

# Multi-Module Decision Fusion in Operational Status Monitoring

Shuai Tan<sup>1</sup>, Fulin Gao<sup>1</sup>, Hongbo Shi<sup>1</sup>, Huaicheng Yan<sup>1</sup>, *Member, IEEE*, and Zheng Mu

**Abstract**—Multimodule and multisensor data with different characteristics can reflect the overall operating status of the equipment from different angles. It is far from enough to monitor the operating status of equipment from a single perspective for this fails to take all valid information into account. However, data integration may face problems such as the curse of dimensionality and scale mismatches. Therefore, decision fusion, which needs to measure and manage evidence conflicts, has attracted extensive attention from scholars. However, most of the state-of-art methods focus on conflict management based on evidence itself and ignore the irrationality of conflict factor  $K$  in measuring conflicts and the reliability of evidence sources in conflict management. In order to solve the above problems, a novel hybrid decision fusion approach is proposed in this article. First, divide data into modules and use the models in the model library to conduct cross-validation, thus obtaining the performance ranking. Then, select the optimal classifier of each module to obtain the evidence for decision fusion. Given the conflict of evidences, the Jensen–Shannon (JS) divergence is used to measure the conflict, and those high-conflict evidences will be revised through sensitivity and support analyses. Finally, the Dempster–Shafer (DS) evidence theory is used to integrate multimodule evidences to assess the status. To prove the feasibility and effectiveness of this approach, a realistic operational shield case in China is used.

**Index Terms**—Conflict management, conflict measurement, decision fusion, Dempster–Shafer (DS) evidence theory, multimodule, operational status monitoring, sensitivity, support.

## I. INTRODUCTION

FROM the perspective of safety, schedule, and economy, it is critical to monitor the operating status of equipment during construction or production [1], [2]. In recent years, developments in sensor and data analysis technology have made operational status monitoring possible, which has attracted widespread attention from scholars [3], [4]. The articles dedicated to review and analyze the state-of-the-art

Manuscript received June 11, 2021; revised November 23, 2021; accepted January 16, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC1522502 and Grant 2020YFC1522505 and in part by the National Natural Science Foundation of China under Grant 62073141. Recommended by Associate Editor M. Abbaszadeh. (*Corresponding author: Shuai Tan.*)

Shuai Tan, Fulin Gao, Hongbo Shi, and Huaicheng Yan are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: tanshuai@ecust.edu.cn; y30190719@mail.ecust.edu.cn; hbshi@ecust.edu.cn; hcyan@ecust.edu.cn).

Zheng Mu is with China Railway 14th Bureau Group Large Shield Engineering Company Ltd., Nanjing, Jiangsu 211800, China (e-mail: muzheng2003@163.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCST.2022.3145648>.

Digital Object Identifier 10.1109/TCST.2022.3145648

data-based status monitoring techniques can be found in [5] and [6]. However, due to the proliferation of sensor types and data volumes, it is still extremely necessary and challenging to conduct timely and accurate operational status monitoring with data-driven methods [7], [8].

There are numerous modules inside large-scale equipment, plant-wide processes, and so on, while the distribution and characteristics of sensors in each module are different. Moreover, some heterogeneous data from different sources are presented in multimodality [9]–[11]. There are two common ways to deal with these issues: feature fusion and decision fusion [12]. Poria *et al.* [13] used feature- and decision-level fusions to conduct sentiment analysis on audio, video, and text. Their study proved that decision fusion is superior to feature fusion in dealing with this issue. Liu *et al.* [14] used the Bayesian fusion and the Dempster–Shafer (DS) evidence theory to make local and global decisions, respectively. In an attempt to achieve a more accurate spectrum sensing, they integrated the information of energy, power spectrum, and signal wave. Zhong *et al.* [15] used minimum redundancy and maximum relevance (mRMR) to establish variable subblocks and then fused distributed monitoring results by the Bayesian inference. As the volume of operational data collected is huge, dimensionality reduction or feature selection is often required when using a single model for status monitoring [16], [17]. During this process, some details will inevitably be ignored. Therefore, it is necessary to prevent the omission of local information through module partitioning and decision-level fusion, thus improving computational efficiency and circumventing laborious feature fusion.

For multimodule, the best classifiers that can capture the module information may be inconsistent due to different characteristics [18]–[20]. Calikus *et al.* [21] point out that there is no single superior model which works perfectly for every case. Yu and Zhao [22] proposed a multimodel exponential discriminant analysis (MEDA) algorithm to deal with the problem that a single model cannot accurately describe the fault information. He *et al.* [23] proposed a fusion strategy based on multicriteria decision-making (MCDM) to integrate different classifiers. Zhou *et al.* [24] compared and verified different classification methods and, finally, found the optimal classifier for rockburst prediction. Therefore, it is necessary to select the optimal classifier for each module to improve the credibility of the evidence used for subsequent decision fusion.

For decision fusion, the DS evidence theory has been widely applied in this field, in which prior knowledge is not strictly

required [25], [26]. However, Wu and Jahanshahi [27] and Liu [28] pointed out two of the outstanding issues: 1) conflict factor  $K$  is not competent to measure conflict and 2) it is prone to counterintuitive results when dealing with conflicting evidence. Many scholars have conducted research on the above issues [7], [29]. For the former, Deng and Wang [30] designed a Tanimoto-based evidence consistency description method. Xiao *et al.* [26] measured conflict by evidence correlation coefficient (ECC). For the latter, Li *et al.* [31] took the consistency of evidence as the weighting index and modified the rules of fusion. Both Xiao [25] and Song and Deng [32] considered the divergence between evidences and information volume to weigh the evidence. Khan and Anwar [33] used entropy-based reward and penalty factors to revise the evidence. Ma and An [34] used fuzzy proximity and correlation coefficient to discount the evidence and discussed the irrationality of the conflict factor  $K$  to measure conflict. However, most state-of-the-art methods consider only part of the above issues, in which case no amount of revision is worthwhile if the evidence is not contradictory or its source is unreliable [35]. Therefore, a complete conflict measurement and management scheme should be developed. Based on achieving accurate identification of conflicting evidence, the reliability of the evidence producer and the relevance between the evidence should be comprehensively considered to manage the high-conflict evidence, thus eliminating the concerns of DS evidence theory.

Overall, the main purpose of this study is to develop a generic multimodule decision fusion framework for solving the problem of operational status monitoring of large-scale equipment, plant-wide processes, and so on. The main contributions of the research are: 1) a module division and model selection strategy is proposed to reduce the complexity of the data during processing, thus ensuring the accuracy and reliability of the primary conclusions from each module and 2) an evidence conflict measurement and management strategy is proposed, which takes the Jensen–Shannon (JS) divergence as the conflict metric and takes evidence support and model sensitivity into consideration to eliminate the inevitable conflict between evidences.

The remainder of this article is structured as follows. Section II provides an introduction to the issues and novel solutions in the acquisition and utilization of evidence. Section III gives a detailed description of the overall framework of the solution. Section IV applies the proposed method to a real shield case in China and produces the experimental results. Section V is the conclusion and research prospect.

## II. EVIDENCE ACQUISITION AND UTILIZATION

### A. Obtaining Evidence

In decision fusion, the decision can be taken as the ultimate decision or the evidence of the next level decision [36]. The acquisition of primary decisions, e.g., evidence, is always the basis of fusion. The evidence in this article exists in the form of the probability distribution, and the classification decision is made according to the maximum probability. The diversified data collected from different modules and perspectives make

the evidence complementary to each other. The integration of diversified evidence through decision fusion helps to accurately access the overall operating state of the equipment. Due to the complexity and diversity of the underground environment, as well as the wear and tear of equipment, the operating state of equipment changes dynamically. Based on expert knowledge, the operating status is divided into normal, medium, and poor (I, II, and III), and thus, the evaluation of the equipment operating status can be regarded as a multiclassification task.

Numerous classification algorithms have their own characteristics, and algorithms that can obtain classification results in the form of probability distributions are also emerging [37]. The common ways to obtain probability distribution are as follows: 1) establish a model for each category and then use the matching degree between the test sample and the model as the classification probability, such as the hidden Markov model (HMM) [38]; 2) use the obtained posterior probabilities as category distributions, such as the quadratic discriminant analysis (QDA) and the probability support vector machine (PSVM) [39]; 3) map categories into one-hot codes and take them as classification probabilities, such as XGBoost and multilayer perceptrons (MLPs) [40], [41]; and 4) obtain probability distribution through soft voting, such as decision tree (DT) and random forest (RF) [42].

The conclusions of these probability distributions can be used as evidence for decision fusion. However, there are multiple segments or modules in complex process or large-scale equipment, which together affects the operating status of the entire system. Due to the proliferation of sensors and differences in location or function, the collected data are often characterized by large volumes and diverse features. In response to the above issues, a module division and model selection strategy is proposed, which partitions the data by mechanism and automatically adapts the optimal classifiers for each module through cross-validation. Then, each module can capture its own unique information through its own optimal classifier and obtain the primary conclusion, i.e., evidence.

Given the input data normalized by Z-score, it is tabulated to  $n$  samples and  $m$  variables  $\mathbf{X} = \{\mathbf{x}_j, j = 1, 2, \dots, m\}$  and output data  $\mathbf{y} = \{y_i, i = 1, 2, \dots, n\}$ . Assume that the multisensor input data are divided into  $N$  modules from variable dimensions according to the type, function, and distribution of sensors: Module<sub>1</sub>, Module<sub>2</sub>, ..., Module <sub>$N$</sub> , and then, the data of each module are split into training set and test set by 8:2.

By cross-validating the models in the model library (models included in the Library: HMM, QDA, PSVM, XGBoost, MLP, DT, and RF) on the training set of each module, the performance ranking matrices can be obtained

$$\text{Rank}^i = \begin{bmatrix} R_{11}^i & \cdots & R_{1j}^i & \cdots & R_{1J}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{g1}^i & \cdots & R_{gj}^i & \cdots & R_{gJ}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{G1}^i & \cdots & R_{Gj}^i & \cdots & R_{GJ}^i \end{bmatrix} \quad (1)$$

where  $i = 1, 2, \dots, N$ ,  $R_{gj}^i$  represents the performance ranking of the  $j$ th algorithm in the  $i$ th module during the  $g$ th cross-validation,  $G$  represents the number of cross-validation, i.e., the number of groups into which the training data is divided, and  $J$  represents the number of models in the model library.

The ranking is based on the average of the  $G$ -fold cross-validated performance metrics for each model. The higher the average score, the higher the ranking. The multicategory metrics used in this study include accuracy, Jaccard similarity coefficient, and area under the curve (AUC), which are common. The accuracy is the percentage of correct predictions in total predictions. The AUC is often used to evaluate the generalization ability of a model. It indicates the probability that the positive case score is higher than the negative case score for two randomly selected heterogeneous samples in the current model. The Jaccard correlation coefficient performs well as a weighted measure of model performance under the class-imbalance problem. The range of AUC, accuracy, and Jaccard correlation coefficient is  $[0, 1]$ . The closer to 1, the better the model performance. For details, please refer to [43] and [44].

According to the performance ranking, the optimal model of each module can be selected

$$\text{Classifier}^i = \arg \min_j \left( \frac{1}{J} \sum_{g=1}^G R_{gj}^i \right), \quad i = 1, 2, \dots, N \quad (2)$$

where  $\text{Classifier}^i$  represents the classifier with the highest performance ranking in the  $i$ th module. When a new test sample arrives, the best classifier selected by each module can obtain its own evidence:  $M_1, M_2, \dots, M_N$ .

### B. DS Evidence Theory

The application of the DS evidence theory in decision fusion has become more widespread. It combines multiple specific evidences to derive abstract fusion conclusions within the framework of discernment [34]. Let  $D$  be the sample space, also known as the recognition frame. There are  $m$  elements  $E$  in  $D$ , and these elements are mutually exclusive, expressed as follows:

$$D = \{E_1, E_2, \dots, E_m\}. \quad (3)$$

The number of subsets of  $D$  is  $2^m$ ,  $\Phi$  is the empty set, and  $2^D$  represents the set of all subsets, as follows:

$$2^D = \{\Phi, \{E_1\}, \{E_2\}, \dots, \{E_m\}, \{E_1, E_2\}, \dots, \{E_1, E_m\}, \dots, \{D\}\}. \quad (4)$$

A basic belief assignment (BBA) is a mapping  $M: 2^D \rightarrow [0, 1]$ , which satisfies

$$\begin{cases} M(\Phi) = 0 \\ \sum_{A_i \in 2^D} M(A_i) = 1 \end{cases} \quad (5)$$

where  $M(A_i)$  represents the degree of trust in proposition  $A_i$ .  $M$  is also called as the mass function.

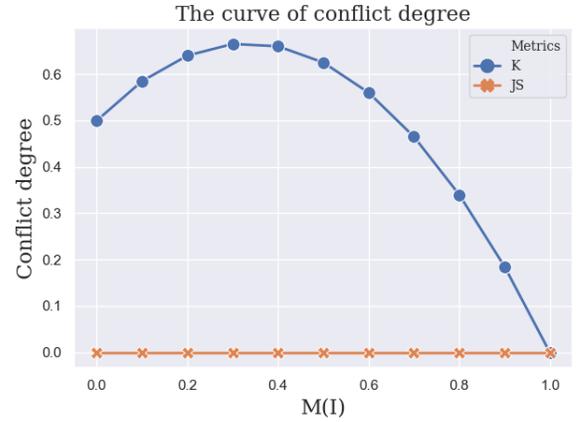


Fig. 1. Conflict measurement curve for BBAs  $M_1$  and  $M_2$ .

Dempster's combination rule is defined as follows:

$$M(A) = M_1 \oplus M_2 = \begin{cases} \frac{1}{1-K} \sum_{B \cap C = A} M_1(B)M_2(C), & A \neq \Phi \\ 0, & A = \Phi \end{cases} \quad (6)$$

where  $\oplus$  represents the operator of combination,  $B$  and  $C$  are subsets of  $2^D$ , and  $K$  is conflict factor in the range of  $[0, 1]$ , denoted as follows:

$$K = \sum_{B \cap C = \Phi} M_1(B)M_2(C). \quad (7)$$

### C. Issue Analysis and Novel Solution of Conflict Metric

The conflict factor  $K$  in (7) defined by DS evidence theory cannot well describe the degree of conflict between evidences [45]. Assume that the following two status assessment evidences have been obtained:

$$\begin{aligned} \{M_1(\text{I}), M_1(\text{II}), M_1(\text{III})\} &= \{0.97, 0.01, 0.02\} \\ \{M_2(\text{I}), M_2(\text{II}), M_2(\text{III})\} &= \{0.01, 0.01, 0.98\} \end{aligned}$$

where I, II, and III represent three different states. The conflict factor  $K$  of the above evidences is 0.9706, indicating high conflict. Also, this result is consistent with the facts. However, when the conflict of evidence is not so significant, the conflict factor  $K$  may fail to measure the conflict. For simplicity, take two identical evidences, i.e.,  $M_1 = M_2$ , and let their probability in state I increase from 0 to 1 at an interval of 0.1. At the same time, take half of the remaining probability in the other two states (the sum of probabilities is 1). The blue markers in Fig. 1 show the degree of conflict measured by  $K$ . It can be seen that, when  $K$  is used to measure conflict, the conflict measurement is related not only to similarity but also to the trust degree of the proposition for two identical evidences. These inconclusive results are obviously unsatisfactory.

JS divergence is widely used to measure the similarity between two probability distributions due to its excellent properties, such as symmetry, boundedness, and satisfying triangular inequalities [25]. The larger the divergence, the

lower the similarity. In view of the above issues, this article takes JS divergence as a new conflict metric. The JS divergence between  $M_1$  and  $M_2$  is formulated as follows:

$$JS(M_1, M_2) = \frac{1}{2} \left[ S \left( M_1, \frac{M_1 + M_2}{2} \right) + S \left( M_2, \frac{M_1 + M_2}{2} \right) \right] \quad (8)$$

$$S(M_1, M_2) = \sum_{l=1}^L M_{1l} \log_2(M_{1l}/M_{2l}) \quad (9)$$

$$\sum_{l=1}^L M_{il} = 1, \quad i = 1, 2 \quad (10)$$

where  $L$  is the length of evidence, which is the total number of categories, replacing with  $10^{-12}$  when  $*$  is 0 in  $\log_2(*)$ , and  $S(M_1, M_2)$  is the Kullback–Leibler (KL) divergence. Since KL divergence has no symmetry, JS divergence is selected as the metric. The orange markers in Fig. 1 show the results of conflict measurement using JS divergence, which are more reasonable than those measured by the conflict factor  $K$ .

#### D. Issue Analysis and Novel Solution of Conflict Management

The DS evidence theory will produce counterintuitive results when dealing with high-conflict evidence [7], [29]. There are two views on the above issue. One is that the DS fusion rules should be modified because the discarding method used in handling conflicts is the culprit for unreasonable results. The other is that evidence conflict should be eliminated before fusion, i.e., conflict management. It is obvious that the latter is more reasonable because the modification of rules usually breaks both commutative and associative laws in fusion rules. In fact, fusion rules are not to blame for counterintuitive results if the evidence conflict is due to sensor failure or inaccurate reports [46].

While most research considers correlations between evidence, the credibility of the sources of evidence is ignored [35]. In this case, no amount of revision is worthwhile if the sources of evidence are unreliable. Therefore, this article proposes a novel revision strategy where both the reliability of evidences and the similarity between evidences are included. Regarding the reliability of the evidences, i.e., the accuracy of the sources, the model sensitivity is used to modify the evidence, as detailed in Section II-D1. As for the similarity between the evidences, i.e., the supporting relationship, the supporting degree is used to modify the evidence, as detailed in Section II-D2.

1) *Sensitivity Analysis*: This article proposes a method of evidence revision based on sensitivity. In binary classification, sensitivity refers to the ratio of correctly identified positive samples to all true positive samples. In this article, the predictive sensitivity in multiclassification is defined as the ratio of correctly identified samples in each prediction category to the total samples predicted to be in this category. The metric can be used to describe the recognition ability of the classifier in each prediction category. When the model obtains a piece of evidence, the credibility of the evidence can be reflected from the sensitivity of this model. The higher the corresponding

sensitivity, the higher the credibility of the evidence. The proposed method modifies the individual evidence according to the sensitivity corresponding to the predicted state of each module. Note that each probability value of the same piece of evidence is multiplied by the same sensitivity.

For evidences,  $M_1, M_2, \dots, M_N$ , the purpose of sensitivity analysis is to obtain the correction weight:  $\hat{w}^0 = [\hat{w}_1^0, \hat{w}_2^0, \dots, \hat{w}_N^0]$ , where  $\hat{w}_i^0$  represents the correction weight of the  $i$ th evidence, the superscript 0 indicates that the weight is obtained by sensitivity analysis, and  $N$  represents the number of evidence to be fused. The sensitivity of each module is obtained by the following average confusion matrix of the  $G$ -fold cross-validation:

$$\text{Con}^i = \begin{bmatrix} C_{11}^i & \cdots & C_{1l}^i & \cdots & C_{1L}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{h1}^i & \cdots & C_{hl}^i & \cdots & C_{hL}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{L1}^i & \cdots & C_{Ll}^i & \cdots & C_{LL}^i \end{bmatrix} \quad (11)$$

where  $i = 1, 2, \dots, N$ , and  $\text{Con}^i$  represents the average confusion matrix obtained by the classifier selected by the  $i$ th module in the cross-validation phase.  $C_{hl}^i$  represents the number of samples with the true sample label of  $h$  and the predicted label of  $l$  in the average confusion matrix of the  $i$ th module, and  $L$  represents the number of categories of operation status.

If the true historical operating status is available before the next monitoring, the above confusion matrices can be updated online. Considering the time cost, the online version no longer performs cross-validation but updates the confusion matrix based on the true and predicted status of the historical data for each module. Taking the  $i$ th module as an example, if the status of the device at moment  $t$  is to be evaluated, and assuming that the true label of the sample at moment  $t - 1$  is  $h$  and the prediction obtained by Classifier <sup>$i$</sup>  is  $l$ , then update  $C_{hl}^i$  to  $C_{hl}^i + 1$ .

Next, the sensitivity matrix can be obtained according to the average confusion matrix of each module

$$\text{Sen} = \begin{bmatrix} S_1^1 & \cdots & S_l^1 & \cdots & S_L^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_1^i & \cdots & S_l^i & \cdots & S_L^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_1^N & \cdots & S_l^N & \cdots & S_L^N \end{bmatrix} \quad (12)$$

where  $S_l^i$  represents the sensitivity of the classifier selected by the  $i$ th module to the  $l$ th state, which can be calculated according to the following equation:

$$S_l^i = e^{\frac{c_{ll}^i}{\sum_{s=1}^L c_{sl}^i}} \quad (13)$$

Finally, the correction weight of the  $i$ th evidence obtained through sensitivity analysis is given as follows:

$$\hat{w}_i^0 = S_l^i, \quad l = \text{State}_i \quad (14)$$

where  $i = 1, 2, \dots, N$ , and  $\text{State}_i$  represents the label corresponding to the maximum predicted probability of evidence obtained in the  $i$ th module.

2) *Support Analysis*: The multiple evidences obtained from different modules can support each other. For a piece of evidence, the greater the number of similar evidences, the higher the degree of similarity. Also, the more support it receives from other evidences, the higher its credibility.

For evidences,  $M'_1, M'_2, \dots, M'_N$ , the purpose of support analysis is to obtain the correction weight:  $\hat{w}^1 = [\hat{w}_1^1, \hat{w}_2^1, \dots, \hat{w}_N^1]$ , where  $\hat{w}_i^1$  represents the correction weight of the  $i$ th evidence, and the superscript 1 indicates that the weight is obtained by support analysis.

The similarity matrix between evidence pairs can be obtained through (8)

$$\text{DMM} = \begin{bmatrix} 0 & \cdots & \text{JS}_{1i} & \cdots & \text{JS}_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{JS}_{i1} & \cdots & 0 & \cdots & \text{JS}_{iN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{JS}_{N1} & \cdots & \text{JS}_{Ni} & \cdots & 0 \end{bmatrix} \quad (15)$$

where  $\text{JS}_{ij}$  represents the JS divergence between the  $i$ th evidence and the  $j$ th evidence. Then, the average difference between the  $i$ th evidence and other evidence bodies is calculated as follows:

$$\tilde{\text{JS}}_i = \frac{\sum_{j=1, j \neq i}^N \text{JS}_{ij}}{N-1}, \quad i = 1, 2, \dots, N. \quad (16)$$

Finally, the correction weight of the  $i$ th evidence can be calculated by normalizing the support degree as follows:

$$\text{Sup}_i = \frac{1}{\tilde{\text{JS}}_i}, \quad i = 1, 2, \dots, N \quad (17)$$

$$\hat{w}_i^1 = \frac{\text{Sup}_i}{\sum_{i=1}^N \text{Sup}_i}, \quad i = 1, 2, \dots, N. \quad (18)$$

### III. SOLUTION APPROACH

To address the problems of the high complexity of diverse data, the irrationality of conflict factor  $K$  in measuring conflict, and counterintuitiveness of the DS evidence theory in dealing with high-conflict evidence, this article proposes a novel hybrid decision fusion method. The proposed method adopts the following hierarchical stepwise strategy.

At Level 1, data are partitioned into modules through mechanism analysis and expert knowledge, such as the function and distribution of sensors. In different modules, HMM, QDA, SVM, XGBoost, MLP, RF, and DT are used for cross-validation on the training set. In this way, the model performance and average confusion matrix can be obtained. Then, the performance of the models in each module is ranked to select the best classifier, and the sensitivity is obtained by the average confusion matrix. After the arrival of the samples, a preliminary status assessment is performed on each module. The results of the assessment, i.e., the evidences for decision fusion, are fed into Level 2 along with sensitivity, i.e., the correction weights.

At Level 2, the degree of conflict between the evidences is measured by JS divergence, and the thresholds are determined by grid search. For those evidences with a high-conflict degree,

the sensitivity analysis and the support analysis are used to modify the evidences. The original conflicting evidences will be replaced by the weighted average evidences. Then, Dempster's combination rule is used to fuse them  $N-1$  times. For those evidences with a low-conflict degree, Dempster's combination rule is directly used for fusion. The overall framework with  $N=3$  is shown in Fig. 2. The processing procedures are given as follows.

Step 1: According to Section II-A, divide the data into modules, and select the optimal classifier for each module to obtain evidence:  $M_1, M_2, \dots, M_N$ .

Step 2: For evidences,  $M_1, M_2, \dots, M_N$ , calculate the JS divergence between the evidence pairs by (8) as the conflict measure. If the JS divergence between each evidence pair does not exceed threshold, the conflict is low, and Dempster's combination rule can be used directly for fusion. Otherwise, go to Step 4.

Step 3: Perform  $N-1$  fusions by (6), the first two pieces of evidence are fused for the first time, and the last fusion result is fused with the next piece of evidence each time. The final fusion result is given as follows:

$$F(M) = M_1 \oplus M_2 \oplus \cdots \oplus M_N. \quad (19)$$

Then, go to Step 7.

Step 4: The sensitivity correction weights can be obtained through Section II-D1, and the correction operations are given as follows:

$$M'_i = \hat{w}_i^0 \times M_i. \quad (20)$$

Recalculate JS divergence between  $M'_i$ . If the JS divergence between each evidence pair does not exceed threshold', go to Step 6. threshold' > threshold is mainly caused by  $\hat{w}_i^0$  greater than 1, which increases the corrected evidences.

Step 5: The support correction weights can be obtained through Section II-D2, and the correction operations are given as follows:

$$M''_i = \hat{w}_i^1 \times M'_i. \quad (21)$$

Step 6: Sum and normalize the revised evidences

$$\tilde{M} = \text{norm} \left( \sum_{i=1}^N M'_i \text{ or } \sum_{i=1}^N M''_i \right). \quad (22)$$

Fuse the weighted average results  $N-1$  times through Dempster's combination rule

$$F(\tilde{M}) = \underbrace{\tilde{M} \oplus \tilde{M} \oplus \cdots \oplus \tilde{M}}_{N-1}. \quad (23)$$

Step 7: State assessment

$$\text{State} = \arg \max \{F(M) \text{ or } F(\tilde{M})\}. \quad (24)$$

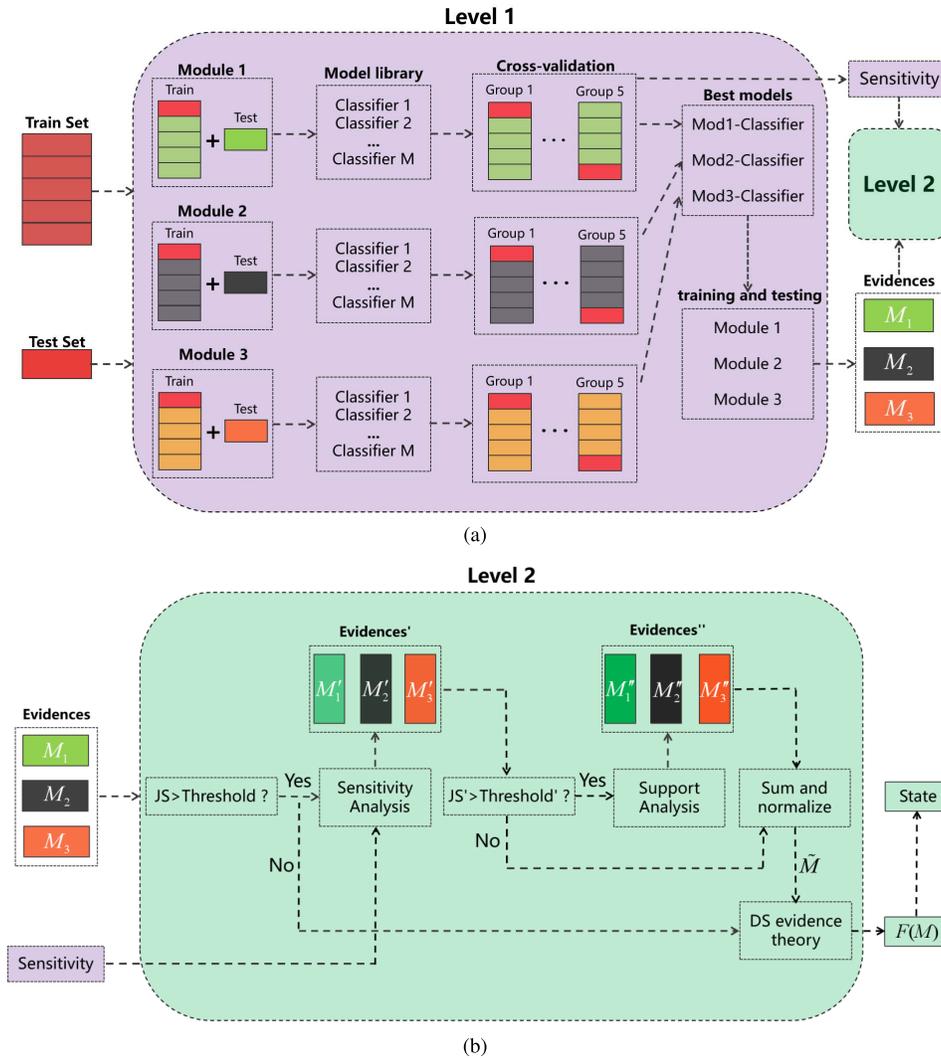


Fig. 2. Overall framework of the proposed method: (a) Level 1 and (b) Level 2.

#### IV. APPLICATION

##### A. Case Background

The application case in this study is the shield tunneling project of the Yellow River Jiluo Road Tunnel in Jinan, China. The tunnel is located on the central axis of the city, connecting two stations A and B from north to south. The total length of the tunnel is 4760 m, including a 3890-m tunnel crossing the Y River and 870-m connecting roads and related ancillary works. The length of the shield section is about 2519.2 m, and the diameter of the shield excavation is 15.76 m. The tunnel stratum is mainly clay and silty clay, so the excavation equipment is a mud-water tunnel boring machine (TBM). Since the operating status of the TBM directly affects the quality and speed of tunnel advancement, an accurate assessment of the current operating status is of great necessity. In this way, timely guidance can be provided to operators in operation adjustments and equipment maintenance.

Based on on-site experience and professional knowledge, the operating status of TBM can be divided into normal,

medium, and poor. TBM includes multiple modules, such as excavation, main drive, propulsion, grouting, hydraulic, and electrical. There are a total of 152 variables, all sampled at a frequency of 0.1 Hz. Mechanism analysis shows that these modules can be classified into three main categories. Module<sub>1</sub> is the main drive electrical module. It is at the heart of the power output of TBM, which is mainly for power conversion and output, and supports the cutter head of TBM for rotating rock breaking. Module<sub>2</sub> is the tunneling module. This module mainly controls the excavation progress of TBM. For example, when the driving direction deviates due to force imbalance, TBM can move forward following the prescribed route through coordination with the advance cylinder. Module<sub>3</sub> is the mud conveying module. It is used to transport the excavated muck to the ground treatment station. Meanwhile, this module can maintain the stability of the excavation surface pressure by adjusting the mud inflow and outflow.

The positions of each module in TBM are shown in Fig. 3, and the information of the sensors in each module is illustrated in Table I.

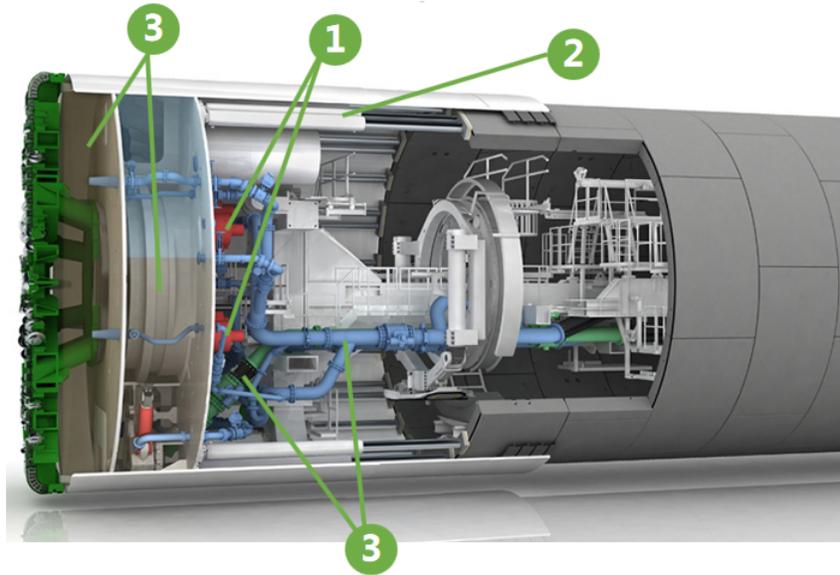


Fig. 3. Module division of the shield machine.

TABLE I  
INFORMATION OF SENSORS

<i>Module<sub>1</sub></i>			<i>Module<sub>2</sub></i>			<i>Module<sub>3</sub></i>		
Number	Types	Units	Number	Types	Units	Number	Types	Units
F1	Torque	MN-m	F8	Penetration	mm/rot	F15	Excavation chamber pressure	bar
F2	Main drive current	A	F9	Thrust	kN	F16	Working chamber pressure	bar
F3	Main drive speed	rpm	F10	Advance rate	mm/min	F17	In and out pulp flow	m <sup>3</sup> /h
F4	Motor torque	MN-m	F11	Propulsion pump pressure	bar	F18	In and out pulp density	t/m <sup>3</sup>
F5	Motor speed	rpm	F12	Propulsion pump swivel angle	deg	F19	Slurry inlet pressure	bar
F6	Motor current	A	F13	Propulsion cylinder pressure	bar	F20	Slurry pump current	A
F7	Motor bearing temperature	°C	F14	Propulsion cylinder thrust	kN	F21	Mud pump speed	rpm
						F22	Mud pump delivery pressure	bar
						F23	Mud pump suction pressure	bar
						F24	Mud pump motor temperature	°C

Note: One type of sensor in the table typically contains multiple sensing devices. For example, there are 12 sensor measuring points for motor speed sampling in *Module<sub>1</sub>*, which are located in different motors or different positions of the same motor.

### B. State Assessment Based on the Proposed Method

As in Section II-A, the models in the model library are first used to perform fivefold cross-validation on the training sets of the three modules to obtain the respective performance ranking matrices:  $\text{Rank}^i$ ,  $i = 1, 2, 3$ . According to (2), the optimal classifier of each module can be determined and used as the state assessment model of the corresponding module for evidence generation. Then, the conflict analysis can be conducted for the evidence obtained from each state assessment model according to Section III, and Dempster's combination rule is used to fuse either the evidence with low conflict without correction or the evidence with high conflict after correction. Ultimately, the evaluation of the operational status is achieved.

1) *Model Selection*: According to Section II-A, an optimal classifier for each module is selected. The weighted average of common multicategory metrics is used to evaluate the performance of the models, such as accuracy, Jaccard similarity coefficient, and AUC. These metrics adopt the concept

of macros in multicategory, i.e., calculating the unweighted average of the indicators for each label. When the three metrics are a weighted average, the appropriate weight can be assigned according to the importance of the indicators. If treated equally, 1/3 of the weight is assigned to each indicator as in this experiment.

First, the whole dataset is divided into three modules, and the data of each module are divided into the training set and test set by 8:2. According to the fivefold cross-validation, the training set is divided into five groups, and the ratio of training to testing in each group is also 8:2.

Then, in each module, the models in the model library are trained and tested on five groups of data. At the same time, the performance ranking and the sensitivity of the optimal classifier for subsequent evidence correction are obtained. The model library includes seven models: HMM, QDA, SVM, XGBoost, MLP, RF, and DT. That is to say, a module can obtain  $5 \times 3 \times 7$  performance indicators, which are respectively from five groups of data, three performance indicators, and seven models to be sorted. Next, the weighted average of the

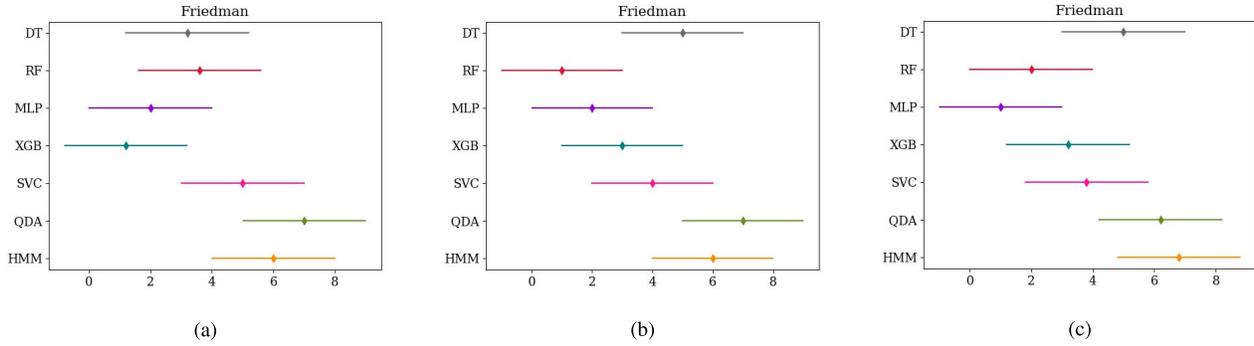


Fig. 4. Friedman test charts: (a) Module<sub>1</sub>, (b) Module<sub>2</sub>, and (c) Module<sub>3</sub>.

three metrics is calculated to get three  $5 \times 7$  performance matrices (here, three represents three modules). The closer the metric is to 1, the better the performance will be. The performance matrix can be converted into three  $5 \times 7$  performance ranking matrices.

Finally, select a state evaluation model for each module, i.e., the optimal model, according to the performance ranking average value, i.e., the average value of the column. The smaller the average order value, the higher the ranking and the better the performance.

Through the above procedure, XGBoost, RF, and MLP are, respectively, at the top of the performance rankings of Modules<sub>1-3</sub>, so they can be used as the state evaluation models of the corresponding modules. The Friedman test and the Nemenyi post-hoc test are also used to measure the performance of the models in the model library. Please refer to [47] for theories about the above test methods. The results of the Friedman test in the three modules are 96, infinity, and 171, respectively, and the critical value of the five datasets and seven algorithms at the 0.05 significance level is 2.508, which is much smaller than the above test results. Thus, the hypothesis that the algorithms are significantly different is accepted.

Nemenyi post-hoc test is used to further distinguish the differences of algorithms, and the Friedman test charts are shown in Fig. 4. The diamond point is the average ordinal value of the algorithm ranking, and the corresponding horizontal line is the critical range calculated by the Nemenyi post-hoc test. The more the overlapping area between the horizontal lines, the smaller the difference between the algorithms. It is obvious that XGBoost, RF, and MLP rank relatively high on Modules<sub>1-3</sub>, and they are significantly different from other algorithms.

It is worth noting that the Friedman test and the Nemenyi test are often used to compare the performance of algorithms on different datasets. Since our purpose is to select the optimal classifier for each module, five cross-validation sets are used to verify differences among models rather than different datasets.

2) *Evidence Revision and Fusion*: The whole dataset has been divided into three modules by the above procedure, and the state evaluation model used in each module has been determined. Thus, the preliminary state evaluation results of different modules can be obtained, i.e., evidences. These

evidences may be conflicting with each other. Since the conflict factor  $K$  cannot accurately measure the degree of conflict, a novel decision fusion method using JS divergence is proposed. It serves as a way to measure conflict and combine sensitivity and support for conflict management. See Section III for details of correction and fusion. The thresholds for judging whether the evidences are highly conflicted are set to 0.65 and 2.53 respectively through grid search. For the revision and fusion of evidence, two actual cases are given, which are, respectively, the 617th and 2467th samples of the test set.

*Case 1*: The conflicts are low, and Dempster's combination rule is directly used to fuse the evidences.

Step 1-1: Obtain the evidence of each module and calculate the JS divergence between evidence pairs by (8)

$$\begin{aligned} \left. \begin{aligned} M_1 &= [0.1422, 0.1885, 0.6694] \\ M_2 &= [0.3292, 0.3452, 0.3256] \\ M_3 &= [0.0184, 0.0000, 0.9816] \end{aligned} \right\} \rightarrow \begin{bmatrix} JS_{12} \\ JS_{13} \\ JS_{23} \end{bmatrix} \\ = \begin{bmatrix} 0.0881 \\ 0.1547 \\ 0.4188 \end{bmatrix} \end{aligned}$$

where  $JS_{12}$ ,  $JS_{13}$ , and  $JS_{23} < \text{threshold} = 0.65$ . Dempster's combination rule is used directly without modification.

Step 1-2: Obtain fusion results by (6)

$$F(M) = M_1 \oplus M_2 \oplus M_3 = [0.0040, 0.0000, 0.9960].$$

Step 1-3: State assessment

$$\text{State} = \text{III.}$$

*Case 2*: The conflicts are high, and the evidences need to be revised and then weighted and summed.

Step 2-1: Obtain the evidence of each module and calculate the JS divergence between evidence pairs by (8)

$$\begin{aligned} \left. \begin{aligned} M_1 &= [0.4749, 0.2903, 0.2347] \\ M_2 &= [0.7500, 0.0000, 0.2500] \\ M_3 &= [0.0016, 0.9983, 0.0001] \end{aligned} \right\} \rightarrow \begin{bmatrix} JS_{12} \\ JS_{13} \\ JS_{23} \end{bmatrix} \\ = \begin{bmatrix} 0.1678 \\ 0.4957 \\ 0.9911 \end{bmatrix} \end{aligned}$$

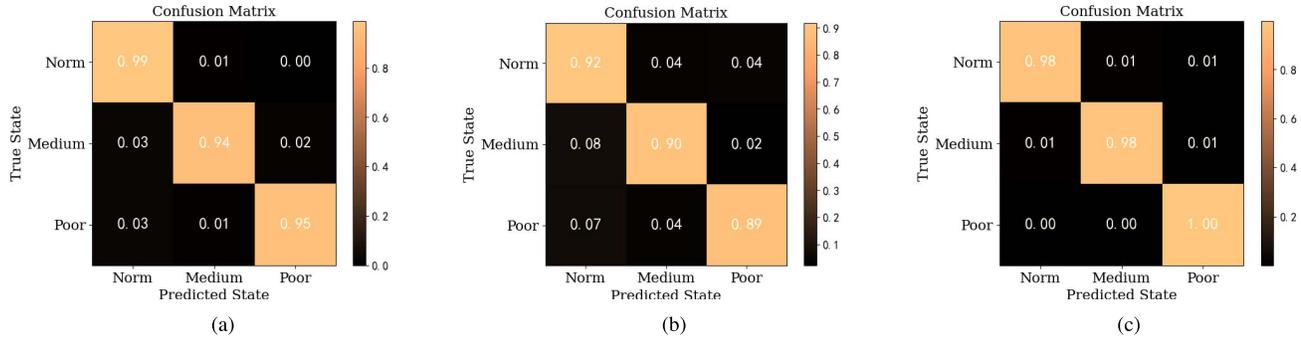


Fig. 5. Confusion matrix for (a) Module<sub>1</sub>, (b) Module<sub>2</sub>, and (c) Module<sub>3</sub>.

where  $JS_{23} > 0.65$ . The correction of evidences is required.

Step 2-2: Sensitivity analysis, see Section II-D1 for details.

Obtain the sensitivity matrix

$$\text{Sen} = \begin{bmatrix} 2.6795 & 2.5828 & 2.5854 \\ 2.5082 & 2.4113 & 2.3972 \\ 2.6599 & 2.6709 & 2.7080 \end{bmatrix}.$$

Get the first correction weights

$$\hat{w}^0 = [2.6795, 2.5082, 2.6709].$$

The results and the JS divergence after the first revision are given as follows:

$$\left. \begin{matrix} M'_1 = [1.2726, 0.7779, 0.6290] \\ M'_2 = [1.8812, 0.0000, 0.6271] \\ M'_3 = [0.0043, 2.6665, 0.0003] \end{matrix} \right\} \rightarrow \begin{bmatrix} JS'_{12} \\ JS'_{13} \\ JS'_{23} \end{bmatrix} \\ = \begin{bmatrix} 0.4316 \\ 1.3254 \\ 2.5659 \end{bmatrix}$$

where  $JS'_{23} > 2.53$ . Evidence revision needs to continue.

Step 2-3: Support analysis, see Section II-D2 for details.

Obtain the distance matrix between evidences

$$\text{DMM} = \begin{bmatrix} 0 & 0.4316 & 1.3254 \\ 0.4316 & 0 & 2.5659 \\ 1.3254 & 2.5659 & 0 \end{bmatrix}.$$

Get the second correction weights

$$\hat{w}^1 = [0.4908, 0.2877, 0.2216].$$

The results after the second correction are given as follows:

$$\left\{ \begin{matrix} M''_1 = [0.6244, 0.3817, 0.3087] \\ M''_2 = [0.5413, 0.0000, 0.1804] \\ M''_3 = [0.0009, 0.5908, 0.0001] \end{matrix} \right.$$

Step 2-4: Evidence summation and normalization

$$\tilde{M} = [0.4439, 0.3700, 0.1861].$$

Step 2-5: Perform Dempster's combination rule twice

$$\begin{aligned} F(\tilde{M}) &= \tilde{M} \oplus \tilde{M} \oplus \tilde{M} \\ &= [0.6049, 0.3504, 0.0446]. \end{aligned}$$

Step 2-6: State assessment

State = I.

3) *Results and Analysis*: The confusion matrix, receiver operating characteristic (ROC) curve, and AUC are often used in the performance analysis of multiclassification models. The performance comparison of the state evaluation algorithm based on decision fusion can also use these three metrics. The confusion matrix is a matching array of predicted labels and true labels. The horizontal and vertical axes of ROC are false and true positive rates, respectively, which is insensitivity to changes in sample distribution. The closer the ROC curve is to the upper left corner, the closer the confusion matrix is to the identity matrix, indicating better model performance. For details, please refer to [48] and [49]. Fig. 5 shows the confusion matrix of the state evaluation results from three modules using the optimal model. Fig. 6 shows the corresponding ROC curve and AUC. Fig. 7 shows the confusion matrix, ROC curve, and AUC of the final fusion result.

Through a comparison between Figs. 5–7, it can be seen that the confusion matrix and AUC after fusion are significantly improved compared to those in Modules<sub>1–3</sub>. Moreover, the performances of Module<sub>1</sub> and Module<sub>2</sub> are relatively lower than that of Module<sub>3</sub>, which is due to the module division. It can be seen from Table I that there are more features in Module<sub>3</sub> than in Modules<sub>1–2</sub>, so this module contains more information, thus possessing better performance.

4) *Comparison of Results and Discussion*: The proposed method first divides the entire dataset into three modules and then uses the corresponding optimal model to conduct the primary state assessment. Finally, the low-conflict evidences without correction or the high-conflict evidences after correction are fused. In order to prove the feasibility and effectiveness of the proposed method, three groups of comparative experiments are carried out. Our experiments are conducted on a Windows Server with a dual –2.80-GHz CPU and a RAM of 4 GB.

a) *Experiment 1*: In order to verify the effectiveness of the module division strategy, the proposed method is compared with the models in the model library. In the process of evaluation, the entire data, instead of the divided data, are divided into the training set and the test set for training and

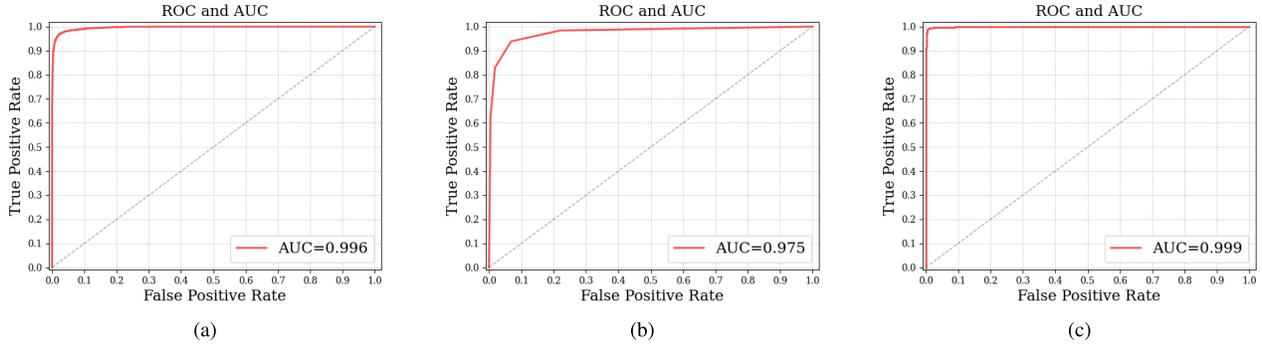
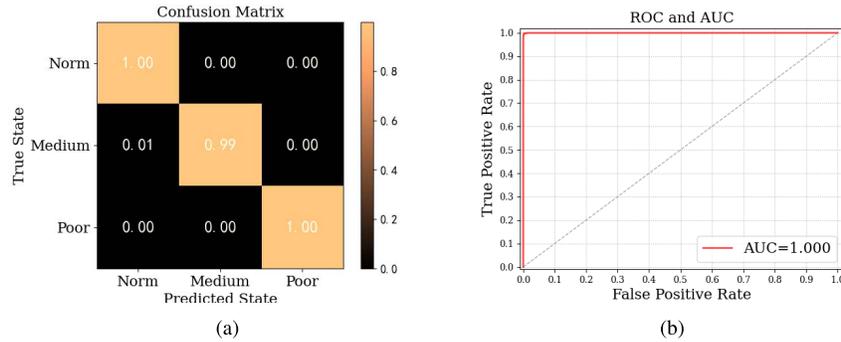
Fig. 6. ROC and AUC for (a) Module<sub>1</sub>, (b) Module<sub>2</sub>, and (c) Module<sub>3</sub>.

Fig. 7. Final fusion result: (a) confusion matrix and (b) ROC and AUC.

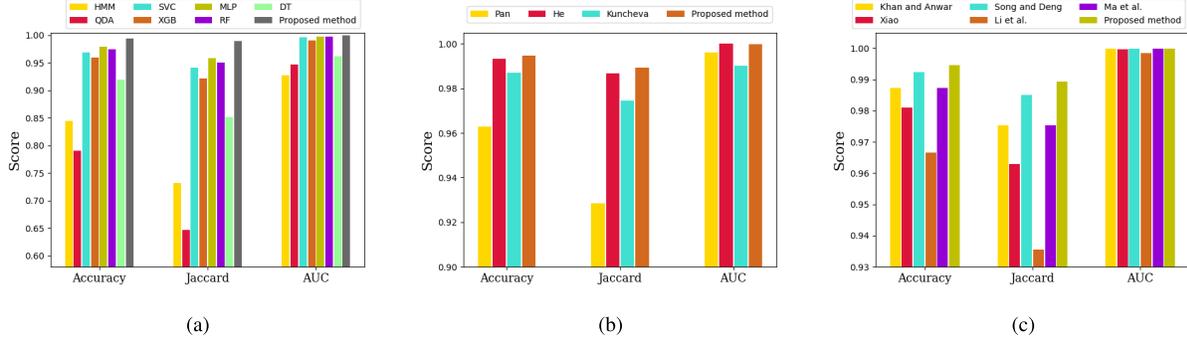


Fig. 8. Compare the results of (a) Experiment 1, (b) Experiment 2, and (c) Experiment 3.

classification, respectively. The comparison results are shown in Fig. 8(a).

The proposed method takes the lead in accuracy, Jaccard similarity coefficient, and AUC. RF, MLP, and XGBoost also have relatively good performance without module partition. This indicates that their information capture ability on this dataset is in a leading position, which is consistent with the model selection results of each module.

Without module partition, the entire dataset would contain too many features, which makes it difficult to fully absorb such enriched information during the training process. However, it means a loss of data information when dimensionality reduction or other data preprocessing methods are adopted. The proposed method is unique in that it uses the selected models to dig into the local details of each module. Through decision fusion, the data complexity can be reduced, and the

information loss can be minimized. This reflects the main advantages of data partitioning.

*b) Experiment 2:* In order to verify the effectiveness of the optimal model selection and fusion strategy, it is compared with other fusion strategies after module division. For a fair comparison, we still use the data of TBM, and the module division is also consistent with Table I. The comparison results are shown in Fig. 8(b) and Table II.

As shown in Fig. 8(b) and Table II, the proposed method performs better than other decision fusion algorithms in terms of accuracy and Jaccard similarity coefficient, while, in the aspect of AUC and time consumption, the performance is just mediocre. The proposed model selection and evidence fusion strategy selects the optimal model for each module and takes the entire probability distribution into consideration. Compared with the method of Pan, which only used SVM,

TABLE II  
COMPARISON OF THE PROPOSED METHOD  
WITH OTHER FUSION METHODS

Methods	Accuracy	Jaccard	AUC	Time
Pan <i>et al.</i> [7]	0.9627	0.9286	0.9962	177.48s
He <i>et al.</i> [23]	0.9934	0.9868	<b>0.9999</b>	306.86s
Kuncheva [50]	0.9870	0.9745	0.9902	<b>18.16s</b>
Proposed method	<b>0.9946</b>	<b>0.9893</b>	0.9997	18.93s

the proposed method can provide more effective evidence for the next level of fusion. Compared with the method of He, where all classifiers are cross-validated both locally and globally, it is more concise and efficient. It should be noted that He used binary classification metrics to measure performance in the original study, but this experiment replaced them with multicategorical metrics, including accuracy, Jaccard's similarity coefficient, and AUC. Compared with Kuncheva's winner-takes-all voting method [50], it takes full account of the trust shown by the evidence for each category.

In terms of time consumption, Pan's method is mainly limited by the long training time of the SVM algorithm, while He's time lost is mainly due to excessive cross-validation. Kuncheva's method is ahead due to its simple voting mechanism. The optimal models selected for the proposed method are XGBoost, RF, and MLP. Although MLP requires more time, its data complexity of its module is relatively low compared with the entire dataset, which makes the overall time consumption acceptable. This further proves the superiority of the proposed model selection and evidence fusion strategy.

*c) Experiment 3:* In order to verify the effectiveness of the proposed evidence correction method, it is compared with other evidence correction methods. The results are shown in Fig. 8(c). The comparison experiments all use the strategy of module partition in Table I and the optimal classifier selection result in Fig. 4.

As shown in Fig. 8(c), the proposed evidence correction method performs better in accuracy and Jaccard similarity coefficient, and is comparable to other methods in AUC. The failure of Li *et al.*'s approach, in this case, mirrors the fact that the approach of modifying DS fusion rules is fragile, which is the reason why the revision strategy is widely accepted. Although the other four methods revised the evidence, none of them took into account the credibility of the evidence sources. If the source of evidence is unreliable due to anomalies in the sensor or model, no amount of correction will help. Compared with the revision strategy that focuses only on the evidence itself, the proposed method considers not only the degree of support between evidences but also the sensitivity of evidence sources, effectively reducing the attention to unreliable evidence. In this way, the higher the sensitivity of the evidence production model and the more support it receives, the higher the credibility and weight of the evidence, and the higher the dominance in the fusion stage, which is in line with intuition. Overall, the feasibility and effectiveness of this correction strategy have been proved in this experiment.

TABLE III  
COMPARISON OF THE ONLINE AND OFFLINE  
VERSIONS OF THE SENSITIVITY ANALYSIS

Versions	Accuracy	Jaccard	AUC	Time
Offline	0.9946	0.9893	0.9997	<b>18.93s</b>
Online	<b>0.9949</b>	<b>0.9899</b>	<b>0.9999</b>	19.97s

*d) Experiment 4:* To illustrate the validity of the online sensitivity analysis, we compared the offline and online versions. It should be noted that the online strategy in Section II-D1 requires that the true historical status is available before the next status monitoring. If this is not the case, it is sufficient to use the sensitivity revision weights derived from the cross-validation phase. Table III shows the results of the comparison between the two. Since the support revision weights are always calculated in real time, the comparison is only for the sensitivity analysis. As shown in Table III, the online version achieves superior performance compared to the offline version. Moreover, the strategy of updating the confusion matrix and recalculating the sensitivity revision weights in real time is not unduly time-consuming. The results show that the performance of the classifier for unseen data is somewhat different from that of the cross-validation phase. However, the offline version is acceptable under the condition that real labels are not available in time. Overall, the feasibility and effectiveness of the proposed online sensitivity analysis strategy were verified in this experiment.

## V. CONCLUSION AND FUTURE WORKS

This research is devoted to developing a generic multimodule decision fusion framework, which is suitable for handling operational status monitoring of large-scale equipment, plant-wide processes, and so on. The decision-making process mainly involves four key steps: 1) divide input data into multiple modules according to the mechanism and professional knowledge; 2) obtain the performance ranking of the models in the model library in each module by performing fivefold cross-validation, and select the corresponding optimal classifier to obtain evidence for the next level of decision fusion; 3) define JS divergence as the conflict metric of evidences, and revise the evidences that exceed the conflict threshold through sensitivity and support; and 4) for low-conflict evidence without revision or high-conflict evidence after revision, the DS evidence theory is used for decision fusion to realize operational status monitoring. Ultimately, the proposed method is applied to a shield machine case of a Chinese tunnel excavation project to verify its feasibility and effectiveness.

As technology evolves, and the variety and functionality of sensors continue to enrich, nonhomogeneous or heterogeneous data will become more prevalent. Information fusion is bound to shine brilliantly. Indeed, further improvements are needed to extend the application of the proposed approach. The method in this study is competent for the fusion of multisource, heterogeneous, and unreliable information, but it cannot directly deal with the unknown or incomplete information prevalent

in the open world, such as emerging unknown patterns and ambiguous information descriptions. Therefore, on the basis of this research, exploring information description, evidence generation and fusion in the open world will be a major part of our future work.

## REFERENCES

- [1] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [2] S. Bagavathiappan, B. Lahiri, T. Saravanan, and J. Philip, "Infrared thermography for condition monitoring—A review," *Infr. Phys. Techn.*, vol. 60, pp. 35–55, Sep. 2013.
- [3] J.-X. Mao, H. Wang, D.-M. Feng, T.-Y. Tao, and W.-Z. Zheng, "Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition," *Struct. Control Health Monitor.*, vol. 25, no. 5, p. e2146, May 2018.
- [4] M. Gao *et al.*, "Dynamic modeling and experimental investigation of self-powered sensor nodes for freight rail transport," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 113969.
- [5] Y. Bao, Z. Chen, S. Wei, Y. Xu, Z. Tang, and H. Li, "The state of the art of data science and engineering in structural health monitoring," *Engineering*, vol. 5, no. 2, pp. 234–242, Apr. 2019.
- [6] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.
- [7] Y. Pan, L. Zhang, X. Wu, and M. J. Skibniewski, "Multi-classifier information fusion in risk analysis," *Inf. Fusion*, vol. 60, pp. 121–136, Aug. 2020.
- [8] Z. Zhao, H. Qi, Y. Qi, K. Zhang, Y. Zhai, and W. Zhao, "Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6080–6091, Sep. 2020.
- [9] S. Wei, Y. Zhao, Z. Zhu, and N. Liu, "Multimodal fusion for video search reranking," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 8, pp. 1191–1199, Aug. 2010.
- [10] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3325–3333, Jun. 2020.
- [11] C. Li, S. Wang, Y. Zhuang, and F. Yan, "Deep sensor fusion between 2D laser scanner and IMU for mobile robot localization," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8501–8509, Mar. 2021.
- [12] Y. Xie, Y. Xia, J. Zhang, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Inf. Fusion*, vol. 42, pp. 102–110, Jul. 2018.
- [13] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," *Neurocomputing*, vol. 174, pp. 50–59, Jan. 2016.
- [14] X. Liu, M. Jia, Z. Na, W. Lu, and F. Li, "Multi-modal cooperative spectrum sensing based on Dempster-Shafer fusion in 5G-based cognitive radio," *IEEE Access*, vol. 6, pp. 199–208, 2018.
- [15] K. Zhong, M. Han, T. Qiu, B. Han, and Y.-W. Chen, "Distributed dynamic process monitoring based on minimal redundancy maximal relevance variable selection and Bayesian inference," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 5, pp. 2037–2044, Sep. 2020.
- [16] K. Taehong, S. A. Ram, and K. Y. Il, "The impact of the PCA dimensionality reduction for CNN based hyperspectral image classification," *Korean J. Remote Sens.*, vol. 35, no. 6, pp. 959–971, 2019.
- [17] S. Duncan and S. Sameer, "Approaches to multisensor data fusion in target tracking: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 12, pp. 1696–1710, Dec. 2006.
- [18] Q. Li, X. Lan, and J. Li, "Information fusion on the two-layer network for robust estimation of multiple geometric structures," *Inf. Sci.*, vol. 530, pp. 148–166, Aug. 2020.
- [19] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1605–1618, May 2018.
- [20] Y. Zhan, H. Leung, K.-C. Kwak, and H. Yoon, "Automated speaker recognition for home service robots using genetic algorithm and Dempster-Shafer fusion technique," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3058–3068, Sep. 2009.
- [21] E. Calikus, S. Nowaczyk, A. Sant'Anna, and O. Dikmen, "No free lunch but a cheaper supper: A general framework for streaming anomaly detection," *Expert Syst. Appl.*, vol. 155, Oct. 2020, Art. no. 113453.
- [22] W. Yu and C. Zhao, "Online fault diagnosis in industrial processes using multimodel exponential discriminant analysis algorithm," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1317–1325, May 2019.
- [23] Q. He *et al.*, "Feasibility study of a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction," *Inf. Fusion*, vol. 55, pp. 207–219, Mar. 2020.
- [24] J. Zhou, X. Li, and H. S. Mitri, "Classification of rockburst in underground projects: Comparison of ten supervised learning methods," *J. Comput. Civil Eng.*, vol. 30, no. 5, Sep. 2016, Art. no. 04016003.
- [25] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Inf. Fusion*, vol. 46, pp. 23–32, Mar. 2019.
- [26] F. Xiao, Z. Cao, and A. Jolfaei, "A novel conflict measurement in decision-making and its application in fault diagnosis," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 186–197, Jan. 2021.
- [27] R.-T. Wu and M. R. Jahanshahi, "Data fusion approaches for structural health monitoring and system identification: Past, present, and future," *Struct. Health Monitor.*, vol. 19, no. 2, pp. 552–586, Mar. 2020.
- [28] W. Liu, "Analyzing the degree of conflict among belief functions," *Artif. Intell.*, vol. 170, no. 11, pp. 909–924, Aug. 2006.
- [29] K. Yuan and Y. Deng, "Conflict evidence management in fault diagnosis," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 121–130, Jan. 2019.
- [30] Z. Deng and J. Wang, "A new evidential similarity measurement based on Tanimoto measure and its application in multi-sensor data fusion," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104380.
- [31] Y. Li, A. Wang, and X. Yi, "Fire control system operation status assessment based on information fusion: Case study," *Sensors*, vol. 19, no. 10, p. 2222, May 2019.
- [32] Y. Song and Y. Deng, "A new method to measure the divergence in evidential sensor data fusion," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 4, Apr. 2019, Art. no. 155014771984129.
- [33] M. N. Khan and S. Anwar, "Paradox elimination in Dempster-Shafer combination rule with novel entropy function: Application in decision-level multi-sensor fusion," *Sensors*, vol. 19, no. 21, p. 4810, Nov. 2019.
- [34] M. Ma and J. An, "Combination of evidence with different weighting factors: A novel probabilistic-based dissimilarity measure approach," *J. Sensors*, vol. 2015, pp. 1–9, Feb. 2015.
- [35] J. Zhao, R. Xue, Z. Dong, D. Tang, and W. Wei, "Evaluating the reliability of sources of evidence with a two-perspective approach in classification problems based on evidence theory," *Inf. Sci.*, vol. 507, pp. 313–338, Jan. 2020.
- [36] D. Fisch, E. Kalkowski, and B. Sick, "Knowledge fusion for probabilistic generative classifiers with data mining applications," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 3, pp. 652–666, Mar. 2014.
- [37] K. Xie, L. Han, M. Jing, J. Luan, T. Yang, and R. Fan, "Review of intelligent data analysis and data visualization," in *Proc. 15th Int. Conf. Broad-Band Wireless Comput., Commun. Appl. (BWCCA)*, Oct. 2020, pp. 365–375.
- [38] Y. Gao, F. Vilecco, M. Li, and W. Song, "Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis," *Entropy*, vol. 19, no. 4, p. 176, Apr. 2017.
- [39] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, Dec. 2004.
- [40] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.
- [41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Aug. 2016, pp. 785–794.
- [42] M. Dumont, R. Maree, L. Wehenkel, and P. Geurts, "Fast multi-class image annotation with random subwindows and multiple output randomized trees," in *Proc. 4th Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, vol. 2, 2009, p. 196.
- [43] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytol.*, vol. 11, pp. 37–50, Feb. 1912.
- [44] S. Jiang, M. Lian, C. Lu, S. Ruan, Z. Wang, and B. Chen, "SVM-DS fusion based soft fault detection and diagnosis in solar water heaters," *Energy Explor. Exploitation*, vol. 37, no. 3, pp. 1125–1146, May 2019.
- [45] A. F. T. Martins, N. A. Smith, E. P. Xing, P. M. Q. Aguiar, and M. A. T. Figueiredo, "Nonextensive information theoretic kernels on measures," *J. Mach. Learn. Res.*, vol. 10, pp. 935–975, Apr. 2009.

- [46] L. Fei, Y. Deng, and Y. Hu, "DS-VIKOR: A new multi-criteria decision-making method for supplier selection," *Int. J. Fuzzy Syst.*, vol. 21, no. 1, pp. 157–175, Feb. 2019.
- [47] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, Feb. 2012.
- [48] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, Nov. 2001.
- [49] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf. Sci.*, vol. 340, pp. 250–261, May 2016.
- [50] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 281–286, Feb. 2002.



**Shuai Tan** received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2005 and 2012, respectively.

She is currently an Associate Professor with the East China University of Science and Technology, Shanghai, China. Her research interests include operation state evaluation for complex industrial processes, fault monitoring and diagnosis, and machine learning of image information.



**Fulin Gao** received the B.S. degree in electrical engineering and automation from Ludong University, Yantai, China, in 2019. He is currently pursuing the master's degree in control science and engineering with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China.

His research interests include operation status assessment of industrial processes, and fault monitoring and diagnosis.



**Hongbo Shi** received the B.E. degree in chemical automation and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 1986 and 2000, respectively.

He is currently a Professor with the East China University of Science and Technology. His research interests include modeling of industrial process and advanced control technology, theory and methods of integrated automation systems, condition monitoring, and fault diagnosis of industrial process.

Dr. Shi was the 2003 Shu Guang Scholar of Shanghai.



**Huaicheng Yan** (Member, IEEE) received the B.Sc. degree in automatic control from the Wuhan University of Technology, Wuhan, China, in 2001, and the Ph.D. degree in control theory and control engineering from the Huazhong University of Science and Technology, Wuhan, in 2007.

In 2011, he was a Research Fellow with The University of Hong Kong, Hong Kong, for three months, and a Research Fellow with the City University of Hong Kong, Hong Kong, in 2012, for six months. He is currently a Professor with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China. His research interests include networked control systems, multiagent systems, and robotics.

Dr. Yan is also an Associate Editor of *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, *International Journal of Robotics and Automation*, and *IEEE OPEN JOURNAL OF CIRCUITS AND SYSTEMS*.



**Zheng Mu** received the B.E. degree in civil engineering from Wuhan University, Wuhan, China, in 2019.

He is currently a Project Leader with China Railway 14th Bureau Group Large Shield Engineering Company Ltd. He has been involved in the construction of metro and large-diameter shield tunnel projects for a long time.

Mr. Mu has won the First Prize of QC achievements of Tunnel Company and the Third Prize of QC achievements in Shanghai Engineering Construction in 2016.