Article:

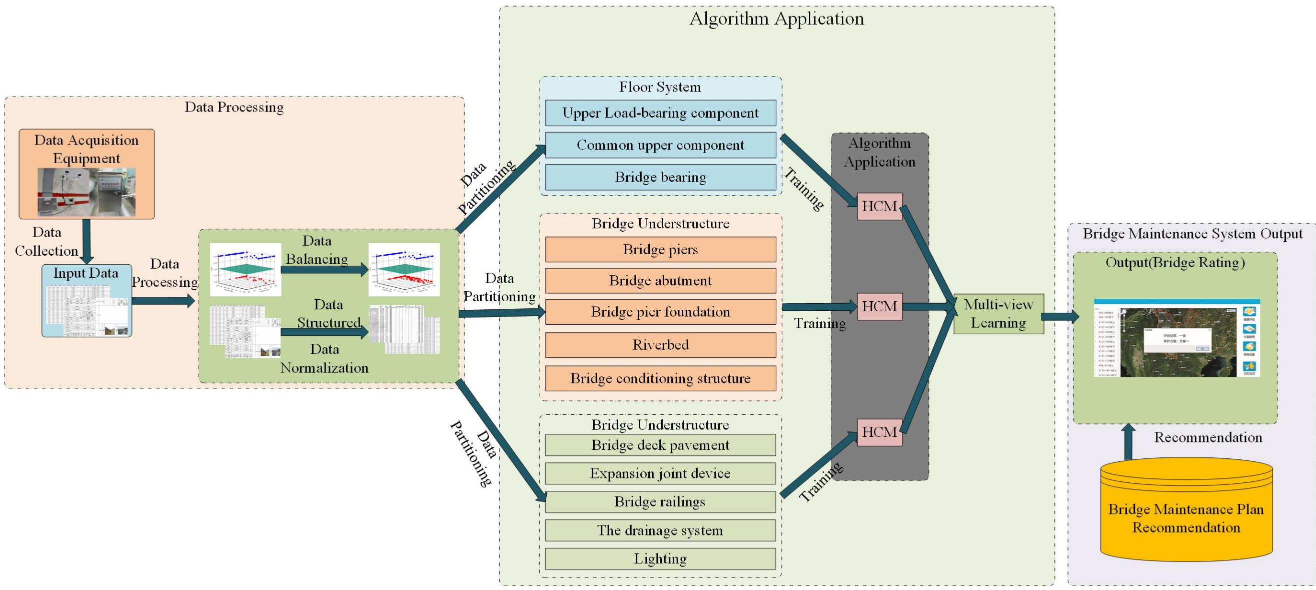
Yang, Y., Nan, F. and Yang, P. orcid.org/0000-0002-8553-7127 (2021) Effective multilayer hybrid classification approach for automatic bridge health assessment on large-scale uncertain data. Journal of Industrial Information Integration, 24. 100234. ISSN 2452-414X

https://doi.org/10.1016/j.jii.2021.100234

Graphical Abstract

Effective Multilayer Hybrid Classification Approach for Automatic Bridge Health Assessment on Large-scale Uncertain Data

Yun Yang,Fengtao Nan,Po Yang



Highlights

Effective Multilayer Hybrid Classification Approach for Automatic Bridge Health Assessment on Large-scale Uncertain Data

Yun Yang,Fengtao Nan,Po Yang

- We design a bridge evaluation system for the classification of Bridges. This is conducive to timely maintenance of the bridge to ensure people's travel safety and reduce the economic losses caused by the bridge safety problems.
- By changing the previous form of manual classification and combining machine learning algorithm with bridge evaluation, automatic bridge status monitoring and evaluation are realized, and the labor cost is reduced.
- A new robust hybrid model based automatic bridge health assessment approach is proposed and development. In this approach, we establish a supervised classifier under the condition of detecting uncertain labels and realize error label correction. Compared to traditional classifiers, this classifier is more feasible to dealing with large-scale bridge data with uncertain labels, where it performs good performance in bridge health evaluation with complex structures.
- A new penalty function analysis method is proposed and integrated in the proposed classifier to measure the final label of the hybrid classification model. This function can iteratively determine the uncertain labels of the labeled data set and evaluate the confidence of the uncertain labels, in which the high confidence data is expanded into the training set and the low confidence data is corrected.
- In practical applications, a comprehensive experimental evaluation and discussion of the proposed method was carried out. Experiments on benchmark data and actual bridge data sets show that the proposed method is superior to alternative solutions and provides potential solutions for practical applications. The results show that when evaluating bridges with complex structures and large-scale uncertain data, our method can greatly outperform other traditional methods.

Effective Multilayer Hybrid Classification Approach for Automatic Bridge Health Assessment on Large-scale Uncertain Data[⋆,⋆⋆]

Yun Yang[a], Fengtao Nan[b] and Po Yang[c,*]

[a]National Pilot School of Software, Yunnan University, Kunming, 650500, China
[b]School of Information Science and Engineering, Yunnan University, Kunming, 650500, China
[c]Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom

ARTICLE INFO

Keywords:
large-scale uncertain label
bridge health evaluation
hybrid classification model
penalty function

ABSTRACT

The health level of the bridge is critical to the safety and maintainability of the bridge. However, with the rapid increase of bridge data with complex information, manual evaluation of bridge health requires high labor and time costs and a lot of related knowledge. Meanwhile, due to the error of sensors and subjective judgment of experts, as well as the influence of the external environment of the bridge, there is great uncertainty in the evaluation of the bridge data. Thence, how to use a large amount of bridge data with large-scale uncertain labels to build a robust classification model for efficient and high-quality bridge health assessment has become an urgent task. In order to better assess the health of the bridge, we adopt a multi-layer hybrid method to iteratively determine the uncertain labels of the target data set, evaluate the confidence of the large-scale uncertain labels, add high-confidence data to the training set, and correct the low-confidence data. Finally, we get an effective classification model with the optimized training data set. This paper studies the learning problem of classification model on labeled data with large-scale uncertain labels, and proposes an effective hybrid classification model (HCM), which can establish a supervised classifier under the condition of detecting uncertain labels and realize error label correction. In order to measure the HCM label assignment problem, we introduce a new penalty function, which can evaluate the label consistency problem of two basic classifiers. Meanwhile, we apply the model to bridge data with uncertain labels for bridge health evaluation. Experiments conducted on synthetic data, benchmark data and real bridge data sets show that the proposed method is superior to other methods and provides an effective and convenient solution for bridge health assessment. At the same time, this method can also be used in other research fields where there are large-scale uncertain labels.

1. Introduction

For industrial applications, how to objective quantify and access the health condition of existing bridges has become an important problem to maintain long-term safety and smooth of traffic flow [31, 29, 27, 19, 36, 23]. The development of industrial information integration promotes the transformation of the industry, and the provision of an automated bridge assessment system has important practical significance for highway traffic [9, 10, 38, 39]. Currently, bridge health evaluation and assessment strategies are mostly manually designed, where experts will extract some key indicators of bridges and further assess their health evaluation. These manual strategies are laborious and time-consuming, especially towards bridges with large-scale and complex structures. Due to the error of sensors and subjective judgment of experts, as well as the influence of the external environment of the bridge, there is great uncertainty in the evaluation of the bridge data.In the industrial applications, how to evaluate the health condition of the existing bridges has become an important problem to ensure the safety and smooth traffic lines [2, 13, 20].

Bridge health evaluation is a complex decision-making process involving multiple levels, indicators and factors [19, 36, 33, 18]. In the process of bridge health evaluation, due to the subjectivity of qualitative evaluation and the limitation of some statistical methods, as well as the interference of many external factors that cannot be quantitatively described. The research of bridge health evaluation theory mainly includes: vibration based identification [8], structural fingerprint and variability [5], system identification [14], optimization of detection points [17]. Now, the technique of integrity assessment has been applied to some simple structures, but it cannot be reliably applied to complex structures. The main reasons that hinder the use of this technology are as follows: [26, 25] 1) The influence of uncertainty and non-structural factors in structure and environment. 2) The measurement information is incomplete. 3) Insufficient measurement accuracy and measurement signal noise. 4) Bridge structure has high redundancy and measurement signal is not sensitive to local damage.

Bridges are an important part of the transportation network and require daily inspection and maintenance activities. Generally, bridge managers pay more attention to the effective management of existing bridges and the development of bridge management systems [24] to assist decision makers in establishing effective repair and maintenance programs. [44] proposed a bridge damage diagnosis and prediction reasoning system based on fuzzy rules, which aims to provide bridge designers with valuable information about the influence of design factors on bridge aging. [21] uses a fuzzy method combining probability theory and fuzzy rea-

*Corresponding author
**Yun Yang and Fengtao Nan contributed equally to this work
yunyang@ynu.edu.cn (Y. Yang); fengtaonan@mail.ynu.edu.cn (F. Nan); Poyangcn@gmail.com (P. Yang)
ORCID(s):

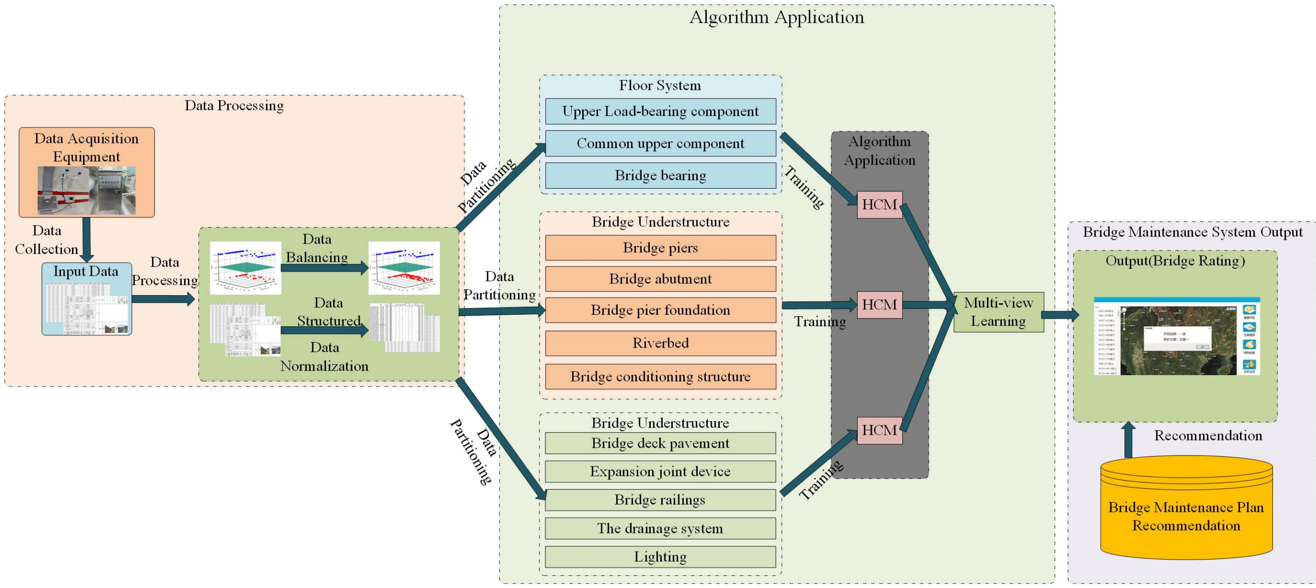


Figure 1: Framework of bridge health evaluation system based on HCM

soning to evaluate the degree of bridge damage and analyze its causes. [32] used Markov model in the bridge management system. [3] describes the Finite Element (FE) and experimental bases of a practical bridge management and maintenance (BMM) system. [35] uses decision tree and Markov process to select the maintenance strategy of the in-service bridge, so that it has the smallest life cycle cost. These systems are efficient and effective to deal with bridges evaluation with simple structures and small-scale data, but suffers from key limitations. Firstly, these methods only consider a single factor in the health evaluation of the bridge. Secondly, the biggest difficulty of the bridge data classification task is that the data label is uncertain, and the single classification algorithm has no label error correction ability. The process of obtaining bridge behavior data from complex bridge engineering, which is known to be faulty, is difficult and costly. In contrast, obtaining behavior data from normally functioning bridges is much easier and cheaper, since most bridges work normally for most of their lifetime. In the process of practical evaluation, due to the subjectivity of qualitative evaluation (i.e., expert subjective experience) and the limitation of some statistical methods (i.e., data dimension is too large, data is presented in multimodal form), as well as the interference of many external factors (i.e., pressure and humidity around the bridge) that cannot be quantitatively described, the evaluation of bridge health has great ambiguity and uncertainty. Therefore, it is highly demanding to explore new classification model based automatic assessment system towards large-scale and uncertain bridge health related data. Meanwhile, those fundamental bridge health evaluation systems do not automatically evaluate bridge health, but are more like a data analysis platform that requires people to participate. There is no maximized reduction in intensity of people's work and the automation of systems is no maximize reduction. These systems do not reduce the intensity of people's work and are highly dependent on people.

In order to solve the complexity of bridge data (i.e., unbalanced, multimodal, noise, less) and improve the accuracy and timeliness of bridge evaluation, this paper attempts to investigate one feasibility of combining the classification algorithm (supervised learning) and clustering algorithm (unsupervised learning) as a hybrid classification model is to solve the problem of classification task and training data label uncertainty. We propose a system for automatic and efficient bridge health evaluation. The main work of the bridge health system assessment is to evaluate (classify) the existing bridge health status based on historical bridge data. In this system, a classifier trained with labeled samples is used, and then a class label is assigned to the existing bridge, and a corresponding maintenance plan is proposed for each evaluated structure. In our implementation, we have designed a hybrid classification model to solve the uncertainty of bridge labeling and successfully applied it to bridge health assessment system. The purpose of this hybrid algorithm is to make it capable of error label correction. Different from other methods to deal with uncertain labels, this hybrid classification model is designed to make it capable of error label correction. The reason for this is the lack of bridge data, which is very precious. This paper introduces a automatic bridge health evaluation system to monitor the bridge health. Our system has a series processes include data collection, data process, etc. In the system, we combine the machine learning algorithm and bridge health evaluation system to build the machine learning model in a data-driven way, so as to minimize the dependence on people and make the system more intelligent. In summary, the main contributions of our work are summarized as follows:

1. We designed a bridge evaluation system for the classification of Bridges. This is conducive to timely maintenance of the bridge to ensure people's travel safety and reduce the economic losses caused by the bridge safety problems.

2. By changing the previous form of manual classification and combining machine learning algorithm with bridge evaluation, automatic bridge status monitoring and evaluation are realized, and the labor cost is reduced.

3. A new robust hybrid model based automatic bridge health assessment approach is proposed and development. In this approach, we establish a supervised classifier under the condition of detecting uncertain labels and realize error label correction. Compared to traditional classifiers, this classifier is more feasible to dealing with large-scale bridge data with uncertain labels, where it performs good performance in bridge health evaluation with complex structures.

4. A new penalty function analysis method is proposed and integrated in the proposed classifier to measure the final label of the hybrid classification model. This function can iteratively determine the uncertain labels of the labeled data set and evaluate the confidence of the uncertain labels, in which the high confidence data is expanded into the training set and the low confidence data is corrected.

5. In practical applications, a comprehensive experimental evaluation and discussion of the proposed method was carried out. Experiments on benchmark data and actual bridge datasets show that the proposed method is superior to alternative solutions and provides potential solutions for practical applications. The results show that when evaluating bridges with complex structures and large-scale uncertain data, our method can greatly outperform other traditional methods.

The remainder of this paper, section II describes the bridge health evaluation system, and section III proposed robust hybrid classification model along with details of the major techniques developed. Section IV reports the experimental results for synthetic data, benchmark datasets and verifies a practical bridge health evaluation model based on our approach as well as the experimental results for bridge datasets. Finally, section V presents the discussion and conclusions.

2. Framework of Bridge Health Evaluation System Based on HCM

In this section, we focus on how our bridge health evaluation system works effectively in the presence of uncertain label data. At the same time, we propose a hybrid classification model (HCM) to deal with uncertain labeling of bridge data. HCM is described in Section III. As illustrated in Fig.1, we extend hybrid classification model (HCM) into a framework of bridge health evaluation system. Meanwhile, the real-time monitoring interface and output results of the system are shown in Fig.2 and Fig.3.

Figure 2: The real-time monitoring interface. This system is applied to a bridge company in Kunming, China, and the interface contents are all in Chinese.The left-hand column of the image is the name of each bridge.The right column is once for bridge health assessment, maintenance program recommendation, special inspection and real-time monitoring functions.

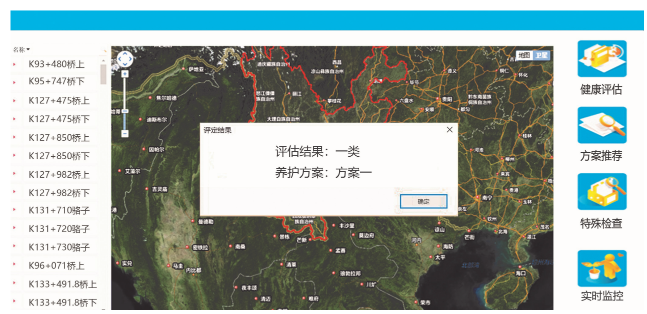


Figure 3: Output of bridge health evaluation system. The main interface is the evaluation result of the bridge, which is shown as a class one bridge, and the maintenance scheme is one.

The system is mainly used to achieve powerful performance on bridge datasets with uncertain label sets. The entire framework can be divided into three modules.

Data Processing: Data acquisition equipment and collected data are shown in Fig.4 and Fig.5. The function of this part is to convert the non-standard bridge data into standard bridge data. The main operations include: 1) unstructured data to structured data; 2) data dimension reduction; 3) data normalization; 4) data balance processing

Algorithm Application: After the data processed in the first step, we analyze the data. Bridge data features are mainly composed of three parts, namely the bridge superstructure, substructure and floor system. The representative features are shown in Fig.6. This module trains our proposed approach HCM on the preprocessed bridge data with both normal and uncertain sets. According to the main evaluation components of the bridge, we introduce multi-view learning [37]. In fact, objects can often be described from different

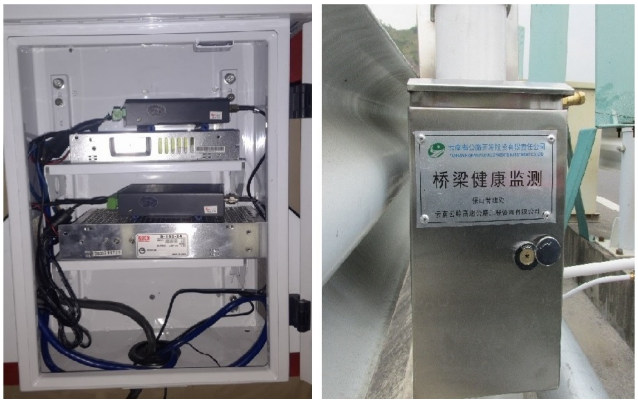


Figure 4: Data acquisition equipment

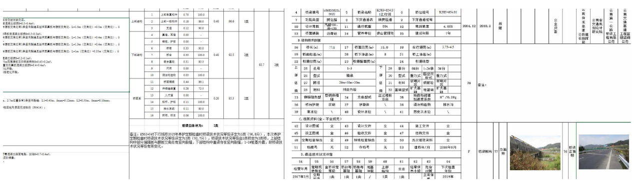


Figure 5: Original bridge data. The data contents are recorded in Chinese, including structured data and unstructured data.

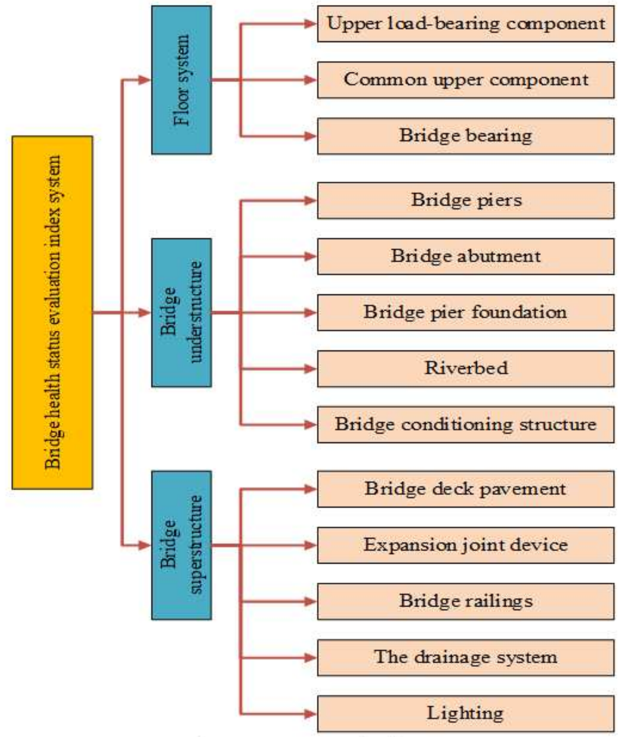


Figure 6: Representative features

views. Usually, a bridge can be described by bridge superstructure, bridge substructure and floor system. We consider the superstructure, substructure, and bridge deck system as auxiliary tasks and the reddest evaluation of the bridge state as the main task. Three classifiers are used to train the processed data obtained from the three components. Then the training results of the three classifiers are used as the characteristics of the main task to train the final bridge evaluation classifier. At the same time, a hybrid classification model is proposed to solve the problem of uncertain labeling in bridge data.

Bridge Health Evaluation System Output: Trained-learners perform classification tasks, where the cases represented by the data (such as bridges) are determined by tags that determine membership in one of the predefined categories of various possible damage levels. Meanwhile, according to the corresponding bridge evaluation grade, we can give the corresponding maintenance plan. In the field of bridge health evaluation, traditional methods mainly adopt manual annotation, and manual analysis is needed for the data collected by equipment, which results in labor cost. At the same time, the analyst needs to have some knowledge of the relevant field. In this case, how to combine machine learning, the establishment of an intelligent system is particularly important. Our bridge health evaluation system can intelligently assess the health status of the bridge without human cost and provide corresponding maintenance plan at the same time.

3. Description of Hybrid Classification Model

In this section, we propose a hybrid classification model (HCM) in order to solve the uncertain labels problem. Our algorithm pseudocode is shown in Algorithm 1. Its first term is completely unsupervised, using only unlabeled instances to measure the difference between the cluster output and the original uncertain label. We use consistent matching to analyze the initial label of the data. We use consistent matching to analyze the initialization label of data. The main purpose of initialization steps is divided into dataset L with normal labels and dataset U with uncertain labels. The second item represents the comparison between the output classification and the clustering output among the monitoring processes. We select the same label output (use the penalty function to analyze the similarity) by the two classifiers as the normal labels, and add the data to the data set L.

3.1. Initialization Procedure

In the beginning, we think that there are uncertain labels in the data and no clear label mechanism is established. We performed cluster analysis on the uncertain label data through unsupervised clustering and use the corresponding matching to measure the clustering results and the original data labels. In order to get the label of each cluster, the data in each cluster is matched consistently with the data in the original data class. That is, the original data and clustering results are intersected, and then normalized to obtain the

Algorithm 1 A robust hybrid classification model with uncertain labels

Input: $D = \{(x_i, y_i)\}_{i=1}^{k}$ ***B_Cluster*** and ***B_Classifier***
Output: L
Initialize: $L = \varnothing; U = \varnothing$

$f(x_i) \leftarrow B\_Cluster(D)$
if $f(x_i) == y_i$ **then**
  Add (x_i, y_i) to L
$U = D - L$
while L is changed **do**
  if $f(x_i) \neq y_i$ **then**
    $f(x) = \sum_{k=1}^{K} p(K = y_k)p(x|K = y_k)$
    $f(x) = \sum_{i=1}^{k} \sum_{j=1}^{K} p(C = y_l|K = y_k)p(x|K = y_k)$
    $f(x) = \sum_{i=1}^{k} \sum_{j=1}^{K} w_{lk}\Pi_k p(x|K = y_k)$
    if $P(f(x_i), y_i) == 1$ **then**
      Add (x_i, y_i) to L
    else
      Add $(x_i, f(x_i))$ to L
  U:U remove (x_i, y_i)
  L:L add $(x_i, y_i or f(x_i))$
return result

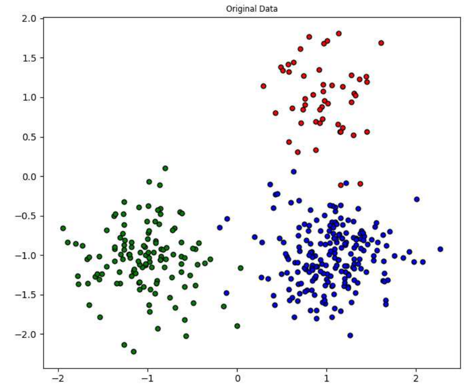


Figure 7: Original data distribution

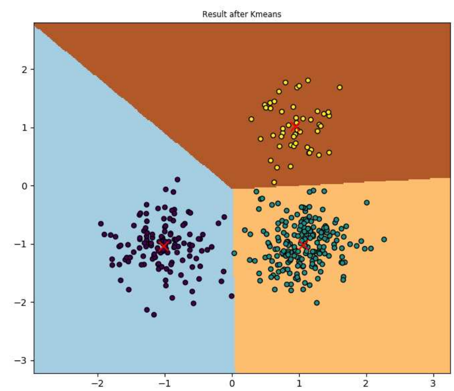


Figure 8: Cluster by K-means

consistency matrix H.

$$H = \begin{bmatrix} \frac{Q_1 \cap O_1}{Q_1}, \frac{Q_1 \cap O_2}{Q_1}, ..., \frac{Q_1 \cap O_n}{Q_1} \\ \frac{Q_2 \cap O_1}{Q_2}, \frac{Q_2 \cap O_2}{Q_2}, ..., \frac{Q_2 \cap O_n}{Q_2} \\ ..., ..., ..., ... \\ \frac{Q_n \cap O_1}{Q_n}, \frac{Q_n \cap O_2}{Q_n}, ..., \frac{Q_n \cap O_n}{Q_n} \end{bmatrix} \tag{1}$$

where Q_i denotes the i-th cluster and O_i denotes the i-th class. In consistency matrix H, we choose the largest value in each row to measure the category of clustering results. The corresponding matching is presented in Fig.7 and Fig.8. In Fig.7, there are 3 clusters corresponding to 3 classes. K-means (Fig.8) is able to deliver a good decision boundary. By our calculations, we obtain the consistency matrix H. From the matrix (2), we can clearly get the similarity between the original data and the clustering results, and we can also assign a label to the unsupervised clustering.

$$H = \begin{bmatrix} 0 & 0.976 & 0.024 \\ 0.01 & 0.005 & 0.985 \\ 0.979 & 0 & 0.021 \end{bmatrix} \tag{2}$$

3.2. Estimation Procedure

Through the initialization step, we divide the original data set into the normal label data set and the uncertain label data set. In the estimation procedure, we think that as long as the given labels in the two evaluation methods are consistent, the uncertain label data can be transformed into the normal label data. In the case that there are differences among the two evaluation criteria, we select the same result (use the penalty function to measure the similarity of the results.) of clustering result And classification result as the new label of the uncertain label data and add it to the uncertain data set again. In this way, our algorithm has error correction ability to a certain extent. At the same time, the normal label data is added to the training set to expand the training data and improve the generalization ability of the model.

Hybrid Classification Model (HCM): We start with the unsupervised clustering where we initially went to a small number of samples that had certain labels. Following the notations in previous section (Initialization Procedure), let $D = \{(x_1, y_1), (x_2, y_2), ..., (x_k, y_k)\}$ denote all the provided data, we use $L = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\}$ and $U = \{(x_{l+1}, y_{l+1}), (x_{l+2}, y_{l+2}, ..., (x_k, y_k)\}$ to denote the certain labeled data and the uncertain labeled data.

We define a hybrid model, which consists of two parts: an unsupervised cluster structure K, which contains $K = (1, 2, ..., k)$ clusters, and a supervised structure C, which contains C classes. According to the traditional hybrid model structure, we assume that the data $D = \{(x_1, y_1), (x_2, y_2), \ldots,$ $(x_k, y_k)\}$ are independent realizations of a random vector

$X \in \mathbb{R}$ with density function:

$$f(x) = \sum_{k=1}^{K} p(K = y_k) p(x|K = y_k) \tag{3}$$

where $p(K = y_k)$is the prior probability of the k-th cluster and $p(x|K = y_k)$ is the corresponding conditional density.

Next, we introduce the supervision information carried by the learning data. Since $\sum_{i=1}^{K} p(K = y_l|C = y_k) = 1$ for all $k = 1, 2, ..., K$,combining formula (3) and $\sum_{i=1}^{K} p(K = y_l|C = y_k) = 1$, we get formula (4).

$$f(x) = \sum_{i=1}^{k} \sum_{j=1}^{K} p(C = y_l|K = y_k) p(x|K = y_k) \tag{4}$$

where $p(C = y_i)|K = y_k$ can be interpreted as the probability that the k-th cluster belongs to the i-th cluster, so the consistency between the cluster and the cluster can be measured. Using the classic symbol of the parametric mixed model and introducing the symbol $\omega_{lk} = p(C = y_k|K = y_k)$, we can reconstruct formula (4) as:

$$f(x) = \sum_{i=1}^{k} \sum_{j=1}^{K} \omega_{lk} \Pi_k p(x|K = y_k) \tag{5}$$

where $\Pi_k = p(K = y_k)$. Therefore, (5) shows both the modeling part of our hybrid model-based approach and the monitoring part through parameters. Since the modeling described in this section is based on a mixed model, we can use any conditional density to model each cluster.

Penalty Function Analysis: According to the labels obtained from the robust HCM, the similarity calculation is carried out between the original labels and HCM labels to determine the confidence of the final labels. Penalty function, presented in (6), measures the similarity between q_n and q_k. The penalty function uses a similarity measure, $r(.,.)$and $s(.,.)$ are calculated as in formulas (7) (8). Then map them to penalty factors in the regular term. The product $r(q_n, q_k) \times (q_n, q_k)$ is normalized in [0,1].

$$P(q_n, q_k) = sign(\frac{r(q_n, q_k) + s(q_n.q_k)}{2}) \tag{6}$$

In this paper, we choose the Pearson correlation coefficient [22] and the Cosine similarity [1] between the probability vectors q_n and q_k. Using Euclidean distance alone may not capture all the information between two vectors. Formally, (8) shows the similarity concerning the correlation

$$r(q_n, q_k) = \frac{\sum_{i=1}^{K} (q_{ni} - \bar{q_n})(q_{nj} - \bar{q_k})}{\sqrt{\sum_{i=1}^{K} (q_{ni} - \bar{q_n})^2} \sqrt{\sum_{i=1}^{K} (q_{ki} - \bar{q_k})^2}} \tag{7}$$

$$s(q_k, q_n) = 1 - \frac{q_k \cdot q_n}{||q_k||_2 ||q_n||_2} \tag{8}$$

where $\bar{q_n}$ is the mean of the vector q_n and $\bar{q_k}$ is the mean of the vector q_k. For the second similarity measure, we compute all the pairwise Cosine similarity (8) between the probability vectors and normalize them in [0, 1].Therefore, similar examples should be close to each other and highly correlated. Since we intend to use the structure generated by the clustering algorithm and the classification algorithm to calculate the similarity in uncertain terms. If the result calculated by Penalty Function Analysis is greater than 0.5, the label of the HCM is consistent with the original label; otherwise, the label is inconsistent. In the case of inconsistency, we select the label result of the HCM as the correct result, while the original label is the sample of the uncertain label. This is a simple explanation of our label correction.

3.3. Optimization

The probability of the occurrence of uncertain label data is introduced into the optimization objective, and it is impossible to estimate the model parameters through theoretical analysis similar to maximum likelihood estimation. In fact, if you take the partial derivative of both sides, you find that the parameters on both sides cancel out. The classification of uncertain label data in the training set can misjudge the hidden value, and (6) is called the estimation problem under incomplete data set. At present, EM algorithm [11] is the classical and mainstream algorithm for estimation problems under incomplete data sets.

4. Experiment Results

In order to verify our method, as described in Sections 2 and 3, we conducted a series of basic experiments and practical applications in this section to demonstrate our HCM method in a more rigorous and fair manner, and recorded detailed experimental results. In the experiment, three groups of representative experiments were conducted. 1) to test the sufficiency of the basic concepts introduced by our method, we propose a robust supervised classification model for uncertain label data, which was initially applied to 2-D-synthetic data sets with different characteristics; 2) to check the effectiveness of our method, the proposed HCM for data with uncertain labels was initially applied to UCI datasets with different characteristics; 3) on this basis, by collecting sufficient and complex bridge data sets, further experimental verification is carried out to compare the test with the existing technology. Finally, we verify the effectiveness of the bridge health assessment system and demonstrate its effectiveness in practical application.

In the process of experimental simulation and verification, the 10-fold cross-validation method was used for verification, and the classifications *Accuracy*, *F-Measure* and *Recall* were recorded respectively. Accuracy [12, 42] is the evaluation standard used by most machine learning algorithms or deep learning algorithms, and is an important indicator to measure the quality of an algorithm [40]. Substantially, *precision* correlated samples indicates the percentage occupied in the retrieved sample, *recall* represents the correlated samples are retrieved relevant percentage of the total samples. The *Precision*, *Recall* and *Accuracy*[41] are de-

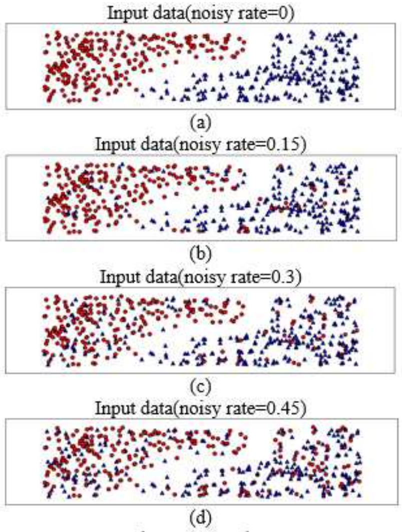


Figure 9: Input data distribution

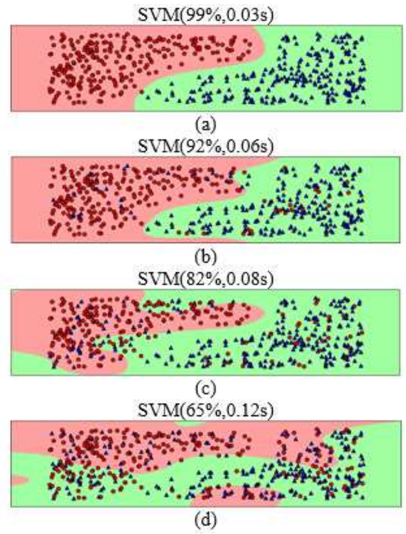


Figure 10: Classification prtfromance on 2-D-synthetic data set with SVM

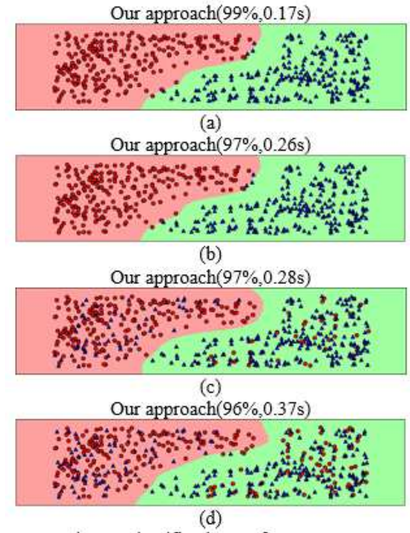


Figure 11: Classification perfromance on 2-D-synthetic data set with our approach

fined as (9), (10) and (11).

$$ACCURACY = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

$$PRECISION = \frac{TP}{TP + FP} \tag{10}$$

$$RECALL = \frac{TP}{TP + FN} \tag{11}$$

where TP is the value assigned to the class correctly, FP is the value assigned to the class incorrectly, and FN is the value that should belong to that class but not assigned to the class. In addition, $F - measure$ combines the accuracy and recall of the two indicators and evaluates the overall performance of the classifier. The most commonly used one is $F - measure$, which is defined as (12).

$$F - Measure = 2 * \frac{PERCISION * RECALL}{PERCISION + RECALL} \tag{12}$$

Uncertain labeling is a pain point in the industrial field, which has a great impact on data-driven models. In the field of machine learning, there are three main strategies for processing uncertain labels: cleaning the data, using robust estimations of model parameters and finally modelling the label noise [6, 15]. **Data cleaning methods**: The removal of data noise can be understood as errors or data deviations from the expected value in the data, that is, the deviation or error of the measured value relative to the real value in the measurement process [34, 7]. However, removing noise samples can reduce the classification bias, but it will also increase the classification variance. Because the cleaned data set is smaller than the original data set, which can easily lead to small samples problems and over-fitting; **Robust estimation of model parameters**: Therefore, other researchers recommend not deleting any learning instances, but building robust supervised classifiers to label noise [4]. However, robust estimation of model parameters generally requires many parameters, and it is difficult to grasp the value of parameters. Although model parameter estimation can solve noise data to a certain extent, it also has complex computational complexity; **Noise modelling**: Noise model is the focus of industrial research in recent years, because the collection of actual data is accompanied by the production of a large number of noise data [6, 28, 30]. Following that, we selected several representative algorithms for uncertain label processing. The idea of Robust Mixture Discriminant Analysis (RMDA) [6] is to use the supervised information carried by the learning data label to assist the modeling of the unsupervised information of the data. This method can build a robust classifier based on the detected label inconsistency. Rank Pruning (RP) [28] obtained consistent noise estimates and equivalent expected risks in learning with undamaged tags under ideal conditions and closed solutions under non-ideal conditions. Probabilistic method (PA) [30] is a supervised learning probabilistic framework with multiple annotators providing labels, but no absolute gold standard. PA iteratively establishes a specific gold standard under which the annotator performance is measured and then the annotator performance is improved. But, there are two strong assumptions that underlie PA. These three methods (RMDA, PA, RP) adopt the idea of eliminating uncertain tags. If the amount of data is insufficient, the ability of these three algorithms is limited.

Table 1
Summary of Parameter Setting

Parameter	SVM	Our approach
GAMMA	2	same
C	1	same
DEGREE	3	same
MAX_ITER	-1	same

4.1. Experiments with Synthetic Datasets

In order to facilitate the understanding how our approach works, we design a synthetic data set of two classes, named noisy *ONE-S*, where two fundamental classification approaches hold. The training set shown in Fig.9 (a) consists of 600 labeled examples marked with *triangle* and *circular*. As depicted in Fig.9 (a), this data set contains one structures separated by a low data density region. For comparison purposes, we apply SVM (kernel is RBF) algorithm on the whole data set. In other words, our approach applies SVM as a base learner on the original data set. For SVM and our approach, parameter settings are shown in Table 1.

In the first experiment, we verify on the synthetic data set, our approach (base learner is SVM, kernel of RBF function) and SVM (kernel of RBF function). As shown in Fig.9 (a), the noisy rate is 0. Compared with Fig.3.3 (a) and Fig.11 (a), the two algorithms are the same in accuracy, but our algorithm consumes a little more time. With the addition of more and more noise(Fig.9 (a)-Fig.9 (d)), the performance of our algorithm (Fig.11 (a)-Fig.11 (d)) is very stable, the classification accuracy is only from 99% to 96%, reduced by 3%. However, the classification accuracy of SVM algorithm is reduced from 99% to 65%(Fig.3.3(a)-Fig.3.3 (d)), which is reduced by 30%. As shown in Fig.3.3 (d) and Fig.11 (d), the time cost of our algorithm is not only higher than that of SVM algorithm, but the classification accuracy of our algorithm is 99 percent, and the classification accuracy of SVM algorithm is 65 percent which is 30 percent higher than that of SVM algorithm. Compare the performance of our method and SVM under different uncertain performance ratios. Compare the performance of our method and the SVM on the *ONE-S* simulation data set with different uncertain label ratios. As expected, our method and SVM give a similar decision boundary without label noise. This experimental result confirms that there is a hidden relationship between the two methods. With the increase in the ratio of uncertain labels, the decision boundary displayed by the support vector machine is very unstable, and our method shows good classification results, with stable decision boundary and good performance, which proves the robustness of our algorithm. Meanwhile, our algorithm has certain error correction ability, which is another reason why our algorithm is better than SVM.

4.2. Experiments with Benchmark Datasets

In the first stage of the simulation experiment, the general classification task is used to evaluate the classification performance of the proposed method. Our benchmark datasets is shown in Table I. The benchmark datasets consist of 13 baseline datasets collected from the UCI database. Table 2 range in size from 106 to 245,057 instances and in dimension from 3 to more than 241 features. Since the benchmark datasets are originally composed of the training set and the test set of the classification task, in order to facilitate us to divide the training set and the test set in the simulation experiment, we combine the two sets into a whole.

Initially, we compare our approach with three classification approaches, including RP, PA and RMDA. Since there is no common implementation code available for the base classification methods, we implement these methods through Python simulations and report average simulation results for a fair comparison with our methods. In terms of experimental parameter setting, these methods adopt Linear Discriminant Analysis (LDA) [16] classifier, and the parameter setting is the default parameter of the LDA algorithm in the *SKLEARN* package.

Table 3 shows the mean and standard deviation of classification accuracy of all datasets with 0.3 nosiy rate, respectively. Tabel II shows that our approach performs highly competitive with RP. Specifically, RP significantly outperforms LDA on 4 of the 13 cases; while HCM&LDA significantly outperforms LDA on 11 cases. Comparing PA algorithm with LDA algorithm, PA wins 9 of 13 case,while HCM&LDA wins 3 more times than the PA algorithm. Meanwhile, we can see that RMDA wins 8 of the 13 cases compared with LDA algorithm, while our algorithm also wins 11 of the 13 cases compared with LDA algorithm.

Table 3 shows the highly competitive performance of RP, PA and RMDA compared with HCM&LDA. Specifically, in terms of wins HCM&LDA outperforms RP approach on 92 percent (12/13) of the cases; And HCM&LDA outperforms RMDA on 70 percent (9/13) of the case; While HCM&LDA outperfroms PA on 77 percent (10/13), respectively. Table II also shows that the performance of our approach may vary when different classification algorithms are used as the combining function. In the experiment, we selected four different classifiers as the basic classifiers, including LDA, CART, KNN and SVM. Specifically, HCM&LDA is significantly superior to LDA in 11 out of 13 patients; HCM&CART significantly outperforms LDA on 9 of the 13 cases; HCM&K NN was significantly superior to LDA in 11 out of 13 cases HCM&SVM significantly outperforms LDA on 13 of the 13 cases; Among these four different basic classifications, the average classification accuracy demonstrates that SVM has better average performance than other candidates. HCM&LDA, HCM&CART, HCM&SVM, HCM&KNN are more accurate than the others. The main reason is that our algorithm adopts a kind of error correction mechanism in uncertain label problem. As a result, we can conclude that the selection of classification algorithm essentially determines the performance of the proposed methods. Finally, Fig.12 shows the CPU time of the RP, PA, RMDA, and our methods (HCM&LDA) on the UCI datasets. In general, RP is always faster than other methods, and PA is always the slowest. However, the CPU time of running our approach is relatively

Table 2
Benchmark Datasets Information

No	Datasets	Attributes	Instances		
			Positive	Negative	Total
1	Australian	14	307	382	689
2	bupa	6	144	200	344
3	cleve	13	136	160	296
4	heart	13	150	120	270
5	appendicitis	7	21	85	106
6	Housevotes	16	123	109	232
7	SkinNonSkin	3	50859	194198	245057
8	mammographic	5	403	427	830
9	Heartstatlog	13	120	150	270
10	Digital1	241	734	766	1500
11	tic-tac-toe	9	625	333	958
12	SPECT	22	110	153	267
13	diabetes	8	268	500	768

Table 3
Mean and standard deviation (%) of 10-fold cross-validation error at 5% of labeled data. •/ ∘ Indicates whether classification is statistically superior/ inferior to the compared method (Decision Tree).Win/ Tie/ Loss denotes the number of datasets where classification is significantly superior/ equal/ inferior to the compared algorithms

Datasets	LDA	RP	PA	RMDA	Our Approach			
					HCM&LDA	HCM&CART	HCM&KNN	HCM&SVM
Australian	73.88±3.94	79.09±5.43•	76.78±4.37•	72.28±4.05∘	78.03±3.61•	71.12±5.91∘	78.61±3.37•	82.44±3.87•
bupa	55.11±6.82	55.80±4.15	62.83±7.22•	52.31±7.39∘	60.17±3.97•	63.18±7.39•	54.33±8.68∘	57.54±9.34•
cleve	60.42±4.50	62.57±7.98•	68.47±8.48•	64.41±5.77•	71.19±4.29•	71.75±5.71•	74.58±2.84•	74.58±4.55•
heart	70.23±7.08	69.12±7.94∘	72.82±8.42•	70.62±6.81	78.43±6.08•	72.11±4.02•	78.07±5.14•	78.42±6.65•
appendicitis	75.38±10.04	58.09±16.03∘	73.05±11.26∘	70.48±10.63∘	86.32±7.07•	87.19±10.06•	89.19±9.07•	90.01±7.09•
housevotes	76.57±6.45	65.84±6.18∘	75.77±4.10∘	84.84±8.26•	85.27±9.87•	81.78±13.04•	88.75±9.04•	90.49±6.06•
SkinNonSkin	76.86±8.43	64.13±10.02∘	76.92±8.51	88.61±5.77•	88.82±6.74•	88.51±7.09•	89.91±4.59•	90.31±4.06•
mammographic	66.51±5.02	69.37±4.68•	70.72±4.39•	67.22±3.33•	79.35±3.45•	74.96±7.83•	74.78±4.22•	77.32±2.45•
Heartstatlog	70.99±3.96	66.16±10.29∘	75.84±3.52•	66.55±5.02∘	79.55±3.01•	69.92±5.55∘	77.69±4.22•	79.17±3.18•
Digital1	86.86±1.71	86.06±11.46	70.17±1.46∘	90.33±1.02•	90.46±1.81•	75.19±7.37∘	88.59±4.55•	88.66±3.94•
tic-tac-toe	43.97±19.06	45.86±17.66•	46.57±21.58•	46.86±22.93•	53.71±14.97•	52.15±17.46•	53.93±16.3•	54.23±12.77•
SPECT	52.99±7.53	49.57±12.47∘	58.28±4.73•	58.26±13.05•	51.84±15.43∘	55.26±3.74•	54.13±8.43•	57.91±12.08•
diabetes	58.58±7.52	56.75±9.28∘	64.84±5.19•	60.55±4.34•	58.84±10.85∘	55.46±9.58∘	56.11±9.65∘	60.91±10.36•
Win/Tie/Loss Against LDA		4/2/7	9/1/3	8/1/4	11/0/2	9/0/4	11/0/2	13/0/0

similar to the RMDA at the middle level of all compared approaches.

Fig.13 shows the performance of RP, PA RMDA and our approach (HCM&LDA) under different ratio of noise data. What can be seen from Fig.3 is highly competitive with RP, PA, RMDA and HCM&LDA for each ratio of noise data (random extend data with uncertain labels). From Fig.13 (a),13 (c), 13 (e), 13 (f) and 13 (g), it can be clearly seen that with the continuous increase of the noise, the classification accuracy of our algorithm is better than the other three algorithms. From Fig.13 (b), 13 (d) and 13 (h), when the noise rate is between 0 and 30%, the classification accuracy of our algorithm is basically the same as that of RP and PA. However, when the noise reaches 40%, our algorithm is obviously better than the other three algorithms. This is because there is an error correction mechanism in our algorithm. When the noise increases from 30% to 50%, the error correction effect is particularly obvious. Finally, we studied the running time of our algorithm (HCM&LDA) and comparison algorithms (RP, PA, and RMDA) on UCI datasets. This experiment was done by performing the settings on a PC running 32 GB of RAM and Windows 10, 4.0GHz Inter(R) Core(TM) i9-9990XE. Fig.12 shows the CPU time of RP, PA, RMDA and our approach (HCM&LDA) on UCI datasets. In general, RP is always faster than other methods, and PA is always the slowest one. However, the CPU time to run our method is relatively similar to the intermediate level RMDA of all the compared methods.

In short, we can get from the Table 3 and Fig. 13, our method usually has higher mean value and lower standard deviation of classification accuracy than other methods, which fully proves that our method has good accuracy

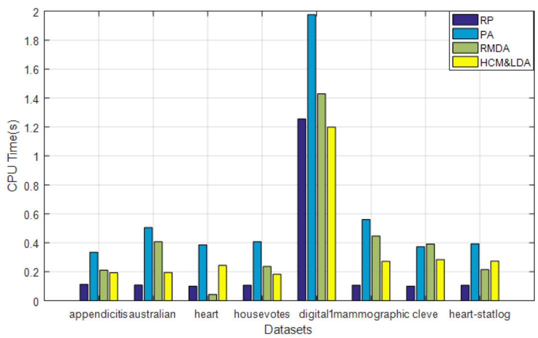


Figure 12: Cpu time (in seconds) of RP, PA, RMDA and HCM&LDA on UCI datasets

and stability. This is because our proposed method can effectively select the most reliable uncertain label data to improve the learning output. As a matter of fact, our approach provides a general framework to solve uncertain label problem, where any conventional classification algorithms can be used as the final stage. The two classification algorithms are just used for demonstration purposes.

4.3. Application to Bridge Health Evaluation System

At present, we have successfully developed the bridge maintenance management platform, and successfully applied our algorithm to the platform. The bridge health evaluation and decision system is mainly a GIS-based bridge engineering data information query system, which solves the visualization management of maintenance data based on geographical rules and can check the information of Bridges, tunnels and other important structures of each highway at any time.

In the bridge health assessment system, sensor equipment can be used to collect a large amount of bridge data, laying a solid foundation for the data-driven decision support function. The above-mentioned decision support system learns bridge knowledge from the collected data and simulates the decision of experts to provide reliable bridge management and protection methods. According to the *Code for Inspection and Test of Highway Bridges and Culverts* and *Code for Evaluation of Highway Bridges and Culverts*, the threshold values of the bridge health evaluation indicators were obtained. Specific bridge health evaluation grades are shown in Table 4. In practice, the third, fourth, and fifth types of bridges are almost non-existent.

4.3.1. Bridge Data

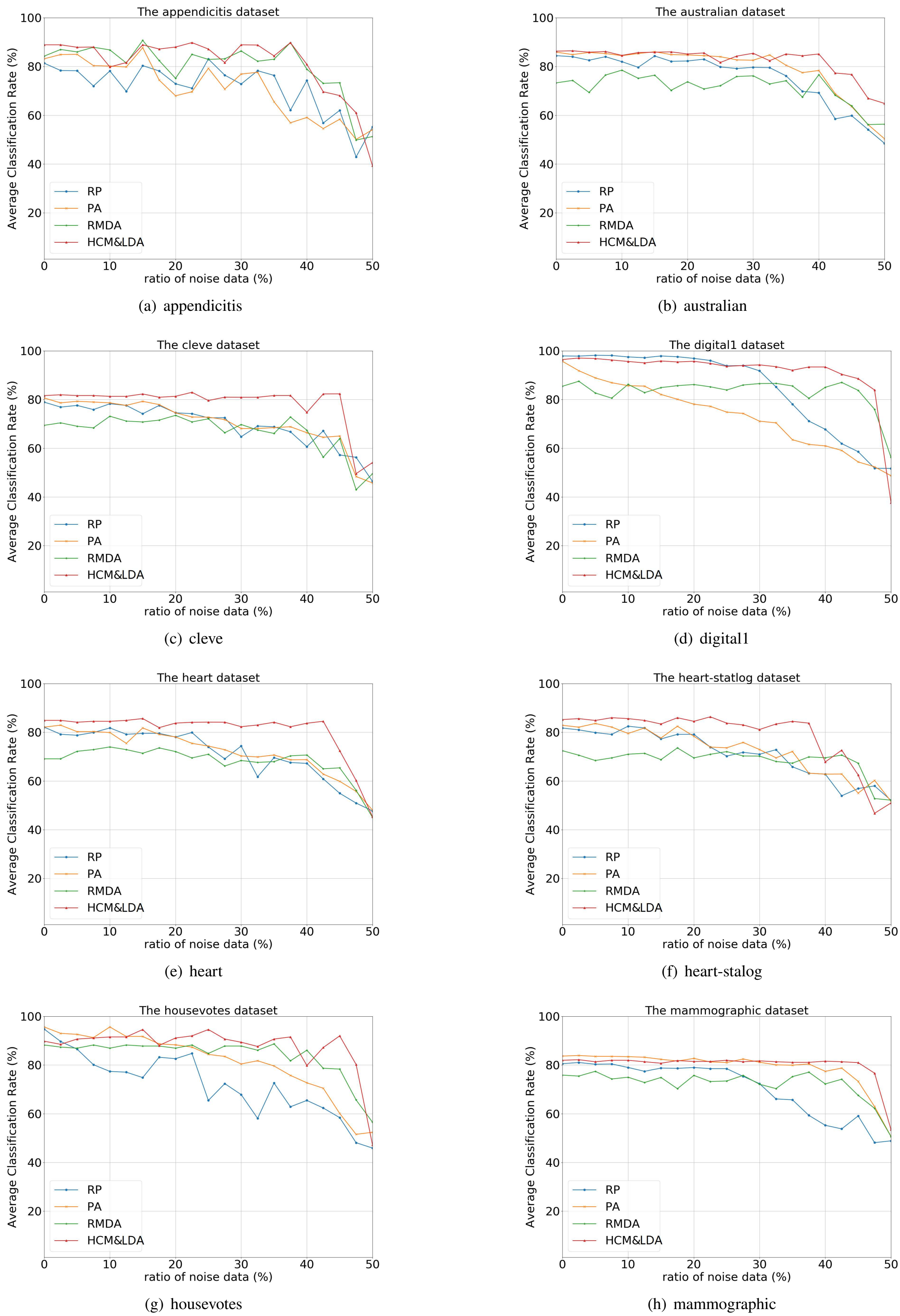
The bridge health evaluation data established in this paper came from a bridge company in Kunming, China. There are a lot of uncertain label data in the original data, and the data structure is very complex. According to the bridge maintenance standards, the bridge is divided into five levels, but in the actual collection of bridge data there are only *Type-1* bridge and *Type-2* bridge. Moreover, the two categories of the bridge data are very unbalanced. As shown in Table 5 above, the dataset contains 1640 bridges with a total of 64 attributes, including 1320 instances of *Type-1* bridge and 320 cases of *Type-2* bridge.

4.3.2. Bridge Health Evaluation Model Based on HCM

We use the *OPENPYXL* package to convert unstructured data into structured data. As shown in Fig.14, in practice, the axial data we collected is unbalanced. In the coaxial system we developed, the SMOTE method is used to solve the problem of unbalanced encoding data distribution and used to construct balanced bridge dataset, as shown in Fig.15. The first type of bridge data points is represented by red points, and the second type of bridge data points is represented by green points. In order to show the balance effect of the SMOTE method, We chose the two most important attributes to display. Then, a feature selection method based on paired constraints is used to perform feature selection on the axial data and select high-impact attributes (or variables) [43].

In order to evaluate the performance of our method, we compare our method (HCM&LDA) with the three methods of RP, PA and RMDA. The above methods follow the same experimental protocol in the simulation. In order to fully evaluate the effectiveness of our algorithm and system, we selected a variety of evaluation criteria, including *Accuracy*, *Recall*, *Precision* and *F-measure*. The basic algorithms of these four algorithms are all LDA algorithms. We run each method ten times and report the average of the categories *Accuracy*, *F-measure*, *Recall*,and *Precision* with the standard deviation in Fig.16-Fig.19, the black line represent the standard deviation.It can be seen that our method achieves the best performance compared with other methods. It can also be explained that our method not only achieves excellent classification *Accuracy*, *F-measure*, *Recall*, *Precision* but also the robustness in a real word application. Meanwhile, the standard deviation of HCM&LDA algorithm is obviously smaller than that of other algorithms, which also reflects that HCM&LDA algorithm has a good stability in the bridge data set.We also compared the computational efficiency of various methods. Fig.20 shows the CPU time of RP, PA, RMDA and our method (HCM&LDA) on the bridged dataset. The results showed that the preprocessing process takes more time than the learning process. In all comparison methods, the CPU time to run our method is relatively similar to RMDA.

Next, in order to evaluation the classification accuracy of our model under different ratios of noise, we set data with different ratios of noise of 10%, 30% and 50%. We run the experiment ten times and report the average *Accuracy* and *AUC* with the standard deviation. To evaluation the classification performance on the different rate of noise bridge dataset, we again compare HCM&LDA, HCM&SVM, HCM
&CART, HCM&KNN with various approaches, including RP, PA and RMDA, the above methods follow the same ex-



(a) appendicitis (b) australian

(c) cleve (d) digital1

(e) heart (f) heart-stalog

(g) housevotes (h) mammographic

Figure 13: Performance of RP, PA, RMDA and HC&LDA for increasing contamination rates on the UCI datasets

Table 4
Bridge Health Evaluation Grade

Technical status rating	Description of bridge technical status
Type-1	Brand new state; Functions in good condition
Type-2	Slight damage; Use function has no effect
Type-3	Secondary damage; Normal operation
Type-4	The bridge function will be seriously affected
Type-5	Out of order

Table 5
Bridge Data Set Information

Datasets	Attribute	Instance		
		Type-1	Type-2	Total
Bridge data	64	1320	320	1640

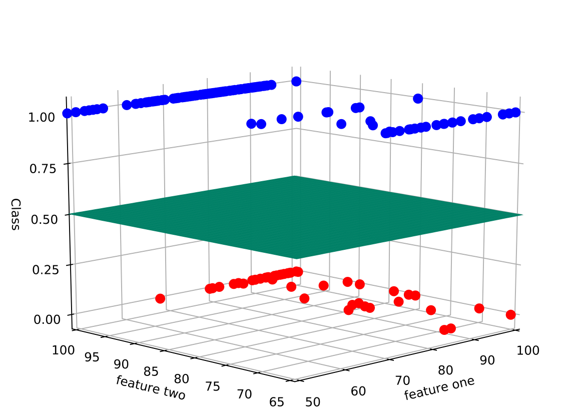


Figure 14: The data distribution of bridge data set with class imbalance

Table 6
Accuracy and AUC of Bridge Data Set for Seven Different Algorithms and the Noise Ratio is 10%

Algorithm	Accuracy	AUC
RP	68.38±6.82	60.322±15.87
PA	72.71±7.03	59.37±9.80
RMDA	85.18±2.71	49.38±2.98
HCM&LDA	81.37±8.78	62.86±8.47
HCM&CART	82.81±7.39	58.17±7.71
HCM&KNN	83.78±9.98	69.34±9.49
HCM&SVM	86.69±7.05	61.72±9.80

perimental protocol in the simulation. We report the average of the classification accuracy and AUC with the standard deviation in Table 6, 7 and 8. As you can see, our method still has the best performance of all the methods compared. The results show that our method not only has good accuracy

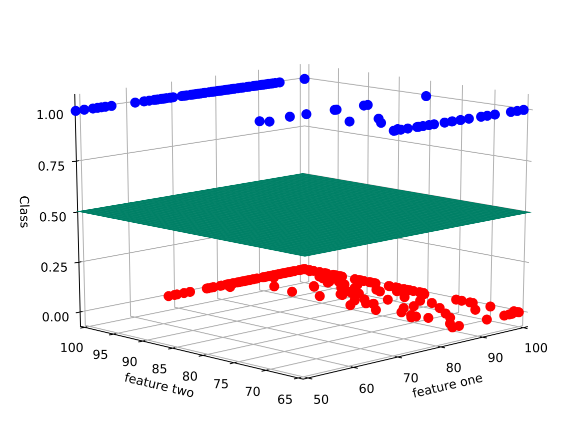


Figure 15: The precision of bridge data set for four different algorithms

Table 7
Accuracy and AUC of Bridge Data Set for Seven Different Algorithms and the Noise Ratio is 30%

Algorithm	Accuracy	AUC
RP	50.69±10.803	53.16±4.62
PA	55.48±5.71	53.92±15.94
RMDA	64.70±18.02	50.60±6.77
HCM&LDA	77.561±14.60	65.28±16.35
HCM&CART	64.80±10.37	52.16±18.05
HCM&KNN	78.52±8.90	59.68±7.64
HCM&SVM	80.44±9.72	57.75±10.73

and AUC, but also has good robustness in practical application. Meanwhile, the standard deviation of HCM&SVM algorithm is obviously smaller than that of other algorithms, which also reflects that HCM&SVM algorithm has a good stability in the bridge data set. The experimental results also show that the standard deviation of our algorithm is relatively small, and it can be analyzed that our algorithm has certain robustness.

The results of RP, PA and RMDA are not ideal, first of all, because the bridge data is not enough. In the absence of samples, we adopt a hybrid model to increase the diversity of classifiers, which is conducive to improving the generalization ability of the initial classifier. Secondly, our algorithm has error correction mechanism. We select the two models

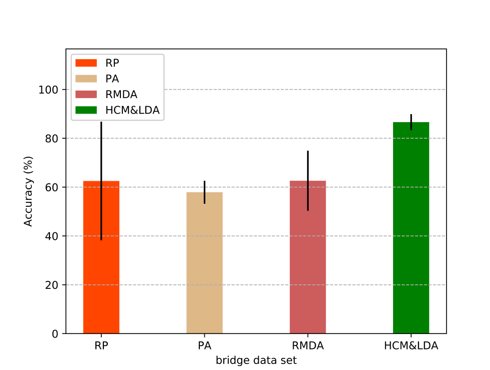


Figure 16: The accuracy of bridge data Set for four different algorithms

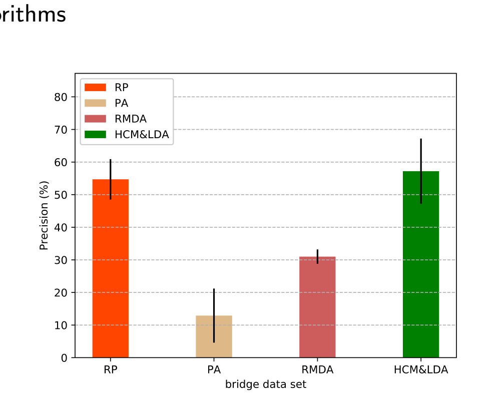


Figure 17: The precision of bridge data set for four different algorithms

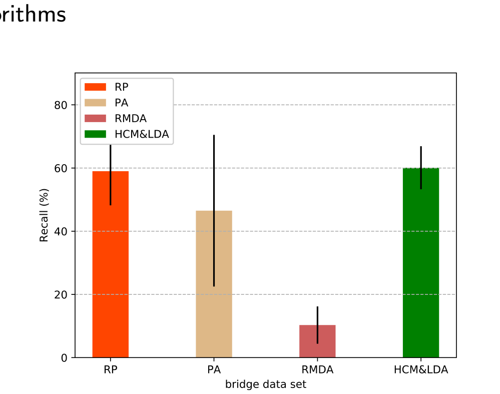


Figure 18: The recall of bridge data set for four different algorithms

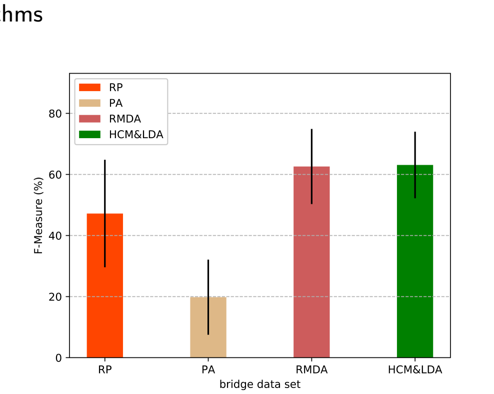


Figure 19: The F-measure of bridge data set for four different algorithms

Table 8
Accuracy and AUC of Bridge Data Set for Seven Different Algorithms and the Noise Ratio is 50%

Algorithm	Accuracy	AUC
RP	53.124±7.96	47.79±7.95
PA	53.74±17.27	58.14±10.36
RMDA	45.77±21.01	49.97±8.32
HCM&LDA	63.74±10.02	52.20±11.42
HCM&CART	57.13±11.71	51.64±14.23
HCM&KNN	75.54±10.78	59.56±8.66
HCM&SVM	77.58±9.17	59.99±8.65

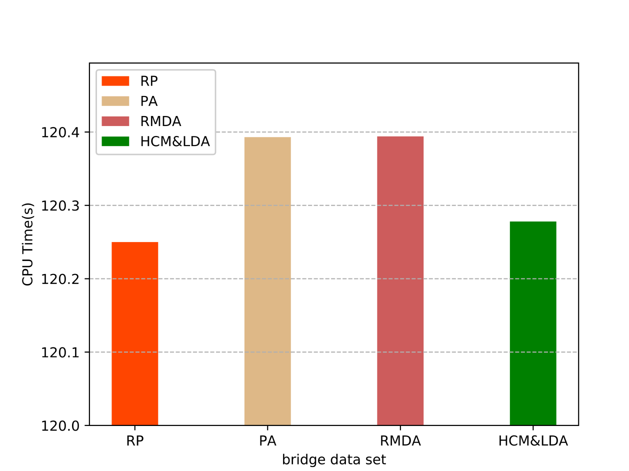


Figure 20: CPU Time (in seconds) of RP, PA, RMDA and HCM&LDA on bridge data set

of clustering (unsupervised) and classification (supervised), and determine its final label by calculating consistency. At the same time, the samples of uncertain labels also participate in the training of the model, which expands the sample set and improves the generalization ability of the classifier.

5. Conclusion and Discussion

In this paper, we propose a hybrid model classification method called robust hybrid classification model (HCM), which is used for classification in the presence of uncertain label. According to our latest research findings, studies on uncertain labels are relatively rare, and this article has to some extent supplemented the research content in the field of uncertain labels. Experimental studies have shown that when uncertain label rate is low, HCM is as effective as fully supervised technology, and even in complex and real situations, HCM is very robust to uncertain labels. In particular, HCM seems to be more robust than existing methods. At the same time, we propose a hybrid classification model for a robust bridge health assessment system. Its main design purpose is to solve the uncertain labeling problem that appears in most collected data sets. Our method adopts a hybrid model, which can make the label prediction more reliable, thereby significantly improving the performance of the supervised learning process. A large number of experimental results show that, compared with the related algorithms (RP, PA and RMDA), our method is very competitive in the benchmark data set and the actual bridge data set collected from the road bridge company. In future, we will further research the complementarity of different hybrid models to make the best choice for basic learners. In the future, we will use semi-supervised learning to solve the problem of uncertain labels in machine learning, and our main idea is that the sample of uncertain labels is more troublesome than the sample of unlabeled ones.

Through the analysis of Table 6, 7 and 8, we can clearly find that both our algorithm and the advanced algorithm have a large standard deviation in bridge evaluation. There may be two reasons for this phenomenon: (1) noise in the data transmission of the device or sensor, and (2) missing values in some data during data integration. In the future work, we will mainly deal with noise reduction processing and data missing value filling of noise data brought by sensors. Since bridge health assessment is a multi-sensor transmission mode, we will mainly consider how to improve the generalization ability of the model under multi-mode to solve the problem of data heterogeneity.

CRediT authorship contribution statement

Yun Yang: Data curation, Writing - review, Visualization, Formal analysis, Conceptualization,Supervision, Writing - original draft. **Fengtao Nan:** Conceptualization of this study, Methodology, Software. **Po Yang:** Data curation, Writing - Original draft preparation.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant 61876166 and Grant 61663046, in part by the Yunnan Applied Fundamental Research Project under Grant 2016FB104, in part by the Yunnan Provincial Young Academic and Technical Leaders Reserve Talents under Grant 2017HB005, in part by the Program for Yunnan High Level Overseas Talent Recruitment, in part of Yunnan Provincial University Key Laboratory Development Project, and in part by the Program for Excellent Young Talents of Yunnan University, in part by the Yunnan provincial major science and technology special plan projects: digitization research and application demonstration of Yunnan characteristic industry, under Grant: 202002AD080001, in part by the Yunnan Applied Basic Research Projects under Grant No.202001BB050034, in part by the Yunnan Applied Basic Research Projects under Grant No.202001BB050034.

References

[1] Abdel-Basset, M., Mohamed, M., Elhoseny, M., Son, L.H., Chiclana, F., Zaied, A.E.N.H., 2019. Cosine similarity measures of bipolar neutrosophic set for diagnosis of bipolar disorder diseases. Artificial Intelligence in Medicine 101, 101735. URL: https://www.sciencedirect.com/science/article/pii/S0933365719303021, doi:https://doi.org/10.1016/j.artmed.2019.101735.

[2] Ai Hong, Guo Shuai, Cai Weisong, 2013. Application of the optimal bp neural network in bridge health assessment, in: Proceedings of 2013 2nd International Conference on Measurement, Information and Control, pp. 921–925. doi:10.1109/MIC.2013.6758110.

[3] Alaylioglu, H., Alaylioglu, A., 1999. Finite element and experimental bases of a practical bridge management and maintenance system. Computers & Structures 73, 281–293. URL: https://www.sciencedirect.com/science/article/pii/S0045794998002594, doi:https://doi.org/10.1016/S0045-7949(98)00259-4.

[4] Bárdossy, A., Singh, S.K., 2008. Robust estimation of hydrological model parameters. Hydrology and Earth System Sciences 12, 1273–1283. URL: https://hess.copernicus.org/articles/12/1273/2008/, doi:10.5194/hess-12-1273-2008.

[5] Bin, Chen, Wang, Dezhang, Sun, Xie, 2014. Integrated system of structural health monitoring and intelligent management for a cable-stayed bridge. TheScientificWorldJournal .

[6] Bouveyron, C., Girard, S., 2009. Robust supervised classification with mixture models: Learning from data with uncertain labels. Pattern Recognition 42, 2649–2658. URL: https://www.sciencedirect.com/science/article/pii/S0031320309001289, doi:https://doi.org/10.1016/j.patcog.2009.03.027.

[7] Brodley, C.E., Friedl, M.A., 1999. Identifying mislabeled training data. J. Artif. Int. Res. 11, 131–167.

[8] Chang, K.C., Kim, C.W., 2016. Modal-parameter identification and vibration-based damage detection of a damaged steel truss bridge. Engineering Structures 122, 156–173. URL: https://www.sciencedirect.com/science/article/pii/S0141029616301845, doi:https://doi.org/10.1016/j.engstruct.2016.04.057.

[9] Chen, Y., 2016. Industrial information integration—a literature review 2006–2015. Journal of Industrial Information Integration 2, 30–64. URL: https://www.sciencedirect.com/science/article/pii/S2452414X16300073, doi:https://doi.org/10.1016/j.jii.2016.04.004.

[10] Chen, Y., 2020. A survey on industrial information integration 2016–2019. Journal of Industrial Integration and Management 05, 33–163. URL: https://doi.org/10.1142/S2424862219500167, doi:10.1142/S2424862219500167, arXiv:https://doi.org/10.1142/S2424862219500167.

[11] Dempster, A.P., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society 39.

[12] Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.

[13] Fang, L., Zhang, L., Yan, S.M., Jia, N., Jing, M., Ma, L., 2011. Design optimization and safety evaluation of cross-sea bridge barrier. Applied Mechanics & Materials 97-98, 100–107.

[14] Farrar, C., James III, G., 1997. System identification from ambient vibration measurements on a bridge. Journal of Sound and Vibration 205, 1–18. URL: https://www.sciencedirect.com/science/article/pii/S0022460X97909779, doi:https://doi.org/10.1006/jsvi.1997.0977.

[15] Frenay, B., Verleysen, M., 2014. Classification in the presence of label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems 25, 845–869. doi:10.1109/TNNLS.2013.2292894.

[16] Guo, Y.R., Bai, Y.Q., Li, C.N., Shao, Y.H., Ye, Y.F., zi Jiang, C., 2021. Reverse nearest neighbors bhattacharyya bound linear discriminant analysis for multimodal classification. Engineering Applications of Artificial Intelligence 97, 104033. URL: https://www.sciencedirect.com/science/article/pii/S0952197620303080, doi:https://doi.org/10.1016/j.engappai.2020.104033.

[17] He, X., Kawatani, M., Hayashikawa, T., Furuta, H., Matsumoto, T., 2011. A bridge damage detection approach using train-bridge interaction analysis and ga optimization. Procedia Engineering 14, 769–776. URL: https://www.sciencedirect.com/science/article/

pii/S187770581101174X, doi:https://doi.org/10.1016/j.proeng.2011.07.097. the Proceedings of the Twelfth East Asia-Pacific Conference on Structural Engineering and Construction.

[18] Jiao, Y., Li, H., Hu, C., Zhang, Z., Zhang, C., 2019. Data-driven evaluation for error states of standard electricity meters on automatic verification assembly line. IEEE Transactions on Industrial Informatics 15, 4999–5010. doi:10.1109/TII.2019.2899465.

[19] Khan, S.M., Atamturktur, S., Chowdhury, M., Rahman, M., 2016. Integration of structural health monitoring and intelligent transportation systems for bridge condition assessment: Current status and future direction. IEEE Transactions on Intelligent Transportation Systems 17, 2107–2122. doi:10.1109/TITS.2016.2520499.

[20] Kim, S.E., Jeon, S.Y., Hong, K.E., Lee, M.C., 2011. Performance evaluation of composite safety barrier for bridge by vehicle impact simulation. Journal of the Korean Society for Advanced Composite Structures 24, 499–506.

[21] Lee, J., Liu, K.F.R., Weiling Chiang, 1999. A fuzzy petri net-based expert system and its application to damage assessment of bridges. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29, 350–370. doi:10.1109/3477.764869.

[22] Ly, A., Marsman, M., Wagenmakers, E.J., 2018. Analytic posteriors for pearson's correlation coefficient. Statistica Neerlandica 72, 4–13.

[23] Malik, H., Zatar, W., 2018. Software agents to support structural health monitoring (shm)-informed intelligent transportation system (its) for bridge condition assessment. Procedia Computer Science 130, 675–682. URL: https://www.sciencedirect.com/science/article/pii/S1877050918304812, doi:https://doi.org/10.1016/j.procs.2018.04.119. the 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.

[24] Miyamoto, A., Kawamura, K., Nakamura, H., 2010. Bridge management system and maintenance optimization for existing bridges. Computer-Aided Civil and Infrastructure Engineering 15, 45–55.

[25] Ngeljaratan, L., Moustafa, M.A., 2020. Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation. Engineering Structures 213, 110551. URL: https://www.sciencedirect.com/science/article/pii/S0141029620300444, doi:https://doi.org/10.1016/j.engstruct.2020.110551.

[26] Ni, Y., Wang, Y., Zhang, C., 2020. A bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data. Engineering Structures 212, 110520. URL: https://www.sciencedirect.com/science/article/pii/S0141029619351478, doi:https://doi.org/10.1016/j.engstruct.2020.110520.

[27] Nie, L., Zhang, L., Yan, Y., Chang, X., Liu, M., Shaoling, L., 2017. IEEE Transactions on Cybernetics 47, 3680–3691. doi:10.1109/TCYB.2016.2577590.

[28] Northcutt, C.G., Wu, T., Chuang, I., 2017. Learning with confident examples: Rank pruning for robust classification with noisy labels. ArXiv abs/1705.01936.

[29] Prasanna, P., Dana, K.J., Gucunski, N., Basily, B.B., La, H.M., Lim, R.S., Parvardeh, H., 2016. Automated crack detection on concrete bridges. IEEE Transactions on Automation Science and Engineering 13, 591–599. doi:10.1109/TASE.2014.2354314.

[30] Raykar, V.C., Yu, S., Zhao, L., Hermosillo, G., Florin, C., Bogoni, L., Moy, L., 2010. Learning from crowds. J. Mach. Learn. Res. 11, 1297–1322.

[31] Rhu, B.J., Song, Y.S., Park, W.T., 2008. Safety evalution of on the cable of extra dosed bridges by fire .

[32] Scherer, W.T., Glagola, D.M., 1994. Markovian models for bridge maintenance management. Journal of Transportation Engineering 120, 37–51.

[33] Su, L., Ma, L., Qin, N., Huang, D., Kemp, A.H., 2019. Fault diagnosis of high-speed train bogie by residual-squeeze net. IEEE Transactions on Industrial Informatics 15, 3856–3863. doi:10.1109/TII.2019.2907373.

[34] Wilson, D.L., 1972. Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics SMC-2, 408–421. doi:10.1109/TSMC.1972.4309137.

[35] Wu, Y., Chang, J., 2011. Study on in-service bridge lcc decision model based on maintenance management system. Advanced Materials Research 255-260, 3933 – 3937.

[36] Xiao, Z., Lim, H.B., Ponnambalam, L., 2017. Participatory sensing for smart cities: A case study on transport trip quality measurement. IEEE Transactions on Industrial Informatics 13, 759–770. doi:10.1109/TII.2017.2678522.

[37] Xu, C., Tao, D., Xu, C., 2013. A survey on multi-view learning. CoRR abs/1304.5634. URL: http://arxiv.org/abs/1304.5634, arXiv:1304.5634.

[38] Xu, L.D., 2020. Industrial information integration – an emerging subject in industrialization and informatization process. Journal of Industrial Information Integration 17, 100128. URL: https://www.sciencedirect.com/science/article/pii/S2452414X20300042, doi:https://doi.org/10.1016/j.jii.2020.100128.

[39] Xu Editor-in-Chief, L.D., 2016. Inaugural issue. Journal of Industrial Information Integration 1, 1–2. URL: https://www.sciencedirect.com/science/article/pii/S2452414X16300115, doi:https://doi.org/10.1016/j.jii.2016.04.001.

[40] Yang, Y., Hu, Y., Zhang, X., Wang, S., 2021. Two-stage selective ensemble of cnn via deep tree training for medical image classification. IEEE Transactions on Cybernetics , 1–14doi:10.1109/TCYB.2021.3061147.

[41] Yang, Y., Jiang, J., 2016. Hybrid sampling-based clustering ensemble with global and local constitutions. IEEE Transactions on Neural Networks and Learning Systems 27, 952–965. doi:10.1109/TNNLS.2015.2430821.

[42] Yang, Y., Jiang, J., 2019. Adaptive bi-weighting toward automatic initialization and model selection for hmm-based hybrid meta-clustering ensembles. IEEE Transactions on Cybernetics 49, 1657–1668. doi:10.1109/TCYB.2018.2809562.

[43] Zhang, D., Chen, S., Zhou, Z.H., 2008. Constraint score: A new filter method for feature selection with pairwise constraints. Pattern Recognition 41, 1440–1451. URL: https://www.sciencedirect.com/science/article/pii/S0031320307004505, doi:https://doi.org/10.1016/j.patcog.2007.10.009.

[44] Zhao, Z., Chen, C., 2002. A fuzzy system for concrete bridge damage diagnosis. Computers & Structures 80, 629–641. URL: https://www.sciencedirect.com/science/article/pii/S0045794902000317, doi:https://doi.org/10.1016/S0045-7949(02)00031-7.