Induction Motor Fault Classification Based on ROC Curve and t-SNE

CHUN-YAO LEE, (Member, IEEE), AND **WEN-CHENG LIN**
Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan
Corresponding author: Chun-Yao Lee (cyl@cycu.edu.tw)

ABSTRACT This paper proposes a novel fault classification method with application to induction motors, which is based on integrating and combining with receiver operating characteristic (ROC) curve and t-distribution stochastic neighbor embedding (t-SNE). According to the feature selection methods of ReliefF, symmetrical uncertainty (SU), and fast correlation-based filter (FCBF), the significant features were verified. Additionally, support vector machine (SVM), k-nearest neighbor (KNN), and decision tree (DT) are also considered as classifiers to identify the simulation results. To begin with, the current signals obtained from distinctive four topologies of working conditions of the motor, which includes healthy, bearing damage, broken rotor bar, and short circuit in stator windings, respectively. The potential feature set is extracted by using Hilbert-Huang transform (HHT) technique. Then, three feature selection methods are adopted to select three optimal feature subsets from the original feature set. Finally, the classification accuracy (ACC) and ROC curve are used to demonstrate the capability of classifiers' recognition. The results showed that the optimal feature subsets significantly reduce the number of selected features and improve the classification ACC and area under the curve (AUC) compared with the original feature set. In conclusion, the proposed method can downgrade the data, demonstrate the scatter plot more intuitively, and identify various types of faults, unlike with other fault diagnosis literature.

INDEX TERMS ROC, t-SNE, motor failure, ReliefF, SU, FCBF, feature selection, HHT.

I. INTRODUCTION

In the industrial age, automated production models have become mainstream. Electric motors are the primary source of power for manufacturing. Its stable operation is considered a desirable part of the production line. Because of downtime, safety considerations, and costly machinery repair early detection of motor's internal faults is highly important [1]. In the age of unattended factories, how to effectively detect and identify any abnormalities, predict potential failures, and implement management to minimize performance degradation and economic costs to avoid dangerous situations is necessary [2]. Data-driven based intelligent fault pattern recognition methods have made fruitful achievements in recent years [3]. As far as induction motors are concerned, they can normally work in harsh environments such as high temperatures, high dust, water (dedicated motors), and frequency converters can change torque and power, which is economical. In harsh conditions, it has been widely used in industrial applications; however, some faults may lead to their

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

failure and economic losses [4]. Therefore, this study will discuss the more common types of faults in induction motors. The measurement includes four conditions: healthy, bearing damage (45%), stator (35%), and rotor (10%) [5], [6]. The signal usually used for measurement is mainly the current signal, that is, the electric signal. Compared with vibration and temperature signals, it is less affected [7]. Therefore, this article chooses to use the current signal for measurement to facilitate subsequent analysis. After that, the three typical classifiers were used to judge the accuracy (ACC) of the features under different sets to obtain relevant recognition results.

The intelligent fault diagnosis method's performance depends on the feature extraction of fault signals, which requires signal processing techniques, human knowledge, and labor [8]. In the last few decades, the types of signal analysis are quite diverse. From the earlier proposed fast Fourier transform (FFT), wavelet transform (WT), and then to Hilbert-Huang transform (HHT), they all have their distinct advantages and drawbacks. However, HHT is well suited to processing non-linear, non-stationary signals and is not constrained by the assumptions of stationarity and

linearity, required for the FFT, and generates both amplitude and frequency information as a function of time [9], [10]. Its advantage lies in using the basis of a posterior definition and has better noise immunity [11]. Through using empirical mode decomposition (EMD) to decompose the signal. Each intrinsic mode function (IMF) can hold specific frequency information to capture useful features. In terms of FFT, data needs to have periodicity. The selection of wavelet functions required for WT conversion is also more complex [12]. FFT and Discrete Wavelet Transform (DWT) are also not suitable for load imbalance and asynchronous sampling, which will result in failure fault identifiers and incorrect fault type [13].

Several works addressed extracting features through signal analysis methods for induction motor fault data. But the type, quantity, and impact of features cannot be quantified. Large-scale features can easily cause system size problems and may use excessive storage space. Also, the existing redundant features can easily cause excessive calculations. In order to solve these problems, the selection of features becomes more and more critical. Using typical methods such as ReliefF, symmetrical uncertainty (SU) and fast correlation-based filter (FCBF) for feature selection can undoubtedly reduce the complexity of calculation and database size so that the algorithm will not change the original features [14], [15]. Compared with the other categories, feature extraction projects important features to facilitate visual observation and can reorganize subspaces and retain the original space's data structure. Feature extraction also plays a vital role in data-driven fault diagnosis and dimensionality reduction for the samples or datasets [16]. Nevertheless, the principal component analysis (PCA) is commonly used to find the main components of the original data and establishes a direct relationship between the high and low dimensional data sets, but it cannot capture the non-linear pattern [17]. Therefore, this research uses the t-distribution stochastic neighbor embedding (t-SNE) to reduce the dimensionality of nonlinear data, which can visualize high-dimensional complex signal patterns. Compared with the original SNE, it uses t-distribution to solve the probability distribution in low-dimensional situations to alleviate the data crowding problem. Simultaneously, this method uses joint probability instead of conditional probability to recalculate KLD to obtain symmetry [18], [19].

For current fault detection, ACC is usually used to compare the results of different races. ACC takes the merit of simple structure, but it has drawbacks of prone to inaccurate models due to data skew. This study also uses the receiver operating characteristic (ROC) curve to compare results. ROC curve is a visual tool for classification models [20], [21]. It originated from the signal detection theory. In recent years, this method distinguishes between negative and positive results by dividing race into two categories and comparing them with ACC. By plotting true positive rate (*TPR*) and false positive rate (*FPR*) curves, a more satisfactory classification can be obtained [22], [23], and the area under the curve (AUC) can be calculated to make it relatively balanced. These two indicators, they also do not depend on the impact of a particular category.

This research shows the effectiveness of traditional methods in dimensionality reduction, feature extraction and feature selection. Create a novel model with the combination of advantages of each method. By comparing these methods and establishing the entire system, the important features of the motor and a better recognition rate can be effectively obtained. Among them, the main contributions of this paper are fourfold. First, HHT has better recognition results than WT and FFT. Secondly, the problem of large-scale features is studied. It turns out that three common feature selection methods can be used to select important features. Third, it is proposed to use the ROC curve as a reference basis and compare AUC with ACC. Moreover, through the support vector machine (SVM), K nearest neighbor (KNN) and decision tree (DT) three machine learning algorithms to compare their classification performance to select the most effective fault diagnosis model. Finally, based on the advantages and disadvantages of feature extraction and feature selection methods, this research achieves the system's robustness through more graphical visualization methods such as t-SNE. Compared with other studies on motor fault classification methods, the new intuitive visualization method can effectively verify the advantages of important features and recognition rate by combining known methods.

II. METHOD OF MEASURING MOTOR
This section will explain the specifications of AC induction motors, and measure and analyze 4 types of current signals including normal, bearing damage, broken rotor bar and short circuit in stator windings. Secondly, introduce the equipment and methods used in the experiment and the overall process of this research to compare the differences between various types of faults and normal motors. Finally, the results of the identification are presented by using the analysis software MATLAB.

A. EQUIPMENT SPECIFICATIONS
The main equipment in this study are four-pole AC induction motors, as shown in Table 1, and the fault types are shown in Fig. 1. By driving the power platform (composed of a torque sensor and a servo motor), and analyzing it with the equipment (NI PXI-1,033) and a computer, then recording the measured data. Through the above equipment, the signal measurement can be completed.

B. EXPERIMENT PROCESS
First of all, this research measures the current signal of AC induction motors in four kinds of fault (normal, bearing, rotor, and stator). Obtain any phase data of the motor U, V, and W through a signal extractor. It is noted that the data sampling time for each measurement is 100 seconds, the sampling frequency is 1,000 Hz, and each signal is measured 100 times for evaluation. The process is shown in Fig. 2.

TABLE 1. Specifications of AC motor.

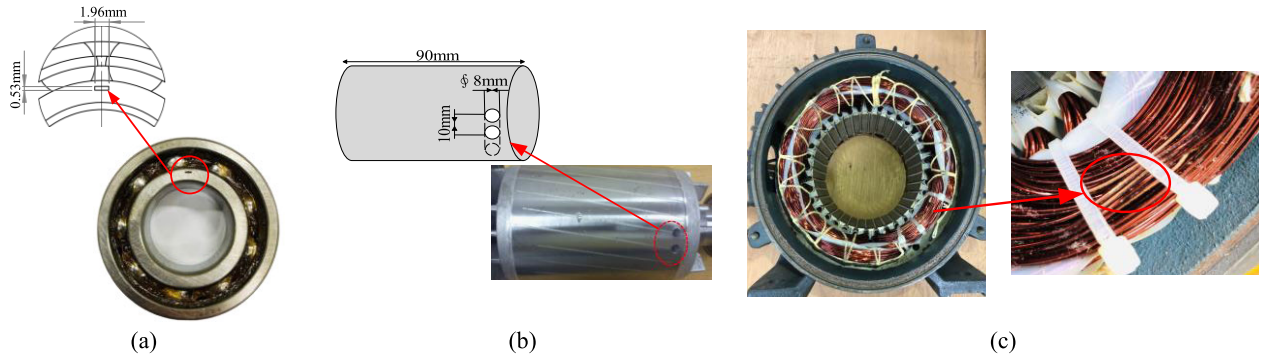| Specification of three-phase four-pole squirrel cage induction motor					
Voltage	220V/380V	Frequency	60Hz	Effectiveness	83.5(100%)
Output	2Hp 1.5kW	Current	5.58A/3.23A		

(a)　　　　　　　　　　　　　(b)　　　　　　　　　　　　　(c)

FIGURE 1. Type of motor fault (a) bearing damage (aperture: 1.96 mm × 0.53 mm) (b) broken rotor bar (2 holes ∮8 mm deep: 10 mm) (c) short circuit in stator windings (2 coil short circuit).

AC servo motor　　　　　Induction motor　　　　　Acquisition of equipment　　　　　PC

FIGURE 2. Flow chart of signal measurement.
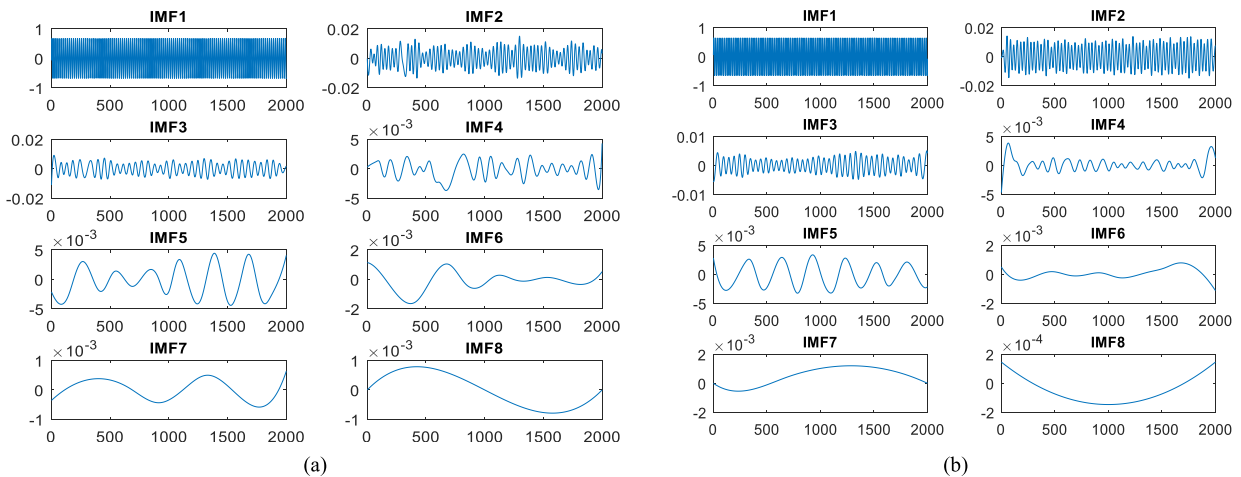

(a)　　　　　　　　　　　　　　　　　　　　　　(b)

FIGURE 3. The illustration of IMF components (a) normal motor (b) bearing damage motor.

Secondly, HHT was used on the personal computer for MATLAB. Among them, the waveform, vibration, and frequency of each IMF were different. Additionally, the purpose of screening each layer of IMF is not only to eliminate the carrier but also can make the waveform more symmetrical. In order for the IMF decomposed by EMD to retain the meaning of its signal, the screening criteria must be set to determine the number of screening levels. This action will stop when the standard deviation (SD) of the two consecutive screening results is less than 0.1. In this research, IMF of 1 to 8 layers can be obtained by EMD. The following takes normal and bearing damage fault motor signals as examples, and the extracted results are shown in Fig. 3(a) and Fig. 3(b). Meanwhile, the instantaneous amplitude and instantaneous

TABLE 2. The feature dataset.

|	Fmax	Fmin	Fmean	Fmse	Fstd	Amax	Amin	Amean	Amse	Astd
IMF1	F1	F2	F3	F4	F5	F41	F42	F43	F44	F45
IMF2	F6	F7	F8	F9	F10	F46	F47	F48	F49	F50
IMF3	F11	F12	F13	F14	F15	F51	F52	F53	F54	F55
IMF4	F16	F17	F18	F19	F20	F56	F57	F58	F59	F60
IMF5	F21	F22	F23	F24	F25	F61	F62	F63	F64	F65
IMF6	F26	F27	F28	F29	F30	F66	F67	F68	F69	F70
IMF7	F31	F32	F33	F34	F35	F71	F72	F73	F74	F75
IMF8	F36	F37	F38	F39	F40	F76	F77	F78	F79	F80
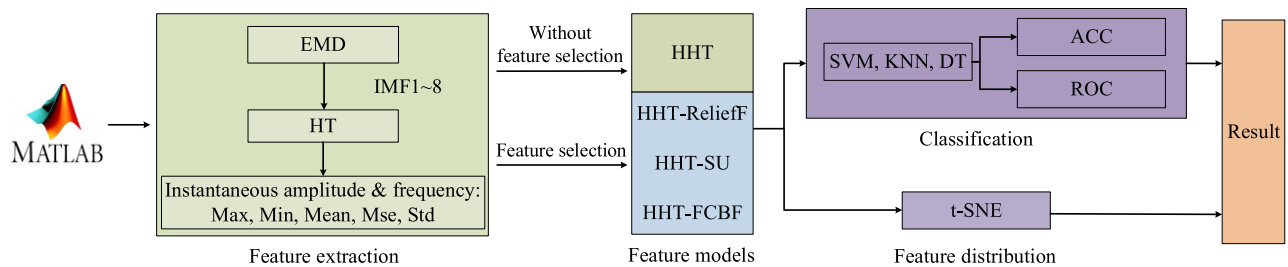
FIGURE 4. The experiment process and result framework diagram.

frequency of each layer can be obtained by HT. After extracting the maximum, minimum, average, standard deviation and root mean square of each layer's instantaneous amplitude and instantaneous frequency, a total of 80 features can be got, as shown in Table 2. Then, the common feature selection methods of ReliefF, SU value, and FCBF are used to generate a total of 4 different sizes of feature sets for identification. In order to prove that the selected feature set can produce better recognition ability under any classifier, the SVM, KNN, and DT are used to generate three classification results for verification. Finally, this study uses t-SNE to transform the features and present them in two and three dimensions. So that the results can be observed in a more intuitive way to prove the feature selection has reference value for the identification of the current signal of the motor.

In short, the entire experiment uses feature extraction and feature selection methods to obtain 4 different size feature sets and then uses common classifiers and feature distributions to present the research results. The process is shown in Fig. 4. The steps of signal processing are listed as follows:

Step 1: Input the current signal of the induction motor, and process the signal through MATLAB software.
Step 2: Through EMD, the signal can be decomposed into 1 to 8 layers of IMF.
Step 3: Use HT for analysis, which can capture the maximum, minimum, average, standard deviation and root mean square of instantaneous amplitude and instantaneous frequency. A total of 80 features (HHT feature set) are available. Steps 1 to Steps 3 are called feature extraction.

Step 4: Use the 3 feature selection methods of ReliefF, SU, and FCBF to screen the feature set of HHT to delete the features that affect identification.
Step 5: After feature selection, four feature sets can be generated by HHT (without feature selection), HHT-ReliefF, HHT-SU, and HHT-FCBF.
Step 6: Import each feature set into SVM, KNN, and DT. They are three classifiers for identification.
Step 7: Use ACC and ROC to present the identification results.
Step 8: Finally, use the t-SNE visualization method to present the distribution of features to verify the research results.

III. SIGNAL ANALYSIS AND CLASSIFICATION METHODS

Induction motors and other related equipment are used in current society. These machines usually run for a long time and require regular maintenance by engineers. If analysis methods can be used to capture the beneficial features of each failure and combining various classifiers, higher classification ACC for fault identification will be expected to solve the fault problem of the motor. However, many signal analysis techniques have been developed in current, and each analysis methods have their advantages and disadvantages. This section will explain the signal analysis method of the HHT. In addition, the ROC algorithm will also be introduced in the study.

A. HILBERT-HUANG TRANSFORM
HHT was jointly proposed by Norden E. Huang *et al.* The analysis is designed based on the mathematical theory of the

mathematician Hilbert in 1998 [24]. This analysis method has better results for unstable or nonlinear signals. It mainly focuses on the following steps [25], [26]:

1) The original signal passes through the EMD to obtain the IMF.
2) Apply Hilbert transform (HT) to the obtained IMF to obtain the instantaneous frequency.

1) EMPIRICAL MODE DECOMPOSITION

Before performing HHT, the original signals need to be decomposed by EMD, and the signals become IMFs compliant state through repeated screening. However, due to HHT's limitation of instantaneous frequency, if this process is omitted, the original signals will not be able to obtain a valid and complete instantaneous frequency. Therefore, by decomposing the original data into EMD, n IMFs and a trend function can be obtained respectively, and then HT can be performed on the obtained IMF for subsequent calculation of signal analysis [27]. For all function types, when the sum of the number of local maximum and local minima is the same as the number of zero crossings or the difference is 1. Then, when the average line of the upper envelope defined by local maximum and the lower envelope of upper envelope approaches zero at any point in time. They can be classified as IMF. The flowchart of the EMD is shown in Fig. 5.

2) HILBERT TRANSFORM

The calculation method of HT is different from the previous analysis of non-linearity and non-steady state. For the combination of IMF, when using HT, the instantaneous amplitude and instantaneous frequency of the required signal can be obtained, as shown in (1). Conjugate complex number is constructed by $x_i(t)$ and $H_i(t)$, as shown in (2). Where $C_i(t)$ is expressed as IMF. After HT operation, $H_i(t)$ can be obtained, where Pv represents the Cauchy principal value, and its purpose is to avoid being the singularity of $\tau = t$ and $\tau = \pm\infty$.

$$H_i(t) = \frac{1}{\pi}Pv \int_{-\infty}^{\infty} \frac{C_i(\tau)}{t-\tau}d\tau \tag{1}$$

$$Z_i(t) = C_i(t) + jH_i(t) = a_i(t)e^{-j\phi_i(t)} \tag{2}$$

As a result of the calculation of formula (2), the instantaneous amplitude $a_i(t)$ and instantaneous phase angle $\phi_i(t)$ can be obtained, which can be converted into formulas (3) and (4) respectively. Then the instantaneous phase $\phi_i(t)$ is differentiated against time to obtain the instantaneous frequency $\omega_i(t)$, as shown in equation (5).

$$a_i(t) = \sqrt{C_i^2(t) + H_i^2(t)} \tag{3}$$

$$\phi_i(t) = arctan\left[\frac{H_i(t)}{C_i(t)}\right] \tag{4}$$

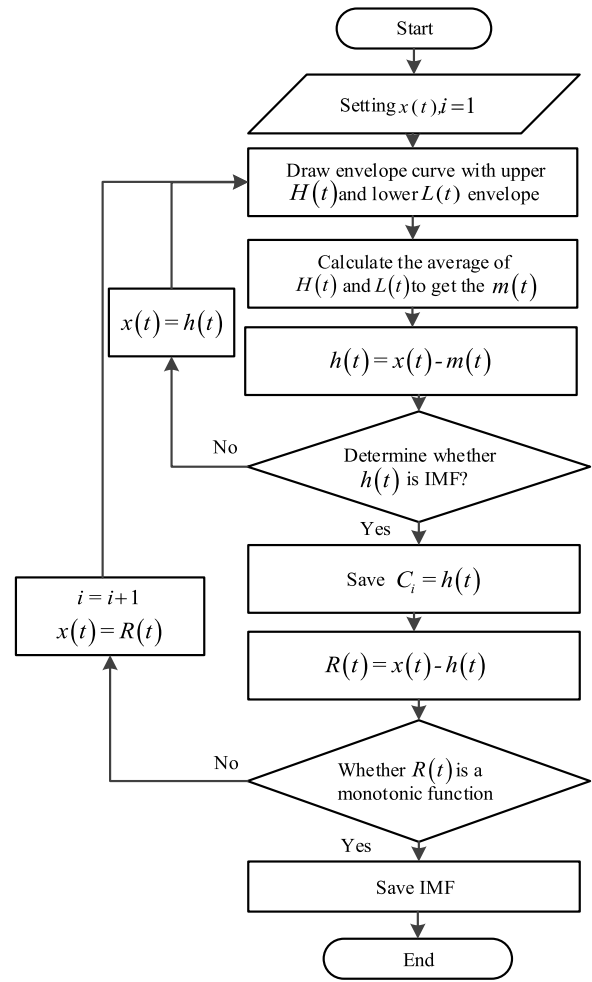$$\omega_i(t) = \frac{d\phi_i(t)}{dt} \tag{5}$$

FIGURE 5. Flowchart of EMD.

Through the above-mentioned correlation calculations, using the instantaneous amplitude $a_i(t)$ and instantaneous frequency $\omega_i(t)$, the time, frequency, and energy distribution can be obtained. This result is called the HT spectrum.

B. RECEIVER OPERATING CHARACTERISTIC CURVES

Compared with the ACC, ROC is a visual tool for the comparison of classification models. Its use was expanded in the 1970s and used in the biomedical field to interpret medical test results. In recent years, its analysis methods have been widely used in machine learning and data mining research [28], [29]. The ROC curve is constructed in a two-dimensional image, and the discrete classifier only predicts the category to which the tested object belongs. There are four possible results: true positive, true negative, false positive, and false negative. If an object is positive and is classified as positive, it will be regarded as a true positive (TP); if it is classified as negative, then It is a false negative (FN). In the same way, if the subject is negative and classified as negative, it will be regarded as true negative (TN); if it is classified as positive, it will be classified as false positive (FP), as shown in Fig. 6 [30]. For fair performance

FIGURE 6. Four elements of the confusion matrix.

evaluation, this study proposes two different evaluation indicators, namely ACC as shown in (6), and area under the ROC curve (AUC), where AUC is drawn by *TPR* and *FPR*, and the equation are shown in (7), (8) and (9).

$$ACC = accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$SP = specificity = \frac{TN}{TN + FP} \tag{7}$$

$$TPR = sensitivity = \frac{TP}{TP + FN} \tag{8}$$
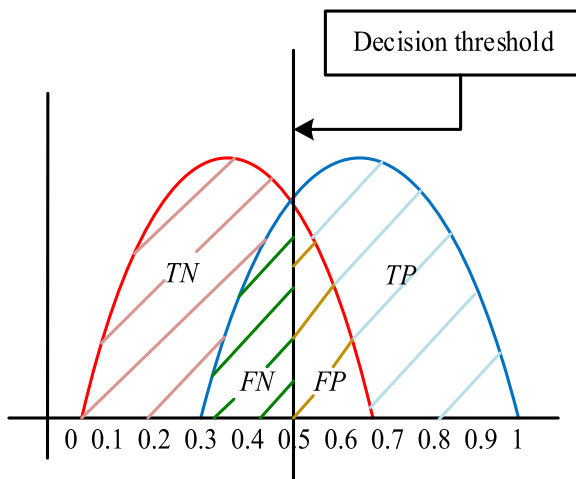
$$FPR = 1 - SP = \frac{FP}{TN + FP} \tag{9}$$

FIGURE 7. *TP* and *FP* will be changed by adjusting the decision-making threshold.

Among them, the ROC curve takes *FPR* as the X-axis and *TPR* as the Y-axis, it is necessary to set different decision thresholds at each point to obtain different *FPR* and *TPR*, as shown in Fig. 7 [31], [32]. Finally, draw a curve to assess the trade-off. The closer the curve is to the top, the higher the TP, and the higher the ACC. AUC is an indicator [33]. The larger the AUC, the better the performance. However, for large-scale screening, in order to minimize

FP, each experiment follows a hierarchical 10-fold cross-validation model, and the results obtained are average scores.

IV. FEATURE SELECTION AND DIMENSIONALITY REDUCTION METHOD

Nowadays, most of data is presented in a high-dimensional way during machine learning, which makes it difficult to observe high-dimensional distribution and features from data. When the number of features is too large, problems such as slower processing speed, overfitting, and difficulty in visualization may occur. Moreover, the curse of data dimensionality poses a severe challenge to many existing feature selection methods with respect to efficiency and effectiveness [34], [35]. Therefore, feature selection and data dimensionality reduction are essential preprocessing techniques in data analysis. Feature extraction is a method of dimensionality reduction [36]. The following research will introduce related applications of methods.

A. FEATURE SELECTION
Feature selection methods aim to reduce the influence of redundant variables by selecting a subset of existing features [37]. Evaluate the importance of each feature by measuring the relationship between each individual function and the output category. Feature selection assigns a weight to each feature, which can be regarded as a ranking to produce the feature list [38]. Therefore, the scope of this paper falls into discussing the commonly used feature selection method ReliefF, SU, and FCBF. Among them, the threshold settings of the following three feature selection methods are all calculated by multiplying the total weight by 0.9 as the standard.

1) RELIEFF
The ReliefF algorithm is more robust and can handle incomplete and noisy data compared with Relief [39]. Start by randomly selecting samples from its training sample set and take *K* neighboring samples *H* of the same category, which can be called Near-Hits. In addition, *K* samples *M* can be found from sample sets of different categories, which can be called Near-Misses. The distance difference is used to assign corresponding weights to features, and the preset threshold is used to effectively remove irrelevant features.

2) SYMMETRICAL UNCERTAINTY
The SU method uses the average amount of information contained in the message as the basis for judgment. This method is a form of information gain normalization, which is based on non-linear related information variables defined by information entropy. Used to reconstruct the degree of correlation between random variables, the value ranges from 0 to 1. When the value is larger, the correlation between *X* and *Y* is greater. When the variable is 0, it means that *X* and *Y* are independent of each other. Otherwise, it means that there is a strong complementarity between each other. The formula is shown in formula (10), where $H(X)$ represents information

entropy, and $I(X)$ is the calculation of information gain.

$$SU(X, Y) = \frac{I(X)}{H(X) + H(Y)} \quad (10)$$

3) FAST CORRELATION-BASED FILTER

FCBF uses symmetric uncertainty to replace information gain and performs the feature selection method [40]. It is an extended solution to SU. The advantage of the method is that it can remove redundant features. During the screening period, FCBF compares the two features and retains the feature that has a higher correlation with the target. Therefore, it uses the features with a higher correlation to complete the screening. This method reduces the time complexity and achieves the efficiency of computing while filtering, which can accelerate the calculation and also improve the recognition rate at the same time. FCBF is regarded as a fast filtering feature selection algorithm [41]. The calculation of this method to delete redundant features is shown in Fig. 8. Among them, F1, F2 and F4 can be regarded as similar features, and F1 is more related to the target, so F2 and F4 are considered redundant features; and on this basis, F6 and F7 can be deleted by F3.

FIGURE 8. Selection method of FCBF.

B. DIMENSIONALITY REDUCTION METHOD
1) PRINCIPAL COMPONENT ANALYSIS

Principal component analysis was proposed by Pearson, K in 1901. It can be regarded as a linear algorithm. It mainly uses the variance of each characteristic variable as a benchmark for measurement [42]. First, this method is to normalize the data to establish a covariance matrix and then use the singular value decomposition (SVD) method to obtain its eigenvector and eigenvalue. The usage is shown in formulas (11) and (12). Finally, the obtained eigenvalues are presented in descending order, and the original data are projected onto the eigenvectors to obtain new eigenvalues.

However, for PCA, the related variables are transformed by orthogonal transformation, and similar data points are placed in a low-dimensional space. This method is easy to cause underfitting of the features. This also means that the features of the data after dimensionality reduction will not be able to

effectively represent the distribution of the original data.

$$A = U \Sigma V^T \quad (11)$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} X^{(i)} \times X^{(i)T} = \frac{1}{m} \times X \times X^T \quad (12)$$

As shown in equation (11), A is an $m \times n$ order matrix, while U and V^T are $m \times m$ and $n \times n$ order matrixes, respectively. This decomposition can be interpreted as the SVD of A. Σ represents the singular value matrix, which is the eigenvalue corresponding to the A matrix. They are generally arranged in the diagonal in descending order, and to obtain the result of A, U is usually multiplied by the common variance matrix of formula (12) to obtain new features.

2) STOCHASTIC NEIGHBOR EMBEDDING

Stochastic Neighbor Embedding (SNE) was proposed by Hinton and Roweis in 2002 [43]. This method uses Euclidean distance to convert to conditional probability, accordingly explaining high-dimensional data through normal distribution and explaining the similarity between points [44], as shown in (13).

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)} \quad (13)$$

$p(j|i)$ is the similarity between x_i and x_j. σ_i is the Gaussian distribution centered on x_i. The low-dimensional y_i and y_j correspond to the high-dimensional x_i and x_j with $q(j|i)$ considered as conditional probability. When setting $\sigma = 1/\sqrt{2}$ and $q(j|i) = 0$, the result can be obtained, and is shown in (14). Finally, the SNE algorithm also applies Kullback-Leibler divergence (KLD) to express the degree of similarity between the two distributions. The objective function is shown in (15). This method is to minimize the loss function between the two distributions. The stochastic gradient descent method is calculated, as shown in (16).

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)} \quad (14)$$

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (15)$$

$$\frac{\delta C}{\delta y_i} = 2 \sum_j (p(j|i) - q(j|i) + p(i|j) - q(i|j)) (y_i - y_j) \quad (16)$$

Because the high-dimensional data cannot be completely retained in the low-dimensional space, a commonly known curse of dimensionality will be produced. The problems of crowding among various ethnic groups also cannot be distinguished. In addition to this, KLD is asymmetric which makes it possible for the SNE method to be optimized.

3) T-DISTRIBUTION STOCHASTIC NEIGHBOR EMBEDDING

In order to improve the curse of dimensionality in SNE method, t-SNE has several distinct features. This method was

proposed by Laurens van der Maaten and Geoffrey Hinton in 2008. It illustrates that this method is a nonlinear visualization method for dimensionality reduction in machine learning [45], [46]. t-SNE is difficult to use the axis/unit of the original high-dimensional data to explain the meaning of the graph, but when reducing the dimensional it is a highly used algorithm as a reference for data dimension reduction studies. In the most diverse applications, this method always presenting good results [47]. This paper will explain the crowding problem and the symmetry problem separately.

a: CROWDING PROBLEM
When the data is projected into a two-dimensional space, the distant points in the high-dimensional space have no position for projection in the low-dimensional space. Therefore, there are often overlap and difficulties to observe phenomena in the low-dimensional space, which is known as the crowding problem. As a solution, t-SNE method will replace the normal distribution used in the low-dimensional space with a t-distribution with 1 of freedom. The probability density function (PDF) of t-distribution is shown in (17) where v is expressed as a degree of freedom. When $v = 1$, it can be simplified to formula (18).

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} \tag{17}$$

$$f(t) = \frac{1}{\pi(1+t^2)} \tag{18}$$

b: SYMMETRIC SNE
This method is aimed at the problem of symmetry, using joint probability instead of conditional probability to recalculate KLD to obtain symmetry [18]. First of all, the equation of q_{ij} is defined as (19). It demonstrates that this equation has a symmetric relation. This representation method will condense the overall algorithm, while there will be situations where outliers are introduced. To solve this problem, the definition of joint probability can be modified as equation (20) and be substituted of the objective function (15). Furthermore, the formula of stochastic gradient descent is shown in (21) to obtain the minimum solution of the loss function which is commonly known as the best solution. Although t-SNE solves the symmetry problem, this algorithm involves quantities of calculations, which can be overwhelming for the system.

$$q_{ij} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq l} \exp\left(-\|y_k - y_l\|^2\right)} \tag{19}$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2} \tag{20}$$

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)\left(1 + \|y_i - y_j\|^2\right)^{-1} \tag{21}$$

t-SNE can not only convert data but also present the data in a two-dimensional or three-dimensional space for visual

observation. The main steps of t-SNE are: First, the algorithm begins by calculating the similarity probability of data points in the high-dimensional space and the similarity probability of the points in the corresponding low-dimensional space. Secondly, to make it easier to project data into low-dimensional space, the algorithm tries to minimize the difference in conditional probability between high-dimensional and low-dimensional data spaces. Finally, in order to evaluate the minimization of the t-SNE conditional probability difference sum, the gradient descent method is used to minimize the sum of the KLD of the original distribution and the corresponding data. This algorithm calculation process is shown in pseudocode 1:

Pseudocode 1 The Simple Version of t-SNE

1: Input Data
2: Calculate cost function using Eq. (15) // *perplexity Perp*
3: Set T, η, // *T = Number of iterations, η = Learning rate*
4: The target result $Y^T = y_1, \ldots, y_n$
5: Calculate the $p_{j|i}$ using Eq. (13)
         //*$p_{j|i}$ is the similarity between x_i and x_j*
6: Calculate the p_{ij} using Eq. (20)
         //*the definition of the joint probability*
7: **for** $t = 1 : T$
8:    Calculate the q_{ij} using Eq. (19)
9:    Calculate the $\frac{\delta C}{\delta y_i}$ using Eq. (21)
         //*the formula of stochastic gradient descent*
10:   $Y^t = Y^{t-1} + \eta \frac{dC}{dY} + \alpha(t)\left(Y^{t-1} - Y^{t-2}\right)$
11: **end**

V. RESULTS OF MOTOR FAULT IDENTIFICATION
This research measures the actual operating signals of induction motors and extracts features by HHT. However, the number and usefulness of features are unknown. To achieve the best performance of the algorithm, the importance of feature selection methods is gradually increasing. Among them, feature selection and feature extraction are more common [34]. Feature selection is identifying behaviors that have a significant contribution to the classifier's ability or finding the best feature subset. This method will let the feature set not be changed, keep important features, and reduce the number of features [35]. For signal analysis, the generated information often has invalid data, and too many repetitive or irrelevant features will cause the classifier to produce over-fitting situations. Therefore, this study compares the feature extraction methods including FFT, WT, and HHT. It can be found that the features obtained by the HHT method have better recognition results in each classifier, as shown in Table 3. Subsequently, this paper uses ReliefF, SU, and FCBF to compare three feature selection methods. By these methods, different feature sets are generated. Among them, FCBF can screen out the most irrelevant features, and the selected features are based on their importance. Finally, through the ROC curve, different feature sets are presented graphically to facilitate ACC's comparison.

TABLE 3. The ACC value of each classifier under different feature extraction sets.

	ACC		
	FFT	WT	HHT
KNN	0.698	0.703	0.713
SVM	0.856	0.865	0.87
DT	0.945	0.979	0.981

TABLE 4. Features by HHT.

Model	Number of Features	Sort by Important Features after Screening
ReliefF	22	F5, F45, F3, F1, F42, F2, F4, F16, F44, F43, F41, F61, F30, F6, F15, F33, F29, F77, F65, F26, F66, F55
SU	19	F41, F43, F44, F42, F45, F5, F1, F2, F4, F31, F49, F11, F3, F38, F67, F56, F48, F30, F15
FCBF	10	F41, F45, F5, F4, F31, F49, F11, F38, F67, F30

In this study, the feature selection methods are used to generate feature sets of different sizes including ReliefF, SU, and FCBF. Compared with HHT, the number of features can be reduced by 72.5%, 76.25%, and 87.5%, respectively, as shown in Table 4.

A. RESULTS OF ROC CLASSIFICATION
This study proposes to use three classifiers including SVM, KNN, and DT to classify the fault conditions of induction motors. Through these four feature sets, the unselected (It means all features are extracted using the HHT method and without using feature selection methods.) and three feature selection methods (ReliefF, SU, and FCBF) in different classification algorithms can be displayed in ROC curve graphs, making it easy to distinguished the performance of each algorithm. If the area under the curve is increased, the corresponding algorithm's classification effect will be better, and vice versa. To evaluate the pros and cons of related features under the classifier after feature selection, this study uses MATLAB software to draw ROC curves of various fault types under different feature sets. Taking KNN as an example, it can be found by drawing the curve that when using this classifier, the ReliefF, SU, and FCBF sets generated by the feature selection methods have advantages compared to the HHT. This result can explain that too many irrelevant features will misjudge the AUC and cause the recognition ability to decline, as shown in Fig. 9.

This study also calculates the area under the surface AUC in detail, as shown in Table 5. First of all, it can be seen that although these classifiers have their distinct standards, they may prefer different algorithms in terms of performance. Secondly, it can be found from the calculation results that if the ranking corresponding to the average AUC is given, the feature set of HHT may have factors that affect the recognition rate. It will also appear weak in rankings. Finally, if different feature selection methods are used with different classifiers to distinguish, the identification ability to bearing faults can reach close to 1 (optimal). This result shows that this fault situation is the most obvious among the 4 fault types.

## B. COMPARISON OF AUC AND ACC RESULTS
To present the results of identifying the types of induction motor faults, this study discusses the identification results by comparing the values of ACC and AUC. Among them, ACC presents the ratio of the classifier's accurate discrimination. Although it can effectively reflect the classifier's performance, it is quite accurate in the face of extreme data such as negative prediction. The value of ACC is substantial. This situation shows that ACC cannot just effectively evaluate the model without the test data; on the contrary, AUC, as a quantitative indicator of ROC, can be drawn into a curve by obtaining different *FPR* and *TPR* to calculate and evaluate the value of the classifier model.

This study uses two kinds of recognition rates for calculation, and the results are shown in Table 6. First of all, we can compare the three typical classifiers and find that using the DT classifier for identification, AUC, and ACC's identification ability is the best in the table. Secondly, this study compared four feature sets of different sizes. The result shows that the feature selection methods used in this research can effectively delete features that are not important to improve the recognition rate. Moreover, And AUC and ACC identification results of the collection after three kinds of feature selection are higher than the feature set of HHT by calculation.

In summary, in order to show the stability of the selected features, this research discusses the results through three kinds of common classifiers of KNN, SVM, and DT. Among them, the recognition results given by different classifiers for different evaluation methods will be different. In the comparison of AUC and ACC, it may also be distinguished between advantages and disadvantages due to their different calculation methods, as shown in the brackets in Table 6. According to the effective evaluation of multiple methods, DT-ReliefF has the best recognition results, which are 99.7% and 99.6% respectively.

C. VERIFICATION OF t-SNE
In order to verify that the feature selection method used produces feature sets of different sizes, and also can effectively screen out the important features of motor fault types. This study uses t-SNE to transform the data and present the results in two or three dimensions. By reducing the dimension,
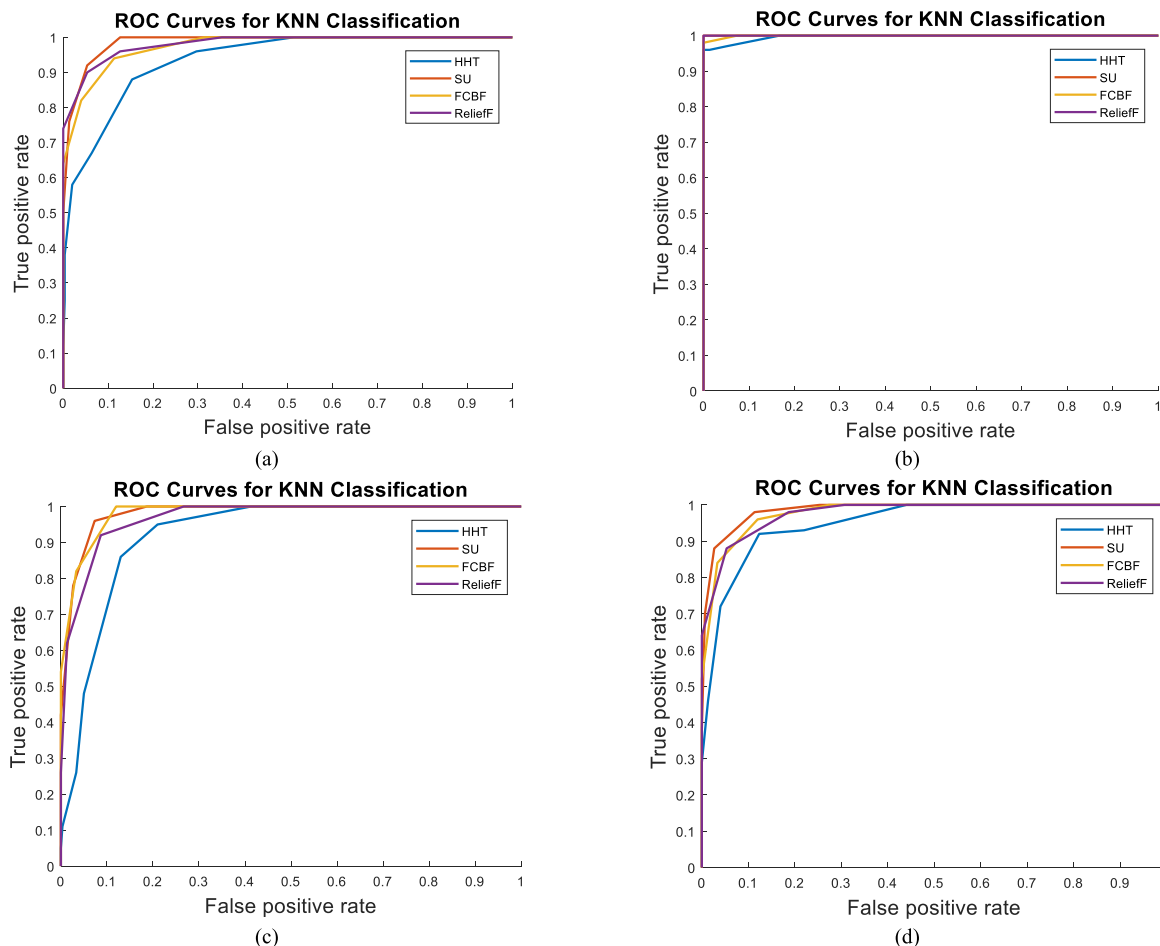
FIGURE 9. Use the KNN to perform ROC mapping on (a) normal (b) bearing damage (c) broken rotor bar (d) short circuit in stator windings.

TABLE 5. The AUC value of each classifier under different feature sets.

Classifiers	Feature model	Normal	Bearing	Rotor	Stator	Average
KNN	HHT	0.936	0.997	0.919	0.951	0.951 (4)
	ReliefF	0.98	1	0.968	0.976	0.981 (3)
	SU	0.985	1	0.981	0.985	0.988 (1)
	FCBF	0.974	0.999	0.981	0.977	0.983 (2)
SVM	HHT	0.852	1	0.998	0.938	0.947 (4)
	ReliefF	0.896	1	1	0.936	0.958 (2)
	SU	0.91	1	1	0.936	0.961 (1)
	FCBF	0.863	1	1	0.936	0.948 (3)
DT	HHT	0.975	1	1	0.989	0.991 (4)
	ReliefF	0.999	1	1	1	0.997 (1)
	SU	0.999	1	1	0.988	0.997 (2)
	FCBF	0.999	1	1	0.983	0.995 (3)

it is possible to visually judge the validity of the algorithm and data collection. The following will discuss two different sizes of feature sets, HHT (not used feature selection) and HHT-ReliefF (combined with feature selection). However,

the distance of t-SNE is meaningless, it is just a concept of the probability distribution.

In addition to creating clusters, t-SNE can also leave a certain distance between them, which simplifies the data
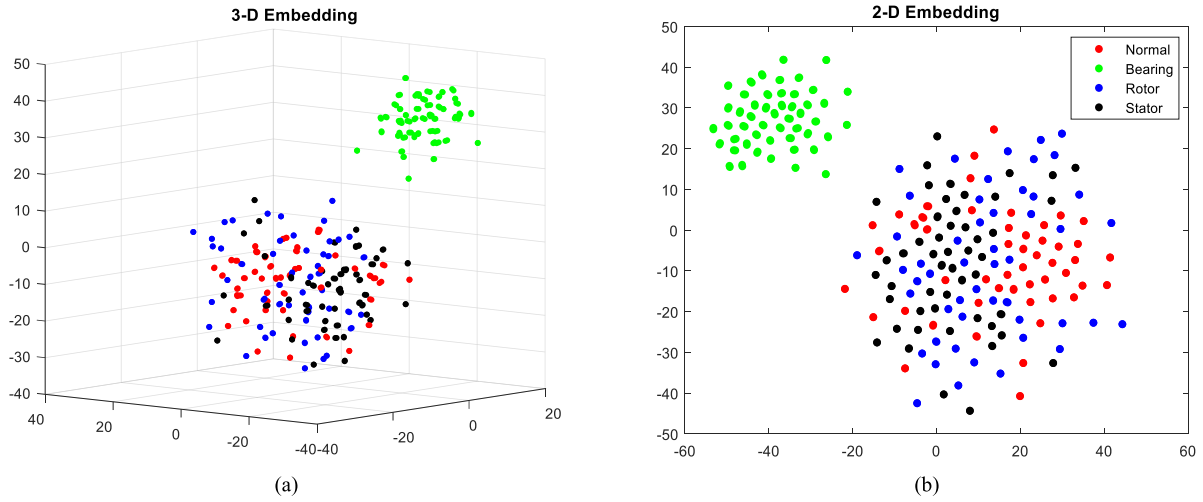
FIGURE 10. Scatter plot of t-SNE events under HHT conditions (a) 3D (b) 2D.
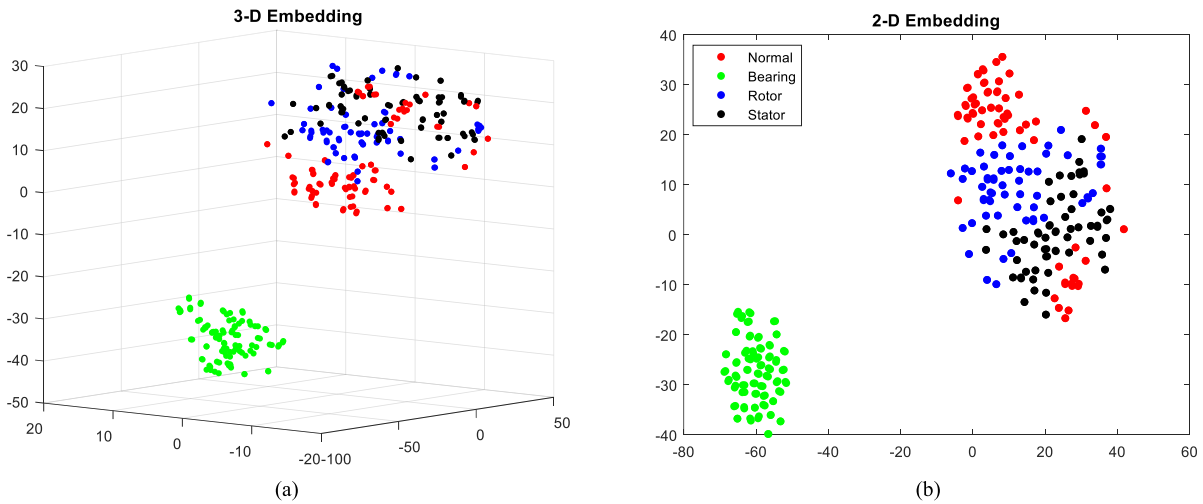

FIGURE 11. Scatter plot of t-SNE events under HHT-ReliefF conditions (a) 3D (b) 2D.

visualization. This t-SNE allows simple visualization of graphs to understand the failure of induction motors under different feature sets. As shown in Figs. 10(a) and (b), HHT feature set is found that no matter in the two or three-dimensional space, only the type of bearing fault can be easily distinguished. The other three types of faults have no noticeable difference in data projection. This model also shows that when feature selection is not used in this study, the features obtained due to the wrong judgment of the motor fault type will reduce ACC and AUC. Then, to verify the results presented by t-SNE with different collection sizes after feature selection, this research also uses the HHT-ReliefF collection for visual observation, as shown in Figs. 11(a) and (b). Results are found that compared with the feature set of HHT in the three-dimensional space, the types of faults are classified more clearly after feature screening. If it is reduced to a two-dimensional space, 4 types of clusters and effective classification scatter plots can be obtained. Finally, compare the visualize ed results with the ACC and AUC recognition rates of the motors in the aforementioned

TABLE 6. Comparison of ACC and AUC.

Classifiers	Feature model	AUC	ACC
KNN	HHT	0.951 (4)	0.713 (4)
	ReliefF	0.981 (3)	0.848 (3)
	SU	0.988 (1)	0.890 (1)
	FCBF	0.983 (2)	0.853 (2)
SVM	HHT	0.947 (4)	0.870 (4)
	ReliefF	0.958 (2)	0.940 (1)
	SU	0.961 (1)	0.935 (2)
	FCBF	0.948 (3)	0.920 (3)
DT	HHT	0.991 (4)	0.981 (4)
	ReliefF	**0.997 (1)**	**0.996 (1)**
	SU	0.997 (2)	0.985 (2)
	FCBF	0.995 (3)	0.984 (3)

to verify the feature selection set. The basic idea of the t-SNE algorithm is to express the similarity of data points by using the joint probability between high-dimensional data points

TABLE 7. The typical faults in the induction motor.

Types of Faults	Signals	Analysis Methods	Identification methods	Advantages	References
Bearing	Vibration	WPT	ANN/SVM	Compared with different classifiers, SVM has higher ACC.	[48]
Rotor	Current	Harmonic frequencies/ corresponding amplitudes	NN	It is quite effective for detecting broken-rotor bar.	[49]
Stator	Current	The information between the phase current signals	ANN	Even in the case of variable loading and phase unbalance.	[50]
Bearing/Rotor	Current	spectrum analysis	New artificial ant clustering method	Better results for unsupervised learning.	[51]
Rotor/Stator	Vibration	Higher order statistics	Fuzzy Logic	ACC can more than 90%	[52]
Rotor/Stator/Phase	Current	FFT/WT	ANN	Effectively predict upcoming failures.	[53]
Bearing/Rotor/Stator/Phase	Vibration/ Current	PCA/ICA	SVM	Use feature extraction technology to make SVM have better results.	[54]
Bearing/Rotor/Stator	Current	HHT-3 kinds of feature selection methods	SVM/KNN/DT	Combining different feature selection methods, the ACC and AUC ability have been improved. Use ROC and t-SNE methods to obtain effective visual experimental results.	The proposed model

NN: Neural Network
ANN: Artificial Neural Network
ICA: Independent Components Analysis

and analog data points corresponding to low-dimensional space. In the most diverse applications, this method always presenting good results. But for the t-SNE graph, the distance is meaningless.

In order to make the advantages of the proposed methodology more prominent, this research also compares the intelligent diagnosis methods of various types of motor faults. As shown in Table 7. Three common failure models are considered to reduce the chance of further damage or complete motor failure due to any specific failure.

VI. CONCLUSION
This research proposes a simple and high-performance asynchronous motor fault diagnosis model based on traditional feature extraction, feature selection, and classifier construction. In this study, the four kinds of fault (a) normal, (b) bearing damage, (c) broken rotor bar, and (d) short circuit in stator windings occurrence in three-phase induction motors' current signal are considered. The original feature extraction methods are highly reliant on the expertise and prior knowledge, also have limited capacities for learning the relationships

between the features and data. Therefore, this research creates a novel model with the combination of advantages of each method. The main contributions of this study are fourfold. Firstly, HHT is very suitable for processing nonlinear and non-stationary signals with better recognition results than WT and FFT. Secondly, most of the data is presented in a high-dimensional manner. Too much quantity may cause problems such as overfitting and difficulty in visualization. However, the ReliefF, SU and FCBF feature selection methods are used to select the important features to generate feature sets of different sizes. Thirdly, to show the stability of the selected features, this study discusses the identification results through three typical classifiers: KNN, SVM and DT. Finally, verify the results of ACC, AUC, ROC, and t-SNE to show. The most advantage of this research is to propose visual ROC and t-SNE methods and combine them with traditional feature extraction and feature selection methods to present important features of induction motor fault identification. Simulation and experimental results show that:

1) This study compares feature selection methods to improve less important or redundant features. The

results show that more than 70% of features can be effectively deleted under the three different screening methods.

2) This study also proposed ROC curves of different feature screening methods, which have advantages compared with HHT. Through a detailed calculation of AUC and ACC, it can be proved that the recognition rate of feature screening can be effectively increased by 2% to 3%.

3) Finally, this study also proved that regardless of the 2D or 3D scatter plot; feature selection sets have better feature distributions to classification by using t-SNE.

In the future, we will also try to add more motor fault types and data and use deep learning for research training. This model will allow a reasonable comparison between the simulation results and actual engine operation. Besides that, multiple published databases, feature selection methods, and other literature to compare motor fault classification have also become necessary.

REFERENCES

[1] O. A. Mohammed, N. Y. Abed, and S. Ganu, "Modeling and characterization of induction motor internal faults using finite-element and discrete wavelet transforms," *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 3434–3436, Oct. 2006.

[2] Z. Gao and X. Liu, "An overview on fault diagnosis, prognosis and resilient control for wind turbine systems," *Processes*, vol. 9, no. 2, p. 300, Feb. 2021.

[3] P. Chen, Y. Li, K. Wang, and M. J. Zuo, "A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition," *Measurement*, vol. 158, Jul. 2020, Art. no. 107739.

[4] M. Ojaghi and R. Akhondi, "Modeling induction motors under mixed radial–axial asymmetry of the air gap produced by oil-whirl fault in a sleeve bearing," *IEEE Trans. Magn.*, vol. 54, no. 11, Nov. 2018, Art. no. 8205105.

[5] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[6] A. J. Bazurto, E. C. Quispe, and R. C. Mendoza, "Causes and failures classification of industrial electric motor," in *Proc. IEEE ANDESCON*, Arequipa, Peru, Oct. 2016, pp. 1–4.

[7] E. Resendiz-Ochoa, R. A. Osornio-Rios, J. P. Benitez-Rangel, R. De J. Romero-Troncoso, and L. A. Morales-Hernandez, "Induction motor failure analysis: An automatic methodology based on infrared imaging," *IEEE Access*, vol. 6, pp. 76993–77003, 2018.

[8] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognit. Syst. Res.*, vol. 53, pp. 42–50, Jan. 2019.

[9] B. Hu, X. Zhang, J. Mu, M. Wu, Z. Zhu, Z. Liu, and Y. Wang, "Spasticity measurement based on the HHT marginal spectrum entropy of sEMG using a portable system: A preliminary study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1424–1434, Jul. 2018.

[10] V. Puliafito, S. Vergura, and M. Carpentieri, "Fourier, wavelet, and Hilbert-huang transforms for studying electrical users in the time and frequency domain," *Energies*, vol. 10, no. 2, p. 188, Feb. 2017.

[11] A. Roy, C.-H. Wen, J. F. Doherty, and J. D. Mathews, "Signal feature extraction from microbarograph observations using the Hilbert–Huang transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1442–1447, May 2008.

[12] D. Borras, M. Castilla, N. Moreno, and J. C. Montano, "Wavelet and neural structure: A new tool for diagnostic of power system disturbances," *IEEE Trans. Ind. Appl.*, vol. 37, no. 1, pp. 184–190, Jan. 2001.

[13] M.-F. Guo, N.-C. Yang, and W.-F. Chen, "Deep-Learning-Based fault classification using Hilbert–Huang transform and convolutional neural network in power distribution systems," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6905–6913, Aug. 2019.

[14] J. Dai, Y. Chen, Y. Yi, J. Bao, L. Wang, W. Zhou, and G. Lei, "Unsupervised feature selection with ordinal preserving self-representation," *IEEE Access*, vol. 6, pp. 67446–67458, 2018.

[15] X. Chen, G. Yuan, W. Wang, F. Nie, X. Chang, and J. Z. Huang, "Local adaptive projection framework for feature selection of labeled and unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6362–6373, Dec. 2018.

[16] Y. Fu, Z. Gao, Y. Liu, A. Zhang, and X. Yin, "Actuator and sensor fault classification for wind turbine systems based on fast Fourier transform and uncorrelated multi-linear principal component analysis techniques," *Processes*, vol. 8, no. 9, p. 1066, Sep. 2020.

[17] A. Chatzimparmpas, R. M. Martins, and A. Kerren, "T-ViSNE: Interactive assessment and interpretation of t-SNE projections," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 8, pp. 2696–2714, Aug. 2020.

[18] W. S. Saif, T. Alshawi, M. A. Esmail, A. Ragheb, and S. Alshebeili, "Separability of histogram based features for optical performance monitoring: An investigation using t-SNE technique," *IEEE Photon. J.*, vol. 11, no. 3, Jun. 2019, Art. no. 7203012.

[19] J. Cheng, H. Liu, F. Wang, H. Li, and C. Zhu, "Silhouette analysis for human action recognition based on supervised temporal t-SNE and incremental learning," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3203–3217, Oct. 2015.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[21] S. Atapattu, C. Tellambura, and H. Jiang, "Analysis of area under the ROC curve of energy detection," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1216–1225, Mar. 2010.

[22] H. Ma, A. I. Bandos, H. E. Rockette, and D. Gur, "On use of partial area under the ROC curve for evaluation of diagnostic performance," *Statist. Med.*, vol. 32, no. 20, pp. 3449–3458, Mar. 2013.

[23] R. Guo, X. Shen, and X. Zhang, "3D ROC histogram: A new ROC analysis tool incorporating information on instances," *IEEE Access*, vol. 7, pp. 183396–183404, 2019.

[24] C. Cai, W. Liu, J. S. Fu, and Y. Lu, "Radar micro-Doppler signature analysis with HHT," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 2, pp. 929–938, Apr. 2010.

[25] H. Zhang, F. Wang, D. Jia, T. Liu, and Y. Zhang, "Automatic interference term retrieval from spectral domain low-coherence interferometry using the EEMD-EMD-based method," *IEEE Photon. J.*, vol. 8, no. 3, Jun. 2016, Art. no. 6900709.

[26] T. Kijewski-Correa and A. Kareem, "Efficacy of Hilbert and wavelet transforms for time-frequency analysis," *J. Eng. Mech.*, vol. 132, no. 10, pp. 1037–1049, Oct. 2006.

[27] M. J. Afroni, D. Sutanto, and D. Stirling, "Analysis of nonstationary power-quality waveforms using iterative Hilbert huang transform and SAX algorithm," *IEEE Trans. Power Del.*, vol. 28, no. 4, pp. 2134–2144, Oct. 2013.

[28] J. Hernández-Orallo, "ROC curves for regression," *Pattern Recognit.*, vol. 46, no. 12, pp. 3395–3411, Dec. 2013.

[29] D. Lin, L. Sun, K.-A. Toh, J. B. Zhang, and Z. Lin, "Twin SVM with a reject option through ROC curve," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1710–1732, Mar. 2018.

[30] P. Sonego, A. Kocsor, and S. Pongor, "ROC analysis: Applications to the classification of biological sequences and 3D structures," *Briefings Bioinf.*, vol. 9, no. 3, pp. 198–209, Jan. 2008.

[31] T. Fawcett, "ROC graphs with instance-varying costs," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 882–891, Jun. 2006.

[32] P. Wan, C. Wu, Y. Lin, and X. Ma, "On-road experimental study on driving anger identification model based on physiological features by ROC curve analysis," *IET Intell. Transp. Syst.*, vol. 11, no. 5, pp. 290–298, Jun. 2017.

[33] P. Zhao, S. C. H. Hoi, R. Jin, and T. Yang, "Online AUC maximization," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, Jun. 2011, pp. 233–240.

[34] N. Gopika and A. M. kowshalaya M. E., "Correlation based feature selection algorithm for machine learning," in *Proc. 3rd Int. Conf. Commun. Electron. Syst. (ICCES)*, Oct. 2018, pp. 692–695.

[35] H. Sawhney and B. Jeyasurya, "A feed-forward artificial neural network with enhanced feature selection for power system transient stability assessment," *Electr. Power Syst. Res.*, vol. 76, no. 12, pp. 1047–1054, Aug. 2006.

[36] C. A. Murthy, "Bridging feature selection and extraction: Compound feature generation," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 757–770, Apr. 2017.

[37] A. Navot, R. Gilad-Bachrach, Y. Navot, and N. Tishby, "Is feature selection still necessary?" in *Subspace, Latent Structure and Feature Selection* (Lecture Notes in Computer Science), vol. 3940. Berlin, Germany: Springer, 2006, pp. 127–138.

[38] G. Taskin, H. Kaya, and L. Bruzzone, "Feature selection based on high dimensional model representation for hyperspectral images," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2918–2928, Jun. 2017.

[39] M. Wu and Y. Wang, "A feature selection algorithm of music genre classification based on ReliefF and SFS," in *Proc. IEEE/ACIS 14th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2015, pp. 539–544.

[40] H. Djellali, S. Guessoum, N. Ghoualmi-Zine, and S. Layachi, "Fast correlation based filter combined with genetic algorithm and particle swarm on feature selection," *2017 5th Int. Conf. Electr. Eng.-Boumerdes (ICEE-B)*, Boumerdes, Algeria, pp. 1–6, Oct. 2017.

[41] C.-Y. Lee and W.-C. Lin, "Induction motor fault classification based on FCBF-PSO feature selection method," *Appl. Sci.*, vol. 10, no. 15, p. 5383, Aug. 2020.

[42] D. Peng and Z. Yi, "Dynamics of generalized PCA and MCA learning algorithms," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1777–1784, Nov. 2007.

[43] G. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proc. NIPS*, vol. 15, 2002, pp. 833–840.

[44] D. Lunga and O. Ersoy, "Spherical stochastic neighbor embedding of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 857–871, Feb. 2013.

[45] N. Pezzotti, J. Thijssen, A. Mordvintsev, T. Hollt, B. Van Lew, B. P. F. Lelieveldt, E. Eisemann, and A. Vilanova, "GPGPU linear complexity t-SNE optimization," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 1172–1181, Jan. 2020.

[46] J. Liu, Q. Li, H. Yang, Y. Han, S. Jiang, and W. Chen, "Sequence fault diagnosis for PEMFC water management subsystem using deep learning with t-SNE," *IEEE Access*, vol. 7, pp. 92009–92019, 2019.

[47] M. A. Da Silva Lopes, A. D. D. Neto, and A. De Medeiros Martins, "Parallel t-SNE applied to data visualization in smart cities," *IEEE Access*, vol. 8, pp. 11482–11490, 2020.

[48] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Fault diagnosis based on optimized node entropy using lifting wavelet packet transform and genetic algorithms," *Proc. Inst. Mech. Eng., I, J. Syst. Control Eng.*, vol. 224, no. 5, pp. 557–573, Aug. 2010.

[49] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Detection of broken rotor bar faults in induction motor at low load using neural network," *ISA Trans.*, vol. 64, pp. 241–246, Sep. 2016.

[50] G. H. Bazan, P. R. Scalassara, W. Endo, A. Goedtel, W. F. Godoy, and R. H. C. Palácios, "Stator fault analysis of three-phase induction motors using information measures and artificial neural networks," *Electr. Power Syst. Res.*, vol. 143, pp. 347–356, Feb. 2017.

[51] A. Soualhi, G. Clerc, and H. Razik, "Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4053–4062, Sep. 2013.

[52] I. Lasurt, A. F. Stronach, and J. Penman, "A fuzzy logic approach to the interpretation of higher order spectra applied to fault diagnosis in electrical machines," in *Proc. 19th Int. Conf. North Amer. Fuzzy Inf. Process. Soc.*, 2000, pp. 158–162.

[53] S. Lee, M. D. Bryant, and L. Karlapalem, "Model- and information theory-based diagnostic method for induction motors," *J. Dyn. Syst., Meas., Control*, vol. 128, no. 3, pp. 584–591, Sep. 2006.

[54] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 241–250, Jul. 2007.

CHUN-YAO LEE (Member, IEEE) received the Ph.D. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2007. Sponsored by the Ministry of Science and Technology of R.O.C., he was a Visiting Scholar with the University of Washington, Seattle, from 2004 to 2005. He was also a Distribution System Designer with the Engineering Division, Taipei Government, from 2001 to 2008. He joined Chung Yuan Christian University as a Faculty Member in 2008. His research interests include power distribution, optimization algorithms, and damage diagnosis.

WEN-CHENG LIN was born in Tainan, Taiwan, in 1996. He received the B.S. degree in 2019. He is currently pursuing the M.S degree in electrical engineering with Chung Yuan Christian University, Taoyuan, Taiwan. His research interests include fault detection and optimization algorithms.

• • •