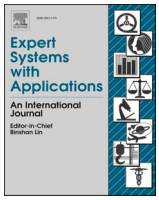
Expert Systems With Applications 240 (2024) 122459

Contents lists available at ScienceDirect

Expert Systems With Applications

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



Safety assessment of tunnel construction based on counterintuitivity detection using multi-profile multi-model ensemble learning

Leilei Chang [a,b], Chenhao Yu [a,b], Limao Zhang [c,*], Xiaobin Xu [a,b], Schahram Dustdar [d]

[a] China-Austria Belt and Road Joint Laboratory on Artificial Intelligence and Advanced Manufacturing, Hangzhou Dianzi University, China
[b] School of Automation, Hangzhou Dianzi University, Hangzhou, China
[c] School of Civil and Hydraulic Engineering, National Center of Technology Innovation for Digital Construction, Huazhong University of Science and Technology, Wuhan, China
[d] Distributed Systems Group, TU Wien, 1040 Vienna, Austria

ARTICLE INFO

Keywords:
Tunnel construction
Safety assessment
Counterintuitive data
Multi-profile ensemble learning
Multi-model ensemble learning

ABSTRACT

Motivated to accurately predict the building tilt rate (BTR) in tunnel construction, a new multi-profile multi-model ensemble learning approach is proposed in this study by distinguishing between counterintuitive data and intuitive data, and handling each type appropriately. The proposed approach consists of three phases, namely the definition, identification, and handling of counterintuitive data, which are the major theoretical contributions of this study. Firstly, counterintuitive data is defined based on the input–output causal relation of safety-related data, introducing a novel concept in this research. Secondly, counterintuitive data is initially identified using a multi-profile ensemble learning approach and subsequently validated through multi-model ensemble learning. Finally, the identified and confirmed counterintuitive data is handled by assigning reduced weights. To validate the practicality of the approach, a case study on predicting the building tilt rate (BTR) in tunnel construction is conducted. The results of the case study demonstrate that the proposed multi-profile multi-model approach yields more accurate predictions of BTR compared to direct prediction and several other machine learning approaches. Furthermore, the proposed approach aids in identifying counterintuitive data in the testing dataset. Additionally, the effectiveness and superiority of the proposed approach are verified by comparing the prediction results when traditional abnormal data or random data is identified and treated as counterintuitive data.

1. Introduction

This study is motivated by defining, identifying, and handling counterintuitive data from intuitive data in safety assessment and management in underground tunnel construction (Singh, Das, Singh, & Racherla, 2023; Chang, Zhang, & Xu, 2023). The building tilt rate (BTR) serves as a crucial safety indicator, reflecting the impact of underground tunneling on nearby buildings (Darroch, Beecroft, & Nelson, 2021; Zou, Moore, Sanayei, Wang, & Tao, 2021). Parameters influencing the BTR include uneven settlements caused by tunnel boring machine (TBM) operations, geological conditions such as soil compression modulus, tunnel conditions like tunnel cover depth, and building conditions such as relative distances from the tunnel (Chang, Zhang, & Xu, 2022; Forsat, Taghipoor, & Palassi, 2021; Conforti et al., 2019). Most of this data consists of numerical and continuous parameters (Pan & Zhang, 2020), e.g., the TBM parameters are directly monitored from the TBM machine, while the geological-related parameters are calculated using data from the total station machine, etc. The BTR is also a numerical value that is calculated using readings of multiple sensors deployed in multiple key locations (Zhang, Wu, Ji, & AbouRizk, 2017).

Multiple sensors have been deployed to gather those data (Chang, Song, & Zhang, 2022). With the abundance of collected data, a natural approach is to develop a prediction model using machine learning (ML) techniques, as adopted in this study. In most cases, the data collected from sensors or experts are directly utilized. However, in some instances where preprocessing is required, anomaly detection (AD) methods are employed to identify and exclude abnormal data, such as high TBM torque, breaches of preset boundaries near buildings, or the BTR exceeding safety thresholds (Chang, Zhang, & Xu, 2022; Pan & Zhang, 2023). While AD-based approaches are easy to implement, they have two major disadvantages. Firstly, abnormal data doesn't necessarily indicate incorrect data, as it can also reflect actual tunneling activities. Secondly, excluding excessive abnormal data may result in an overly

* Corresponding author.
E-mail addresses: 42538@hdu.edu.cn (C. Yu), zlm@hust.edu.cn (L. Zhang), xuxiaobin1980@163.com (X. Xu), schahram.dustdar@tuwien.ac.at (S. Dustdar).

https://doi.org/10.1016/j.eswa.2023.122459
Received 10 May 2023; Received in revised form 13 August 2023; Accepted 3 November 2023
Available online 7 November 2023

Nomenclature

Acronyms

BTR	building tilt rate
TBM	tunnel boring machine
AD	anomaly detection
ML	machine learning
I	the input
$x_1, x_2, \ldots, x_M$	the parameters of the input
M	the number of attributes in the input
y	the single output
$rg(x)$ min	lower limit of input safety range
$rg(x)$ max	upper limit of input safety range
$rg(y)$ min	lower limit of output safety range
$rg(y)$ max	upper limit of output safety range
f	the mapping relationship $I \rightarrow y$
I^{normal}	the input with all of the attributes are normal attributes
$I^{abnormal}$	the input with a subset of its attributes are abnormal attributes
y^{normal}	normal output
$y^{abnormal}$	abnormal output
f^{causal}	the causal relationship of intuition: $I^{normal} \rightarrow y^{normal}$ or $I^{abnormal} \rightarrow y^{abnormal}$
$d^{intuitive}$	intuitive data which satisfied the causal relationship f^{causal}
$d^{counterintuitive}$	counterintuitive data which is with a normal input and an abnormal output, i.e., $I^{normal} \rightarrow y^{abnormal}$, or an abnormal input and a normal output, i.e., $I^{abnormal} \rightarrow y^{normal}$.
$d^{abnormal}$	abnormal data whose input I or output y is abnormal
D_T	training dataset
D_V	testing dataset
S	the number of partial-datasets
Bs T	S partial datasets from the original training dataset
P	the number of training data sets
$es\ p$	absolute error of the pth set of data from the sth sub-model
$Ds\ T$	S equal partial-datasets from training dataset
yp	the actual output of the pth set of training data
$ys\ p$	the output from the sth sub-model of the pth set of training data
MAPE	mean absolute percentage error
w_s	the weights of S sub-models
$e\ p$	average error of the pth set of data
s^*	the sub-model that the pth set of data is not used to construct
Ω	ordered data list
$\Omega_{intuitive}$	the datasets with intuitive data
$\Omega_{counterintuitive}$	the datasets with counterintuitive data
Λ	a list of ascending ordered data
$\triangle ep,p+1$	the marginal errors
f_p	the frequency of the pth set of data being confirmed as counterintuitive data by M machine learning approaches
$w_{p,0}$	weight of the pth set of data before identifying the type of data
w_p	weight of the pth set of data after identifying the type of data
μ	the updating weight for different types of data
MAE	mean absolute error
Q	the number of testing data sets
$ya\ q$	the actual output of the qth set of testing data
$ye\ q$	the predicted output of the qth set of testing data
BPNN	back-propagation neural network
GPR	gradient process regression
$w_{counterintuitive}$	the weight of the counterintuitive data
$w_{intuitive}$	the weight of the intuitive data
SVM	support vector machine
ANFIS	Adaptive-Network-Based Fuzzy Inference System

"clean" dataset, leading to overfitting and an inability to recognize newly collected data. Although AD-based approaches are sufficient as practical safety control measures, they do not contribute to the training of a more accurate BTR prediction model.

The more pressing challenge in constructing an accurate BTR prediction model is to effectively utilize data with a causal input–output relation, i.e., intuitive data. Specifically, A BTR model is only designed to produce normal (abnormal) output whenever there is a normal (abnormal) input, i.e., recognize intuitive data. For counterintuitive data, i.e., data with a normal input but abnormal output or abnormal input but normal input, they should be identified and properly handled.

Therefore, a multi-profile multi-model approach is proposed in this study, specifically addressing the (1) **definition**, (2) **identification**, and (3) **handling** of counterintuitivity for the safety assessment of tunnel

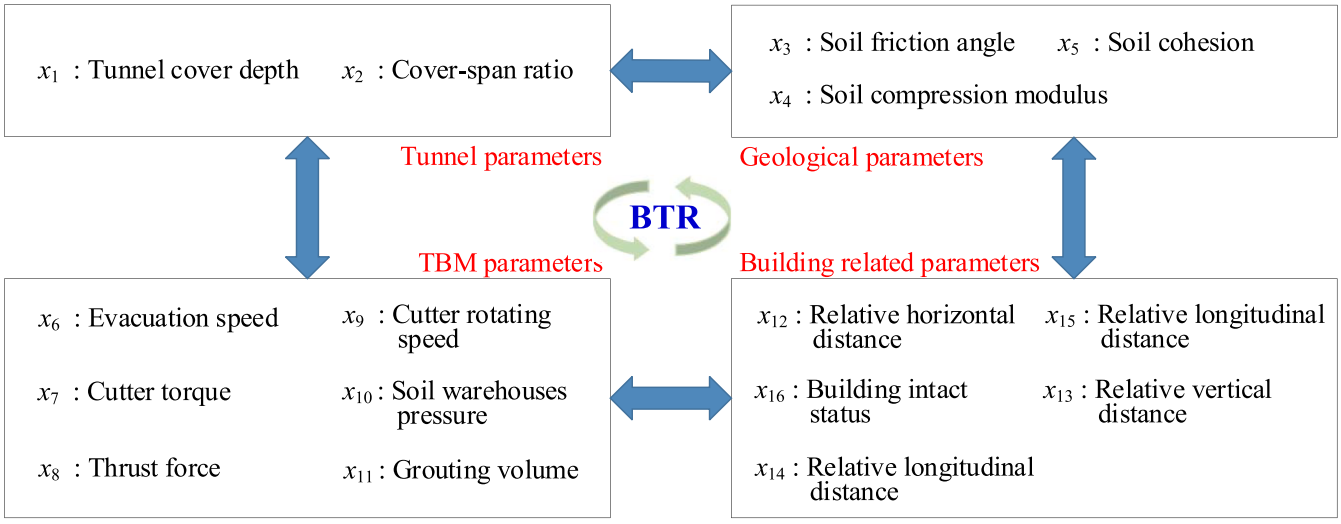


Fig. 1. Tunnel construction data with sixteen parameters in four categories as the input and the BTR as the output.

construction. The major steps involved are as follows. First, counterintuitive data is **defined** based on the causal logic relation between input and output. Next, multi-profile ensemble learning is utilized to initially **identify** counterintuitive data, which is then confirmed by multi-model ensemble learning. Finally, counterintuitive data is **handled** by assigning it a reduced weight in the safety assessment model training process. To validate the efficiency of this approach, a practical case study on BTR prediction in metro tunnel construction in Wuhan City is conducted.

The theoretical contributions of this study are as follows, namely the definition, identification, and handling of counterintuitive data. Firstly, counterintuitive data is defined as data where the input and output do not adhere to a causal relation, distinguishing it from the traditional classification of normal/abnormal data, which focuses solely on input or output. Secondly, a novel multi-profile multi-model ensemble learning approach is employed for the identification of counterintuitive data. Multi-profile ensemble learning generates multiple profiles of the original training data for cross-examine each set, while multi-model ensemble learning eliminates interference from inferior machine learning approaches. Lastly, counterintuitive data is handled by assigning it a smaller weight for constructing a more accurate safety assessment model.

The rest of this paper is organized as follows. A review of the relevant literature is provided in Section 2. Then, the counterintuitive data and related concepts are defined in Section 3, and the new approach is proposed in Section 4. Case study and computational results are provided in Section 5. Section 6 presents further validation of the motive of the proposed approach. Lastly, the final section concludes the paper and provides some future research directions.

2. Related works

2.1. Tunnel construction data for building tilt rate reduction

Tunneling is one of the main activities in metro construction. Ensuring the safety in tunneling is very important, particularly through the monitoring of the BTR of nearby buildings within a safety range (Patrucco, Pira, Pentimalli, Nebbia, & Sorlini, 2021; Okudan, Budayan, & Dikmen, 2021). To predict the BTR, (Feng & Zhang, 2021; Zhang, Wu, Ji, & AbouRizk, 2017) identified a total of 16 parameters categorized into four groups, as depicted in Fig. 1. These include (i) geological parameters (with three sub-parameters), (ii) tunnel-related parameters (with two sub-parameters), (iii) TBM-related parameters (with six sub-parameters), and (iv) building-related parameters (with five sub-parameters). Consequently, these sixteen parameters from the four categories are utilized as input variables, while the BTR serves as the output for developing the safety assessment model for tunnel construction.

The following introduces how the parameters are derived as well as the types.

(1) The tunnel-related parameters are calculated using coordinates from total station instrument;

(2) The geological parameters are pre-measured in the lab in the pre-geological survey;

(3) The TBM-related parameters are directly gathered from the TBM;

(4) For the building-related parameters, the relative horizontal/vertical/longitudinal distance is measured using the total station instrument, and the building foundation integrity and the building intact status are scores produced by the engineers.

2.2. Traditional anomaly detection in building tilt rate reduction prediction

Anomaly detection (AD) has traditionally been a prevalent approach for safety assessment in tunnel construction (Bamaqa, Sedky, Bosakowski, Bastaki, & Alshammari, 2022; Liu et al., 2021). This method involves close monitoring of safety-related data to promptly identify anomalies, such as high torque in the TBM (Rong, Lu, Wang, Wen, & Rong, 2019), breach of preset boundaries between the underground tunnel site and buildings (Kang, 2019), or exceeding the safety threshold of the BTR (Chang, Zhang, Fu, & Chen, 2022; Zhang, Wu, Ji, & AbouRizk, 2017). Whenever an anomaly is detected, construction activities are halted until the risks are eliminated and the data returns to normal. The AD-based approach is straightforward to implement, as safety thresholds are predefined without ambiguity. Hence, it is effective for practical purposes, serving as a means to identify potential risks that could lead to construction accidents and require immediate elimination through predetermined protocols (Zhang, Wu, Ji, & AbouRizk, 2017). However, the AD-based approach cannot function as a preventive measure to anticipate and prevent hazards from occurring initially (Quatrini, Costantino, Gravio, & Patriarca, 2020; Ruff et al., 2021).

Recent advancements in machine learning (ML) have introduced a new approach to safety assessment in tunnel construction (Bai, Cheng, & Li, 2021). The common practice of the ML-based approach involves constructing a safety assessment model using archived data and using this model to make predictions when new data is collected (Lin, Shen, & Zhou, 2021; Jordan & Mitchell, 2015). With an accurate ML-based safety assessment model, it becomes possible not only to identify immediate dangers in tunnel construction but also to predict hidden risks and take preventive measures in advance, thereby avoiding potential economic losses and even saving lives (Ye, Jin, & Chen, 2022). Moreover, the ML-based safety assessment model can be implemented automatically, reducing the resources required compared to traditional anomaly detection methods (Bai, Cheng, & Li, 2021). However, a major drawback of ML-based approaches is that they require comprehensive training to ensure accurate safety assessment (Bejani & Ghatee, 2021), as ML is inherently data-driven (Bergen, Johnson, de Hoop, & Beroza, 2019). Therefore, there is a significant need for accurate data, typically obtained through highly reliable sensors. However, the harsh working environment in underground tunnel construction, characterized by high temperatures and humidity, can lead to accelerated sensor degradation, subsequently affecting data accuracy (Zhang, Wu, Ji, & AbouRizk, 2017).

To summarize, there are both advantages and disadvantages to the traditional AD-based approaches and the ML-based approaches for the safety assessment of tunnel construction. Comparatively, the traditional AD-based approach is more pragmatic and relatively easy to implement (Bamaqa, Sedky, Bosakowski, Bastaki, & Alshammari, 2022; Liu et al., 2021) whereas the ML-based approach can make predictions beforehand to provide beforehand warning (Bai, Cheng, & Li, 2021). Hence, a natural approach is to differentiate abnormal data from normal data using anomaly detection to train the ML-based approach and enhance its accuracy. However, there is an inconsistency between the AD- and the ML-based approaches. Specifically, the AD-based approach by definition detects anomalies if any out-of-the-normal range is found in safety-related data of tunnel construction, e.g., over speeding of the TBM machine (Rong, Lu, Wang, Wen, & Rong, 2019) or a too-high BTR (Zhang, Wu, Ji, & AbouRizk, 2017). Comparatively, the ML-based approach predicts the BTR as a causal reaction to multiple geological-, building-, and TBM-related parameters (Jordan & Mitchell, 2015). In one word, the AD-based approach focuses on the normal/abnormal condition of data, either the input or the output part, whereas the ML-based approach focuses on the input–output causal relation of data. Building on this, if the detected abnormal data is excluded from training the ML-based safety assessment model, only normal data would be left that is all within the safety (normal) range. Then, the trained ML-based safety assessment model would no longer be able to recognize any abnormal data, i.e., the model has become overfit, because it has never "seen" any abnormal data.

Be reminded that the ML-based safety assessment model produces the BTR as the output that is the causal output of multiple geological-, building-, and TBM-related parameters as the input. To further explain,

Table 1
Comparison between traditional ensemble learning and multi-profile multi-model ensemble learning.

Type	Procedures	Main feature and focus
Traditional ensemble learning	Including bagging and boosting; All sub-models are related.	Final output is one model; Does not focus on data.
Multi-profile* ensemble learning	Mainly rooted from bagging; Each profile (each sub-model) is independent, and all sub-models are not related; Each profile makes the majority of the training dataset, yet all profiles are not identical.	Does not produce a final model; Primary focus is on data: a set of data is counterintuitive data if there are relative bigger sub-errors from multiple sub-models.
Multi-model ensemble learning	To repeat multi-profile ensemble learning using multiple machine learning approaches.	To rule out the interference of an inferior machine learning approach.

*Profile: By randomly selecting a large proportion, e.g., 80%, of data from the training dataset, a profile is created that guarantees (1) each profile is a representee of the original training dataset, and (2) all profiles are not identical.

when the geological-, building-, and TBM-related parameters are abnormal, the model should predict the BTR as abnormal, or when those parameters are normal, the BTR should be predicted as normal, which are normal data that makes the majority of all data in tunnel construction. These intuitive data exhibit a causal relation between input and output. However, there are also counterintuitive data where the parameters are abnormal, but the BTR is normal, or vice versa. The safety assessment model is not expected to recognize such counterintuitive data. If both intuitive and counterintuitive data are considered by the model, a conflict arises in determining the BTR prediction when all parameters are normal. This conflict forms the fundamental motive of this study.

2.3. Ensemble learning

Initially proposed by Leo Breiman in 1996 (Breiman, 1996), ensemble learning combines predictions from multiple individual models to enhance prediction accuracy. It has been demonstrated to possess higher generalization ability, robustness against overfitting, and the ability to capture diverse aspects of the data by leveraging multiple models (Zhou, 2021). However, ensemble learning also presents certain disadvantages. It demands additional computational resources and training time due to the need to train multiple models. The interpretability of ensemble models may be compromised as they involve combining outputs from different models. Ensemble learning is also sensitive to noise and outliers in the data, potentially impacting the performance of individual models and consequently the ensemble (Kato, Mao, Tang, Kawamoto, & Liu, 2020). Bagging and boosting are among the most popular variants of ensemble learning. Bagging involves training multiple models on different bootstrapped samples of the training data and averaging their predictions (Ribeiro & Coelho, 2020). Boosting iteratively updates models to produce a final unified model (Bentéjac, Csörgő, & Martínez-Muñoz, 2021). Either way, they both belong to ensemble learning with the objective of minimizing the error of the training dataset and improving the accuracy of the testing dataset regardless of the specific baseline model, or employed as bagging or boosting. Note that ensemble learning belongs to supervised learning in a generic sense (Chiang, Shih, Lin, & Shih, 2014; Arican & Aydin, 2022) where the validation metric is to minimize the modeling error, which is different from unsupervised learning such as clustering (Borlea, Precup, & Borla, 2022) where the metrics include the Davies-Bouldin Index, the Silhouette Index, etc.

However, in ensemble learning, all data is treated equally without regard for the input–output relationship. In this study, the objective is to accurately predict the BTR in metro tunneling, with a specific focus on differentiating counterintuitive data from intuitive data. To achieve this objective, a multi-profile multi-model ensemble learning approach is proposed, incorporating the following aspects:

(1) A multi-profile ensemble learning procedure is designed for initially identifying counterintuitive data from the complete training datasets. Specifically, "multi-profile" is employed by training multiple sub-models using multiple sub-datasets which are sampled from the original complete datasets. In this sense, (i) each sub-model represents a profile and (ii) each one is different from anyone else. By comparing and integrating multiple profiles, the counterintuitive data is thus identified.

(2) A multi-model ensemble learning procedure is designed for confirming counterintuitive data. As only one machine learning approach is adopted as the baseline approach in multi-profile ensemble learning, it could suffer from possible inefficiency of the specific baseline approach. Thus, multiple machine learning approaches should be used for cross-examining the results. As a result, counterintuitive data is confirmed when identified in the multi-model ensemble learning.

Table 1 summarizes the difference between the intended multi-profile multi-model ensemble learning approach and traditional ensemble learning approach.

3. Definition of counterintuitive data in metro tunnel construction

As described in Section 2.1, there are sixteen parameters used as the input for predicting the BTR (the output), which is also the data

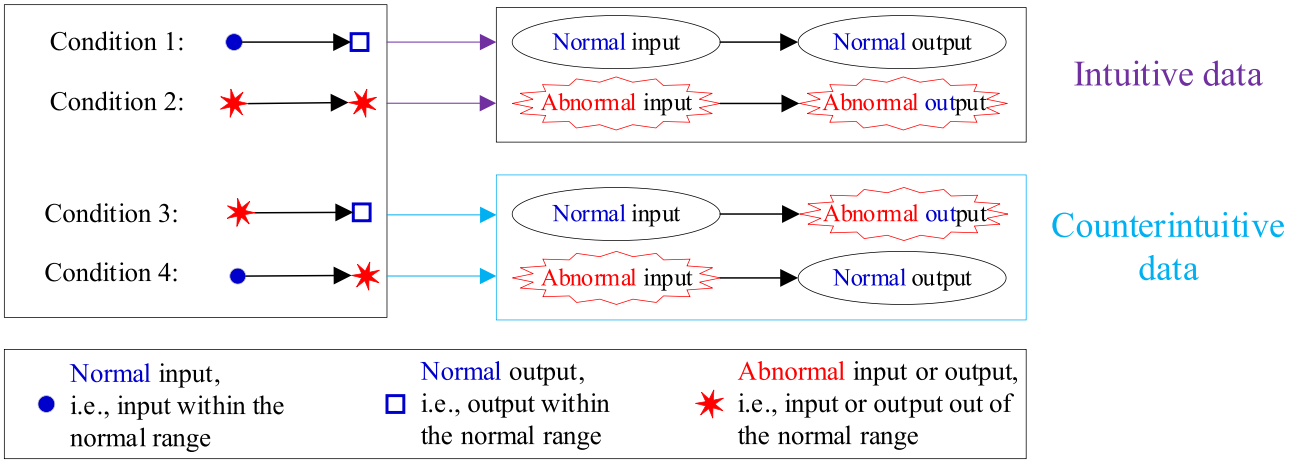


Fig. 2. Definition of Intuitive data and counterintuitive data.

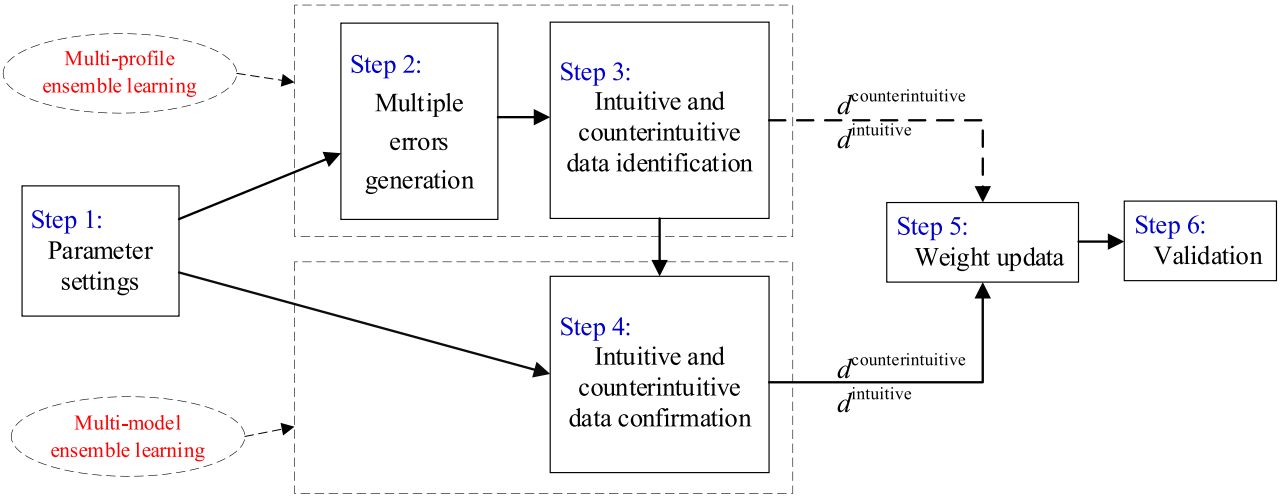


Fig. 3. The framework of the proposed multi-profile multi-model approach.

structure in this study. Nonetheless, the definitions of counterintuitive data, intuitive data, and abnormal data are still presented in a generic sense. Suppose that a data pair (I, y) in metro tunneling is with two parts, namely the input part $I = [x_1, x_2, \ldots, x_m, \ldots, x_M]$ with M attributes and the output part y, i.e., the BTR. Each attribute x_m in the input and the output y are with a respective safety range, i.e., $x_m \in [rg(x_m)\ \min, rg(x_m)\ \max]$ and $y \in [rg(y)\ \min, rg(y)\ \max]$. The following gives Definitions 1–4 (see Fig. 2).

Definition 1 Causal relation: it refers to the relation between the input and output of data that is consistent with the causal logic. Specifically, the mapping relationship f: I → y is a causal relation, where a normal input leads to a normal output or an abnormal input leads to an abnormal output, i.e., $I^{normal} \rightarrow y^{normal}$ or $I^{abnormal} \rightarrow y^{abnormal}$ where the input is normal I^{normal} if $x_m \in [rg(x_m)\ \min, rg(x_m)\ \max]$, m = 1,…,M, and it is $I^{abnormal}$ if otherwise, and the output is normal y^{normal} if $y \in [rg(y)\ \min, rg(y)\ \max]$, and it is $y^{abnormal}$ if otherwise.

Definition 2 **Intuitive data** (see Fig. 2): Data pair (I, y) is called intuitive data and is marked as $d^{intuitive}$, if it satisfies the causal relationship f^{causal}: I → y. Note that there is $d^{intuitive}$: $I^{normal} \rightarrow y^{normal}$ or $d^{intuitive}$: $I^{abnormal} \rightarrow y^{abnormal}$. Data in Conditions 1–2 are recognized as intuitive data. The majority of all data is intuitive data, especially data in Condition (1).

Definition 3 **Counterintuitive data** (see Fig. 2): Data pair (I, y) is called counterintuitive data and is marked as $d^{counterintuitive}$. Specifically, counterintuitive data is with a normal input and an abnormal output, i.e., $d^{counterintuitive}$: $I^{normal} \rightarrow y^{abnormal}$, or an abnormal input and a normal output, i.e., $d^{counterintuitive}$: $I^{abnormal} \rightarrow y^{normal}$. Data in Conditions 3–4 are recognized as counterintuitive data. Comparatively, counterintuitive data occupies a minority of all data.

Definition 4 Abnormal data: Data pair (I, y) is called abnormal data whose input I or output y is abnormal, i.e., $I = [x_1, x_2, \ldots, x_m, \ldots, x_M]$ where $xm \notin [rg(x_m)\ \min, rg(x_m)\ \max]$, m = 1,2,…,M, or $y \notin [rg(y)\ \min, rg(y)\ \max]$. Thus, there is, $d^{abnormal}$: $I^{abnormal}$ or $d^{abnormal}$: $y^{abnormal}$. Abnormal data refers to Conditions 2–4 in Fig. 2.

According to above **Definitions 1–4**, the following points can be summarized:

(1) Intuitive data and counterintuitive data are exclusive, i.e., $d^{intuitive} \cap d^{counterintuitive} = \emptyset$.

(2) Abnormal data is irrelevant to intuitive data and counterintuitive data as it does not focus or reflect the relation between the input and output. Abnormal data only depicts the condition of the input or the output of data. Thus, abnormal data could be either intuitive data or counterintuitive data (see Conditions 1–4 in Fig. 2, and Definitions 2–4).

(3) Outlier data in some studies is referred to as data with an out-of-the-normal-range input or output (Alghushairy, Alsini, Soule, & Ma, 2020; Blázquez-García, Conde, Mori, & Lozano, 2021), which would make outlier data and abnormal data equivalent concepts according to Definition 4.

4. Approach

4.1. Framework

Fig. 3 gives the framework of the proposed multi-profile model-model ensemble learning approach where the pseudocodes and programming codes (in MATLAB) of Steps 2–6 are given in XXXXX.

Step 1: Parameter settings;

Set the parameter settings, which encompass determining the size of the training dataset and testing dataset. Select multiple machine learning approaches with their specific parameter settings. Set the marginal error threshold for counterintuitive data identification.

Step 2: Multi-profile ensemble learning to generate multiple sub-errors.

Create multiple profiles of the original training dataset by (i) sampling with replacement to generate multiple sub-datasets, (ii) constructing multiple sub-models using multiple sub-datasets, and (iii) producing multiple sub-errors for the sub-datasets. More details are given in Section 4.2.

Step 3: Multiple sub-errors-based initial counterintuitive data identification.

Identify initial intuitive and counterintuitive data according to the marginal error threshold (set in Step 1), the sub-errors (calculated in Step 2), and the weights of sub-models (calculated using the sub-errors). More details are given in Section 4.3.

Step 4: Multi-model ensemble learning for further counterintuitive data confirmation.

Confirm intuitive and counterintuitive data by comparing the results from multiple machine learning approaches. When multiple machine learning approaches identify a dataset as counterintuitive, its status as counterintuitive data is further affirmed. Thus, a set of data is intuitive data if and only if it is recognized by all machine learning approaches. More details are given in Section 4.4.

Step 5: Data handling by updating weights for counterintuitive data.

The weights for counterintuitive data should be updated according to how many times it has been confirmed by multiple machine learning approaches. Specifically, the weight for counterintuitive data should be much reduced if it has been confirmed by more machine learning approaches. More details are given in Section 4.5.

Step 6: Final model construction and validation of the testing dataset.

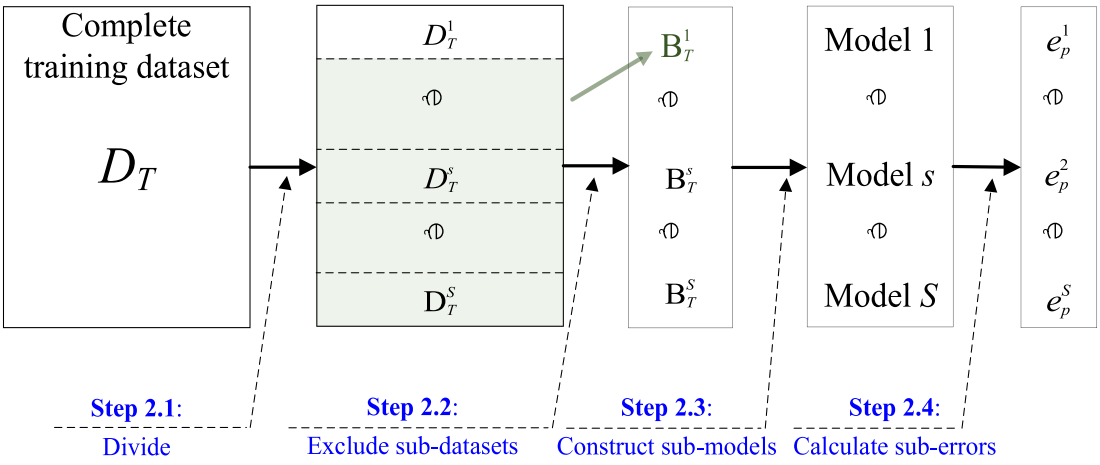


Fig. 4. Steps of multi-profile ensemble learning.

Table 2
Calculation of multiple sub-errors following Step 2.

Data Nos.*	Sub-errors			
	Sub-model 1	Sub-model 2	Sub-model 3	Sub-model 4
1	/	2	1	3
2	10	/	3	8
3	1	1	/	1
4	1	1	1	/

*There should be more data and sub-models being constructed in practice to meet the statistical requirements.

With the updated weight, the complete training dataset is used to construct a final model and conduct validation of both the prediction accuracy and data type identification on the testing dataset. More details are given in Section 4.6.

4.2. Multi-profile ensemble learning for multi-sub-error generation

Step 2: Multi- profile ensemble learning for multi-sub-error generation.

In multi-profile ensemble learning, multiple profiles are created by sampling a majority of the original dataset to create multiple sub-datasets which are then used to construct multiple sub-models, i.e., sampling with replacement. Fig. 4 shows the sub-steps.

Step 2.1: Sample the original complete training dataset D_T into S equal partial-datasets *Ds T* by guaranteeing that any set of data only belongs to one and only one partial-dataset.

Step 2.2: Construct S sub-datasets *Bs T* by excluding the *s*th partial-dataset from D_T, i.e., B1 T={ *D*1 *T*, …, *Ds T*, …, *DS T* }.

Step 2.3: Construct S sub-models using S sub-datasets. Select a machine learning approach to construct S sub-models using S-1 sub-datasets *Bs T*. Note that since all of S sub-datasets are not completely different, S sub-models are also different.

Step 2.4: Calculate the respective absolute error from the *s*th sub-model as the *s*th sub-error.

The absolute error *es p* of the *p*th set of data from the *s*th sub-model is calculated by Eq. (1),

$$e_p^s = |y_p - y_p^s| \tag{1}$$

where y_p denotes the actual output of the *p*th set of data, and *ys p* denotes the output from the *s*th sub-model of the *p*th set of data. Note that there would be no *ys p* if the *p*th set of data is not used in constructing the *s*th sub-model. If so, there is *es p* = 0.

Example 1 Suppose that there is an original dataset is with four sets of data, namely Data Nos. 1, 2, 3, and 4, i.e., d_1, d_2, d_3, and d_4. Next, the implementation of Step 2 is given as follows.

(1) Following Step 2.1, they are naturally divided into four "equal partial-datasets" and each is with a set of data.

(2) Following Step 2.2, four sub-datasets are created, namely B1 T={d_2, d_3, d_4}, B2 T={d_1, d_3, d_4}, B3 T={d_1, d_2, d_4}, and B4 T={d_1, d_2, d_3}.

(3) Following Step 2.3, four sub-models are constructed using respective sub-datasets.

(4) Following Step 2.4, the sub-errors are calculated from sub-models, as presented in Table 2.

By illustrating the implementation details of Step 2, four sub-datasets/models have helped create four profiles of the original dataset which are not revealed by the original dataset. For example, it is improbable to directly tell any original data may be counterintuitive. Yet, the sub-errors in Table 2 implicitly indicate that Data No. 2 may be counterintuitive as it can not be recognized by most sub-models. The next Step 3 would further explain how counterintuitive data is identified.

4.3. Multi-error-based initial counterintuitive data identification

Step 3: Multi-errors-based initial counterintuitive data identification.

Step 3.1: Calculate the weights of S sub-models based on the results of S sub-models.

The mean absolute percentage error (MAPE) of the sub-models is calculated to denote their ability to recognize the respective sub-training dataset by Eq. (2).

$$MAPE_s = \frac{1}{P}\sum_{p=1}^{P}\frac{e_p^s}{y_p} \tag{2}$$

where there are P sets of data to construct the *s*th sub-model, and *es p* is calculated by Eq. (1). There is MAPEs ∈ [0, 100 %] with $MAPE_s = 0$ or $MAPE_s = 100$ % denoting the highest and lowest modeling ability of the *s*th sub-model. Then, MAPE is used to calculate the weights of sub-models.

$$w_s = 1 - MAPE_s \tag{3}$$

where there is $w_s \in [0, 1]$. For example, $w_s = 1$ denotes that the *s*th sub-model can recognize all training data at 100 % accuracy since $MAPE_s = 0$, and it should be granted with the highest weight, and vice versa.

Step 3.2: Calculate the weighted average error of data from S sub-models.

Calculated the average error e_p of the *p*th set of data from S-1 sub-models (any data is only used to construct S-1 sub-models) by considering the original errors and the weights of the sub-models.

Table 3
Identification of counterintuitive data following Step 3.

Data Nos.*	Original y**	Sub-errors				Weighted avg. error	Data type
		Sub-model 1	Sub-model 2	Sub-model 3	Sub-model 4		
1	20	/	2	1	3	0.09	Intuitive
2	20	10	/	3	8	0.29	Counterintuitive
3	20	1	1	/	1	0.04	Intuitive
4	20	1	1	1	/	0.04	Intuitive
MAPEs		0.20	0.07	0.08	0.20	The marginal error is set as 0.10.	
Weights		0.80	0.93	0.92	0.80		

*There should be more data and sub-models being constructed in practice to meet the statistical requirements.
**y is normally different from different data. This is only for the simplicity of demonstration.

$$e_p = \frac{1}{S-1}\sum_{s=1,s\neq s^*}^{S}(w_s e_p^s) \tag{4}$$

where *es p* denotes the original error from Eq. (1) of Step 2.4, and w_s denotes the weight of the *s*th sub-model from Eq. (3) of Step 3.1, and s^* denotes the sub-model that the *p*th set of data is not used to construct. Nonetheless, since Step 2.4 has required that $es^* p = 0$ if the *p*th set of data is not used to construct the s^* sub-model, Eq. (4) can also be transferred into the following Eq. (5).

$$e_p = \frac{1}{S-1}\sum_{s=1}^{S}(w_s e_p^s) \tag{5}$$

Step 3.3: Identify intuitive data and counterintuitive data according to a marginal error threshold.

Rearrange the data according to the weighted errors in ascending order to form an ordered list Ω,

$$\mathbf{\Lambda} = [d_1, d_2, \cdots, d_{p'}, d_{p'+1}, \cdots, d_{P-1}, d_P] \tag{6}$$

where there is $e_1 < e_2 < \ldots < e_p < e_{p+1} < \ldots < e_{P\text{-}1} < e_P$.

Calculate the marginal errors by the following Eq. (7),

$$\Delta e_{p,p+1} = e_{p+1} - e_p \tag{7}$$

Identify intuitive data and counterintuitive data by comparing the marginal errors calculated in Eq. (7) with a preset marginal error threshold ν, as indicated by the following Eq. (8)

$$\begin{cases} \Omega_{\text{intuitive}} = [d_1, d_2, \cdots, d_{p^*}] \\ \Omega_{\text{counterintuitive}} = [d_{p^*+1}, \cdots, d_{P-1}, d_P] \end{cases} \text{if } \Delta e_{p^*,p^*+1} > \nu \tag{8}$$

where $\Omega_{\text{intuitive}}$ and $\Omega_{\text{counterintuitive}}$ denote the datasets with intuitive and counterintuitive data, respectively.

Example 2 Building upon Example 1, still suppose that there is an original dataset is with four sets of data, namely Data Nos. 1, 2, 3, and 4, i.e., d_1, d_2, d_3, and d_4. Next, the implementation of Step 3 is given as follows and also in Table 3. (1) Following Step 3.1, the MAPEs and the weights are calculated, as presented in Table 3. (2) Following Step 3.2, the weighted average errors are calculated. (3) Following Step 3.3, the counterintuitive data is identified by setting the marginal error as 0.10. According to Table 3, it is clear that the weighted average error of Data No. 2 is significantly higher than the marginal error. Thus, Data No. 2 is identified as counterintuitive data according to Step 3.

To further elaborate, traditional approaches treat all data equally without distinction, which indirectly "favors" data with a higher error as the mean error is normally used as the training objective. Example 2 (also this study) demonstrates that this should be corrected: some data, e.g., Data No. 2, does not need to be favored because it represents a counterintuitive input–output relation. This is also the motive and major theoretical contribution of this study.

4.4. Multi-model ensemble learning for further counterintuitive data confirmation

Step 4: Multi-model ensemble learning for further counterintuitive data confirmation;

Step 4.1: Repeat Steps 2 and 3 for *M* machine learning approaches to produce *M* identification results for counterintuitive data.

Step 4.2: Confirm counterintuitive data according to *M* identification results from *M* machine learning approaches by the following Eq. (9).

$$\text{the } p\text{th set of data is} = \begin{cases} d^{\text{counterintuitive}} \text{ if it is identified by at least one ML approach} \\ d^{\text{intuitive}} \text{ if it is NOT identified by all } M \text{ approaches} \end{cases} \tag{9}$$

where it indicates that the *p*th set of data is (1) confirmed as counterintuitive data if it has been identified as counterintuitive data by at least one machine learning approach, and it is (2) confirmed as intuitive data if and only if it has NOT been identified by all *M* approaches.

4.5. Data handling by updating weights for counterintuitive data

Step 5: Data handling by weight updating for counterintuitive data.

Step 5.1: Calculate the frequency f_p of the *p*th set of data being confirmed as counterintuitive data by *M* machine learning approaches.

Step 5.2: Set weight penalty coefficients for counterintuitive data according to problem requirements, e.g., $\mu^{-1} = 10^{fp} = [10^1, 10^2, 10^3, \ldots]$ for $f_p = 1, 2, 3, \ldots$.

Step 5.3: Update the weight for counterintuitive data.

For the confirmed counterintuitive data, their weights should be updated accordingly by Eq. (10),

$$w_p = \mu w_{p,0} \tag{10}$$

where $w_{p,0}$ and w_p denote the weight of the *p*th set of data before and after identifying the type of data, respectively. μ denotes the updating weight for different types of data. If the intuitive data is also included in Eq. (10) for comprehensiveness, Eq. (10) can be translated into the following Eq. (11),

$$\begin{cases} w_p = w_{p,0} \textit{ if the pth set of data is intuitive data} \\ w_p = \mu_{p,0} \textit{ if the pth set of counter is intuitive data} \end{cases} \tag{11}$$

where it indicates that the more often a set of data is confirmed as counterintuitive data, the more its weight should be reduced.

4.6. Final model construction and validation on the testing dataset

Step 6: Final model construction and validation on the testing dataset.

Step 6.1: Final model construction using the complete training dataset.

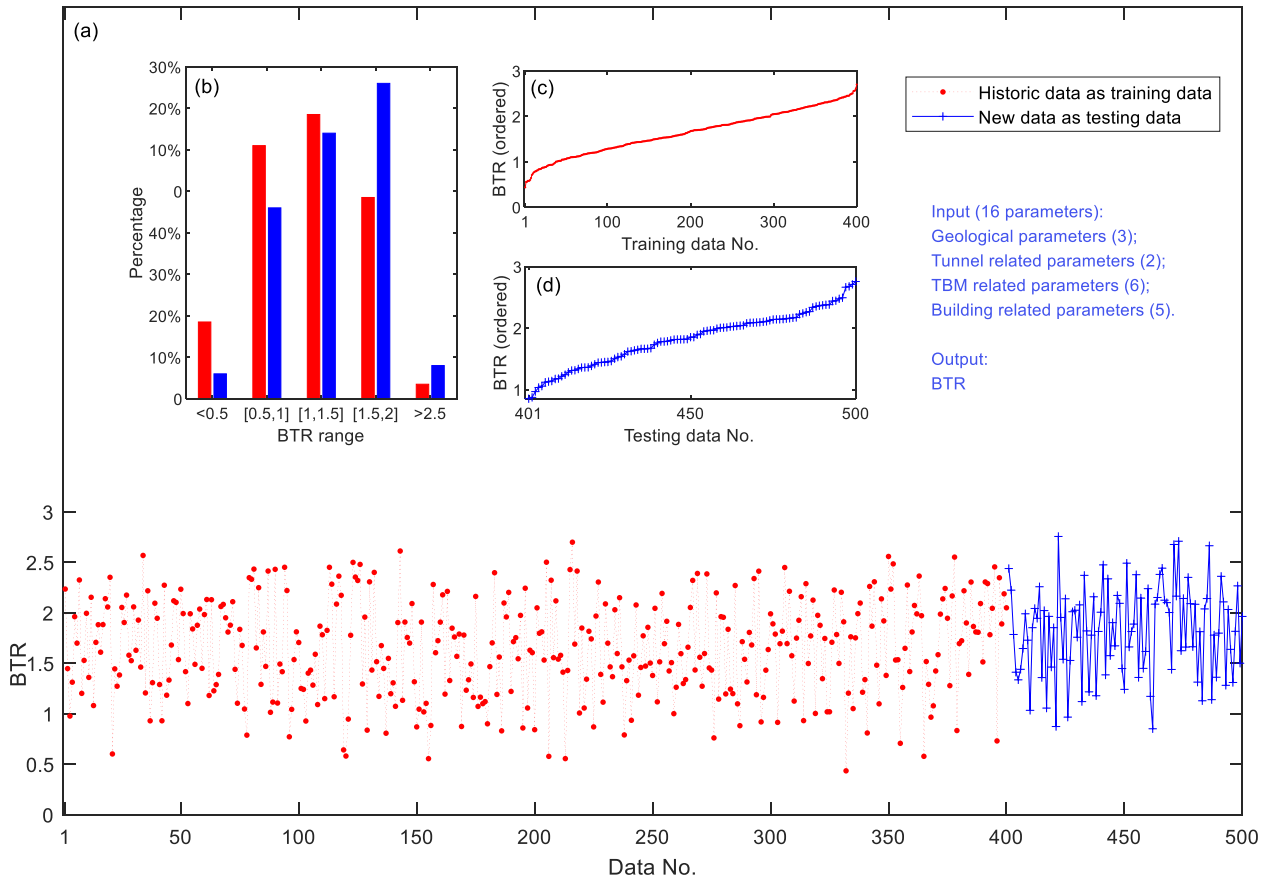


Fig. 5. Data distribution. (a) BTRs of all data in original order, (b) BTR distribution of all data in varied ranges, (c) training data by BTR in ascending order, and (d) testing data by BTR in ascending order.

With the updated weight, the complete training dataset D_T with P sets of data is used to construct a final model. As multiple machine learning approaches are used in the proposed approach, each machine learning approach should be tested separately to construct the final model. The objective function in modeling training for the final model is given in Eq. (13),

$$MAE_s = \frac{1}{P}\sum_{p=1}^{P} w_p |y_p - y_p^s| \tag{13}$$

where *ys p* denotes the predicted result produced by the *s*th machine learning approach for the *p*th set of data, y_p denotes the actual output of the *p*th set of data, and w_p denotes the weight of the *p*th set of data determined by Step 5 in Section 3.5 according to the data type.

Step 6.2: Validation of the prediction accuracy on the testing dataset.

For the testing dataset D_V with Q sets of data, the mean absolute error (MAE) is used as the performance indicator and it is calculated by the following Eq. (14),

$$MAE = \frac{1}{Q}\sum_{q=1}^{Q} |y_q^a - y_q^e| \tag{14}$$

where *ya q* and *ye q* denote the actual and predicted output of the *q*th set of testing data.

Step 6.3: Validation of the data type identification on the testing dataset.

According to the prediction errors by the machine learning approaches with and without the proposed approach (with and without identifying varied types of data), the data type of the testing dataset can also be identified.

5. Practical case study

5.1. Background

A practical case of Metro Line No. 4 in the city of Wuhan, Hubei, China, is studied (Chang, Zhang, & Xu, 2022; Zhang, Wu, Zhu, & AbouRizk, 2017a,b). A total of 500 sets of data have been systematically collected from three neighboring metro stations (see Fig. 5(a)). All data is with sixteen parameters of four categories (see Section 2.1) and a BTR. Thus, the sixteen parameters can be used as input to predict the BTR (range of BTR is given in Fig. 5(b)). As all data is sequenced according to the time they are collected, the first 400 sets of data are used as the training dataset while the rest 100 sets of data are used as the testing dataset. Fig. 5(c/d) illustrate the ascending order distribution of the BTR within the training and testing dataset. The proximity of the BTR distribution between these datasets can be attributed, in part, to the relatively uniform geological and urban development conditions across the three stations.

5.2. Identification of counterintuitive data

Two machine learning approaches are employed: the back-propagation neural network (BPNN) (Ding, Wang, Han, & Wei, 2018) and the gradient process regression (GPR) (Park et al., 2017). The parameter settings are as follows:

(1) For BPNN, it is implemented using the *nntool* of Matlab. The number of layers is 3, the number of neurons is 4, epochs is 1000, the goal is 5E-4, the transfer function is *trainlm*.

Table 4
Procedures of identifying and confirming counterintuitive data.

Contents	Steps	Implementation	Identified or confirmed counterintuitive data*	Intuitive data	Figs
Initial identification of counterintuitive data	Steps 2–3	BPNN initially identifies 9 sets of data as counterintuitive data.	Data Nos. **370**, 30, **251**, 59, **101**, **228**, **281**, 25, **2**	All the rest 391 sets of data	Fig. 6(b)
		GPR initially identifies 12 sets of data as counterintuitive data.	Data Nos. 3, **370**, 139, **251**, 105, **101**, **228**, **281**, **2**, 396, 108, 332	All the rest 388 sets of data	Fig. 6(c)
Further confirmation of counterintuitive data	Step 4	6 sets of data are confirmed by both BPNN and GPR as counterintuitive data.	Data Nos. **370**, **251**, **101**, **228**, **281**, and **2**	All the rest 385 sets of data	Fig. 6(b), 6(c), and lower part of Fig. 6 (a)
		The rest 9 sets of data are confirmed by either BPNN or GPR.	Data Nos. 30, 59, and 25 by BPNN; Data Nos. 3, 139, 105, 396, 108, and 332 by GPR		

*The data number refers to the number of data in the original data training dataset.

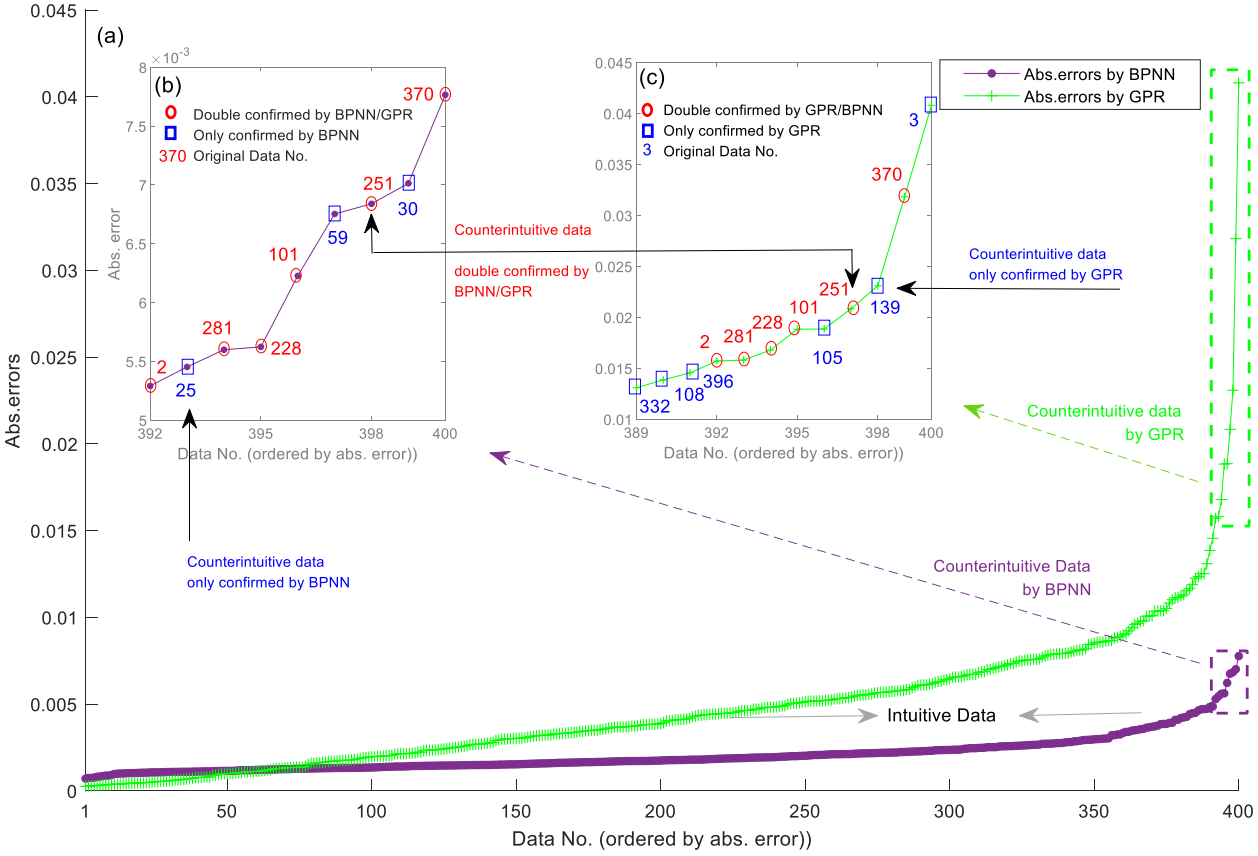


Fig. 6. Identification and confirmation of counterintuitive data: (a) Absolute errors of training data by BPNN and GPR, (b) Counterintuitive data identified by BPNN, and (c) Counterintuitive data identified by GPR.

(2) For GPR, it is implemented using the *fitrgp* function in Matlab. The fit method is *exact*, the predict method is *exact*, the explicit basis is *constant*.
(3) BPNN and GPR are first separately used to construct models for identifying intuitive and counterintuitive data. Initially, the original training dataset is divided into 10 sub-training datasets to separately construct sub-models, i.e., each sub-training dataset is composed of 400*(9/10) = 360 sets of data. In other words, each set of data is selected nine times to participate in constructing nine sub-models.
(4) The average errors are calculated based on the modeling accuracy of each sub-model, and the marginal error thresholds of 4.0000E-04, and 5.0000E-04 are set for BPNN and GPR to identify the counterintuitive data, respectively.

Table 4 and Fig. 6 shows the procedures for identifying and confirming counterintuitive data according to the proposed approach. Specifically for the result, (i) Data Nos. 370, 251, 101, 228, 281, and 2 are by both BPNN and GPR as counterintuitive data, (ii) Data Nos. 30, 59, and 25 are only identified by BPNN while Data Nos. 3, 139, 105, 396, 108, and 332 are only identified by GPR as counterintuitive data, and (iii) all the rest 385 sets of data are confirmed as the intuitive data.

5.3. Prediction results on the testing dataset

The weights for the intuitive and counterintuitive data are updated by the following Eq. (15),

$$\begin{cases} w_p = w_{p,0} if the p th set of data is intuitive data \\ w_p = \mu w_{p,0} if the p th set of data is counterintuitive data \end{cases} \quad (15)$$

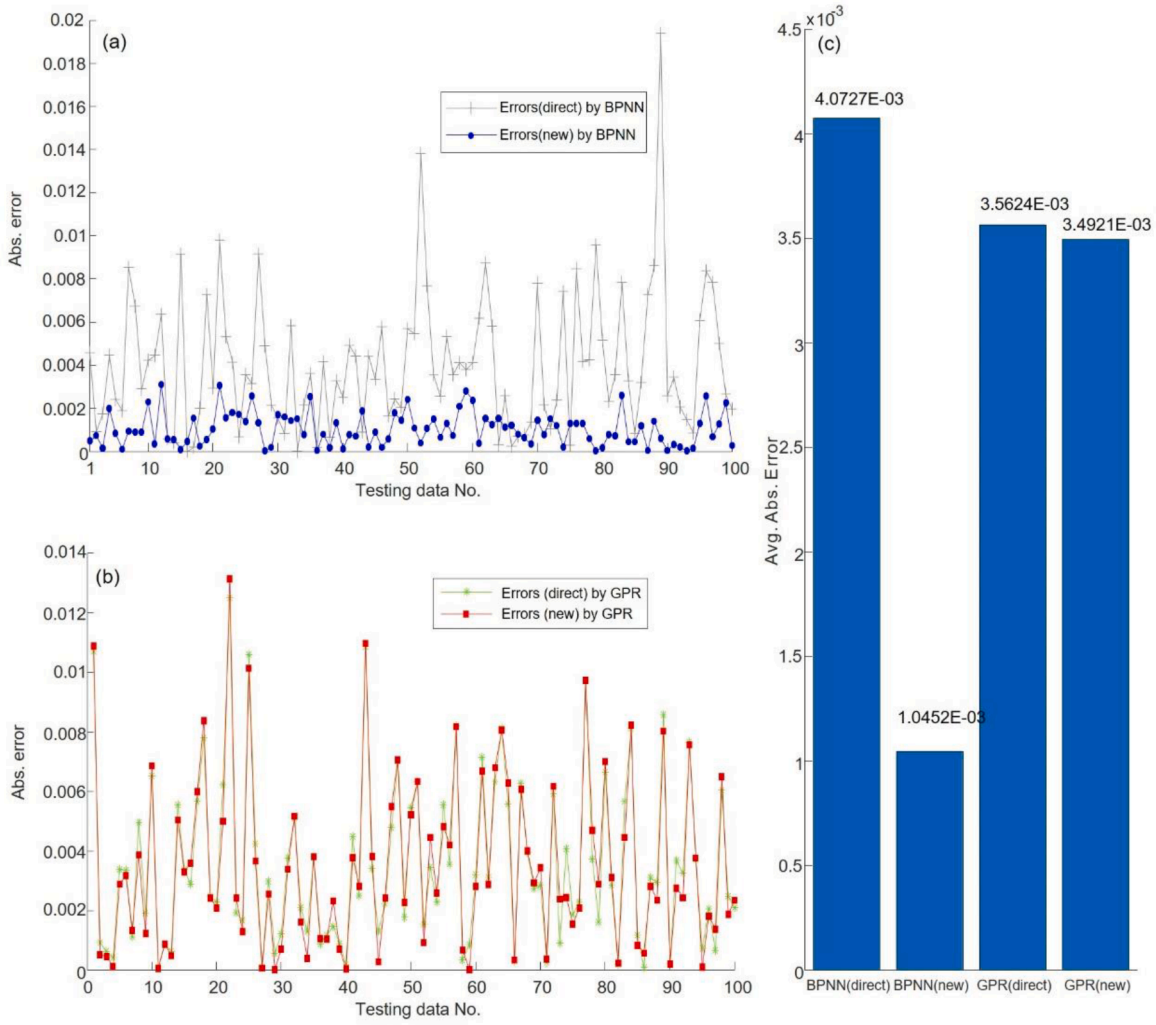


Fig. 7. Absolute errors on the testing dataset: (a) BPNN, (b) GPR, and (c) average results.

where the coefficient for weight update is set as $\mu = 0.1$, denoting that the weight is reduced to the tenth of the original weight if identified by either BPNN or GPR and $\mu = 0.01$ if confirmed by both BPNN and GPR.

The efficiency of the proposed approach is validated using a newly collected dataset consisting of 100 sets of data. Comparative results are presented in Fig. 7, which includes the original data, the predicted results by BPNN and GPR using the original 400 sets of data directly, and the predicted results by BPNN and GPR using the proposed approach. The results indicate that the proposed approach leads to enhance prediction accuracy when contrasted with the direct utilization of the original 400 historic data sets. For BPNN, the prediction results on the 100 sets of testing data show an improvement of 74.34 %, with a prediction result of 1.0452E-03 compared to 4.0727E-03 obtained by directly using the training dataset (see Fig. 7(a)). However, for GPR, there is no significant improvement observed, as the MAEs for directly using the training dataset and the new approach are 3.5624E-03 and 3.4921E-03, respectively (see Fig. 7(b)).

According to the results presented in Fig. 7, the following two conclusions can be drawn:

(1) BPNN as the baseline model has produced superior results to GPR. This can be attributed to BPNN's ability to effectively handle problems in high dimensions, such as the practical case with 16 parameters in the input.

(2) Consistency and logical results of BPNN vs. GPR: BPNN has consistently produced reliable and logical results, whereas GPR has not shown significant improvement. Despite varying results achieved by diverse machine learning approaches, it is anticipated that the proposed approach, which identifies and manages distinct data types, would surpass the direct utilization of the entire training dataset without such identification and management. Fig. 7 confirms this expectation, as BPNN achieved notably superior results using the proposed approach compared to the direct use of the complete training dataset (see Fig. 7 (a)), while GPR's results were indistinguishable (see Fig. 7(b)).

The convergence ability of the proposed approach is tested by measuring the number of iterations. Specifically, the proposed approach with BPNN, i.e., BPNN (new), requires 163 (100 %) rounds of iterations compared with 61 (37.42 %) rounds of iterations by directly adopting BPNN, i.e., BPNN (direct). This is most owing to the fact that counterintuitive data has been identified and handled by the proposed approach. In other words, the proposed approach has presented a "cleaner" data set for BPNN which saves the computational power. In other words, BPNN is sensitive to the quality of the dataset. The convergence investigation concerning GPR is nonapplicable as it is not with the iterative feature.

5.4. Identify intuitive and counterintuitive data in the testing dataset

Be reminded that the weight of the identified counterintuitive data is much reduced while the weight of the intuitive data remains stable by the proposed approach, i.e., $w_{\text{counterintuitive}} \ll 1$, $w_{\text{intuitive}} = 1$. Consequently, during the pretraining process, the safety assessment model

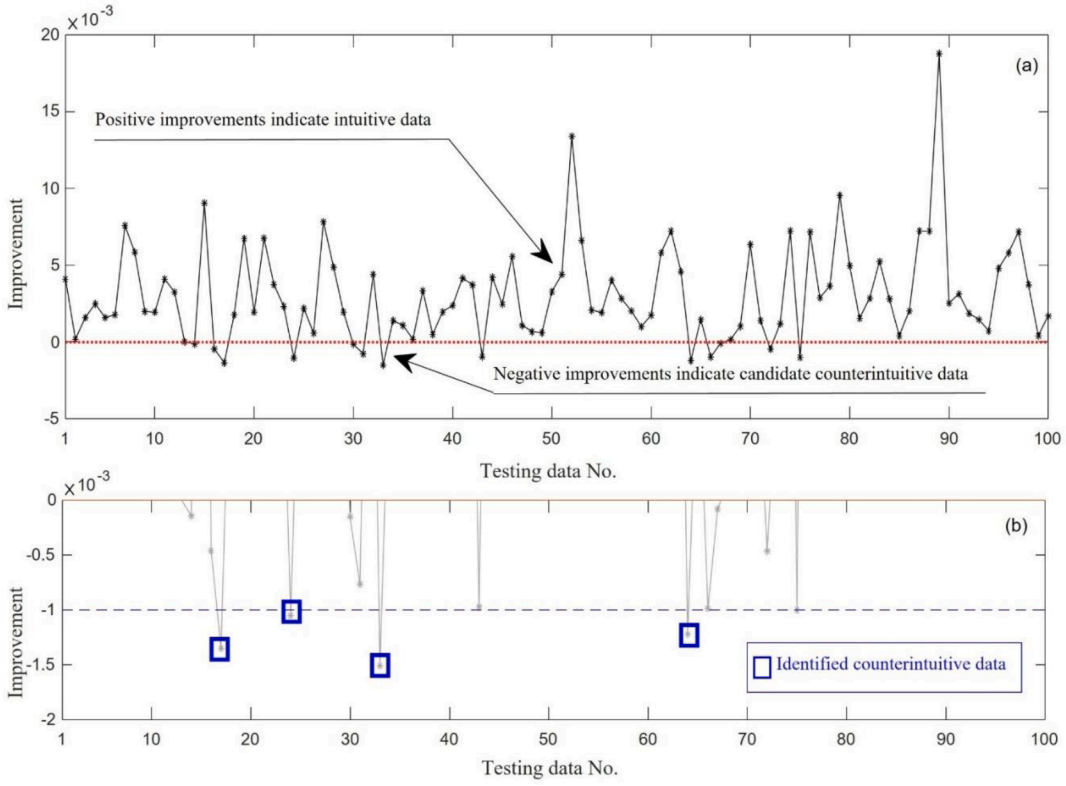


Fig. 8. Identification of counterintuitive data in the testing dataset: (a) results with the threshold = 0, and (b) results with the threshold = -1×10^{-4}.

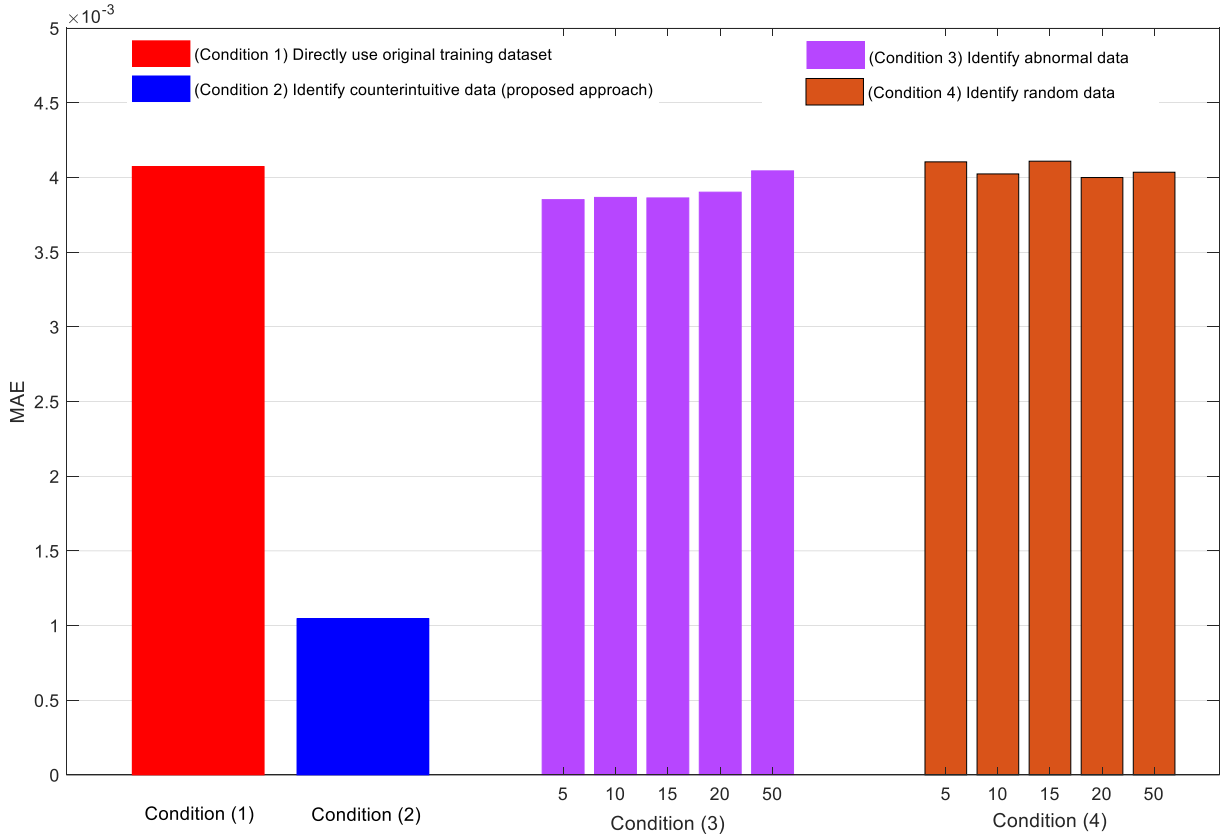


Fig. 9. MAEs comparison of four conditions for validation of the motive of the proposed approach.

Table 5
MAEs comparison using different approaches.

No.	Approaches	Parameter settings	MAE
1	BPNN	5 layers, 10 neurons, 1000 epochs, goal is 0, *trainlm*	4.0727E-03
2	GPR	fit method: *exact*, predict method: *exact*, explicit basis: *constant*.	3.5624E-03
3	SVM	kernel function: RBF, $C = 0.2$, $\sigma^2 = 10$	1.2254E-02
4	ANFIS	membership function: *gaussmf* (training/validation/testing: 300/100/100)	2.1764E-01
5	This study	Detailed procedures are given in Section 3.	1.0452E-03

should yield more accurate results for intuitive data compared to counterintuitive data. Considering that both intuitive and counterintuitive data exist in the testing dataset, it is also expected that more accurate results will be obtained for intuitive data compared to counterintuitive data in the testing phase. Given the superior and consistent performance of BPNN over GPR in previous sections, only BPNN is utilized in this analysis. Fig. 9 illustrates the comparative results, and a detailed analysis is provided below.

(1) If the threshold to differentiate intuitive and counterintuitive data is set as "0", as presented in Fig. 8 (a), the proposed approach identifies 87 sets of testing data as intuitive and the remaining as counterintuitive. However, the presence of counterintuitive data in the testing dataset (13 %) is significantly higher than that in the training dataset (3.75 %), as discussed in Section 4.2 and depicted in Fig. 6.
(2) Further investigation reveals that out of the 13 counterintuitive data sets, 9 exhibit very small differences (in the range of 10^-4 or even 10^-5), while the remaining 4 (Data Nos. 17, 24, 33, and 64) demonstrate relatively larger differences (see Fig. 8(b)). Considering this, if a threshold of "10^-3" or 0.001 is used to differentiate intuitive and counterintuitive data, the four testing data sets, equivalent to 4 % of the total (4/100), would be identified as counterintuitive data. This percentage closely aligns with the 3.75 % of counterintuitive data in the training dataset, indicating a similar distribution between the training and testing data.

5.5. Comparison with other approaches

BPNN and GPR had been separately used as the machine learning approaches in previous sections, and their results are again given in Table 5. Furthermore, two other approaches are used for comparison, namely the support vector machine (SVM) (Ouchen et al., 2020), which yielded an MAE of 1.2254E-02, and the Adaptive-Network-Based Fuzzy Inference System (ANFIS) (Karaboga and Kaya, 2019), resulting in an MAE of 2.1764E-01, respectively. Table 5 also gives the results and corresponding parameter settings. According to Table 5, the approach proposed in this study has produced superior results. Comparatively, SVM and ANFIS may be suitable for other cases but they are NOT selected as the machine learning approaches in this study owing to inferior performance in this case study.

6. Further validation of the motive by comparing with excluding abnormal data and randomly excluding data

Be reminded that the motive of this study is the inadequacy of traditional normal/abnormal data classification in supporting safety assessment in tunnel construction. In order to overcome this constraint, the study formulates, identifies, and manages intuitive and counterintuitive data. While the proposed approach has demonstrated more accurate predictions of BTRs compared to direct usage of BPNN/GPR, as well as SVM and ANFIS in Section 4, it is crucial to validate the motivation by testing the following two hypotheses:

Table 6
Four conditions for validation in the numerical case.

No.	Condition	Features	Specific procedure	No. of runs*
Condition (1)	Direct modeling	/	All data is assigned with equal weight, i.e., $w = 1$ for all data	1
Condition (2)	Proposed approach	Counterintuitive data	15 sets of counterintuitive data, $w = 0.1$ for 9 confirmed by one model, and $w = 0.01$ for 6 confirmed by double models, and the rest 385 as intuitive data, i.e., $w = 1$.	1
Condition (3)	Hypothesis 1	Abnormal data	5/10/15/20/50** sets of **abnormal** data, $w = 0.1$, and the rest as intuitive data, i.e., $w = 1$.	1
Condition (4)	Hypothesis 2	Random data	5/10/15/20/50** sets of **random** data, $w = 0.1$, and the rest as intuitive data, i.e., $w = 1$.	30*

*Since the size of the training dataset is fixed at 400 sets of data, there is no randomness in Conditions (1)-(3). Since multiple sets of data are randomly selected in Condition (4), the experiment is to be conducted 30 times to test its robustness;
**Since a total of 15 sets of data have been identified as counterintuitive data by the proposed approach (Condition (2)), varied sets of data, namely 5, 10, 15, 20, and 50, are tested in Conditions (3) and (4).

Hypothesis 1 Identify abnormal data and treat them as counterintuitive data. Identify abnormal data from the training dataset according to the BTR of data, and assign them with weights of w = 0.1 in constructing the safety assessment model.

Hypothesis 2 Randomly select data and treat them as counterintuitive data. Randomly select data from the training dataset, and assign them with weights of w = 0.1 in constructing the safety assessment model.

Table 6 summarizes more specifics of direct modeling (only using BPNN), the proposed approach, Hypothesis 1, and Hypothesis 2. Note that since BPNN has produced superior and consistent results over GPR in previous sections, only BPNN is adopted in Hypothesis 1 and Hypothesis 2 (Table 7).

Table 5 and Fig. 8 show the results with varied parameter settings in different conditions.

According to Table 5 and Fig. 8, the following conclusions can be drawn.

(1) The proposed approach (Condition (2) has exhibited superior outcomes in comparison to the other three conditions, thereby confirming the rationale behind the approach to distinctively identify and manage counterintuitive data.

(2) Condition (3), which identifies abnormal data, has produced inferior results compared to Condition (2) (identifying counterintuitive data) but is similar to Condition (1) (no identification of data type). This suggests that identifying abnormal data is ineffective in improving the accuracy of the safety assessment model. Even the smallest mean absolute error (MAE) in Condition (3), which identifies 5 sets of abnormal data, is over three times higher than that of the proposed approach.

(3) There is a steady but slow increasing trend in the MAEs in Condition (3). This is attributed to the overfitting of the safety assessment model, as more data with higher BTRs are identified as abnormal and assigned a weight of 0.1. Consequently, data with smaller weights are undervalued in training the BPNN-based safety assessment model. As

Table 7
MAEs of four conditions for validation in the practical case.

Conditions		MAEs			Runs
Condition (1)	directly using all training data	4.0727E-03			1
Condition (2)	15 sets of counterintuitive data	1.0452E-03			1
Condition (3) (Hypothesis 1)	5 sets of abnormal data	3.8202E-03			1
	10 sets of abnormal data	3.8343E-03			
	15 sets of abnormal data	3.8310E-03			
	20 sets of abnormal data	3.9673E-03			
	50 sets of abnormal data	4.0213E-03			
Statistics of Condition (4)		min	avg.	var.	30
Condition (4) (Hypothesis 2)	5 sets of random data	3.9886E-03	4.0930E-03	1.3974E-09	
	10 sets of random data	3.9199E-03	4.0139E-03	2.6397E-09	
	15 sets of random data	4.0925E-03	4.1249E-03	1.1489E-09	
	20 sets of random data	3.9995E-03	4.0045E-03	4.3151E-09	
	50 sets of random data	3.9296E-03	4.0146E-03	7.0754E-09	

the testing data conforms to the same distribution as the original training dataset, encompassing data with both higher and lower BTRs, the excessively fitted model encounters challenges when dealing with data possessing elevated BTRs. Hence, the prediction errors for BTRs on the testing datasets in Condition (3) increase.

(3) Condition (4), which randomly selects data, yields results similar to Condition (1), as expected. There are no discernible trends in the MAEs of Condition (4) when 5, 10, 15, 20, or 50 sets of data are randomly selected. Randomly selecting data does not significantly affect the data distribution.

Overall, the findings support the efficacy of the proposed approach in handling intuitive and counterintuitive data, while highlighting the limitations of identifying abnormal data or randomly selecting data for constructing a more accurate safety assessment model.

7. Conclusions and future works

Motivated to define, identify and handle counterintuitive data, a new approach is proposed to construct a more accurate safety assessment in metro tunnel construction. The theoretical contributions of this study are as follows. (1) **Definition of counterintuitive data**: Unlike traditional approaches that focus solely on the input or output part of data, counterintuitive data is defined as data that deviates from the causal relation between input and output. This definition provides a more comprehensive understanding of data characteristics. (2) **Identification of counterintuitive data** by multi-profile multi-model ensemble learning: The study introduces a new approach for identifying counterintuitive data. It utilizes multi-profile ensemble learning to create multiple profiles of the original training data, enabling cross-examination of each dataset. Furthermore, multi-model ensemble learning is employed to mitigate the influence of inferior machine learning approaches. (3) **Handling of counterintuitive data**: To handle counterintuitive data, researchers assign it a smaller weight in the construction of the safety assessment model. This approach acknowledges the distinctive nature of counterintuitive data and contributes to the development of a more accurate model.

The proposed approach is validated through a practical case study involving the prediction of BTR in metro tunnel construction in Wuhan City. The results of the case study demonstrate the efficacy of the proposed approach. (1) Accurate Identification of Counterintuitive Data: The proposed approach successfully identifies counterintuitive data, showcasing its ability to capture and differentiate data that deviates from the expected causal relation. (2) Superior Performance: The proposed approach demonstrates better performance compared to the method of directly using the complete training dataset without identifying counterintuitive data. This highlights the effectiveness of the approach in improving prediction accuracy. (3) Suitability of BPNN: The study reveals that BPNN is more suitable as the baseline model in comparison to GPR, SVM, and ANFIS. This finding emphasizes the importance of selecting an appropriate machine learning approach for specific problem domains. (4) Validation of Motive: The superior performance of the proposed approach, when compared to the results obtained when abnormal data and random data are treated as counterintuitive data, validates the motivation behind the study.

For future studies, more investigations into the parameter settings should be studied to produce superior performances, including how intuitive/counterintuitive data should be divided, and how the parameters of the machine learning approaches should be determined. More practical cases should be tested as well to further validate the efficiency of the proposed approach.

CRediT authorship contribution statement

Leilei Chang: Methodology, Software, Writing – original draft, Writing – review & editing. **Chenhao Yu:** Methodology, Writing – original draft, Validation. **Limao Zhang:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Xiaobin Xu:** Validation. **Schahram Dustdar:** Supervision, 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.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Key R&D project 2022YFE0210700, National Natural Science Foundation of China (72271101), Fundamental Research Funds for the Provincial Universities of Zhejiang (GK239909299001-010), Zhejiang Province Public Welfare Technology Application Research Project (LTGG23F030003), and the Zhejiang Province Outstanding Youth Fund (R21F030001).

References

Arican, E., & Aydin, T. (2022). An RGB-D descriptor for object classification. *Romanian Journal of Information Science and Technology (ROMJIST), 25*, 338–349.
Alghushairy, O., Alsini, R., Soule, T., & Ma, X. G. (2020). A review of local outlier factor algorithms for outlier detection in big data streams. *Big Data and Cognitive Computing, 5*(1), 1.
Bai, X. D., Cheng, W. C., & Li, G. (2021). A comparative study of different machine learning algorithms in predicting EPB shield behaviour: A case study at the Xi'an metro, China. *Acta Geotechnia, 16*(12), 4061–4080.
Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of grandient boosting algorithms. *Artificial Intelligence Review, 54*, 1937–1967.
Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys (CSUR), 54*(3), 1–33.

Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review, 54*(8), 6391–6438.
Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science, 363*(6433), aau0323.
Borlea, I., Precup, R., & Borla, A. (2022). Improvement of K-means cluster quality by post processing resulted clusters. *Procedia Computer Science, 199*, 63–70.
Bamaqa, A., Sedky, M., Bosakowski, T., Bastaki, B. B., & Alshammari, N. O. (2022). SIMCD: SIMulated crowd data for anomaly detection and prediction. *Expert Systems with Application, 203*, Article 117475.
Chiang, H., Shih, D., Lin, B., & Shih, M. (2014). An APN model for Arrhythmic beat classification. *Bioinformatics, 30*, 1739–1746.
Chang, L. L., Song, X. T., & Zhang, L. M. (2022). Uncertainty-oriented reliability and risk-based output control for complex systems with compatibility considerations. *Information Sciences, 606*, 512–530.
Conforti, A., Trabucchi, I., Tiberti, G., Plizzari, G. A., Caratelli, A., & Meda, A. (2019). Precast tunnel segments for metro tunnel lining: A hybrid reinforcement solution using macro-synthetic fibers. *Engineering Structure, 199*, Article 109628.
Chang, L. L., Zhang, L. M., & Xu, X. B. (2022). Randomness-oriented multi-dimensional cloud-based belief rule base approach for complex system modeling. *Expert Systems With Applications, 203*, Article 117283.
Chang, L. L., Zhang, L. M., & Xu, X. B. (2023). Causality-based multi-model ensemble learning for safety assessment in metro tunnel construction. *Reliability Engineering & Systems Safety, 234*, Article 109168.
Chang, L. L., Zhang, L. M., Fu, C., & Chen, Y. W. (2022). Transparent digital twin for output control using belief rule base. *IEEE Transactions on Cybernetics, 52*, 10364–10378.
Darroch, N., Beecroft, M., & Nelson, J. D. (2021). A qualitative analysis of the interfaces between urban underground metro infrastructure and its environment in London. *Tunnelling and Underground Space Technology, 114*, Article 103930.
Ding, D., Wang, Z., Han, Q. L., & Wei, G. L. (2018). Neural-network-based output-feedback control under round-robin scheduling protocols. *IEEE Transactions on Cybernetics, 49*(6), 2372–2384.
Forsat, M., Taghipoor, M., & Palassi, M. (2021). 3D FEM model on the parameters' influence of EPB-TBM on settlements of single and twin metro tunnels during construction. *International Journal of Pavement Research and Technology, 2*, 1–14.
Feng, L., & Zhang, L. M. (2021). Reliability-based multi-objective optimization in tunneling alignment under uncertainty. *Structural and Multidisciplinary Optimization, 63*(6), 3007–3025.
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260.
Kang, C. D. (2019). Spatial access to metro transit villages and housing prices in Seoul, Korea. *Journal of Urban Planning and Development, 145*(3), 05019010.
Karaboga, D., & Kaya, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey. *Artificial Intelligence Review, 52*(4), 2263–2293.
Kato, N., Mao, B., Tang, F., Kawamoto, Y., & Liu, J. (2020). Ten callenges in advancing maching learning technologies toward 6G. *IEEE Wireless Communications, 27*(3), 96–103.
Liu, R., Liu, W., Li, H., Wang, H., Geng, Q., & Dai, Y. (2021). Metro anomaly detection based on light strip inductive key frame extraction and MAGAN Network. *IEEE Transactions on Instrument Measurement, 71*, Article 500021.
Lin, S. S., Shen, S. L., Zhou, A., et al. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. *Automation in Construction, 122*, Article 103490.
Okudan, O., Budayan, C., & Dikmen, I. (2021). A knowledge-based risk management tool for construction projects using case-based reasoning. *Expert Systems with Application, 173*, Article 114776.
Ouchen, S., Steinhart, H., Benbouzid, M., & Blaaberg, F. (2020). Robust DPC-SVM control strategy for shunt active power filter based on H∞ regulators. *International Journal of Electrical Power & Energy Systems, 117*, Article 105699.
Pan, Y., & Zhang, L. (2020). Multi-classifier information fusion in risk analysis. *Information Fusion, 60*, 121–136.
Pan, Y., & Zhang, L. (2023). Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions. *Archives of Computational Methods in Engineering, 30*, 1081–1110.
Park, J., Lechevalier, D., Ak, R., Ferguson, M., Law, K. H., Lee, Y. T., et al. (2017). Gaussian process regression (GPR) representation in predictive model markup language (PMML). *Smart Structures and Systems, 1*(1), 121–141.
Patrucco, M., Pira, E., Pentimalli, S., Nebbia, R., & Sorlini, A. (2021). Anti-collision systems in tunneling to improve effectiveness and safety in a system-quality approach: A review of the state of the art. *Infrastructures, 6*(3), 42.
Quatrini, E., Costantino, F., Gravio, G. D., & Patriarca, R. (2020). Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. *Journal of Manufacture System, 56*, 117–132.
Ribeiro, M. H. D. M., & Coelho, L. S. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing, 86*, Article 105837.
Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., et al., (2021). A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE, 109*(5), *756–796*.
Rong, X., Lu, H., Wang, M., Wen, Z., & Rong, X. (2019). Cutter wear evaluation from operational parameters in EPB tunneling of Chengdu Metro. *Tunnelling and Underground Space Technology, 93*, Article 103043.
Singh, S. K., Das, A. K., Singh, S. R., & Racherla, V. (2023). Prediction of rail-wheel contact parameters for a metro coach using machine learning. *Expert Systems with Application, 215*, Article 119343.
Ye, X. W., Jin, T., & Chen, Y. W. (2022). Machine learning-based forecasting of soil settlement induced by shield tunneling construction. *Tunnelling and Underground Space Technology, 124*, Article 104452.
Zhou, Z. (2021). Ensemble learning. *Machine Learning*, 181–210.
Zou, C., Moore, J. A., Sanayei, M., Wang, Y., & Tao, Z. (2021). Efficient impedance model for the estimation of train-induced vibrations in over-track buildings. *Journal of Vibration & Control, 27*(7–8), 924–942.
Zhang, L. M., Wu, X. G., Ji, W., & AbouRizk, S. M. (2017). Intelligent approach to estimation of tunnel-induced ground settlement using wavelet packet and support vector machines. *Journal of Computing Civil Engineering, 31*(2), 04016053.
Zhang, L. M., Wu, X. G., Zhu, H., & AbouRizk, S. M. (2017a). Performing global uncertainty and sensitivity analysis from given data in tunnel construction. *Journal of Computing Civil Engineering, 31*(6), 04017065.
Zhang, L. M., Wu, X. G., Zhu, H., & AbouRizk, S. M. (2017b). Perceiving safety risk of buildings adjacent to tunneling excavation: An information fusion approach. *Automation in Construction, 73*, 88–101.