IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 20, 2019, accepted September 8, 2019, date of publication September 24, 2019, date of current version October 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2943380

GMM Clustering-Based Decision Trees Considering Fault Rate and Cluster Validity for Analog Circuit Fault Diagnosis

JUNYOU SHI, QINGJIE HE, AND ZILI WANG
School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
Corresponding author: Qingjie He (heqingjie92@163.com)

ABSTRACT Traditional decision trees for fault diagnosis often use an ID3 construction algorithm. For promoting the accuracy and efficiency of decision trees, considering the cluster validity and fault rates, this paper proposes two improved trees, CV-DTs and FR-DTs. This paper mainly has two highlights. The first highlight is to propose a CV-DT which is constructed by an improved ID3 algorithm considering the cluster validity index. A new cluster validity index which can compare the cluster validities of different attributes is proposed to modify the information gain. This method selects the splitting attributes with higher classification credibility and increases the diagnostic accuracy. The second highlight is to propose an FR-DT which is constructed by an improved ID3 algorithm considering the fault rates. This algorithm not only considers the partitioning ability of each attribute, but also considers the isolation priority of faults with higher fault rates. This method decreases the average diagnostic steps and promotes the diagnostic efficiency. Through a simulation case and a real board case, these decision trees are proved to be effective diagnostic tools which have higher accuracies or efficiencies in analog circuit.

INDEX TERMS Cluster validity, decision tree, fault rate, GMM clustering, ID3 algorithm.

I. INTRODUCTION

Due to the basic characteristics of analog circuits such as non-linearity and tolerance of components, inefficient fault models, inadequate accessible nodes, and uncertainty in the measurements, advanced fault diagnoses in analog circuits have attracted a lot of research attention [1]–[7]. The most popular fault diagnostic method for analog circuits is the fault dictionary based approach as it reduces the diagnostic complexity and the computation cost [8]. The fault dictionary is the look-up table which consists of the fault-free and faulty cases of the circuit. In the fault dictionary based approach, the fault diagnosis depends on comparing the faulty values with the knowledge values. This approach has three main advantages, 1) it can be applied at both the component level and system level, 2) it can allow users comprehend the whole inference process, 3) it has a low computation cost.

The associate editor coordinating the review of this manuscript and approving it for publication was Youqing Wang.

A decision tree is a typical fault dictionary based classification approach, which can reveal the entire process of classification by interpreting the rules and constructing an appropriate framework to quantify the values of outcomes and the possibilities of achieving them. The best-known construction algorithms of decision trees are ID3 and C4.5. ID3 uses information gain to select split points [9] and C4.5 uses information gain ratio to select split points [10]. In addition, there are many novel construction methods, such as new splitting criteria [11] and the use of metaheuristic searching strategy [12].

Compared with other classification algorithms, such as the kNN [13], PCA [14], SVM [15] and ANN [16] algorithms, a decision tree has these advantages. 1) Compared with a SVM classifier, a decision tree can solve the multi-class classification problem. The multiple classification problem is only solved by integrated SVMs. 2) Compared with an ANN classifier, a decision tree is a kind of rule based classifier, which has a simple reasoning structure: 'if symptom then class'. Based on this reasoning logic, we can easily comprehend and interpret the classification results. 3) Compared with a PCA or kNN classifier, a decision tree can deal with the noisy data well, and it does not calculate the matrix eigenvalues and a lot of statistics as PCA. 4) Compared with an SVM classifier and an ANN classifier, a decision tree does not need to set prior hypotheses or presumptions about parameters. It does not need to learn the parameters and thus decreases the computation time and complexity of the training process.

Fault diagnoses based on decision trees have been widely applied in many fields, such as electric power apparatus [17], transmission lines [18]–[20], rotating machinery, and solar photovoltaic arrays. However, a decision tree also has two main drawbacks, overfitting and instability [21]. Overfitting is the result of an improper stopping condition in the construction of a decision tree and it will result in low classification or prediction accuracy. Instability is the result of an improper splitting condition in the selection process of split points and will result in a high level of noise or disturbance sensibility.

Many researchers are devoted to the optimization of decision trees. Kim [21] proposed a semi-supervised decision tree which splits nodes by utilizing both target variables and the structural characteristics of data to improve the accuracy of classifiers. The metric of the best split point is measured by impurity and inhomogeneity simultaneously. Sok *et al.* [22] proposed a multivariate alternating decision tree which allows boosting within a single decision tree. This method retains high comprehension and accuracy simultaneously. Liu *et al.* [23] proposed a private decision tree algorithm based on the noisy maximal vote and an effective privacy budget allocation strategy. The ensemble model under differential privacy adopts a better impurity metric for evaluating attributes and thus boosts the accuracy and improves the stability. Focusing on the monotonic classification, where the objects with better feature values should not be assigned to a worse decision class, the present solution is to optimize the splitting conditions. Hu *et al.* [24] introduced a new measure of feature quality, called rank mutual information (RMI), which combines the advantage of robustness of Shannon's entropy with the ability of dominance rough sets in extracting ordinal structures from monotonic data sets. Based on the RMI, Pei *et al.* [25] constructed multivariate decision trees with monotonicity constraints (MMT). The classification model partitions via oblique hyperplane in the input space and uses improved splitting criteria with rank mutual information (RMI) or rank Gini impurity (RGI). This method can improve the monotonic classification accuracy and stability. Segatori *et al.* [26] proposed a distributed fuzzy decision tree learning scheme shaped according to the MapReduce programming model, which relies on a novel distributed fuzzy discretizer based on fuzzy information entropy. In our previous work, we proposed a quantum clustering based multi-valued quantum fuzzification decision tree (QC-MQFDT) [27]. This decision tree uses adaptive fuzzification method to discretize continuous-valued data and constructs a quantum fuzzy entropy to evaluate the information of each attribute, which makes the decision tree more stable and robust. We also proposed a multi-valued Fisher's fuzzy decision tree (MFFDT) [28]. This decision tree uses Fisher's linear discriminant principles to obtain optimal attributes which can make the tree more concise and accurate.

The establishment of a decision tree is mainly affected by two aspects. The first one is the selection of the splitting attributes, which has been extensively studied in above literatures. The second one is the partition ability of a splitting attributes, which is determined by the clustering results. Different clustering algorithms have different cluster numbers and different degrees of discrimination.

Clustering algorithms are widely used in history data processing and analysis. Xu and Tian [29] summarized 19 clustering algorithms as two types. The traditional methods contain partition based, fuzzy theory based, distribution based, density based algorithms, and so on. The modern methods contain kernel based, ensemble based, quantum theory based, affinity propagation based, spatial data based algorithms, and so on. Gaussian mixture model (GMM) clustering belongs to the distribution based algorithm. It uses the weighted sum of several Gaussian distribution functions to estimate the probability density distribution of samples and the clustering result is to maximize the probability density of samples. GMM is a soft classification in which the classification result is the value of Gaussian probability density instead of the determined cluster [31], [32]. Therefore the GMM has the advantages of interpretation and robustness against observation noise in the low-dimensional data representation [33].

Compared with partition based clustering like K-means clustering, GMM clustering is regarded as an Mahalanobis distance based approach to measure the distance between the data and clusters. Mahalanobis distance has two main advantages. Firstly, it is independent of the measurement scale, which can change the contribution of each coordinate value to the distance according to its fluctuation or uncertainty. Secondly, it can eliminate the effect of deviations, which can handle the noisy data or measurement certainty. Therefore, GMM clustering is widely used in the engineering applications, such as tomography [34], image analysis and recognition [35], [36], bioinformatics pattern classification [37], urban traffic flow prediction [38], industrial processes monitor [39], and sparse reconstruction [40].

In general, cluster algorithms cannot determine the optimal cluster number automatically. Thus, the usual approach is to compare the clustering performances with different cluster numbers. For selecting the optimal value, the cluster validity index (CVI) is proposed, which measures the degree of each cluster number value fitting the input data. CVI estimates the quality of a partition by measuring the compactness within a cluster and the separation among the clusters. The commonly used CVIs contains Dunn index, Calinski-Harabasz index, Davies-Bouldin index, S_Dbw index, and so on [41].

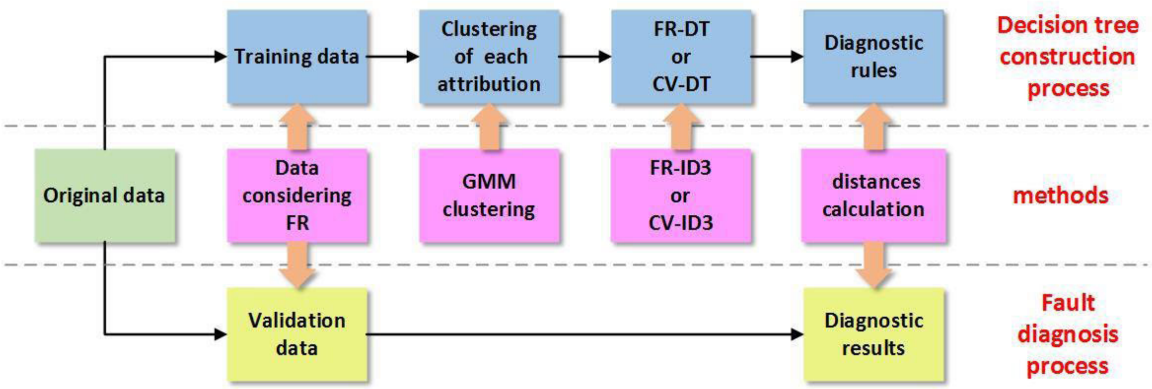


FIGURE 1. The implementation routine of the proposed decision trees for fault diagnosis.

Traditional fault diagnosis based on decision trees mainly has two drawbacks. The first one is that the splitting points are selected without considering the degree of separation among the clusters. The second one is that the decision tree is constructed without considering the fault rates. The essence of ID3 and C4.5 algorithms is to focus on the partitioning ability of each attribute and to construct the concise and efficient decision trees. In other words, ID3 and C4.5 algorithms ignore the influences of the cluster validities of different attributes and the fault rates. The cluster validity mainly influences the decision accuracy. When two clusters are close to each other, the wrong decisions may happen frequently and thus influence whether to follow decisions further. The fault rates mainly influence the decision efficiency. In fact, different faults have different fault rates and the faults with higher rates will happen more frequently. Therefore, the faults with high fault rates should be diagnosed as early as possible.

For the purpose of promoting the diagnostic accuracy and efficiency of decision trees, this paper proposes the fault diagnosis using the GMM clustering based decision trees considering the fault rate and cluster validity separately. These methods mainly contain two highlights.

Focusing on the first drawback, we propose an improved ID3 algorithm considering cluster validity and construct a CV-DT. This algorithm uses a new cluster validity to measure the degree of compactness in a cluster and the separation among the clusters, then the new validity is used to modify the information gain of ID3. This new cluster validity index is not the validity measurement of different cluster numbers, but the validity measurement of optimal cluster numbers of different attributes. It is expressed as the average ratio of the nearest distance between two neighboring clusters to the distance between the means of two neighbor clusters. The higher cluster validity index represents the higher classification credibility and has a positive influence on information gain. Overall, it is verified that CV-ID3 algorithm will have higher diagnostic accuracy than traditional ID3 algorithm.

Focusing on the second drawback, we propose an improved ID3 algorithm considering the fault rate and construct a FR-DT. This algorithm firstly determines the number of fault data based on the fault rate. The faults with higher fault rates are allocated more data numbers. Then we define a new information gain to demonstrate the diagnostic ability of each attribute. This new entropy consists of two information gain entropies. One of the information gain entropies is based on the number of clusters and considers the partitioning ability of each attribute. The other is based on the fault rates and considers the isolation priority of faults with a higher fault rates. Through the synthesis of two information gain entropies, we will construct a more efficient decision tree for fault diagnosis. It is verified that the FR-ID3 algorithm will have a higher diagnostic efficiency than the traditional ID3 algorithm.

The rest of this paper is organized as follows. In section 2, the implementation routine of fault diagnosis using CV-DT and FR-DT is presented. In section 3, the principle of a decision tree using the ID3 algorithm is introduced. In section 4, the principle of GMM clustering and CH validity index are introduced. In section 5, the CV-ID3 and FR-ID3 algorithms are presented to construct the CV-DT and FR-DT. In section 6, two cases of circuit fault diagnosis are provided, in which these two decision trees method are verified.

There are several necessary abbreviations for this paper.

ADS	average diagnostic steps
CH	Calinski-Harabasz
CV-ID3	ID3 algorithm considering cluster validity
CV-DT	decision tree constructed by CV-ID3 algorithm
CVI	cluster validity index
FR-ID3	ID3 algorithm considering fault rates
FR-DT	decision tree constructed by FR-ID3 algorithm
GMM	Gaussian Mixed Model

II. IMPLEMENTATION ROUTINE

The implementation routine of the proposed method of electronic circuit fault diagnosis is illustrated in Fig. 1. The whole figure consists of three parts. The first part is the construction process of the CV-DT and FR-DT, the second part is the fault diagnosis process using these two decision trees, and the third part is the methods which are used in the first and second parts.

In the construction process of CV-DT and FR-DT, different faults contain different data numbers based on the fault rates in the training data. Then these data are partitioned using the GMM clustering algorithm. Meanwhile, the means and variances of clusters are obtained. Based on the traditional ID3 algorithm, we study the effect of cluster validity and fault rates on the selection of splitting points respectively and propose two new improved ID3 algorithms called CV-ID3 and FR-ID3 algorithms. In the last step, we use the standardization distances which depends on the means and variances of clusters as the diagnostic rules.

In the fault diagnosis process, using CV-DT and FR-DT, we first calculate the standardization distances between the validation data and each cluster, and then we select the decision branch with the smallest distance. We will firstly compare the diagnostic efficiency of traditional DT and FR-DT. Then we compare the diagnostic accuracy of traditional DT, CV-DT, and other commonly used multiple classification methods.

III. DECISION TREE USING THE ID3 ALGORITHM

A decision tree is a decision support tool which has a multi-valued tree-like structure. The best-known construction algorithms of decision trees are ID3 and C4.5. However, in this paper, because each fault only contains the single class of value in each test point, the information gain ratio of C4.5 is 1 for any splitting attributes. It means that we cannot select the optimal splitting attributes and construct a decision tree through C4.5 algorithm in this paper. The detailed argument is shown in appendix.

ID3 algorithm uses information gain as the criterion of attribute selection in the splitting nodes. Because the result of the attribute selection based on information gain is partial to the multi-valued attribute, the ID3 algorithm can be regarded as a kind of greedy search strategy. In other words, the essence of ID3 algorithm is to construct a decision tree that is as efficient as possible.

Let $A = \{A_1, A_2, \cdots, A_M\}$ denote the attributes set and M be the number of attributes, $C = \{C_1, C_2, \cdots, C_N\}$ denote the class set and N is the number of classes. Let $|C_n|\ (n = 1, 2, \cdots, N)$ be the number of samples belonging to the class C_n. Let $X = \left\{x_{ij}, 1 \le i \le N_T, 1 \le j \le M\right\}$ denote the sample set, where x_{ij} is the value in the ith sample corresponding to attribute A_j. N_T is the total number of samples.

The information entropy of the sample set X is shown in the form

$$Entropy\,(X) = -\sum_{n=1}^{N} \frac{|C_n|}{N_T} \log \frac{|C_n|}{N_T} \tag{1}$$

If the sample set X is divided by attribute A_j to the clusters $S_1, S_2, \cdots, S_{r_j}$, where r_j is the number of subsets divided by attribute A_j. Let $|S_k| \left(k = 1, 2, \cdots, r_j\right)$ be the number of samples belonging to the subset S_k, $|C_{kn}|$ be the number of samples belonging to class C_n in the cluster S_k. Then the information entropy of the sample set X in the attribute A_j condition is shown in the form

$$\begin{aligned} Entropy\left(X, A_j\right) &= \sum_{k=1}^{r_j} \frac{|S_k|}{N_T} Entropy\,(S_k) \\ &= -\sum_{k=1}^{r_j} \frac{|S_k|}{N_T} \left(\sum_{n=1}^{N} \frac{|C_{kn}|}{|S_k|} \log \frac{|C_{kn}|}{|S_k|} \right) \end{aligned} \tag{2}$$

The information gain in the attribute A_j condition is shown in the form

$$IG\left(X, A_j\right) = Entropy\,(X) - Entropy\left(X, A_j\right) \tag{3}$$

IV. GMM CLUSTERING AND CH VALIDITY INDEX

GMM uses the combination of finite Gaussian probability density functions to estimate the distribution of samples. The purpose of GMM clustering is to maximize the probability density of the samples. Each Gaussian function is a cluster and each sample will be clustered into the Gaussian function which has the largest probability density. The GMM is defined as

$$p\,(x; \theta) = \sum_{k=1}^{n} \pi_k N_k\,(x; \mu_k, \Sigma_k) \tag{4}$$

where x is a d-dimension data, n is the number of Gaussian functions, θ is the parameters set, $\pi_k\ (k = 1, 2, \cdots, n)$ is the weight corresponding to each Gaussian function and $\sum_{k=1}^{n} \pi_k = 1$, $N_k\,(x; \mu_k, \Sigma_k)$ is the kth Gaussian probability density function, μ_k and Σ_k are the average vector and covariance matrix, respectively.

To obtain the maximum likelihood of probability density, we often use the EM algorithm to estimate the GMM parameters. The EM algorithm contains two steps.

(1) Supposing that the average vector and covariance matrix of each Gaussian probability density function are known, we can estimate the probability $\varpi\,(x_i, k)$ that the sample x_i is generated by the kth Gaussian function.

(2) We uses $\varpi\,(x_i, k)$ to estimate the average vector μ_k, the covariance matrix Σ_k and the weight π_k of each Gaussian probability density function.

Repeat the two steps above until the GMM parameters converge to stable values. Then we will obtain n clusters corresponding to n Gaussian distributions $N_k\,(x; \mu_k, \Sigma_k)\ (k = 1, 2, \cdots, n)$.

Because the GMM clustering algorithm need to be preset a cluster number, a CVI is used to select an optimal cluster number. The CH validity index uses the sum of squared distances between a clustering center and the data belonging to this cluster to measure the degree of data compactness in the cluster, then uses the sum of squared distances between the center of the whole data set and each cluster center to measure the degree of separation among clusters. The CH validity index is the ratio of the degree of separation among the clusters to the degree of data compactness of all the clusters, and is computed as

$$CH(NC) = \frac{\frac{1}{NC-1}\sum_{i=1}^{NC} n_i d^2(c_i, c)}{\frac{1}{n-NC}\sum_{i=1}^{NC}\sum_{x\in C_i} d^2(x, c_i)} \tag{5}$$

where n is the total amount of data, c is the center of the whole data, NC is the number of clusters, C_i is the ith cluster, c_i is the center of C_i, n_i is the amount of data in C_i, and $d(x_1, x_2)$ is the distance between two data.

We run the GMM clustering algorithm several times with possible cluster numbers, and we select the cluster number NC corresponding to the largest CH validity index as the optimal cluster number.

V. CONSTRUCTION OF CV-DT AND FR-DT

The traditional ID3 algorithm does not consider the effect of fault rates and cluster validity. In general, attributes with higher cluster validity should be selected as early as possible, so they can promote the diagnostic accuracy of decision trees. On the other hand, faults with higher fault rates should also be diagnosed as early as possible, thus they can promote the diagnostic efficiency of decision trees. Considering the cluster validity and fault rates, this section presents two improved ID3 algorithms, CV-ID3 and FR-ID3 algorithms.

A. CV-ID3 ALGORITHM

The traditional ID3 algorithm selects splitting attributes based on the number of clusters and the data number distribution in each cluster. However, it ignores the effect of the distribution of clusters and the distribution of data in space on the classification accuracy. An optimal partition should have the largest compactness in a cluster and the separation among the clusters. Thus we can univocally classify the faults characterized by the highest probability. From the viewpoint of GMM clustering, the Gaussian distributions of two clusters are shown in Fig. 2.

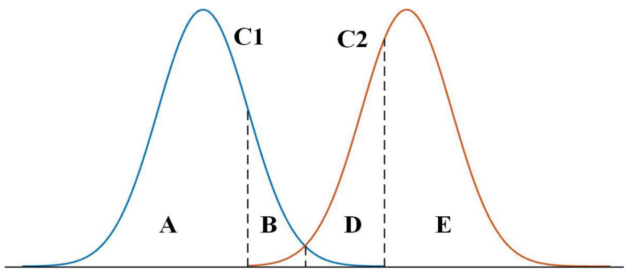


FIGURE 2. The intersection relationship of two cluster curves.

C1 and C2 are two clusters generated by the GMM clustering algorithm. When the data fall into areas A or E, they can be correctly classified into C1 or C2 separately. When the data fall into the areas B and D, they have the risk of being misclassified. In the area B, data may be misclassified into C1, and in the area D, data may be misclassified into C2. Obviously, the areas B and D are wider and the range of misclassified data is larger. For decreasing the misclassified data, we expect to obtain the clusters with smaller areas B and D. Therefore, besides the number of clusters, we should select the splitting attribute that has the clusters with a higher degree of distinction for higher classification accuracy.

Referring to the definition of cluster validity [38], we expect to use a CVI to compare the cluster validity of different attributes. However, the existing CVIs are only suitable for measuring the different partition qualities inside a same data space or inside an attribute. For different attributes and their different data spaces, they have different optimal cluster numbers, different numerical ranges, and so on, therefore these CVIs are not applicable.

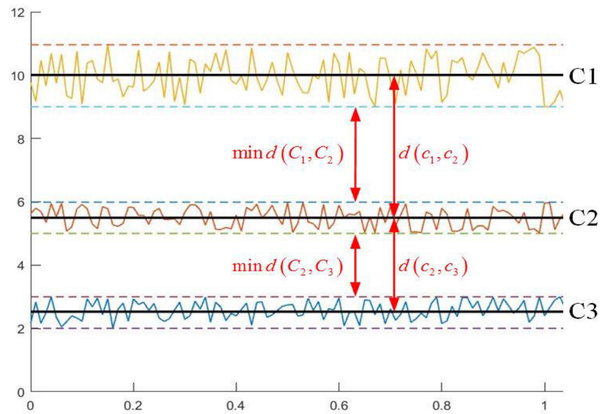


FIGURE 3. The graphical description of the proposed CVI.

For measuring and comparing the clustering validity of different attributes objectively, we propose a new CVI which eliminates the effect of the optimal cluster numbers and numerical ranges. The formula of the new CVI is as follows

$$CVI_{new} = \frac{1}{NC_{opt}-1}\sum_{i=1}^{NC_{opt}-1}\frac{\min d(C_i, C_{i+1})}{d(c_i, c_{i+1})} \tag{6}$$

where NC_{opt} is the optimal cluster number, C_i is the ith cluster, c_i is the mean of C_i, $d(c_i, c_{i+1})$ is the distance between the mean of C_i and the mean of C_{i+1}, $\min d(C_i, C_{i+1})$ is the shortest distance between C_i and C_{i+1}. The new CVI calculation consists of two steps. The first step is the CH indexes calculation of different attributes. In this step, we find the optimal cluster number of each attribute NC_{opt}. The second step is the new CVI calculation of different attributes. In this step, we calculate the new CVI of each attribute in the condition of optimal cluster number and find the max value of CVI_{new}. The detailed description of the formula is shown in Fig. 3.

As shown in Fig. 3, the dimensionless value $\frac{\min d(C_i, C_{i+1})}{d(c_i, c_{i+1})}$ measures the degree of difference between two neighboring clusters and eliminates the effect of different numerical ranges. $\frac{1}{NC_{opt}-1}$ is used to average the total differences and eliminate the effect of cluster numbers.

The proposed CVI helps us select the splitting attribute with higher credibility. However, when two attributes have the same CVI, we tend to select the attribute with more clusters. Therefore, considering the cluster validity and information gain simultaneously, we propose a CV-ID3 algorithm. The formula of information gain considering cluster validity is as follow.

$$IG^{CV}\left(X, A_j\right) = CVI_{new}\left(A_j\right) * IG\left(X, A_j\right) \tag{7}$$

where $IG\left(X, A_j\right)$ is the information gain of attribute A_j with the traditional ID3 algorithm, and $IG^{CV}\left(X, A_j\right)$ is the modified information gain of attribute A_j with CH-ID3 algorithm.

Similar to the traditional ID3 algorithm, the proposed CV-ID3 algorithm selects the attribute with the largest $IG^{CV}\left(X, T_j\right)$ as the splitting point. If the CVI of attribute A_j is equal to 1, it can be regarded that the information provided by A_j is completely true and valid. However, if CVI of attribute A_j approaches 0, it can be understood that the information provided by A_j is completely confused and invalid. In this respect, the proposed CVI can be regarded as the measurement of information credibility.

B. FR-ID3 ALGORITHM

In the rest part of the paper, we mainly discuss the application of the decision tree in analog circuit diagnosis. We will replace the attribute with the test point, and replace the classification object with the fault mode in the decision trees.

In general, we use the ADS to measure the efficiency of a decision tree. There are two ADS definitions which are computed as

$$ADS_1 = \frac{\sum_{i=1}^{N} s_i}{N} \tag{8}$$

$$ADS_2 = \frac{\sum_{i=1}^{N} n_i s_i}{n} \tag{9}$$

where n is the total number of fault data, N is the number of faults, n_i is the number of ith fault data, s_i is the diagnostic step of the ith fault in the decision tree, i.e. the number of splitting attributes that we use to determine ith fault.

The ADS_1 is defined as the average diagnostic steps of all kinds of faults, which assumes that all faults happens with a same probability. The ADS_2 is defined as the average diagnostic steps of all the faults that happened over a period of time, which means that different faults may happen with different frequencies. From the viewpoint of practical application, the second definition is more significant.

The essence of the traditional ID3 algorithm is that it is a kind of greedy algorithm, which means that it expects to select the attribute which has the most clusters at each splitting point. In this way, the traditional ID3 algorithm can construct an optimal decision tree which minimizes (8). As for (9), we not only need more clusters in each splitting process, but also need to isolate the fault which has the larger n_i as early as possible. The traditional ID3 algorithm cannot find the optimal solution which satisfies the isolation requirement.

In (9), the larger n_i means that the ith fault happens with a higher frequency, thus we can replace n_i with the fault rate λ_i. Obviously, different faults have different fault rates and we must consider the number of clusters and fault rates simultaneously. On the one hand, if we select a splitting attribute based only on the number of clusters, although the depth of the decision tree may be shortened, the fault with a higher rate may be diagnosed by using more diagnostic steps and thus increases the ADS. On the other hand, if we select a splitting attribute based only on the fault rates, i.e. give priority to the diagnosis of faults with higher fault rates, the depth of the decision tree may increase and thus the diagnostic steps of other faults will increase. Therefore, for the purpose of obtaining the general optimal solution of ADS, we propose a FR-ID3 algorithm which considers the number of clusters and the fault rates as the selection rules of splitting attributes simultaneously.

Firstly, in the traditional ID3 algorithm, the data number of each fault is commonly not definite, which means it cannot reflect the frequency of occurrence of each fault. In the FR-ID3 algorithm, we must determine the fault data number based on the fault rates. If the fault rate of fault A is twice that of fault B, then the data number of fault A is two times more than fault B.

Secondly, in the traditional ID3 algorithm, we use one information gain to measure the amount of information that each attribute provides. In the FR-ID3 algorithm, we use two information gain entropies to measure the amount of information that each attribute provides. One of the information gain entropies is based on the number of clusters as the traditional ID3 algorithm. The other is based on the fault rates and considers the isolation priority of faults with a higher fault rate.

Let $T = \{T_1, T_2, \cdots, T_M\}$ denote the test points set and M be the number of attributes, $F = \{F_1, F_2, \cdots, F_N\}$ denote the fault modes set, N be the number of fault modes, and $\lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ denote the N fault rates. Let $|F_n|\ (n = 1, 2, \cdots N)$ be the number of samples belonging to the fault F_n considering the fault rate. Let $X = \left\{x_{ij}, 1 \le i \le T, 1 \le j \le M\right\}$ denote the sample set, where x_{ij} is the value in the ith sample corresponding to test point T_j. N_T is the total number of samples.

(1) The information gain based on the number of clusters is the same as the ID3 algorithm, which is written as

$$IG_{NC}\left(X, T_j\right) = Entropy_{NC}\left(X\right) - Entropy_{NC}\left(X, T_j\right) \tag{10}$$

$$Entropy_{NC}\left(X\right) = -\sum_{n=1}^{N} \frac{|F_n|}{N_T} \log \frac{|F_n|}{N_T} \tag{11}$$

$$\begin{aligned} Entropy_{NC}\left(X, T_j\right) &= \sum_{k=1}^{r_j} \frac{|S_k|}{N_T} Entropy\left(S_k\right) \\ &= -\sum_{k=1}^{r_j} \frac{|S_k|}{N_T} \left(\sum_{n=1}^{N} \frac{|F_n|}{|S_k|} \log \frac{|F_n|}{|S_k|} \right) \end{aligned} \tag{12}$$

where $IG_{NC}\left(X, T_j\right)$ is the information gain based on the number of clusters, $Entropy_{NC}(X)$ is the information entropy based on the fault modes, $Entropy_{NC}\left(X, T_j\right)$ is the conditional entropy based on the partitioning ability of test point T_j.

(2) The information gain based on the fault rates is written as

$$IG_{FR}\left(X, T_j\right) = Entropy_{FR}(X) - Entropy_{FR}\left(X, T_j\right) \tag{13}$$

$$Entropy_{FR}(X) = -\sum_{n=1}^{N_T} \frac{1}{N_T} \log \frac{1}{N_T} \tag{14}$$

$$Entropy_{FR}\left(X, T_j\right) = -\sum_{i=1}^{m_j} \frac{1}{N_T} \log \frac{1}{N_T}$$

$$m_j = \sum_{F_n \notin F^*} |F_n| \tag{15}$$

where $IG_{FR}\left(X, T_j\right)$ is the information gain based on the fault rates, $Entropy_{FR}(X)$ is the information entropy based on the fault data number, $Entropy_{FR}\left(X, T_j\right)$ is the conditional entropy based on the isolation ability of a single fault of test point T_j, F^* denotes the fault modes set which can be isolated by test point T_j, m_j denotes the number of fault data which cannot be isolated by test point T_j.

In the formula of $Entropy_{FR}(X)$, because each fault data needs to be isolated, we regard each fault data as a different item and the probability of each fault data is $1/N_T$. In the formula of $Entropy_{FR}\left(X, T_j\right)$, because the isolated faults have no effect on this conditional entropy, we consider the faults data that cannot be isolated singly. They can also be regarded as different and the same probability $1/N_T$. In (15), the larger $Entropy_{FR}\left(X, T_j\right)$ means m_j is larger, and the larger m_j means the number of faults that cannot be isolated is larger. On the contrary, if $|F_n|$ which satisfied $F_n \in F^*$ is larger, i.e. the fault with the higher fault rate can be isolated, then $Entropy_{FR}\left(X, T_j\right)$ is smaller and the isolation ability of test point T_j is better.

(3) The new information gain in FR-ID3

As mentioned above, we cannot use only $IG_{NC}\left(X, T_j\right)$ or $IG_{FR}\left(X, T_j\right)$ as the selection rule of test points. If we select test points only based on $IG_{NC}\left(X, T_j\right)$, the ADS_2 may become larger because the faults with a higher fault rates may be isolated using more diagnostic steps. If we select test points only based on $IG_{FR}\left(X, T_j\right)$, the ADS_2 may become larger because the other faults may be isolated using more diagnostic steps. For considering the number of clusters and fault rates simultaneously, we combine $IG_{NC}\left(X, T_j\right)$ and $IG_{FR}\left(X, T_j\right)$ together to form a new information gain.

$$IG^{FR}\left(X, T_j\right) = IG_{NC}\left(X, T_j\right) + IG_{FR}\left(X, T_j\right) \tag{16}$$

Similar to the traditional ID3 algorithm, the proposed FR-ID3 algorithm selects the test point with the largest $IG^{FR}\left(X, T_j\right)$ as the splitting point. In this respect, the FR-ID3 algorithm also can be regarded as a kind of greedy algorithm. However, in this greedy algorithm's case, the largest $IG_{NC}\left(X, T_j\right)$ or largest $IG_{FR}\left(X, T_j\right)$ cannot ensure the largest $IG^{FR}\left(X, T_j\right)$ either, thus $IG^{FR}\left(X, T_j\right)$ balances the importance relationship between the partitioning ability and the fault isolation priority of test point T_j.

C. COMPARISON BETWEEN CV-ID3 AND FR-ID3 ALGORITHMS

We have discussed two improved ID3 algorithms. The CV-ID3 algorithm considers the effect of cluster validity and improves the fault diagnostic accuracy, and the FR-ID3 algorithm considers the effect of fault rate and improves the fault diagnostic efficiency. In general, the CV-DT and FR-DT are different. The CV-DT has the highest accuracy and the FR-DT has the highest efficiency. Therefore, we cannot construct a decision tree with the highest accuracy and efficiency simultaneously. If we combine these two algorithms together, such as replacing $IG\left(X, A_j\right)$ in (7) by $IG^{FR}\left(X, A_j\right)$, the decision tree constructed by this combined algorithm has lower accuracy than the CV-DT and lower efficiency than the FR-DT.

Based on the different requirements of the fault diagnostic application, fault diagnosis is divided into online diagnosis and offline diagnosis. For online diagnosis, it is important to locate the faults timely and quickly, especially when a timely decision is needed. Using the FR-ID3 algorithm, the average diagnostic time of online diagnosis is significantly reduced. For offline diagnosis, what we care about is how to locate the fault accurately. For this reason, the CV-ID3 algorithm is used to ensure the highest diagnostic accuracy.

In general, the FR-ID3 algorithm prefers to select the test point with more clusters. However, in many cases, more cluster numbers the partition has, the higher compactness the clusters have. Thus the CVI of the selected test point may be small and the accuracy of decision tree will be decreased. On the other hand, the CV-ID3 algorithm prefers to select the test point with a larger CVI and in many cases this test point has few clusters. It will result in a larger ADS and lower efficiency. In other words, high accuracy will be at the expense of lower efficiency and high efficiency will be at the expense of lower accuracy.

For CV-ID3 and FR-ID3 algorithms, the calculation process of the information gains are the same as traditional ID3 algorithm in the definition. Therefore, the time complexity and the space complexity of CV-DT and FR-DT are the same as traditional DT. The time complexity is $O(N * M * D)$, where N is the number of samples, M is the number of splitting points, and D is the depth of the tree. The space complexity is $O(N + M^* Split)$, where $Split$ is the average number of the partitioning numbers of each splitting point.

VI. CASE STUDY

A. CASE 1

Case 1 is a simulation model of an op amp active filter circuit. The circuit is shown in Fig.5. This circuit contains

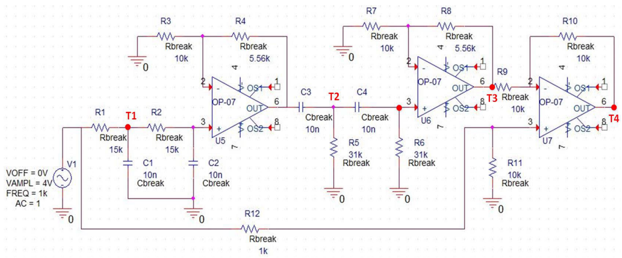


FIGURE 4. The op amp active filter circuit.

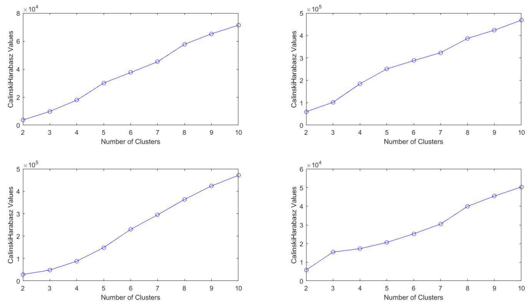


FIGURE 5. CH validity index curves of K-means clustering algorithm.

12 resistances and 4 capacitors, and we assume each resistance and capacitor has an open-circuit fault and a short-circuit fault, thus there are a total of 32 faults that can be inserted. In the simulation process, an excitation sinusoidal source with the frequency of 1 kHz and the amplitude of 4 V is loaded as the input. The component deviation tolerance is set as 5% for both the resistances and capacitors, and the lot tolerance is also set as 5%. A total number of 100 Monte Carlo simulations are carried out for each state. In each simulation, we only select the last two periods of waveform as the output voltages, which can confirm that the circuit has already reached a steady state, and gather the maximum value as the feature data of each test point.

As this paper aims to improve the ID3 algorithm and construct new decision trees, the selection of test points are not discussed specially. We finally select 4 test points in this op amp active filter circuit [42], which are marked in the Fig. 4.

We obtained 100 data for each fault state through simulation. We selected 70 data from each fault state randomly as a training dataset and the remaining 30 data as the validation dataset. Firstly, we compare 4 clustering methods, GMM clustering algorithm, K-means clustering algorithm, hierarchical clustering algorithm, Affinity Propagation (AP) clustering algorithm. We calculate CH validity indexes of different cluster numbers and select the cluster number with the largest CH validity index for each test point. The CH validity index curves of each clustering method are shown in Fig. 5 to 8.

We calculate the CH validity indexes of cluster numbers from 1 to 10. Because K-means clustering, hierarchical clustering and AP clustering are Euclidean distance-based method, the CH validity index curves of K-means clustering,

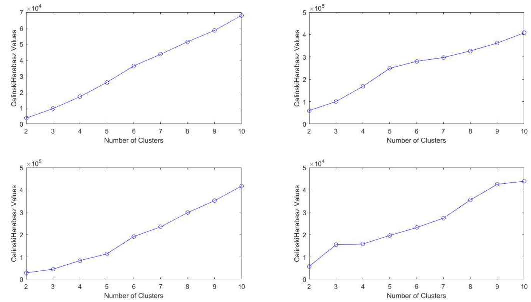


FIGURE 6. CH validity index curves of hierarchical clustering algorithm.

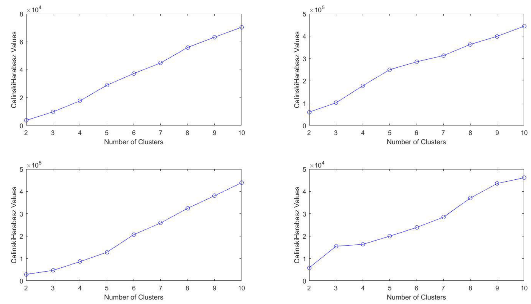


FIGURE 7. CH validity index curves of AP clustering algorithm.

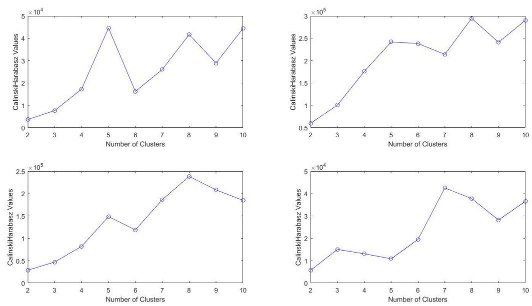


FIGURE 8. CH validity index curves of GMM clustering algorithm.

hierarchical clustering and AP clustering are monotone increasing, which results in the invalid clustering results, i.e. one sample is one cluster. For GMM clustering, the standardization distance-based method makes the CH validity index curves are ups and downs, thus we can find that the optimal cluster numbers of 4 test points are 5, 8, 8, and 7. The training dataset and the GMM clustering results of 4 test points are shown in Fig. 9. Then through the GMM clustering algorithm, we automatically obtain the mean and variance of each cluster which are then used as decision criterions. We will first compare the diagnostic efficiency of traditional DT and FR-DT. Then we will compare the diagnostic accuracy of traditional DT, CV-DT, kNN, SVM, and ANN approaches.

Firstly, we need to determine the data number ratio of 32 faults based on fault rates. We assume that the fault rates of F3 to F24, and F27 to F32 are λ. The fault rates of F1 and F2 are 6λ for unstable solder joints, and the fault rates of F25 and F26 are 8λ for high temperature. For convenience, in the calculation of information entropy and ADS, we set

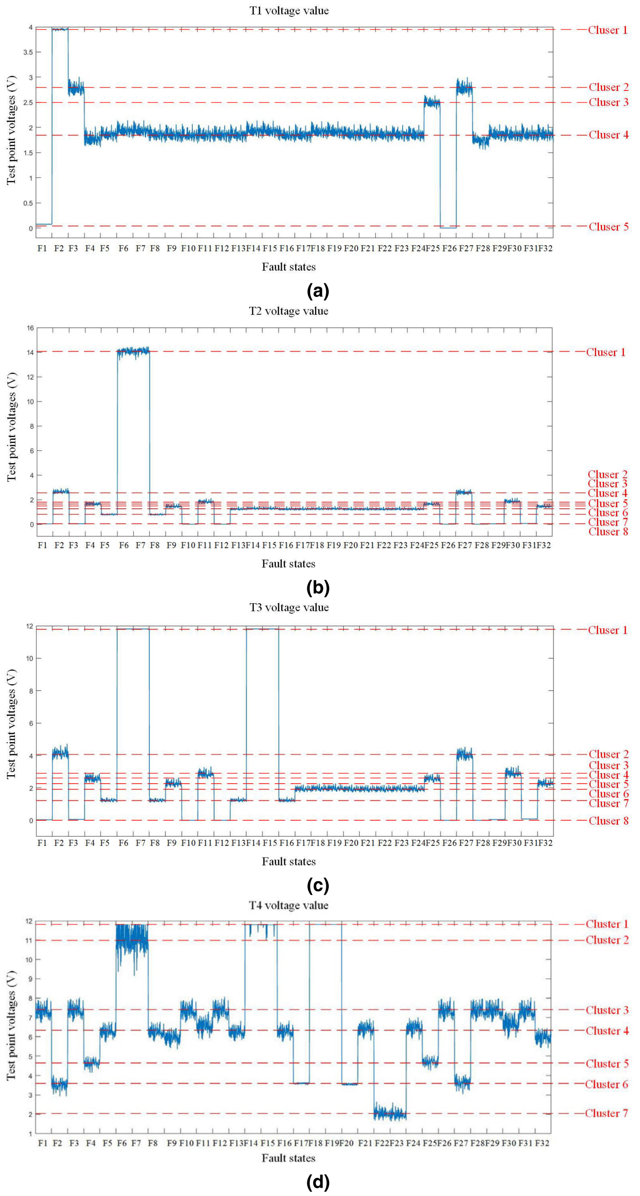


FIGURE 9. The training data and clustering results of 4 test points. (a) T1, (b) T2, (c) T3, (d) T4.

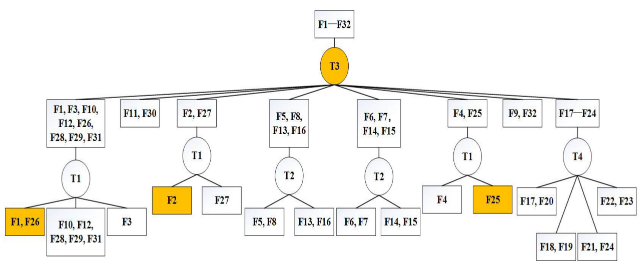


FIGURE 10. The traditional DT using the traditional ID3 algorithm.

10 fault data for F3 to F24, and F27 to F32, 60 fault data for F1 and F2, and 80 fault data for F25 and F26.

Then we use the traditional ID3, FR-ID3, and CV-ID3 algorithms respectively to construct the decision trees. These decision trees are shown in Fig. 10 to Fig. 12.

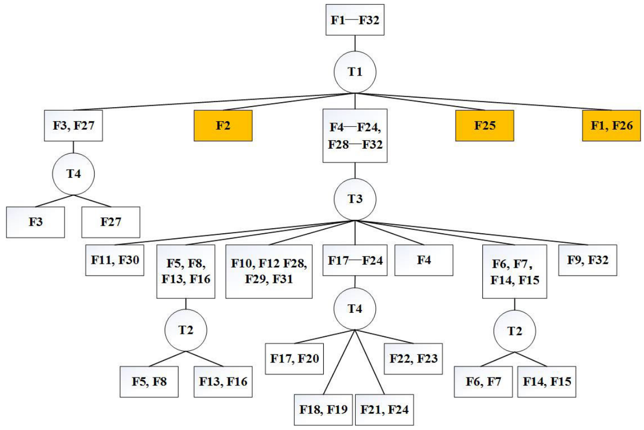


FIGURE 11. The FR-DT using the FR-ID3 algorithm.

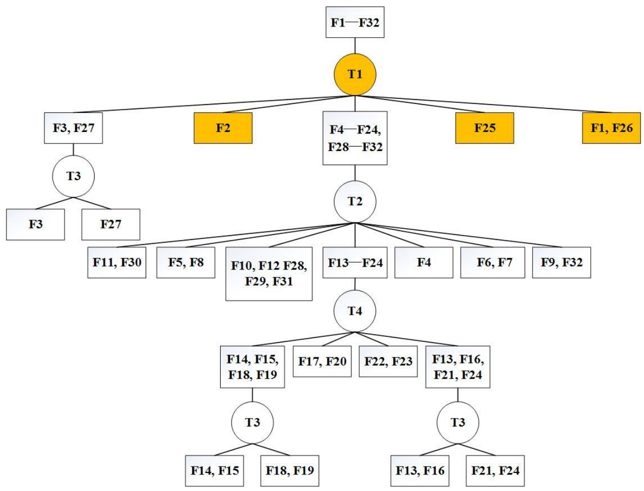


FIGURE 12. The CV-DT using the CV-ID3 algorithm.

Because traditional ID3 algorithm uses greedy strategy, the decision tree with ID3 has the largest breadth and the smallest depth. As shown in Fig. 10, the first splitting point is T3 which has the most clusters and the largest depth of this decision tree is only 2. The F1, F2, F25 and F6 which have larger fault rates are all diagnosed in second step, thus it will result in larger ADS and lower diagnostic efficiency.

Although the largest depth of FR-DT is 3 and larger than traditional DT, it avoids the local optimal in greedy strategy because the modification of information gain considers the isolation priority of the fault with a higher fault rate. As shown in Fig. 11, although the first splitting point is T1 which has only 5 clusters, F1, F2, F25 and F6 are diagnosed using only one step. Thus it will result in a smaller ADS and higher diagnostic efficiency. The ADS results of traditional DT and FR-DT are shown in Table 1. The FR-DT has the smallest ADS 100/56. The traditional DT has the largest ADS 108/56.

Because the CV-DT selects the splitting point with higher credibility, it means that this decision tree may have lower diagnostic efficiency and higher diagnostic accuracy. As shown in Fig. 11, although the information gains of T3 is the largest, since the proposed CVI of T1 is much larger than

TABLE 1. The ADS of traditional DT and FR-DT.

Steps	Traditional DT	FR-DT
1 step	F9, F11, F30, F32	F1, F2, F25, F26
2 steps	F1-F8, F10, F12-F29, F31	F3, F4, F9, F10, F11, F12, F27-F32
3 steps	null	F5-F8, F13-F24
4 steps	null	null
Total steps	4×1+52×2=108	28×1+12×2+16×3=100
ADS	108/56	100/56

TABLE 2. The different fault classes.

Class	Circuit states	Class	Circuit states
C1	F1, F26	C2	F2
C3	F3	C4	F4
C5	F5, F8	C6	F6, F7
C7	F9, F32	C8	F10, F12, F28, F29, F31
C9	F11, F30	C10	F13, F16
C11	F14, F15	C12	F17, F20
C13	F18, F19	C14	F21, F24
C15	F22, F23	C16	F25
C17	F27		

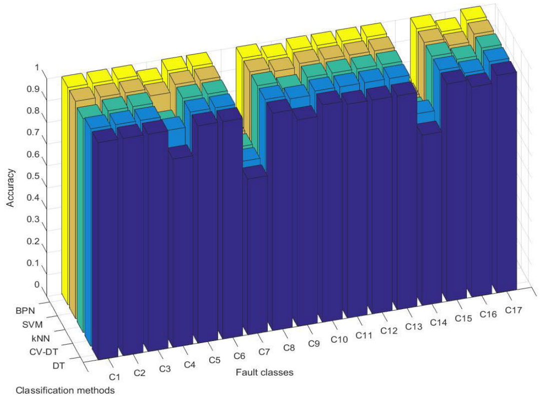


FIGURE 13. The diagnostic accuracy results of DT, CV-DT, kNN, SVM and BPN.

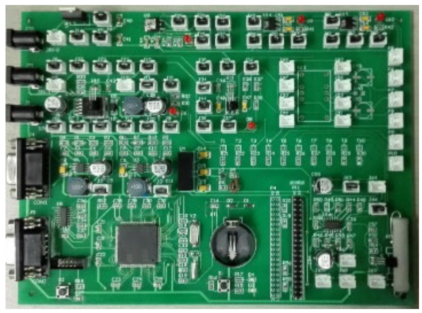


FIGURE 14. The power supply circuit board.

others, the first splitting point is T1. In addition, the largest depth of CV-DT is 4, which also means that the improvement of accuracy may be at the expense of efficiency. Thus, this decision tree has the smallest breadth and the largest depth.

To verify the high diagnostic accuracy of CV-DT, we also compare with commonly used multiple classification methods, such as kNN, SVM, and ANN approaches. In kNN classification, we use the cross validation method to determine the optimal k value. In SVM multiple classification, we select one-versus-one classification method and use the cross validation method to determine the parameters of the Gaussian kernel. In ANN classification, we use the BP network with three layers and also use the cross validation method to determine the number of hidden nodes.

Then we use a validation dataset which consists of 30 pieces data from each fault state to compare the diagnostic accuracy, we count the number of data which are classified correctly in each fault and calculate the diagnostic accuracy, as shown in Fig. 13.

where C1 to C17 are the decision results, which represent the single faults or fault ambiguity groups, as shown in Table 2.

The average diagnostic accuracies of traditional DT, CV-DT, kNN, SVM, and BPN are 95.73%, 96.77%, 96.14%, 96.56%, and 96.56% respectively. Obviously, CV-DT not only has a higher diagnostic accuracy than traditional DT, but also has the highest diagnostic accuracy among commonly used multiple classification approaches.

As shown in Fig. 13, the common diagnostic errors of DT and CV-DT are mainly located in C4, C7, and C14, which resulted from a high degree of data intersection in T2 and T3 under the condition of component tolerances. In C11, C12 and C13, CV-DT avoids the misclassification resulted from the first splitting point T3 which has lower CH index.

For kNN, SVM, and BPN, they consider that 4 test points provides information with equal importance, thus the test point which has confused information will decrease the diagnostic accuracy. However, because the order of the test points are optimized by CH index in CV-DT, the information importance of each test point is set different. The first selected splitting test point has the more importance and larger weight and the last selected splitting test point has the least importance and smallest weight. In this condition, the effect on the misclassification of the test point that has confused information can be weakened.

B. CASE 2

Case 2 is a power supply board, as shown in Fig. 14. The board consists of 6 power supply conversion circuit, including a 28V to 12V circuit, a 12V to 5V circuit, a 5V to 3.3V circuit, a 3.3V to 2.5V circuit, a 3.3V to 1.8V circuit, and a 3.3V to 0.9V circuit. We set 6 test points to measure 6 voltage outputs corresponding to 6 power supply conversion circuits. We used the pluggable type fault injection method to inject 15 open circuit faults containing single faults and compound faults. The schematic circuit diagrams are shown in Fig. 15, and the 15 faults are described in Table 3.

Because the effect of a fault may cover the effects of other faults, many compound faults are excluded. For example,

TABLE 3. The 15 injected faults.

Fault	Circuit states
F1	Input voltage open in the 5V to 3.3V circuit
F2	ground terminal open in the 12V to 5V circuit
F3	ground terminal open in the 28V to 12V circuit
F4	Input voltage open in the 3.3V to 1.8V circuit
F5	Output voltage open in the 5V to 3.3V circuit
F6	ground terminal open in the 5V to 3.3V circuit
F7	Inhibit terminal open in the 3.3V to 0.9V circuit
F8	Voltage output adjust terminal open in the 3.3V to 0.9V circuit
F9	F1 and F2 happen simultaneously
F10	F2 and F4 happen simultaneously
F11	F2 and F5 happen simultaneously
F12	F2 and F6 happen simultaneously
F13	F4 and F5 happen simultaneously
F14	F4 and F6 happen simultaneously
F15	F5 and F6 happen simultaneously

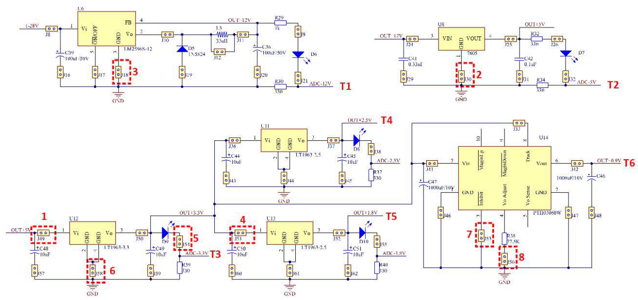


FIGURE 15. The schematic circuit diagrams, faults injection locations, and 6 test points.

when F1 happens, i.e. input voltage is open in the 5V to 3.3V circuit, then the 3.3V to 2.5V circuit, the 3.3V to 1.8V circuit, and the 3.3V to 0.9V circuit will not work and we cannot observe the circuits states, whether normal states or faulty states. Therefore, we obtain 15 groups of fault data and each group contains 1000 data. We select 700 data from each group randomly as the training dataset and the remaining 300 data as the validation dataset. The training dataset and the GMM clustering results of 6 test points are shown in Fig. 16. The optimal cluster numbers of 6 test points are 2, 3, 4, 3, 2, and 3.

We assume that the fault rates of F1, F2, F4, F5, F6, F7, and F8 are λ, and the fault rate of F3 is 4λ for the high temperature in the 28V to 12V circuit. We assume that each single fault is independent and the compound fault happens only when two single faults happen simultaneously. Therefore, according to the reliability parallel model, the fault rate of each compound fault λ_C satisfies $\frac{1}{\lambda}+\frac{1}{\lambda}=\frac{1}{\lambda_C}$, i.e. $\lambda_C=\frac{\lambda}{2}$. For convenience, we set 10 fault data for F9 to F15, 20 fault data for F1, F2, F4, F5, F6, F7, and F8, and 80 fault data for F3.

Then we use the traditional ID3, FR-ID3, and CV-ID3 algorithms respectively to construct the decision trees. These trees are shown in Fig. 17 to Fig. 19.

The ADS results of the traditional DT and FR-DT is shown in Table 4 and the diagnostic accuracy results of the traditional DT, CV-DT, kNN, SVM, and BPN are shown in Fig. 20. In this case, we also use cross validation method to find the optimal parameters of each classification model.

TABLE 4. The ADS of Traditional DT and FR-DT.

DT	Traditional DT	FR-DT
ADS	80/29	76/29

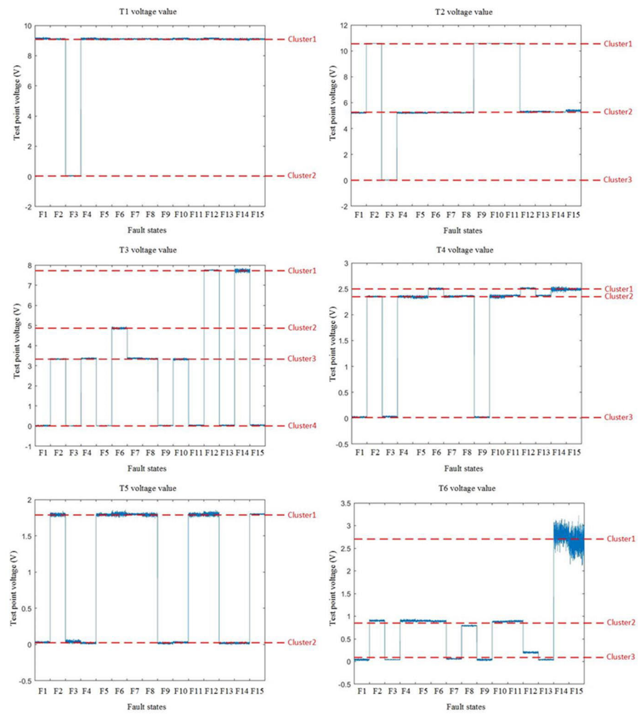


FIGURE 16. The training data and clustering results of 6 test points.

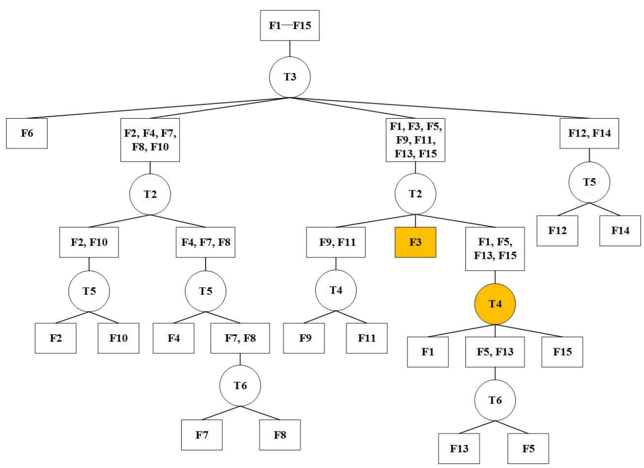


FIGURE 17. The traditional DT using the traditional ID3 algorithm.

The average diagnostic accuracies of traditional DT, CV-DT, kNN, SVM, and BPN are 98%, 100%, 98.89%, 98.89%, and 99.56% respectively. Obviously, CV-DT has the highest diagnostic accuracy 100%. Although the diagnostic accuracy of CV-DT is only 2% higher than DT and 0.44% higher than BPN, we think the accuracy increasing from 99.56% to 100% is more meaningful.

As shown in Fig. 20, the common diagnostic errors of these multiple classification approaches are mainly located

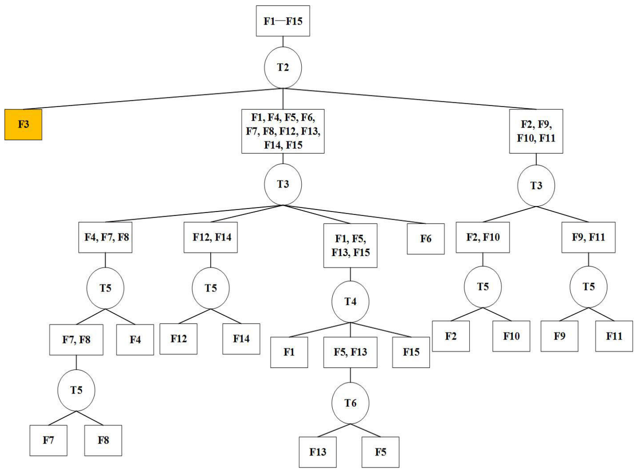


FIGURE 18. The FR-DT using the FR-ID3 algorithm.

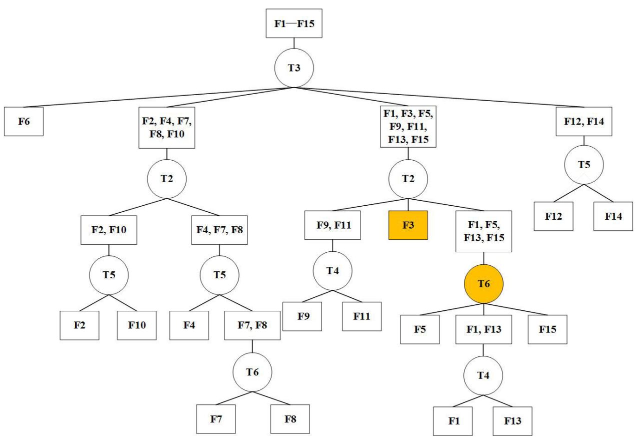


FIGURE 19. The CV-DT using the CV-ID3 algorithm.

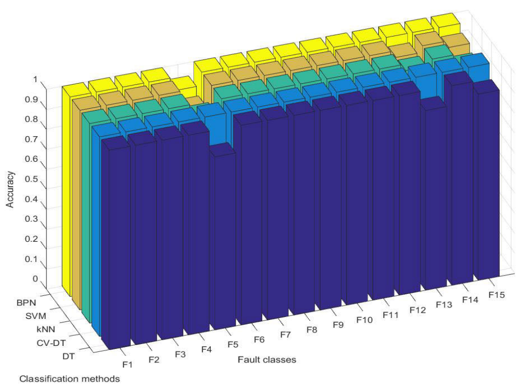


FIGURE 20. The diagnostic accuracy results of DT, CV-DT, kNN, SVM and BPN.

in F5, F13, and F15, which resulted from a high degree of data intersection in T4 under the condition of component tolerances. The CV-DT only uses the T4 information as the splitting point of F1 and F13 in the last step. On the other hand, F5, F13, and F15 are distinguished by T6 before the use of T4. Therefore, CV-DT can avoid the misclassification by the confused information of T4 completely.

VII. CONCLUSION

Focusing on the two drawbacks of traditional decision trees with the ID3 algorithm, this paper proposes two improved decision trees: CV-DT and FR-DT for analog circuit fault diagnosis. These two decision trees have two highlights. The first one is to propose a CV-DT which is constructed by an improved ID3 algorithm considering the cluster validity index. We propose a new CVI which can compare the cluster validities of different attributes. The larger CVI demonstrates that the partition has the higher credibility. We use CVI to modify the information gain, and this method increases the diagnostic accuracy. The second one is to propose an FR-DT which is constructed by an improved ID3 algorithm considering the fault rates. This algorithm divides the information gain into two aspects. One of the information gain entropies is based on the number of clusters and considers the partitioning ability of each attribute. The other is based on the fault rates and considers the isolation priority of faults with a higher fault rate. This method decreases the ADS and promotes the diagnostic efficiency.

A simulation case and a real board case are firstly given to compare the diagnostic efficiency of traditional DT and FR-DT, then compare the diagnostic accuracy among traditional DT, CV-DT, kNN, SVM, and BPN approaches. The two cases both show that the FR-DT promotes the diagnostic efficiency than traditional DT, and the CV-DT promotes the diagnostic accuracy compared with other multiple classification methods.

Our methods work well only when fault data can be distinguished by the hyperplanes that are vertical or parallel to any axis in space. When the fault data have a complex space structure, we may need to use oblique hyperplanes to distinguish them as our proposed methods are invalid. In our future work, we need to study the selection of splitting points which combine two or more attributes, or we need to study data preprocessing methods and extract attributes with a higher classification performance.

In addition, based on the decision tree method, the random forest algorithm is integrating a lot of decision trees for classification. Based on the proposed modified FR-DT and CV-DT, when each decision tree has more accurate or efficient diagnostic results, the diagnostic performance of the random forest is also increasing. In our future work, we will study the random forest algorithm considering fault rate or cluster validity.

APPENDIX

For C4.5 algorithm, the information gain ratio is shown in the form

$$IGR\left(X,A_j\right)=\frac{IG\left(X,A_j\right)}{SI\left(X,A_j\right)}=\frac{Entropy\left(X\right)-Entropy\left(X,A_j\right)}{SI\left(X,A_j\right)}$$

$$=\frac{\sum\limits_{n=1}^{N}\frac{|C_n|}{N_T}\log\frac{|C_n|}{N_T}-\sum\limits_{k=1}^{r_j}\frac{|S_k|}{N_T}\left(\sum\limits_{n=1}^{N}\frac{|C_{kn}|}{|S_k|}\log\frac{|C_{kn}|}{|S_k|}\right)}{\sum\limits_{k=1}^{r_j}\frac{|S_k|}{N_T}\log\frac{|S_k|}{N_T}}$$

We unfold the expression of the information gain, we will obtain

$$\begin{aligned} &IG\left(X, A_j\right) \\ &= \sum_{n=1}^{N} \frac{|C_n|}{N_T} \log \frac{|C_n|}{N_T} - \sum_{k=1}^{r_j} \frac{|S_k|}{N_T} \left(\sum_{n=1}^{N} \frac{|C_{kn}|}{|S_k|} \log \frac{|C_{kn}|}{|S_k|} \right) \\ &= \sum_{n=1}^{N} \log \frac{|C_n|}{N_T} \sum_{k=1}^{r_j} \frac{|C_{kn}|}{N_T} - \sum_{k=1}^{r_j} \sum_{n=1}^{N} \frac{|C_{kn}|}{N_T} \log \frac{|C_{kn}|}{|S_k|} \\ &= \sum_{k=1}^{r_j} \sum_{n=1}^{N} \frac{|C_{kn}|}{N_T} \left(\log \frac{|C_n|}{N_T} - \log \frac{|C_{kn}|}{|S_k|} \right) \end{aligned}$$

In general, one category may contain more than one value in an attribute, thus $|S_k| = \sum_{n=1}^{N} |C_{kn}| \leq \sum_{n_k} \left|C_{n_k}\right|$ and $\left|C_{kn_k}\right| \leq \left|C_{n_k}\right|$, obviously, the information gain (IG) is different from the split information (SI). However, in this paper, we only consider that the circuit is working in the single condition and each fault only contains the single class of value in a test point, thus $|S_k| = \sum_{n=1}^{N} |C_{kn}| = \sum_{n_k} \left|C_{n_k}\right|$, $\left|C_{kn_k}\right| = \left|C_{n_k}\right|$, and $|C_{kn}| = 0\,(n \neq n_k)$ in this case,

$$\begin{aligned} IG\left(X, A_j\right) &= \sum_{k=1}^{r_j} \sum_{n=1}^{N} \frac{|C_{kn}|}{N_T} \left(\log \frac{|C_n|}{N_T} - \log \frac{|C_{kn}|}{|S_k|} \right) \\ &= \sum_{k=1}^{r_j} \sum_{n_k} \frac{\left|C_{n_k}\right|}{N_T} \left(\log \frac{\left|C_{n_k}\right|}{N_T} - \log \frac{\left|C_{n_k}\right|}{|S_k|} \right) \\ &= \sum_{k=1}^{r_j} \sum_{n_k} \frac{\left|C_{n_k}\right|}{N_T} \log \frac{|S_k|}{N_T} \\ &= \sum_{k=1}^{r_j} \frac{|S_k|}{N_T} \log \frac{|S_k|}{N_T} = SI\left(X, A_j\right) \end{aligned}$$

Obviously, the information gain (IG) is equal to the split information (IG) for each splitting test points.

ACKNOWLEDGMENT

The authors would like to thank the editor and reviewers for their insightful comments and suggestions, which helped improve the paper significantly. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] J. Cui and Y. Wang, "A novel approach of analog circuit fault diagnosis using support vector machines classifier," *Measurement*, vol. 44, no. 1, pp. 281–289, Jan. 2011.
[2] H. Luo, Y. Wang, H. Liu, and Y. Jiang, "Module level fault diagnosis for analog circuits based on system identification and genetic algorithm," *Measurement*, vol. 45, no. 4, pp. 769–777, May 2012.
[3] H. Luo, W. Lu, Y. Wang, L. Wang, and X. Zhao, "A novel approach for analog fault diagnosis based on stochastic signal analysis and improved GHMM," *Measurement*, vol. 81, pp. 26–35, Mar. 2016.
[4] A. Zhang, C. Chen, and B. Jiang, "Analog circuit fault diagnosis based UCISVM," *Neurocomputing*, vol. 173, pp. 1752–1760, Jan. 2016.
[5] P. Chen, L. Yuan, Y. He, and S. Luo, "An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis," *Neurocomputing*, vol. 211, pp. 202–211, Oct. 2016.
[6] H. Luo and C. Y. W. Jiang, "A SVDD approach of fuzzy classification for analog circuit fault diagnosis with FWT as preprocessor," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10554–10561, Aug. 2011.
[7] G. Xu-Sheng, G. Wen-Ming, D. Zhe, and L. Wei-Dong, "Research on WNN soft fault diagnosis for analog circuit based on adaptive UKF algorithm," *Appl. Soft Comput.*, vol. 50, pp. 252–259, Jan. 2017.
[8] D. Binu and B. S. Kariyappa, "A survey on fault diagnosis of analog circuits: Taxonomy and state of the art," *AEU-Int. J. Electron. Commun.*, vol. 73, pp. 68–83, Mar. 2017.
[9] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
[10] J. R. Quinlan, *Programs for Machine Learning*, San Mateo, CA, USA: Morgan Kaufmann, 1993.
[11] M. Jaworski, P. Duda, and L. Rutkowski, "New splitting criteria for decision trees in stationary data streams," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2516–2529, Jun. 2018.
[12] R. Rivera-Lopez and J. Canul-Reich, "Construction of near-optimal axis-parallel decision trees using a differential-evolution-based approach," *IEEE ACCESS*, vol. 6, pp. 5548–5563, 2018.
[13] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.
[14] N. Li, S. Guo, and Y. Wang, "Weighted preliminary-summation-based principal component analysis for non-Gaussian processes," *Control Eng. Pract.*, vol. 87, pp. 122–132, Jun. 2019.
[15] Z. Sun, K. Hu, T. Hu, J. Liu, and K. Zhu, "Fast multi-label low-rank linearized SVM classification algorithm based on approximate extreme points," *IEEE Access*, vol. 6, pp. 42319–42326, 2018.
[16] L. Wang, B. Yang, Y. Chen, X. Zhang, and J. Orchard, "Improving neural-network classifiers using nearest neighbor partitioning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2255–2267, Oct. 2017.
[17] H. Hirose, M. Hikita, S. Ohtsuka, S.-I. Tsuru, and J. Ichimaru, "Diagnosis of electric power apparatus using the decision tree method," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 15, no. 5, pp. 1252–1260, Oct. 2008.
[18] S. R. Samantaray, "Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line," *IET Generation, Transmiss. Distribution*, vol. 3, no. 5, pp. 425–436, May 2009.
[19] A. Jamehbozorg and S. M. Shahrtash, "A decision tree-based method for fault classification in double-circuit transmission lines," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2184–2189, Oct. 2010.
[20] A. Swetapadma and A. Yadav, "A novel decision tree regression-based fault distance estimation scheme for transmission lines," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 234–245, Feb. 2017.
[21] K. Kyoungok, "A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree," *Pattern Recognit.*, vol. 60, pp. 157–163, Dec. 2016.
[22] H. K. Sok, M. P.-L. Ooi, Y. C. Kuang, and S. Demidenko, "Multivariate alternating decision trees," *Pattern Recognit.*, vol. 50, pp. 195–209, Feb. 2016.
[23] X. Liu, Q. Li, T. Li, and D. Chen, "Differentially private classification with decision tree ensemble," *Appl. Soft Comput.*, vol. 62, pp. 807–816, Jan. 2018.
[24] Q. Hu, X. Che, L. Zhang, D. Zhang, M. Guo, and D. Yu, "Rank entropy-based decision trees for monotonic classification," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2052–2064, Nov. 2012.
[25] S. Pei, Q. Hu, and C. Chen, "Multivariate decision trees with monotonicity constraints," *Knowl.-Based Syst.*, vol. 112, pp. 14–25, Nov. 2016.
[26] A. Segatori, F. Marcelloni, and W. Pedrycz, "On distributed fuzzy decision trees for big data," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 174–192, Feb. 2018.
[27] Y. Cui, J. Shi, and Z. Wang "Analog circuit fault diagnosis based on quantum clustering based multi-valued quantum fuzzification decision tree (QC-MQFDT)," *Measurement*, vol. 93, pp. 421–434, Nov. 2016.
[28] Y. Cui, J. Shi, and Z. Wang "Analog circuits fault diagnosis using multi-valued Fisher's fuzzy decision tree (MFFDT)," *Int. J. Circuit Theory Appl.*, vol. 44, no. 1, pp. 240–260, Jan. 2016.
[29] D. Xu and Y. A. Tian "A comprehensive survey of clustering algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, Jun. 2015.

[30] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, Vol. 2, Jun. 1999, pp. 246–252.

[31] Y.-M. Cha and J.-H. Han, "High-accuracy retinal layer segmentation for optical coherence tomography using tracking kernels based on Gaussian mixture model," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 2, pp. 32–41, Mar./Apr. 2014.

[32] B. A. Akram, A. H. Akbar, and O. Shafiq, "HybLoc: Hybrid indoor Wi-Fi localization using soft clustering-based random decision forest ensembles," *IEEE Access*, vol. 6, pp. 38251–38272, 2018.

[33] Y. Zhao, A. K. Shrivastava, and K. L. Tsui, "Regularized Gaussian mixture model for high-dimensional clustering," *IEEE Trans. Cybern.*, vol. 49, no. 10, pp. 3677–3688, Oct. 2019.

[34] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 769–779, May 2014.

[35] X. Hou, T. Zhang, G. Xiong, Z. Lu, and K. Xie, "A novel steganalysis framework of heterogeneous images based on GMM clustering," *Signal Process., Image Commun.*, vol. 29, no. 3, pp. 385–399, Mar. 2014.

[36] J. Zhang, Z. Yin, and R. Wang, "Pattern classification of instantaneous cognitive task-load through GMM clustering, Laplacian eigenmap, and ensemble SVMs," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 947–965, Jul./Aug. 2017.

[37] S.-D. Oh, Y.-J. Kim, and J.-S. Hong, "Urban traffic flow prediction system using a multifactor pattern recognition model," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2744–2755, Oct. 2015.

[38] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, Jan. 2013.

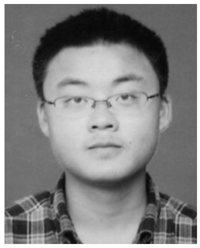
[39] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, Oct. 2018.

[40] H. Li, J. Sun, D. Meng, and Q. Zhang, "A multiobjective approach based on Gaussian mixture clustering for sparse reconstruction," *IEEE Access*, vol. 7, pp. 22684–22697, 2019.

[41] R. C. Barros, M. P. Basgalupp, A. C. P. L. F. de Carvalho, and M. G. Quiles, "Clus-DTI: Improving decision-tree classification with a clustering-based decision-tree induction algorithm," *J. Brazilian Comput. Soc.*, vol. 18, no. 4, pp. 351–362, Apr. 2012.

[42] Y. Cui, J. Shi, and Z. Wang, "Analog circuit test point selection incorporating discretization-based fuzzification and extended fault dictionary to handle component tolerances," *J. Electron. Test.*, vol. 32, no. 6, pp. 661–679, Dec. 2016.

JUNYOU SHI received the Ph.D. degree from the School of Reliability and Systems Engineering, Beihang University, Beijing, China, in 2004, where he is currently an Associate Professor. His main research interests include system testability, system reliability, and prognostics and health management (PHM).



QINGJIE HE received the B.E. degree from the School of Reliability and Systems Engineering, Beihang University, Beijing, China, in 2010, where he is currently pursuing the Ph.D. degree. His main research interests include system testability, fault diagnosis, and prognostics and health management (PHM).

ZILI WANG received the M.S. degree from the School of Reliability and Systems Engineering, Beihang University, Beijing, China, in 1988, where he is currently a Professor. His research interests include reliability theory and practices, system engineering, and prognostics and health management (PHM).