC values (enhanced databases).

are very similar. This indicates that the learning model may struggle to gain knowledge about these rare events (Hu and Jiang, 2019; Gu et al., 2021). Meanwhile, various historical accidents (Scawthorn and Johnson, 2000; Krausmann and Cruz, 2013; Chakraborty et al., 2018; Sezen and Whittaker, 2006; Zhou and Zhao, 2021; Girgin, 2011b) indicate that these rare events are highly possible to lead to catastrophic chemical accidents.

The proposed imbalance modification process is applied to modify $\mathscr{O}_e$, a relatively balanced database $\mathscr{O}_{em}$ can be obtained. The predefined rarity threshold is one of the key hyper-parameters. If the value of predefined rarity threshold is too high, artificially generated instances may contaminate the information contained in the original database (Zhang and Qiu, 2022; Gan et al., 2022; Sawant and Prabukumar, 2020). If the value of predefined rarity threshold is too low, the imbalance of the database cannot be modified adequately. After cross validation, the predefined rarity threshold is set to be $0.7\frac{\sum_{x\in X_e} m(x)}{N_e}$. Through the imbalance modification process, ten modified databases are derived from the corresponding enhanced database. The composition of the obtained modified databases is shown in Table 6. As shown in Fig. 7, compared to $\mathscr{O}_e$, the number of instances with minority damage states have all increased to varying degrees. On the premise of ensuring data reliability, a relatively balanced database $\mathscr{O}_{em}$ can be obtained. Then, the LP algorithm is adopted to transform $\mathscr{O}_{em}$ into the corresponding LDL data set $\boldsymbol{\Theta} = \{\boldsymbol{X}_{em} \in \mathbb{R}^{N_{em}\times 13}, \boldsymbol{D} \in \mathbb{R}^{N_{em}\times 5}\}$.

3.2. Algorithm performance analysis

In this section, the ten enhanced databases are first adopted to be the input of DBN, respectively. The training process can be regarded as a multiclass classification problem (Hu and Jiang, 2019). Theoretically, the increase of hidden layer can enhance the model accuracy but also inevitably increase the time complexity of the training process, and may lead to over-fitting (Hu and Jiang, 2019; Low et al., 2020; Tian et al., 2022). Three widely used conventional neural networks that have shallow architectures, namely neural networks with a single hidden layer (SNN), support vector machine (SVM), and BPNN, are adopted for comparison to demonstrate the superiority of the DBN. In practical applications, the settings of model hyperparameters have always been an inevitable problem in machine learning. The training dataset consists of 60% instances in the enhanced database, the validation dataset consists of 20% instances, and the rest of instances are regarded as the test dataset. Both the validation dataset and the test dataset are not involved in the training process (Hu and Jiang, 2019; Low et al., 2020; Tian et al., 2022), of which the validation dataset is adopted to adjust model hyperparameters and the test dataset is adopted to evaluate model generalization ability. According to some preliminary test, a DBN with 3 hidden layers is considered and the neurons in three hidden layers are set as $\{27, 15, 10\}$. We cannot guarantee that the hyperparameters settings are the optimal combination. However, for most cases, the above settings can make the DBN obtain the data-driven model with great accuracy within the acceptable time.

As shown in Fig. 8, the error rates of the DBN-driven models are

Table 7
Error rate of each damage state (DBN, enhanced databases).

$\mathscr{O}_e$	DS_1	DS_2	DS_3	DS_4	DS_5	Error rate
1	0.033	0.253	0.387	0.600	0.714	0.104
2	0.053	0.222	0.483	0.563	0.667	0.118
3	0.090	0.286	0.387	0.500	0.353	0.148
4	0.070	0.280	0.593	0.600	0.538	0.140
5	0.023	0.315	0.517	0.429	0.667	0.101
6	0.050	0.266	0.382	0.462	0.615	0.119
7	0.082	0.218	0.333	0.583	0.273	0.125
8	0.068	0.186	0.333	0.462	0.813	0.125
9	0.089	0.146	0.516	0.500	0.538	0.135
10	0.052	0.299	0.531	0.833	0.533	0.131

Table 8
Error rate of each damage state (DBN, modified databases).

$\mathscr{O}_{em}$	DS_1	DS_2	DS_3	DS_4	DS_5	Error rate
1	0.085	0.135	0.182	0.188	0.210	0.135
2	0.083	0.131	0.177	0.198	0.212	0.134
3	0.091	0.146	0.184	0.197	0.200	0.140
4	0.113	0.128	0.159	0.191	0.178	0.138
5	0.107	0.142	0.161	0.214	0.199	0.143
6	0.098	0.132	0.162	0.207	0.201	0.137
7	0.088	0.145	0.167	0.186	0.202	0.134
8	0.091	0.131	0.175	0.198	0.213	0.135
9	0.101	0.127	0.174	0.221	0.199	0.139
10	0.089	0.132	0.167	0.178	0.195	0.130

obviously lower than the counterparts of the shallow networks. The AUC value (0.5–1) (Hu and Jiang, 2019; Xu et al., 2021) is adopted to further analyze the performance of classifiers obtained by the four algorithms. The larger the AUC is, the better the classification performance is. As shown in Fig. 9, for majority damage states (DS_1 and DS_2), the DBN-driven model presents obvious advantages. In this work, the DBN adopts layer-by-layer strategy of which the values of the all latent variables in each layer can be implicit by a single, bottom-up pass. The superposition of the underlying multilayer RBM makes the characteristics of the sample data set more obvious and effective, and it can be used directly in BPNN (Gan et al., 2022; Cao et al., 2022; Li et al., 2022; Tian et al., 2022). The stacked structure allows the DNN compactly represent highly non-linear and highly-varying functions. Thus, the DBN outperforms the shallow neural networks when handling the complex learning task due to its strong ability in feature learning and data expression. However, all the classifiers are not effective to identify minority damage states.

As shown in Table 7, the obtained models failing to assign the same attention to minority damage states as the majority and further leads to the lack of generalization ability, although the overall performance is considerable (Hu and Jiang, 2019; Debowski et al., 2012; Gu et al., 2021). Due to the low frequency, instances with high damage states are often difficult to estimate. To handle the imbalanced database, the proposed imbalance modification process is adopted to obtain ten modified databases. With the test dataset unchanged, the model error rate is state in Table 8. To further verify the imbalance modification process, results shown in Table 7 and Table 8 are compared in Fig. 10.

As shown in Fig. 10 (a), the overall error rate of the models driven by $\mathscr{O}_{em}$ is slightly worse than that driven by $\mathscr{O}_{e}$. Moreover, for the instances in damage state DS_1, the models driven by $\mathscr{O}_{em}$ may also has higher error rate than the models driven by $\mathscr{O}_{e}$. This is because the involvement of artificially generated instances inevitably changes the knowledge model contained in the original data (Hu and Jiang, 2019; Zhang and Qiu, 2022; Gan et al., 2022; Sawant and Prabukumar, 2020). For $\mathscr{O}_{e}$, the instances in damage state DS_1 are sufficient, and the overall error rate is largely determined by these instances in majority damage state. However, the models driven by $\mathscr{O}_{e}$ cannot accurately evaluate the instances in minority damage state, which shows poor generalization ability. On the contrary, as shown in Fig. 10 (c-f), the models driven by $\mathscr{O}_{em}$ can achieve a balanced estimation for all damage states. This indicates that the proposed imbalance modification process can effectively deal with the imbalanced data.

3.3. Seismic damage estimation

The above data-driven model still suffers from the limitations associated with crisp models, which is incompetent to handle the uncertainties arising from the label ambiguity (Gao et al., 2017; Xu et al., 2021). To handle the label ambiguity, the LP algorithm is adopted to recover the damage state probability distribution from the logical labels, which utilizes the topological information of the feature space and the correlation among the damage states (Xu et al., 2021). Through the label enhancement process, the LDL database Θ can be obtained for learning the damage state probability distribution. The label distribution learning can be regarded as a regression problem (Geng, 2016; Xu et al., 2021). Through the label distribution learning process, the proposed HDBN-LDLS can obtain a robust seismic damage estimation model. In this section, the reliability of the obtained robust seismic damage estimation model is validated by an actual damaged tank in the 2006 Silakhor earthquake (Miladi and Razzaghi, 2019).

3.3.1. Instance description

The 2006 Silakhor earthquake $M_w = 6.1$ was occurred at a focal depth of 7 km. The maximum S_p record was 4.32 m/s^2 (horizontal) and 5.24 m/s^2 (vertical). The 2006 Silakhor earthquake damaged many above-ground cylindrical tanks. One of the most damaged tanks (Tank 3) was analyzed by Miladi and Razzaghi (2019), instance features of Tank 3 are stated in Table 9.

According to the Supplemental Material. B, the seismic damage parameters can be obtained: $F_u = 2.639e + 5N/m$; $F_w = 3.065e + 4N/m$; $J = 6.764$; $\sigma_C = 70.775MPa$; $F_C = 71.558MPa$. The calculation results show that the maximum longitudinal shell compression stress σ_C is very close to the allowable longitudinal shell-membrane compression stress F_C. Since $F_u \gg F_w$ & $J > 1.54$, the tank is not stable and cannot be self-anchored. In fact, the 2006 Silakhor earthquake caused shell uplift, cracking of foundation and buckling (Miladi and Razzaghi, 2019). The actual seismic damage characteristics are basically consistent with the seismic response analysis results, of which the reliability of the proposed data complement process can be effectively guaranteed.

3.3.2. Estimation results analysis

The instance features are regarded as the input of the robust seismic damage estimation model. Then, the seismic damage state probability distribution can be obtained. The estimation result is $\{d^{DS_1} = 0.021, d^{DS_2} = 0.271, d^{DS_3} = 0.486, d^{DS_4} = 0.162, d^{DS_5} = 0.060\}$. According to the definitions of five seismic damage states, the damage state of Tank 3 can be identified as DS_3. This is consistent with the estimation result. The membership degree of DS_3 to the instance is 0.486, which is consistent with the actual physical damage of Tank 3. In addition, the explicit relative importance of other damage states can also be obtained from the seismic damage state probability distribution. As shown in Fig. 11, compared with traditional crisp models, the proposed robust methodology can provide a more natural mapping relation of physical damage to probability space.

3.4. Seismic damage state probability curves

After the training process, the obtained seismic damage estimation model can quickly estimate the seismic damage state probability distribution, which provides a powerful tool for the seismic design of liquid storage tanks. Through changing the value of S_p (0.05 g~1.5 g), the seismic damage state probability curves of Tank 3 can be obtained

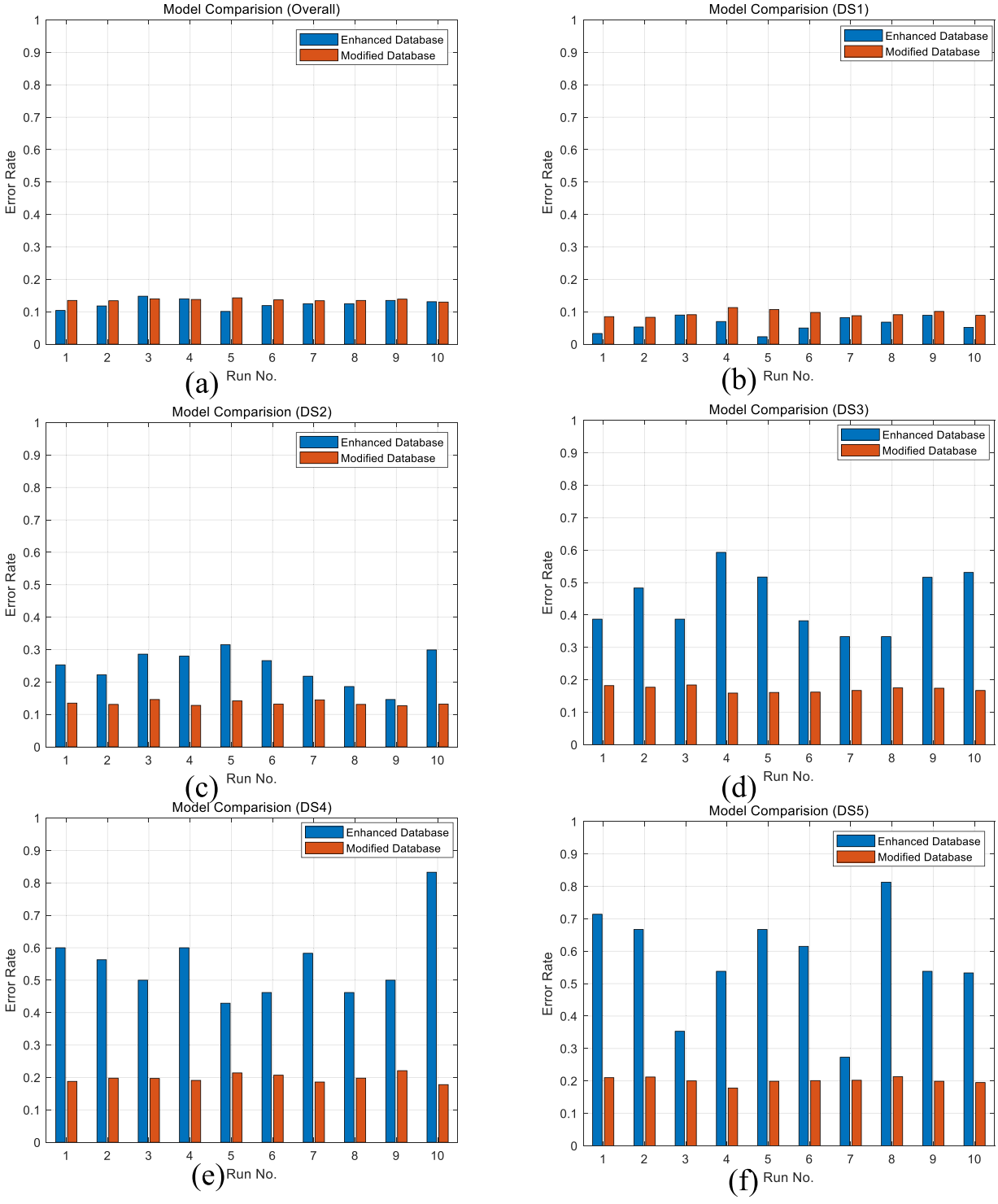

Fig. 10. Model Comparison ($\mathcal{O}_e$ vs $\mathcal{O}_{em}$).

Table 9
Instance Features of Tank 3 (Miladi and Razzaghi, 2019).

Parameters	Values
D	3.8 m
H	4.5 m
t_s	6 mm
t_r	6 mm
t_b	6 mm
t_a	Not equipped with annular ring
A_S	1
R_t	2. Fixed Roof Tank
Steel Properties	Elastic Modulus: 210 kPa; Yield Stress: 240 MPa; Density: 7850 kg/m^3
h	4.2 m
ρ_L	fuel oil, 920 kg/m^3
S_p	0.4 g
SC	1

(Fig. 11). The proposed methodology deals with label ambiguity by allowing the answer to consist of more than one damage state. The damage state with highest membership degree is denoted as the dominant state of seismic damage. As show in Fig. 12, with the increase of earthquake intensity, the membership degrees of $DS_1 \sim DS_5$ to Tank 3 changes dynamically.

The intersection of the probability curve corresponding to each seismic damage state is defined as the *critical damage point*. When the seismic intensity exceeds the corresponding critical damage point, the dominant seismic damage state of the tank will change accordingly. Thus, the *critical damage points* provide a more convenient way to characterize the seismic performance of liquid storage tanks. As shown in Fig. 11, Tank 3 has four *critical damage points,* $\{0.145\text{g}, 0.328\text{g}, 0.587\text{g}, 1.128\text{g}\}$. This indicates that the dominant seismic damage state of Tank 3 is DS_1 with the seismic intensity $S_p \leq 0.145g$; the dominant seismic damage state of Tank 3 is DS_2 with the seismic intensity $0.145g < S_p \leq 0.328g$; the dominant seismic damage state of Tank 3 is DS_3 with the seismic intensity $0.328g < S_p \leq 0.587g$; the dominant seismic damage state of Tank 3 is DS_4 with the seismic intensity $0.587g < S_p \leq 1.128\text{g}$; the dominant seismic damage state of Tank 3 is DS_5 with the seismic intensity $S_p > 1.128\text{g}$.

As mentioned by many studies (D'Amico and Buratti, 2019; Gabbianelli et al., 2022; Lee et al., 2019; Miladi and Razzaghi, 2019), anchoring technology can greatly improve the seismic performance of storage tanks. (Miladi and Razzaghi (2019) observed the damage state of 28 above-ground cylindrical tanks that experienced the 2006 Silakhor earthquake. None of the mechanically anchored tanks was damaged. Thus, we assume that Tank 3 is mechanically anchored, the corresponding seismic damage state probability curves are shown in Fig. 13. The mechanically-anchored Tank 3 has two *critical damage points,* $\{0.691\text{g}, 0.989\text{g}\}$. This indicates that the dominant seismic damage state of mechanically-anchored Tank 3 is DS_1 with the seismic intensity $S_p \leq 0.691\text{g}$; the dominant seismic damage state of mechanically-anchored Tank 3 is DS_2 with the seismic intensity $0.691\text{g} < S_p \leq 0.989\text{g}$; the dominant seismic damage state of

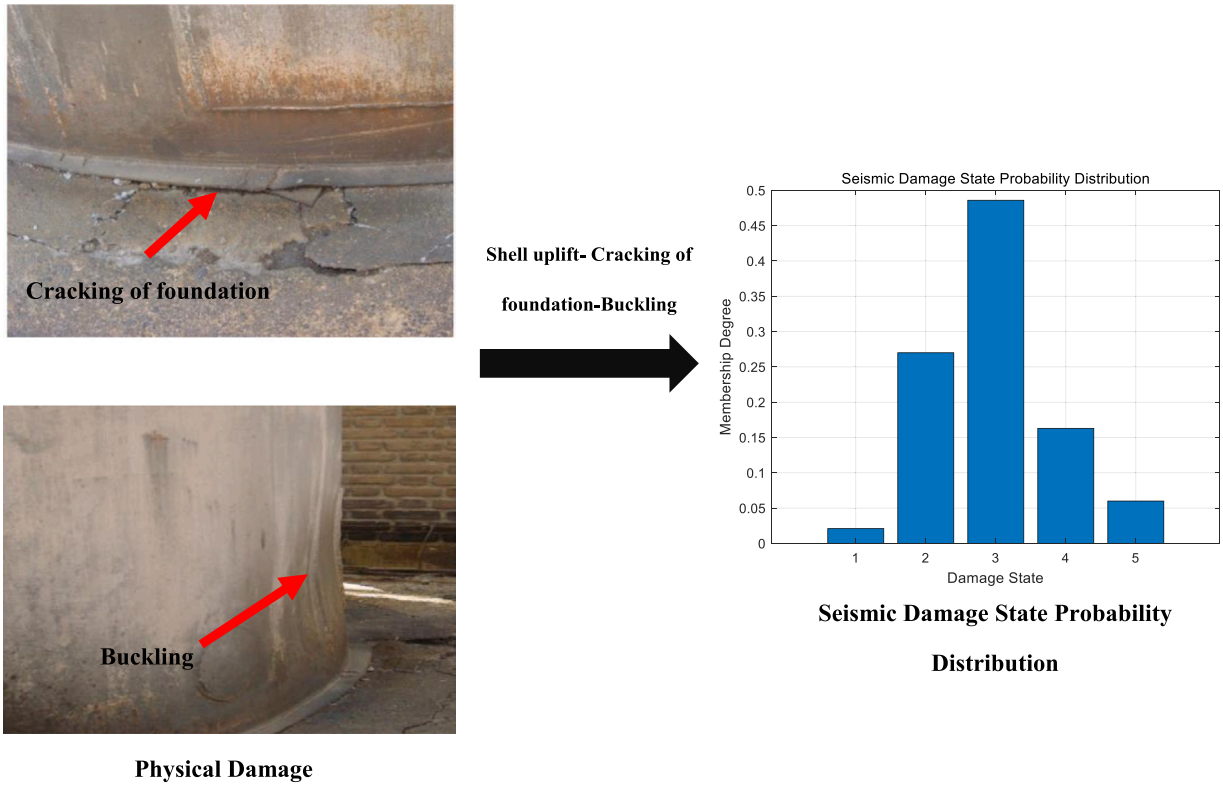

Fig. 11. Mapping relation of physical damage to probability space.

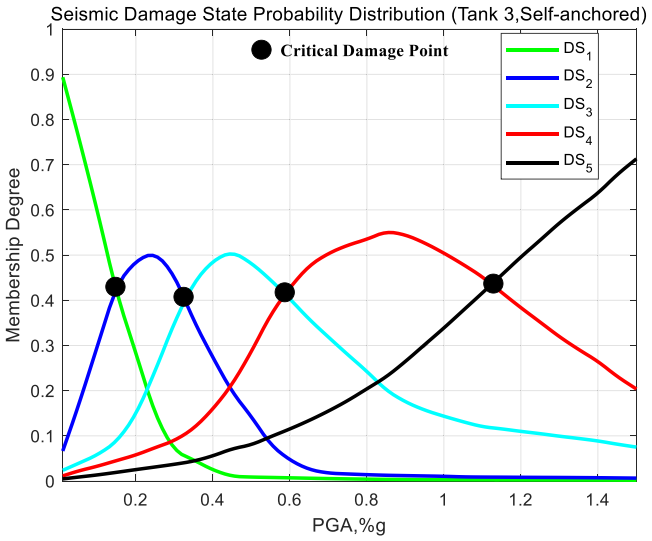

Fig. 12. Seismic damage state probability curves (Tank 3, Self-anchored).

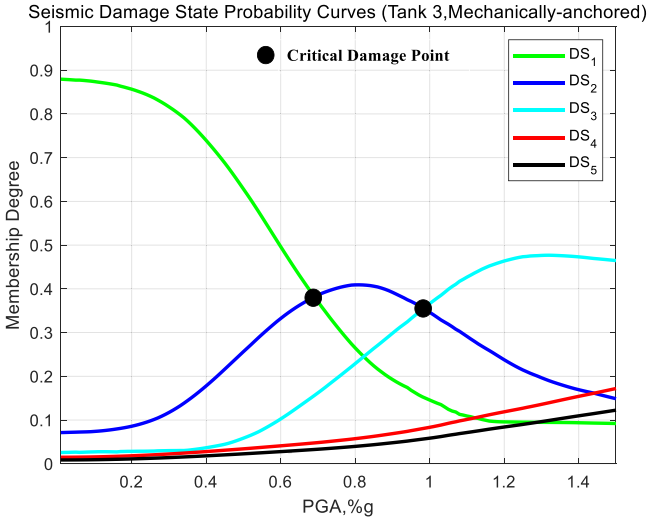

Fig. 13. Seismic damage state probability curves (Tank 3, Mechanically-anchored).

mechanically-anchored Tank 3 is DS_3 with the seismic intensity $0.989\text{g} < S_p \leq 1.5g$. According to the accident statistical analysis (D'Amico and Buratti, 2019; Salzano et al., 2003; Vílchez et al., 2001; Alliance, 2001b; Alliance, 2001a), seismic events with the seismic intensity $S_p > 1.5g$ are not considered. We found that the seismic performance of the mechanically-anchored Tank 3 is much better than that of self-anchored storage tank. Through comparing the critical damage points of the tanks, the effects of various parameters on the seismic performance of liquid storage tanks can be quantitatively analyzed. In practical application, the proposed methodology can quickly identify the tank parameters that satisfy the specific anti-seismic requirements.

4. Conclusions

In this work, label ambiguity and damage data imbalance associated with seismic damage estimation of liquid storage tanks were especially concerned. A novel data-driven methodology was developed, which integrated a data complement process for filling in the missing information from database, an imbalance modification process for generating specific instances with minority damage states, a label enhancement process for recovering the damage state distribution from the logical label, and a label distribution learning for obtaining a robust seismic damage estimation model.

Simulation results indicate that the obtained seismic damage estimation model can achieve a balanced estimation for all damage states while maintaining sufficient robustness to cope with label ambiguity. Compared with traditional crisp models, the proposed robust methodology provided a more natural mapping relation of physical damage to probability space. Even though seismic damage estimation of liquid storage tanks is the main motivation of our work, the proposed methodology is not limited to this, as many fragility analysis problems have analogous uncertain attributes.

Compared with the traditional fragility curves, the seismic damage state probability curves provide a new and comprehensive perspective

for fragility analysis. The seismic damage state probability curves can be used to analyze the topological information of the feature space (equipment parameters and hazard intensity) and the correlation among the damage states. The proposed methodology provides a promising tool for seismic design of liquid storage tank.

There are still some issues in the current study that have not yet been resolved. The selection of the intensity measure is important for fragility analysis. From the perspective of the proposed data-driven methodology, this work is not limited to a specific seismic intensity measure. The seismic intensity measure actually depends on the original post-earthquake damage databases. However, most databases still use the Peak Ground Acceleration as the intensity measure. The Peak Ground Acceleration alone cannot fully capture the complex response of a liquid storage tank under seismic excitation. Moreover, for vector-valued intensity measures, specific data mapping methods may be required for dimension reduction. These issues will be the key points of our future research.

CRediT authorship contribution statement

Jinkun Men: Conceptualization, Methodology, Software, Validation, Writing - Original Draft, Writing - Review & Editing. **Guohua Chen**: Conceptualization, Resources, Supervision, Project Administration, Funding Acquisition, Writing - Review & Editing. **Genserik Reniers**: Conceptualization, Writing - Review & Editing. **Xiaohui Rao**: Data Extraction. **Tao Zeng:** Writing - Review & Editing.

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 (22078109), the Key-Area Research and Development Program of Guangdong Province (2019B111102001), Ministry of Science and Technology of the People's Republic of China (G2022163006L) and the China Scholarship Council (202206150061/202006150080).

Appendix

Table A1
Overview of representative studies for the seismic fragility of liquid storage tanks.

Authors (Year)	Probabilistic models	Evidence	Main work
O'Rourke and So (2000)	Logistic Regression	Post-earthquake Damage Data	● The seismic behavior of cylindrical on-grade, steel liquid storage tanks was investigated; ● The seismic damage data of 397 tanks in nine seismic events was collected for fragility analysis; ● According to the HAZUS methodology (Kircher Charles et al., 2006), five damage states (no damage to complete failure) were adopted to describe the physical damage.
American Lifeline Alliance (ALA, 2001a,b)	Least Square Regression	Post-earthquake Damage Data	● The seismic damage data of 532 tanks in 19 seismic events was collected for fragility analysis; ● Five damage states were defined to describe the physical damage.
Salzano et al. (2003)	Probit Regression	Post-earthquake Damage Data	● The ALA seismic damage data (Alliance, 2001b; Alliance, 2001a) reorganized in terms of risk states with reference to the loss of content; ● Based on the framework of probit analysis, a simple and useful statistic tool was developed for seismic QRA.
Berahman and Behnamfar (2009)	Bayesian Analysis	Finite Element Analysis (FEA)	● The fragility of un-anchored steel storage tanks in petroleum complexes was estimated by a probabilistic seismic demand model; ● The observation data of two failure modes, elephant foot buckling and welding failure, were obtained by FEA.
Buratti and Tavano (2014) (Buratti and Tavano, 2014)	Regression analysis	Added Mass Modeling	● The secondary buckling occurring in the top part of the tank was investigated; ● Incremental nonlinear time-history analyses were performed to identify the critical buckling loads.
Saha et al. (2016)	Monte Carlo Simulation	Response Surface Model (RSM)	● The uncertainty associated with isolator was investigated; ● The RSM was adopted to observe the non-linear seismic response of the base-isolated liquid storage tanks.
Zuluaga Mayorga et al. (2019)	Monte Carlo Simulation	Limit State Equation (LSE)	● The LSEs of buckling and overturning were extracted from the American standard API-650 (Spritzer and Guzey, 2017); ● The large-scale sample space obtained by Monte Carlo Simulation was divided by the LSEs into two categories: safety and failure.
(D'Amico and Buratti, 2019)	Bayesian Analysis	Post-earthquake Damage Data	● The seismic damage data of 3026 tanks in 24 seismic events was collected for fragility analysis; ● Two sets of damage states were adopted to describe the severity of structural damage and the quantity of liquid releases.
Bezir et al. (2022)	Logit, Probit and Cumulative Lognormal Model and Maximum Likelihood Method	Post-earthquake Damage Data/ Finite Element Analysis	● The seismic damage data of 4509 tanks were extracted from previous studies, another 101 damage data was obtained from the FEA. ● Different statistical methods were used for curve fitting to the data called "observation frequency" of the damage.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psep.2023.02.079.

References

A.L. Alliance, Seismic Fragility Formulation for Water Systems: Part 2—Appendices, in, American Society of Civil Engineers, Reston, 2001b, pp. 1–231.

A.L. Alliance, Seismic Fragility Formulation For Water Systems—Part 1: Guideline, in, American Society of Civil Engineers, Reston, 2001a, pp. 1–96.

Amin, M.T., Khan, F., Amyotte, P., 2019. A bibliometric review of process safety and risk analysis. Process Saf. Environ. Prot. 126, 366–381.

Antonioni, G., Spadoni, G., Cozzani, V., 2007. A methodology for the quantitative risk assessment of major accidents triggered by seismic events. J. Hazard. Mater. 147, 48–59.

API, Welded Tanks for Oil Storage (API 650 13th Edition), in, American Petroleum Institute, Washington, 2020.

Arunthavanathan, R., Khan, F., Ahmed, S., Imtiaz, S., 2021. A deep learning model for process fault prognosis. Process Saf. Environ. Prot. 154, 467–479.

Bakalis, K., Karamanos, S.A., 2021. Uplift mechanics of unanchored liquid storage tanks subjected to lateral earthquake loading. Thin-Walled Struct. 158, 107145.

Bakalis, K., Kohrangi, M., Vamvatsikos, D., 2018. Seismic intensity measures for above-ground liquid storage tanks. Earthq. Eng. Struct. Dyn. 47, 1844–1863.

Berahman, F., Behnamfar, F., 2009. Probabilistic seismic demand model and fragility estimates for critical failure modes of un-anchored steel storage tanks in petroleum complexes. Probab. Eng. Mech. 24, 527–536.

Bezir, F., Öztürk, S., Sarı, A., Akgül, K., 2022. Fragility analysis of atmospheric storage tanks by observational and analytical data. Int. J. Steel Struct. 192–205.

Brus, D.J., de Gruijter, J.J., van Groenigen, J.W., 2006. Chapter 14 – Designing spatial coverage samples using the k-means clustering algorithm. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Developments in Soil Science. Elsevier, pp. 183–192.

Buratti, N., Tavano, M., 2014. Dynamic buckling and seismic fragility of anchored steel tanks by the added mass method. Earthq. Eng. Struct. Dyn. 43, 1–21.

Cao, M., Zhang, T., Wang, J., Liu, Y., 2022. A deep belief network approach to remaining capacity estimation for lithium-ion batteries based on charging process features. J. Energy Storage 48, 103825.

Chakraborty, A., Ibrahim, A., Cruz, A.M., 2018. A study of accident investigation methodologies applied to the Natech events during the 2011 Great East Japan earthquake. J. Loss Prev. Process Ind. 51, 208–222.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

B. Debowski, S. Areibi, G. Grewal, J. Tempelman, A Dynamic Sampling Framework for Multi-class Imbalanced Data, in: 2012 11th International Conference on Machine Learning and Applications, 2012, pp. 113–118.

D'Amico, M., Buratti, N., 2019. Observational seismic fragility curves for steel cylindrical tanks. J. Press. Vessel Technol. 141, 010904.

Fabbrocino, G., Iervolino, I., Orlando, F., Salzano, E., 2005. Quantitative risk analysis of oil storage facilities in seismic areas. J. Hazard. Mater. 123, 61–69.

Gabbianelli, G., Perrone, D., Brunesi, E., Monteiro, R., 2022. Seismic acceleration demand and fragility assessment of storage tanks installed in industrial steel moment-resisting frame structures. Soil Dyn. Earthq. Eng. 152, 107016.

Gan, Y., Zhu, H., Guo, W., Xu, G., Zou, G., 2022. Deep semi-supervised learning with contrastive learning and partial label propagation for image data. Knowl. -Based Syst. 245, 108602.

Gao, B.-B., Xing, C., Xie, C.-W., Wu, J., Geng, X., 2017. Deep label distribution learning with label ambiguity. IEEE Trans. Image Process. 26, 2825–2838.

Geng, X., 2016. Label distribution learning. IEEE Trans. Knowl. Data Eng. 28, 1734–1748.

Ghosh, S., Ghosh, S., Chakraborty, S., 2017. Seismic fragility analysis in the probabilistic performance-based earthquake engineering framework: an overview. Int. J. Adv. Eng. Sci. Appl. Math. 13, 1–14.

Girgin, S., 2011a. The natech events during the 17 August 1999 Kocaeli earthquake: aftermath and lessons learned. Nat. Hazards Earth Syst. Sci. 11, 1129–1140.

Girgin, S., 2011b. The natech events during the 17 August 1999 Kocaeli earthquake: aftermath and lessons learned. Nat. Hazards Earth Syst. Sci. 11, 1129–1140.

D.E. Goldberg, J. Richardson, Genetic algorithms with dynamic niche sharing for multimodal function optimization, In: Proceedings of IEEE International Conference on Evolutionary Computation, IEEE, Nagoya, Japan, 1987, pp. 786–791.

Gu, X., Zhao, Y., Yang, G., Li, L., 2021. An imbalance modified convolutional neural network with incremental learning for chemical fault diagnosis. IEEE Trans. Ind. Inform. 18, 1–10.

Han, H., Wang, W.-Y., Mao, B.-H., 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (Eds.), Advances in Intelligent Computing. Springer, Berlin Heidelberg, pp. 878–887 (Berlin, Heidelberg).

Hernandez-Hernandez, D., Larkin, T., Chouw, N., 2021. Shake table investigation of nonlinear soil–structure–fluid interaction of a thin-walled storage tank under earthquake load. Thin-Walled Struct. 167, 108143.

Housner, G.W., 1957. Dynamic pressures on accelerated fluid containers. Bull. Seismol. Soc. Am. 47, 15–35.

Housner, G.W., 1963. The dynamic behavior of water tanks. Bull. Seismol. Soc. Am. 53, 381–387.

Hu, Z., Jiang, P., 2019. An imbalance modified deep neural network with dynamical incremental learning for chemical fault diagnosis. IEEE Trans. Ind. Electron. 66, 540–550.

Huang, K., Chen, G., Khan, F., 2022a. Vulnerability assessment method for domino effects analysis in chemical clusters. Process Saf. Environ. Prot. 164, 539–554.

Huang, M., Chen, G., Yang, P., Hu, K., Zhou, L., Men, J., Zhao, J., 2022b. Multi-hazard coupling vulnerability analysis for buckling failure of vertical storage tank: floods and hurricanes. Process Saf. Environ. Prot. 161, 528–541.

Kircher Charles, A., Whitman Robert, V., Holmes William, T., 2006. HAZUS earthquake loss estimation methods. Nat. Hazards Rev. 7, 45–59.

Krausmann, E., Cruz, A.M., 2013. Impact of the 11 March 2011, Great East Japan earthquake and tsunami on the chemical industry. Nat. Hazards 67, 811–828.

Krawinkler, H., 2000. Progress and challenges in seismic performance assessment. PEER Cent. N. 3, 1–4.

Lee, S., Kim, B., Lee, Y.-J., 2019. Seismic fragility analysis of steel liquid storage tanks using earthquake ground motions recorded in Korea. Math. Probl. Eng. 2019, 6190159.

Li, X., Liu, Q., Wu, Y., 2022. Prediction on blockchain virtual currency transaction under long short-term memory model and deep belief network. Appl. Soft Comput. 116, 108349.

Low, C.-Y., Park, J., Teoh, A.B.-J., 2020. Stacking-based deep neural network: deep analytic network for pattern classification. IEEE Trans. Cybern. 50, 5021–5034.

Men, J., Chen, G., Zeng, T., 2022a. Multi-hazard coupling effects in chemical process industry – Part I: preliminaries and mechanism. IEEE Syst. J. 17, 1626–1636.

Men, J., Chen, G., Yang, Y., Genserik, R., 2022b. An event-driven probabilistic methodology for modeling the spatial-temporal evolution of natural hazard-induced domino chain in chemical industrial parks. Reliab. Eng. Syst. Saf. 226, 108723.

Men, J., Chen, G., Zeng, T., 2022c. Multi-hazard coupling effects in chemical process industry – Part II: research advances and future perspectives on methodologies. IEEE Syst. J. 17, 1637–1647.

Meng, Y., Lu, C., Yan, Y., Shi, L., Liu, J., 2015. Method to analyze the regional life loss risk by airborne chemicals released after devastating earthquakes: a simulation approach. Process Saf. Environ. Prot. 94, 366–379.

Miladi, S., Razzaghi, M.S., 2019. Failure analysis of an un-anchored steel oil tank damaged during the Silakhor earthquake of 2006 in Iran. Eng. Fail. Anal. 96, 31–43.

Mohan, A., Nakano, A., Ferrara, E., 2021. Graph signal recovery using restricted Boltzmann machines. Expert Syst. Appl. 185, 115635.

Ning, L., Pittman, R., Shen, X., 2018. LCD: a fast contrastive divergence based algorithm for restricted Boltzmann machine. Neural Netw. 108, 399–410.

NZSEE, Seismic Design of Storage Tanks: November 2009, in, New Zealand Society for Earthquake Engineering Inc, New Zealand, 2009.

O'Rourke, M.J., So, P., 2000. Seismic fragility curves for on-grade steel tanks. Earthq. Spectra 16, 801–815.

Ozdemir, Z., Souli, M., Fahjan, Y.M., 2010. Application of nonlinear fluid–structure interaction methods to seismic analysis of anchored and unanchored tanks. Eng. Struct. 32, 409–423.

P. Ministry of Housing and Urban-Rural Development, Code for design of vertical cylindrical welded steel oil tanks (GB 50341–2014), in, China Planning Press, Beijing, 2014.

Phan, H.N., Paolacci, F., Alessandri, S., 2019. Enhanced seismic fragility analysis of unanchored steel storage tanks accounting for uncertain modeling parameters. J. Press. Vessel Technol. 141.

Pourbahrami, S., Balafar, M.A., Khanli, L.M., Kakarash, Z.A., 2020. A survey of neighborhood construction algorithms for clustering and classifying data points. Comput. Sci. Rev. 38, 100315.

Ricci, F., Moreno, V.C., Cozzani, V., 2021. A comprehensive analysis of the occurrence of Natech events in the process industry. Process Saf. Environ. Prot. 147, 703–713.

Saha, S.K., Matsagar, V., Chakraborty, S., 2016. Uncertainty quantification and seismic fragility of base-isolated liquid storage tanks using response surface models. Probab. Eng. Mech. 43, 20–35.

Salzano, E., Iervolino, I., Fabbrocino, G., 2003. Seismic risk of atmospheric storage tanks in the framework of quantitative risk analysis. J. Loss Prev. Process Ind. 16, 403–409.

Sawant, S.S., Prabukumar, M., 2020. A review on graph-based semi-supervised learning methods for hyperspectral image classification. Egypt. J. Remote Sens. Space Sci. 23, 243–248.

Scawthorn, C., Johnson, G.S., 2000. Preliminary report: Kocaeli (Izmit) earthquake of 17 August 1999. Eng. Struct. 22, 727–745.

Sezen, H., Whittaker Andrew, S., 2006. Seismic performance of industrial facilities affected by the 1999 Turkey earthquake. J. Perform. Constr. Facil. 20, 28–36.

Spritzer, J.M., Guzey, S., 2017. Review of API 650 Annex E: design of large steel welded aboveground storage tanks excited by seismic loads. Thin-Walled Struct. 112, 41–65.

Tian, J., Liu, Y., Zheng, W., Yin, L., 2022. Smog prediction based on the deep belief - BP neural network model (DBN-BP). Urban Clim. 41, 101078.

Vílchez, J.A., Montiel, H., Casal, J., Arnaldos, J., 2001. Analytical expressions for the calculation of damage percentage using the probit methodology. J. Loss Prev. Process Ind. 14, 193–197.

Wang, J., Wang, M., Yu, X., Zong, R., Lu, S., 2022. Experimental and numerical study of the fire behavior of a tank with oil leaking and burning. Process Saf. Environ. Prot. 159, 1203–1214.

Xu, N., Liu, Y.-P., Geng, X., 2021. Label enhancement for label distribution learning. IEEE Trans. Knowl. Data Eng. 33, 1632–1643.

Yang, Y., Chen, G., Reniers, G., 2020. Vulnerability assessment of atmospheric storage tanks to floods based on logistic regression. Reliab. Eng. Syst. Saf. 196, 106721.

Zhang, S., Qiu, T., 2022. Semi-supervised LSTM ladder autoencoder for chemical process fault diagnosis and localization. Chem. Eng. Sci. 251, 117467.

Zhou, J., Zhao, M., 2021. Impact of earthquake direction and liquid sloshing on column supported tank. Eng. Struct. 247, 113037.

Zuluaga Mayorga, S., Sánchez-Silva, M., Ramírez Olivar, O.J., Muñoz Giraldo, F., 2019. Development of parametric fragility curves for storage tanks: a Natech approach. Reliab. Eng. Syst. Saf. 189, 1–10.