Article

Improved Random Forest Algorithm Based on Decision Paths for Fault Diagnosis of Chemical Process with Incomplete Data

Yuequn Zhang ⓘ, Lei Luo, Xu Ji ⓘ and Yiyang Dai *ⓘ

Department of Chemical Engineering, Sichuan University, Chengdu 610065, China;
zhangyuequn@stu.scu.edu.cn (Y.Z.); dreisteine@stu.scu.edu.cn (L.L.); jixu@scu.edu.cn (X.J.)
* Correspondence: daiyy@scu.edu.cn

Abstract: Fault detection and diagnosis (FDD) has received considerable attention with the advent of big data. Many data-driven FDD procedures have been proposed, but most of them may not be accurate when data missing occurs. Therefore, this paper proposes an improved random forest (RF) based on decision paths, named DPRF, utilizing correction coefficients to compensate for the influence of incomplete data. In this DPRF model, intact training samples are firstly used to grow all the decision trees in the RF. Then, for each test sample that possibly contains missing values, the decision paths and the corresponding nodes importance scores are obtained, so that for each tree in the RF, the reliability score for the sample can be inferred. Thus, the prediction results of each decision tree for the sample will be assigned to certain reliability scores. The final prediction result is obtained according to the majority voting law, combining both the predicting results and the corresponding reliability scores. To prove the feasibility and effectiveness of the proposed method, the Tennessee Eastman (TE) process is tested. Compared with other FDD methods, the proposed DPRF model shows better performance on incomplete data.

Keywords: random forest; decision path; fault diagnosis; incomplete data; reliability scores

1. Introduction

There is a large number of variables and parameters in a chemical process, and the complex relationships between them make the chemical process high-dimensional, nonlinear, and strongly coupled. In addition, most chemical processes have long operating cycles, and ensuring that the process is in a normal and safe state of operation while producing qualified products is a challenging issue. With an extended operating time, equipment aging and fouling, and other factors, will lead to a slow decline in the process equipment performance, with the process operation state gradually approaching the security boundary. When a fault occurs, fault detection and diagnosis (FDD) is required to detect and handle the process faults. FDD has always been a hot issue in the field of chemical process safety, and many methods have been proposed and applied [1,2].

Such methods can be divided into the following two categories: model-based methods and data-driven methods [3]. Initially, the FDD approaches were mostly based on mechanism models, using the principle of first nature to model processes and diagnose faults [4,5]. Although these methods are robust and reliable, they are difficult to apply to more complex dynamic processes, due to the high requirements of process experience and knowledge [6]. In contrast, data-driven methods have better applicability to process mechanisms, higher accuracy, and lower requirements for mechanisms, and have become a research hotspot in recent years. Early data-driven FDD methods are mainly qualitative methods, including expert systems (ES), qualitative trend analysis (QTA) methods, and signed directed graphs (SDG), which are difficult to ensure the accuracy of diagnosis, and they cannot efficiently process large amounts of data. Besides, due to the increasing scale and complexity of modern industrial processes, more and more historical data are available, thus quantitative methods based on processing historical data have great advantages in applicability over

qualitative methods [7]. Quantitative methods, such as principal component analysis (PCA), independent component algorithm (ICA), and Gaussian mixture model (GMM), are used in FDD. PCA is a powerful tool for dimension reduction, which allows the most important variable information to be retained. It has been extensively used in feature extraction, data compression, image processing, and pattern recognition [8,9]. ICA can process high-intensity, high-noise and related data by extracting independent statistical variables hidden in the process to achieve process dimensionality reduction [10–12]. GMM uses different Gaussian components to describe multiple working modes in the general process. The prior probability of each Gaussian component represents the possibility of the process operating under each specific operating condition [13,14]. In recent years, with the popularization of neural networks, scholars have proposed various non-statistical methods; for example, the convolutional neural network (CNN) realizes image recognition and compressed sensor signal processing by automatic feature extraction and deep learning through hidden layers [15–17]. The artificial immune system (AIS) [18–20] performs well in self-improvement, and can improve the accuracy and efficiency of diagnosis through self-learning in the diagnosis process. These non-statistical methods have become the latest research highlights.

However, although great progress has been made in the theoretical research of the FDD algorithm, and high accuracy has been obtained in toy models, such as the TE process, there are still some difficulties in practical industrial applications. With the rapid development of technologies such as the Internet of Things, and its widespread application in industry, the amount of process data has increased rapidly. Many FDD algorithms may have problems such as low efficiency and an inability to apply the algorithm to multimodal analysis when exposed to massive amounts of data. In addition, due to the frequent mechanical failures or human errors in recording process data, there is often a large number of abnormal or missing values in the data samples collected [21,22], which destroys the continuity of timing data and invalidate statistical methods such as PCA. This brings challenges to the practical industrial application of the FDD algorithm [23]. Therefore, there is a need for a new FDD method that is adapted for handling large amounts of data and is able to handle abnormal or missing data. To handle problems of missing data, four types of approaches are frequently used in the classification and discriminant analysis of incomplete data [24], as follows:

- Deletion of the incomplete feature vectors, and classification of the complete data portion only;
- Development of a multi-classifier corresponding to all the combinations of feature subsets, and classification of incomplete data using the model trained by the same available features;
- Imputation or estimation of missing data, and classification using the edited set;
- Implementation of exploiting procedures for which classification can be still accomplished in the presence of missing variables.

The first method increases bias and may lead to significant information loss [22,25]. The second method can only achieve good results with algorithms such as PCA, when dealing with a very limited number of faults and features, and explodes, in terms of complexity, when numerous features exist in chemical processes [26,27]. The third method uses specific values, such as means and random values, to fill in the missing data, artificially increasing the noise of the data. It is necessary to wait until enough observations are obtained to carry out the imputation, which is time consuming and undermines its effectiveness [28]. For the fourth method, previous studies have shown that some algorithms, such as decision trees, support vector machines, fuzzy algorithms, etc., can effectively process missing values [29–31]. Among them, the RF model stands out due to its applicability in the case of missing data [32]. The random forest model is composed of several weak classifiers. Therefore, compared with other statistical methods, it has a strong anti-interference ability. It can operate effectively in large datasets, without dimensionality reduction, and is suitable for high-dimensional input samples [33]. The RF algorithm can also be used in multimodal

processes, in which most of the above-mentioned FDD methods are difficult to apply [34,35], and can be classified without feature transformation. In particular, when there is a small amount of missing data in the sample, RF can still maintain good stability [36,37]. However, as the amount of missing data continues to increase, the accuracy of RF will inevitably decrease. Even if the missing values are filled in the way described in the third method above, the subjectivity of the filling method selection and the problem of noise introduction cannot be avoided. Therefore, the question becomes how to use the data information retained in the sample in an objective way to reason and restore the missing values as accurately as possible, so as to alleviate the impact of missing data on the accuracy of the data-driven modelling process, such as FDD.

Therefore, using the mean values to fill the missing dataset, this paper proposes an improved RF based on decision paths, namely, DPRF, to further modify the results of the RF algorithm to minimize the impact of filling the missing values on the accuracy of the model. According to the importance of each node in the classification process, the algorithm defines a new parameter—reliability score (RS)—to characterize the impact of missing data on the classification, and uses it to correct the original prediction results on the edited dataset to obtain the final classification results. This algorithm will be discussed in detail later in this paper. The main innovations of this paper are listed as follows:

(1). A new FDD algorithm based on random forest is proposed, which is applicable to different missing data patterns.
(2). The concept of decision tree reliability is put forward, based on a decision path to quantify the impact of missing data on the RF model results.
(3). The test performance of the proposed algorithm on the benchmark model is proved to be better than that of other classical FDD algorithms for data loss.

The rest of the paper is arranged as follows. Section 2 reviews the basic mechanism of the random forest classification algorithm and proposes an improved random forest algorithm, defining local importance and reliability scores to modify the prediction results. Section 3 shows how to use the famous Tennessee Eastman process to verify the performance of the improved algorithm. For comparison, the diagnostic properties of traditional RF algorithms and other FDD methods are also tested. Section 4 summarizes this work.

2. Materials and Methods

2.1. Traditional Random Forest Algorithm

Random forest is a kind of ensemble learning whose core lies in random sample selection and random feature selection. It uses multiple decision trees as the base learner for learning, and applies voting laws to integrate the results of each learner to complete the learning task [38,39].

When the RF model is trained, the structure of the T decision trees is determined, and the optimal characteristics of each split node and the sample size belonging to each category on each node can be obtained. Next, a single sample is input to get the split node that the sample has passed through in a certain decision tree, and the leaf node represents the final classification result of the sample in the decision tree. By arranging these split features in order, the decision path of a specific sample in a specific decision tree can be obtained.

For the features included in the decision path, the amount of information contained in each feature is different, resulting in different importance of each node for target prediction. Node importance is introduced to quantify it.

It is assumed that for a particular sample x, the label predicted by the tth decision tree in the random forest is y. It is defined that the training samples belonging to the y category at the kth node of the decision tree are positive samples. The proportion of positive samples in all training samples contained in this node is denoted as $r_k^t(y)$, which can also be considered as the probability that the training sample contained in node k belongs to the predicted sample category y. The difference in proportion of positive samples in child node and its corresponding parent node can be viewed as the node importance of the child node [40–42]. The larger the difference, the higher the purity of

the sample split to the child node compared to that of the parent node, thus the higher the importance of the child node for the classification problem. Referring to the definition of local increment [42], local importance (LI) of each node for the predicted label can be defined as follows:

$$LI_k^t(y) = r_k^t(y) - r_{parent(t,k)}^t(y) \tag{1}$$

where $parent(t, k)$ represents the parent node of the kth node in the decision path of sample x in the tree t.

More specifically, the iris dataset (https://www.kaggle.com/arshid/iris-flower-dataset, accessed on 14 September 2021) is used as a numerical example to better explain what the decision path is and to give the calculation process of local importance. The iris dataset contains 150 samples, which can be divided into Setosa (class 0), Versicolour (class 1), Virginica (class 2), each containing 50 samples. Each sample contains the following four characteristics: sepal length (f_1), sepal width (f_2), petal length (f_3), and petal width (f_4). For better training effects, 50 trees are trained and the minimum sample size for each category is set to 5 to prevent overfitting. Take the 12th tree in the RF model for example, whose internal structure is shown in Figure 1 below. Class size on each node represents the number of training samples belonging to each category; for example, class size at the root node equals [32,34,39], which means that 34 samples belong to class 0, 32 samples belong to class 1, and 39 samples belong to class 2. At the root node, f_3 is the optimal split feature based on the Gini coefficient, and if the value of the third feature of the input sample is no more than 2.35, it moves to the left node. Otherwise, it will move to the right node. Thus, all input samples are divided into two categories. Similarly, the left child nodes and right child nodes then select the optimal split characteristics to split collectively, and the split process will continue repeatedly until the classification pruning requirements with a minimum sample number of 5 are reached. The specific information for the test sample x_6 is shown in Table 1.
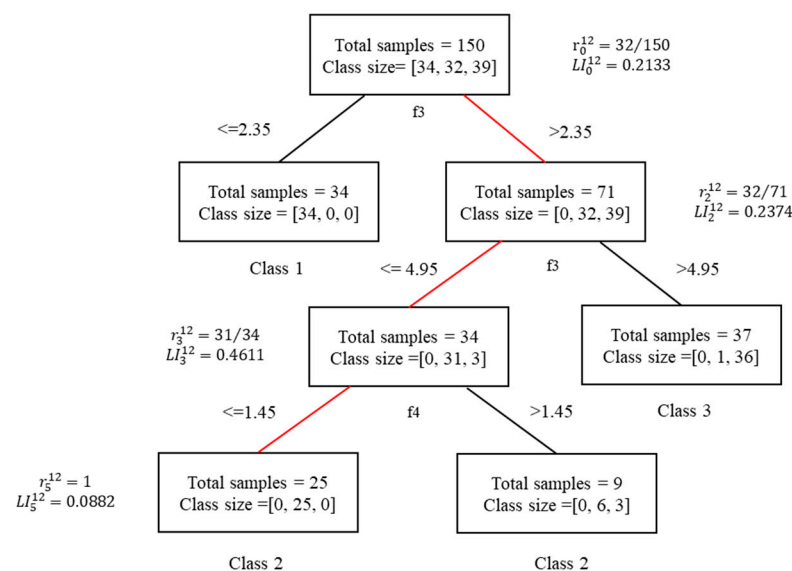

Figure 1. Example of classification of the iris dataset.

Table 1. The feature values of test sample x_6 in the iris dataset.

	f_1	f_2	f_3	f_4
x_6	5.6	2.5	3.9	1.1

x_6 satisfies the following:

$$f_3 = 3.9 > 2.35, \; f_3 = 3.9 \le 4.95, \; f_4 = 1.1 \le 1.45$$

Therefore, the decision path will be recorded as $n_0 \to n_2 \to n_3 \to n_5$ as shown in the red path in Figure 1. The corresponding decision nodes set is denoted as $\mathcal{DP} = \{n_0, n_2, n_3, n_5\}$. The final predicted sample category of sample x_6 is 2, and the local importance of each node in the decision path of the sample x_6 can be calculated according to Equation (1).

$$
\begin{aligned}
n_0 &: LI_0^{12}(y = 2) = r_0^{12} = \tfrac{32}{150} = 0.2133 \\
n_2 &: LI_2^{12}(y = 2) = r_2^{12} - r_0^{12} = \tfrac{32}{71} - \tfrac{32}{150} = 0.2374 \\
n_3 &: LI_3^{12}(y = 2) = r_3^{12} - r_2^{12} = \tfrac{31}{34} - \tfrac{32}{71} = 0.4611 \\
n_5 &: LI_5^{12}(y = 2) = r_5^{12} - r_3^{12} = \tfrac{25}{25} - \tfrac{31}{34} = 0.0882
\end{aligned}
\tag{2}
$$

where $LI_k^t(y = 2)$ represents the local importance of the node k on the decision path of x_6 in the decision tree t with predicted label $y = 2$. All the results are shown in Figure 1 where $LI_3^{12} > LI_2^{12} > LI_0^{12} > LI_5^{12}$ indicates that in the classification process of sample x_6 in the 12th decision tree, node three is the most important with the largest increase in sample purity compared to the parent node. Node two has the second highest importance, and node five has the worst importance. It can be inferred that the data missing on node three will lead to the most significant error in the prediction result.

The corresponding classification result can be obtained on each decision tree similarly to the above procedure for each sample. The idea of the ensemble algorithm is to obtain the final classification results from the classification results of all weak sub-classifiers according to certain voting rules. In the RF model, there are the following two voting rules: soft voting rule and hard voting rule. A brief introduction is given.

Hard voting

Under the hard voting rule, each decision tree can give the classification result of a specific sample. According to the majority voting law, the result with the most occurrences is the final classification result. Its expression is as follows:

$$
H_{\text{RF,hard}}(x) = \underset{y \in \mathcal{Y}}{\arg\max} \left\{ \sum_{t=1}^{T} v_t(x, y) \right\}
\tag{3}
$$

$$
v_t(x, y) = \begin{cases} 1, f_t(x) = y \\ 0, f_t(x) \neq y \end{cases}
\tag{4}
$$

where x represents the test sample. Label value y belongs to $\mathcal{Y}$, and $\mathcal{Y}$ represents a collection of all possible classification results. $f_t(x)$ represents the classification result of decision tree t for x. $v_t(x, y)$ is the voting variable and $H_{\text{RF,hard}}(x)$ represents the final classification results based on the hard voting.

Soft voting

The average value of the probabilities of all sub-classifiers predicting samples to be a certain category is used as the deciding criterion, and the category with the highest probability is selected as the final classification result. Its expression is as follows:

$$
H_{\text{RF,soft}}(x) = \underset{y \in \mathcal{Y}}{\arg\max} \left\{ \frac{1}{T} \sum_{t=1}^{T} P_t(x, y) \right\}
\tag{5}
$$

where $P_t(x, y)$ represents the probability of the prediction label belonging to y in the decision tree t.

2.2. Improved RF Algorithm Based on Decision Path

The traditional RF classification algorithm uses decision trees such as CART as the base learner and applies a certain voting rule to summarize the results of all base learners, so RF has good stability. In this way, even if there is a small amount of missing data in the sample, it will only affect the decision-making process and results of several specific decision trees, and the fault diagnosis results obtained generally remain accurate. Therefore,

RF has strong robustness to a small amount of missing data caused by machine failure or bad operating conditions. When a large number of data are missing, it is common to fill the dataset with default or mean values. However, this processing method may change the distribution of data, introduce system bias, and thus reduce the classification accuracy of the RF algorithm.

Therefore, in order to make the most of the information available, to alleviate the adverse effect of the common filling methods on the RF when a large number of data go missing, this paper utilizes the decision path information to improve the accuracy of RF on the sample containing missing data.

According to the discussion of the decision path above, when sample x lacks several features, which are in the decision path corresponding to sample x in decision tree t, the classification results predicted by the decision tree are considered unreliable, that is, the missing feature nodes included in the decision path will reduce the credibility of the corresponding prediction results. To describe the adverse influence, the reliability score (RS) for the prediction results of each decision tree for a specific sample is defined. As is shown in the concept of local importance, the higher the local importance of a node containing missing data, the greater the adverse effect on the result. Since missing data occurs randomly, a decision path may consist of multiple split nodes that contain the same or different missing data, and the classification bias gets larger as the number of problem nodes increases. The sum of the local importance of all problem nodes in a decision path can indicate the degree of data loss that is unfavorable to the classification result. Considering that the length of the decision path of different samples in different decision trees is not the same, in order to compare the relative weakening effect on the classification accuracy, the ratio of the sum of the local importance of the reliable nodes to that of all nodes in the decision path can be calculated in the case of missing data. The reliability score of decision tree t can be denoted with respect to sample x as follows:

$$RS_t(x,y) = 1 - \frac{\sum_{k' \in \mathcal{DP}_{\text{miss}}(t,x)} LI_{k'}^t(y)}{\sum_{k \in \mathcal{DP}(t,x)} LI_k^t(y)} \tag{6}$$

where $\mathcal{DP}(t,x)$ and $\mathcal{DP}_{\text{miss}}(t,x)$ are a collection of all the nodes of the decision path in the decision tree t, and a collection of all the corresponding missing nodes, respectively.

The iris data are still used as an example to illustrate how RS is calculated. By analysis in Section 2.1, the test sample x_6 in the 12th decision tree passes through $n_0 \to n_2 \to n_3 \to n_5$, and the predicted result is 2. If f_3 is missing in x_6, then the nodes n_0 and n_2 containing the missing feature f_3 are unreliable, as shown in Figure 2.

In the RF, the length of the decision path of the same sample x is not the same in different decision trees. In order to measure the RS of the decision trees with different decision path lengths, all the RS in the RF are normalized in this paper as follows:

$$RS_t^{\text{norm}}(x,y) = \frac{RS_t(x,y) - RS_{\min}(x,y)}{RS_{\max}(x,y) - RS_{\min}(x,y)} \tag{8}$$

where $RS_{\min}(x,y)$ and $RS_{\max}(x,y)$ represent the minimum and maximum values of RS values for all decision trees in a random forest, respectively, as follows:

$$RS_{\min}(x,y) = \min_{t' \in \{1,\dots T\}} RS_{t'}(x,y) \tag{9}$$

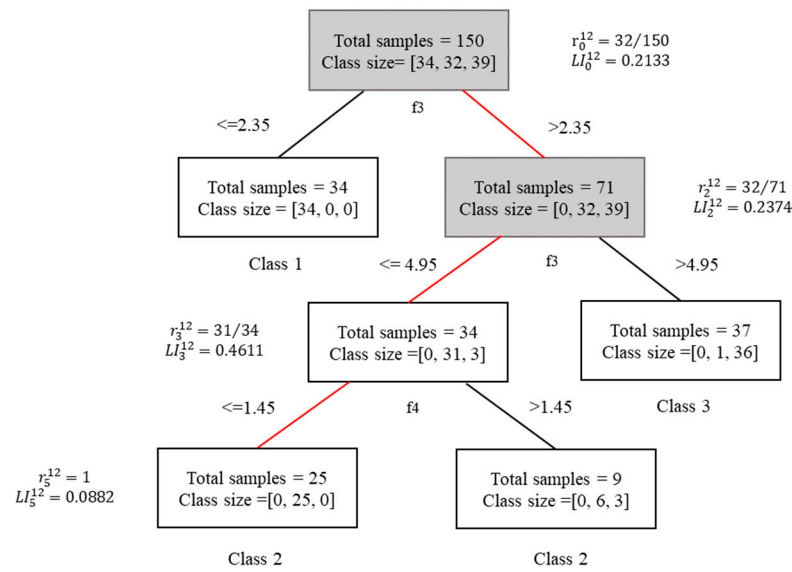$$RS_{\max}(x,y) = \max_{t' \in \{1,\dots T\}} RS_{t'}(x,y) \tag{10}$$

Figure 2. The process of classifying iris dataset with missing data.

According to the definition, the RS value of the 12th decision tree for sample x_6 is as follows:

$$RS_{12}(x = x_6, y = 2) = \frac{(0.2133 + 0.2374 + 0.4611 + 0.0882) - (0.2133 + 0.2374)}{(0.2133 + 0.2374 + 0.4611 + 0.0882)} = 0.5493 \tag{7}$$

For cases with missing data, this paper retains the original prediction results of each classifier because although some data in the test samples go missing, remaining data are still valid and retain important information. Besides, this paper uses reliability scores to revise the prediction results according to missing data contained in the decision path. The specific amendment process will be discussed according to two different voting rules.

When applying hard voting, the test sample x is inputted, and each decision tree can obtain the predicted label value y and the corresponding reliability scores, denoted as $RS_t^{norm}(x, y)$. For different classifiers with the same prediction label, the sum of their reliability scores is used to obtain the total reliability predicted in the category. The label with the highest total reliability scores in all categories is taken as the final prediction result.

$$H_{DPRF,hard}(x) = \underset{y \in \mathcal{Y}}{argmax} \left\{ \sum_{t=1}^{T} v_t(x, y) \cdot RS_t^{norm}(x, y) \right\} \tag{11}$$

When applying soft voting, the test sample x is inputted, and each decision tree outputs the predicted probability of each category and the reliability scores of the classifier under the condition of missing data. The prediction probability of each category is multiplied by the reliability scores to obtain the revised predicted probability of each category, and the revised prediction probabilities of all sub-classifiers in the RF are added to obtain the total prediction of each category probability. The category with the highest predicted probability is taken as the final classification result of the sample x.

$$H_{DPRF,hard}(x) = \underset{y \in \mathcal{Y}}{argmax} \left\{ \frac{1}{T} \sum_{t=1}^{T} P_t(x, y) \cdot RS_t^{norm}(x, y) \right\} \tag{12}$$

The soft voting rule summarizes the probabilities of all categories in all decision trees before voting, while the hard voting rule first votes in each decision tree and then aggregates the results. The former rule can better retain the classification probability information of each decision tree to the final result, while the latter one first votes in the decision tree, losing the probability information too early. Therefore, the diagnosis accuracy is theoretically higher under the soft voting rule.

By repeating the above procedures, the classification results for all the samples can be obtained.

The proposed DPRF algorithm is summarized as follows:

Stage 1: data collection and model training

(1). Collecting data: collect datasets for X and Y: $X_{N \times D} = \{(x_{1,1}, \dots, x_{1,D}), \dots, (x_{N,1}, \dots, x_{N,D})\}$ and $Y_{N \times 1} = \{y_1, \dots, y_N\}$;

(2). Generating training and testing sets: Split the datasets of X and Y into training and testing sets: (X_{train}, Y_{train}) and (X_{test}, Y_{test}), of which the sample sizes correspond to N_{train} and N_{test}, respectively. Note that X_{train} and Y_{train} do not contain any missing data;

(3). Training RF model: train the RF with complete data (X_{train}, Y_{train});

(4). Obtaining nodes' LI scores: calculate the LI scores for all the nodes on the trees in the RF according to the above Equation (1).

Stage 2: classification with incomplete data

(1). Identifying decision paths for test sample: consider a test sample x_{test} in X_{test} with incomplete data, for each tree f_t in the RF, identify the corresponding decision path and decision nodes set by $\mathcal{DP}(t, x)$ and $\mathcal{DP}_{\text{miss}}(t, x)$;

(2). Calculating reliability scores for decision trees: for all the trees in the RF, calculate their normalized RS scores according to Equations (6) and (8);

(3). Obtaining final classification result for x_{test}: calculate the final weighted result according to the hard or soft voting procedures in Equations (11) and (12).

In short, the DPRF model firstly uses the complete industrial process dataset to train the RF model. Then, for the actual process samples that may contain missing data, DPRF automatically determines the location of the missing features of the sample, and completes the calculation of the decision paths, reliability scores and final FDD results on all decision trees. DPRF uses the information retained in the dataset to correct the judgment of RF, and alleviates the influence of subjective factors.

3. Case Study of the Tennessee Eastman Process

This section tests DPRF, together with other classic FDD algorithms in the famous Tennessee Eastman process, to examine the performance of DPRF and compare it with other methods [43]. At the same time, other fault diagnosis methods, such as the back propagation neural network (BP), deep belief neural network (DBN), and radial basis function (RBF) neural network, are also used to compare diagnostic performance. In this paper, we use overall accuracy (OACC) [44,45] and label-based accuracy (LACC) [45,46] to evaluate the performance of each classification algorithm in detecting and correctly classifying faults. The expressions of OACC and LACC are as follows:

$$LACC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{13}$$

$$OACC = \frac{1}{|\mathcal{Y}|} \sum_{y=1}^{|\mathcal{Y}|} LACC_y \tag{14}$$

where $|\mathcal{Y}|$ is the number of elements in collection $\mathcal{Y}$, that is, the total number of labels. TP_i is the number of data that are predicted as fault i, which is actual fault i, while TN_i is the number of data that are predicted as fault i, which is not fault i, indeed. FP_i is the number of data that are predicted as normal or other faults, which is actual fault i, while FN_i is the number of data that are predicted as normal or other faults, which is not that case, indeed.

It can be observed that LACC represents the classification accuracy rate on each classification label, while OACC represents the average classification accuracy rate on all the labels.

3.1. Tennessee Eastman Process Description

The Tennessee Eastman (TE) benchmark chemical process was first introduced by Downs and Vogel in 1993 [47], to evaluate the performance of all kinds of process monitoring and fault diagnosis methods [48–50]. The TE process is constructed with the following five major operation units: a reactor, a product condenser, a vapor–liquid separator, a recycle compressor, and a product stripper. A total of 11 manipulated variables and 41 measured variables are contained in the TE process, and the data can be attributed to 21 fault types. These 21 fault types are detailed in Table 2. The TE process is shown in the flowchart in Figure 3.

Table 2. Process faults for Tennessee Eastman process. Reprinted with permission from ref. [51]. Copyright 2019 Elsevier.

Fault	Process Variable	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV (2)	B composition, A/C feed ratio constant (stream 4)	Step
IDV (3)	D feed temperature (stream 2)	Step
IDV (4)	Reactor cooling water inlet temperature	Step
IDV (5)	Condenser cooling water inlet temperature	Step
IDV (6)	A feed loss (stream 1)	Step
IDV (7)	C header pressure loss-reduced availability (stream 4)	Step
IDV (8)	A, B, and C feed compositions (stream 4)	Random variation
IDV (9)	D feed temperature (stream 2)	Random variation
IDV (10)	C feed temperature (stream 4)	Random variation
IDV (11)	Reactor cooling water inlet temperature	Random variation
IDV (12)	Condenser cooling water inlet temperature	Random variation
IDV (13)	Reaction kinetics	Slow shift
IDV (14)	Reactor cooling water valve	Sticking
IDV (15)	Condenser cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown
IDV (21)	Valve position constant (stream 4)	Constant position

3.2. Results

The training set of the standard dataset for the TE process (http://web.mit.edu/braatzgroup/links.html, accessed on 14 September 2021) contains 500 samples that are generated from the normal state, and 480 samples that are generated from each fault state. For the testing set, there are 960 samples in the normal state and each fault state. For the testing set of each fault type, the fault is introduced at the 160th sample. Thus, the first 160 samples are in the normal state and the last 800 samples are in the corresponding fault state [53]. Because many FDD algorithms have been tested using a standard dataset by other researchers, it is convenient to compare the results to the general methods by experimenting on the same standard dataset.

In recent fault diagnosis studies, methods based on neural networks and deep learning have emerged [54–57]. BP, which was developed in 1986, by scientists led by Rumelhart and McClelland, is a multi-layer feed-forward neural network that is trained by error reverse propagation algorithms and is the most widely used neural network. DBN consists of several restricted Boltzmann machine (RBM) layers and a BP layer, and its unsupervised pre-training can extract high-level abstract representations from the input data [58–60]. RBF is often used to approximate multi-dimensional surfaces by a linear combination of radial symmetric functions, based on Euclidean distance. In recent years, RBF has been used in FDD [61–64]. This section compares the proposed DPRF with these methods.
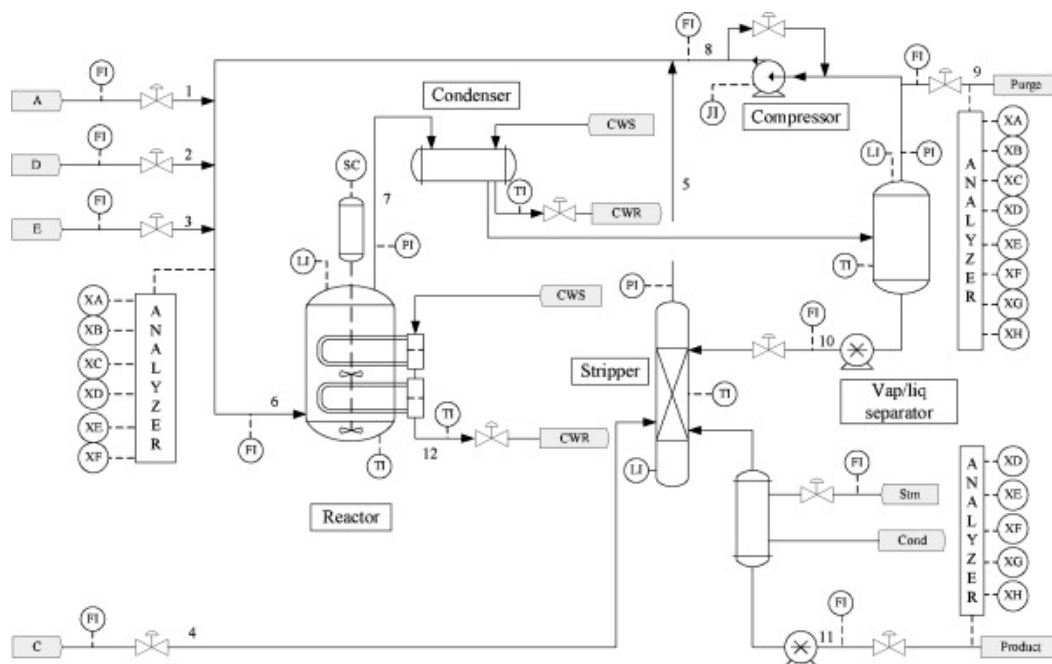
Figure 3. Process flowsheet of TE process. Reprinted with permission from ref. [52]. Copyright 2012 Elsevier.

Some previous research on missing data modelling created loss based on the following three missing patterns: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [65,66]. However, it is sometimes difficult to distinguish the missing mechanisms. This paper proposes a random data loss construction method in the entire test set, randomly setting missing positions on the test dataset matrix $D_{16,800 \times 52}$ at the missing rates $mr = 0.1, 0.2, 0.3, 0.4$, with corresponding amounts of missing data 1680, 3360, 5040, and 6720, respectively. This method does not rely on the three common patterns of missing data. It has better compatibility in various situations. Such a setting also makes the position and number of missing features differ among different samples, making it easier to examine different FDD algorithms under various missing situations. Besides, the amount of missing data in this setting is larger, in which case many FDD algorithms tend to fail. From the discussion in the previous section, DPRF can flexibly search for the missing locations sample by sample, automatically calculate the decision paths and corresponding missing nodes of each sample on all decision trees, and then obtain the RS of each decision tree based on the importance of the missing nodes. Therefore, DPRF can automatically adapt to situations where different locations and different numbers of sample data go missing, without manual intervention. DPRF has better universality compared with other FDD algorithms.

The OACC values of BP, DBN, RBF, the traditional RF algorithm, and its improved version DPRF at each missing rate are plotted in Figure 4, and the details are listed in Table 3. When the test set is complete, the RS of all the decision trees in RF and DPRF are equal to one, that is, there is no need to modify the prediction results. In this case, the OACC and LACC of DPRF are the same as traditional RF, and both algorithms perform well, with OACC values as high as 66.07% under the soft voting rule and 65.64% under the hard voting rule. From the experimental results, the OACC value obtained by the soft voting rule is higher than that obtained by the hard voting rule at the same missing data rate in either RF or DPRF. This experimental result is consistent with the conclusion in the theoretical analysis.
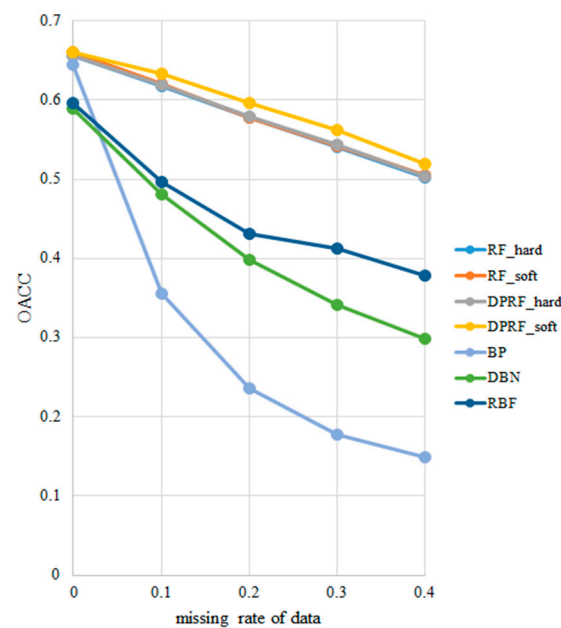
Figure 4. Comparison of OACC values at different missing rates.

Table 3. OACC values of the seven FDD algorithms.

Missing Rate	RF$_{hard}$	RF$_{soft}$	DPRF$_{hard}$	DPRF$_{soft}$	BP	DBN	RBF
0	0.6564	0.6607	0.6564	0.6607	0.6452	0.5892	0.5967
0.1	0.6179	0.6201	0.6187	0.6340	0.3561	0.4809	0.4964
0.2	0.5784	0.5783	0.5791	0.5963	0.2361	0.3986	0.4308
0.3	0.5411	0.5422	0.5429	0.5616	0.1777	0.3410	0.4130
0.4	0.5027	0.5052	0.5043	0.5199	0.1485	0.2981	0.3780

DPFR based on hard voting, and RF based on both hard and soft voting have approximately the same test results at missing rates of 0.1, 0.2, 0.3, and 0.4 in the standard dataset. The curves of the three algorithms roughly coincide. However, it can be observed from Table 3 that DPRF is better than RF under the hard voting rule. In contrast, the DPRF algorithm based on the soft voting rule is significantly better than other algorithms, with OACC values that are 1.4%, 1.8%, 1.94%, and 1.47%, respectively, higher than that of the RF algorithm under the same rule. The OACC values of BP, DBN, and RBF fluctuate significantly when missing data occurs.

4. Discussion

In order to compare the stability of each algorithm when data are lost, the robustness evaluation function is defined as follows:

$$Robustness_{i,mr} = \begin{cases} 4, \ if \ |ratio_{i,mr}| \leq 0.05 \\ 3, \ if \ 0.05 < |ratio_{i,mr}| \leq 0.1 \\ 2, if \ 0.1 < |ratio_{i,mr}| \leq 0.2 \\ 1, if \ |ratio_{i,mr}| > 0.2 \end{cases} \tag{15}$$

$$Robustness_{i,mr} = \frac{OACC_{i,mr} - OACC_{i,mr-0.1}}{OACC_{i,mr}} \tag{16}$$

$$i \epsilon \{RF_{hard}, RF_{soft}, DPRF_{hard}, DPRF_{soft}, BP, DBN, RBF\}, mr = 0.1, 0.2, 0.3, 0.4$$

where $OACC_{i,mr}$ represents the overall accuracy rate of algorithm i at the missing rate mr. $ratio_{i,mr}$ represents the relative change rate of OACC for algorithm i, between the

missing rate *mr* and the previous missing rate $(mr - 0.1)$. When the missing rate gradually increases, the relative change rate of the OACC of an algorithm increases, indicating that the robustness of the algorithm is poor at this stage, as shown by the lower robustness score obtained. Otherwise, the smaller the relative change rate of the overall accuracy rate, the better the robustness of the algorithm at this stage, thus the higher the robustness score obtained. Therefore, the robustness scores are set, ranging from 1 to 5, as shown in the above formula.

Based on the function, we evaluate the stability of all seven FDD algorithms in this paper, for samples containing missing data. The results are listed in Table 4.

Table 4. Robustness scores for the seven algorithms.

Missing Rate	RF$_{hard}$	RF$_{soft}$	DPRF$_{hard}$	DPRF$_{soft}$	BP	DBN	RBF
0.1	3	3	3	4	1	2	2
0.2	3	3	3	3	1	2	2
0.3	3	3	3	3	1	2	4
0.4	3	3	3	3	2	2	3
Average	3	3	3	3.25	1.25	2	2.75

As can be observed from Table 4, DRPF based on soft voting rules has the best robustness and the strongest adaptability to data loss, while BP is the most sensitive to data loss. The average robustness scores of DPRF based on the hard voting rule is three, which is the same as that of RF based on both soft and hard voting rules. These algorithms have good stability for data loss. RBF's average robustness score is up to 2.75, and its performance is the best among the three neural network models, but is inferior to that of the DRPF and RF models. This feature is qualitatively revealed in Figure 4, where BP's OACC curve is the steepest, with the lowest robustness score, while the OACC curves of RF and DPRF are relatively flat, with the robustness scores ranking high.

As mentioned above, OACC represents the average classification accuracy rate on all labels. In order to further compare the detection performance of the FDD algorithms on various fault labels, it is necessary to compare the LACC values of each fault. Since RF and DPRF both perform well in terms of OACC, this section only compares the LACC values of the two algorithms on various faults. Table 5 shows the specific results of LACC at the missing rate 0.1, under the soft voting rule. The other results are plotted in Figures 5 and 6. Figure 5 shows that for the hard voting rule, the diagnostic accuracy rate of DPRF in each category is generally higher than that of RF at the same missing rate. As the missing rate increases, the accuracy of DPRF and RF in each category decreases, but DPRF still has an advantage in diagnostic accuracy compared to RF. Similarly, under the soft voting rule in Figure 6, the LACC of DPRF is improved in most categories, compared to that of RF. Moreover, in Figure 5, it is clear that the sensitivity of each label to the missing data is different. As the missing rate increases for some categories, such as 7, 14, and 17, the accuracy rate sharply decreases, while for other categories, such as 6, 16, and 20, the accuracy rate only fluctuates slightly.
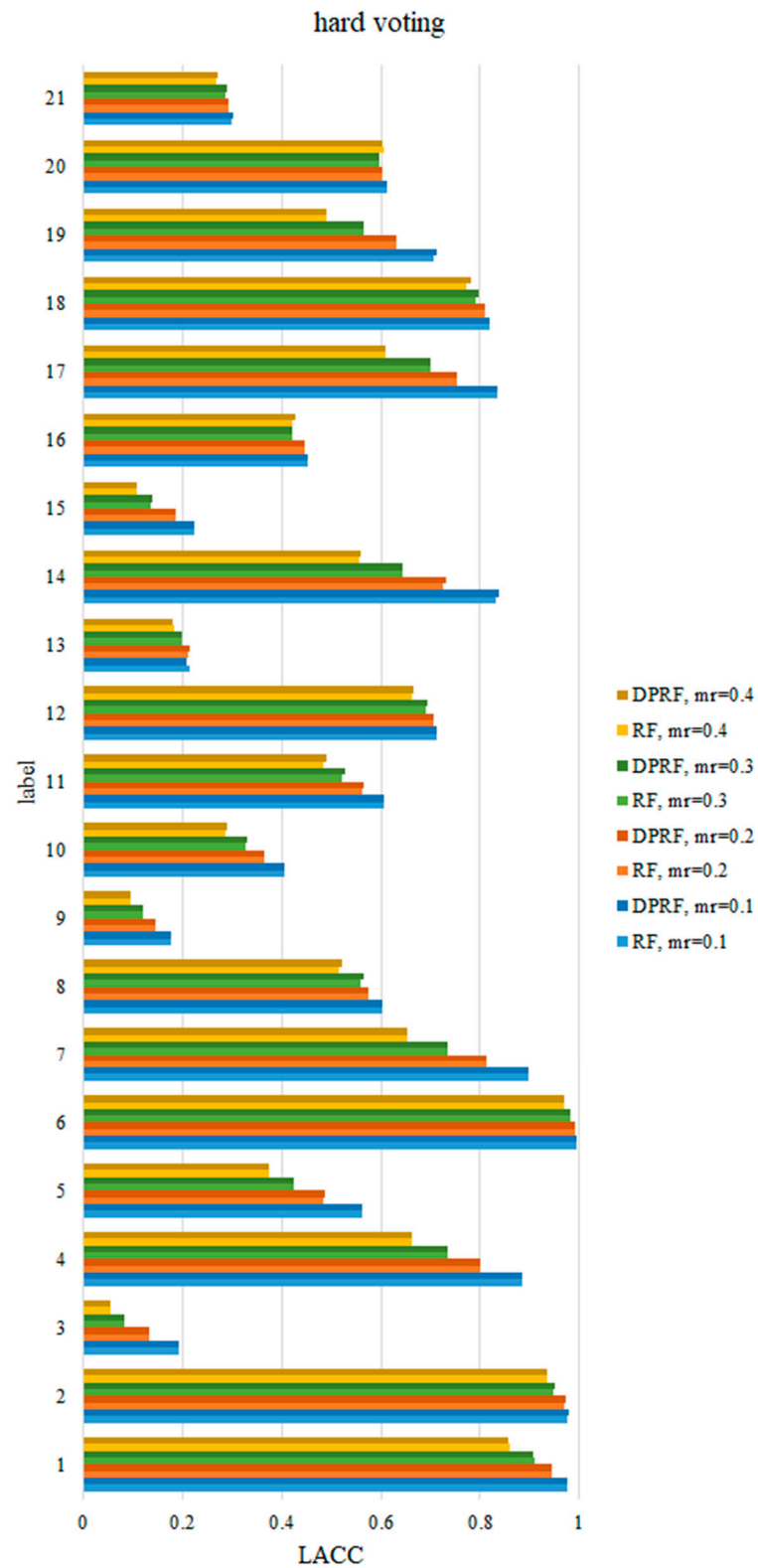
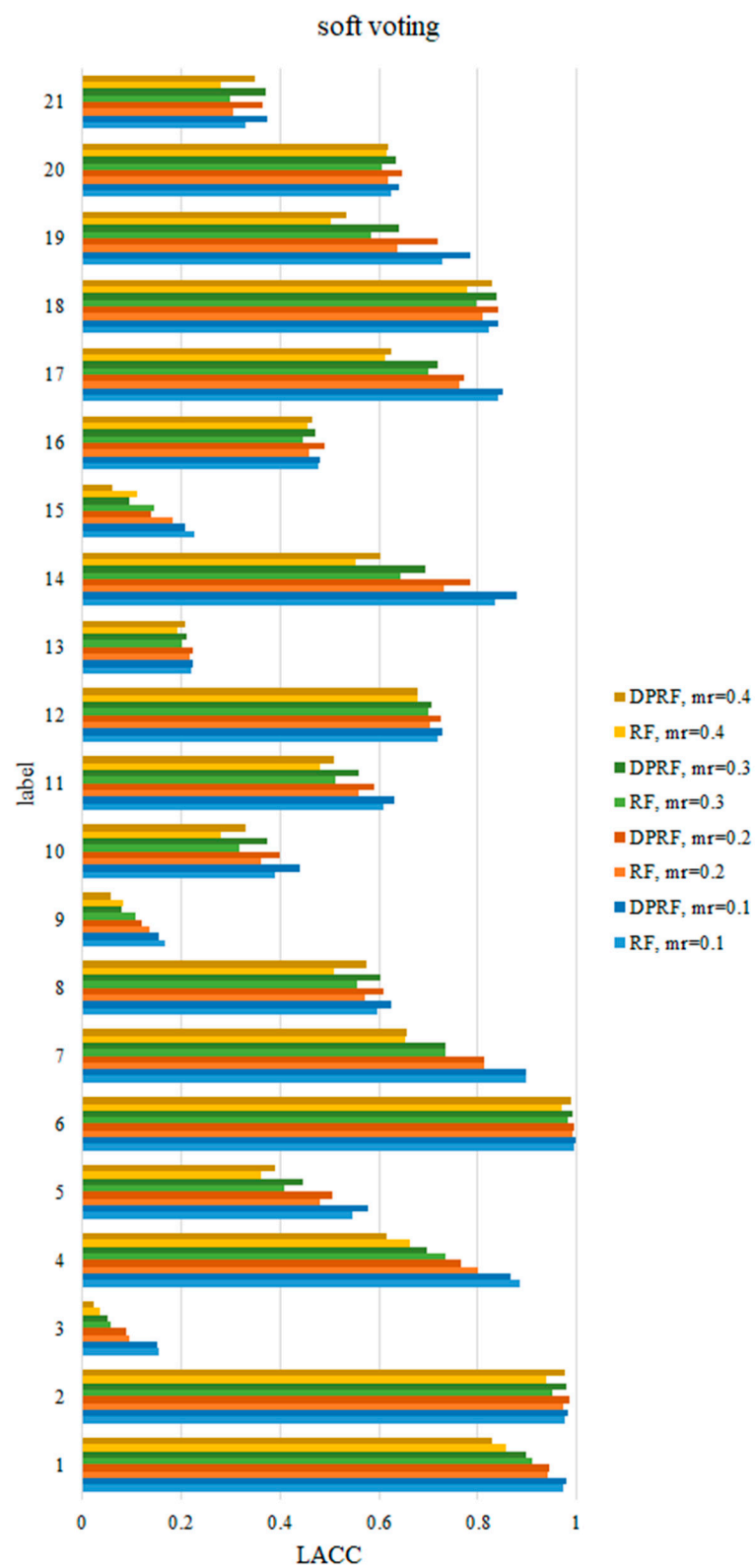**Figure 5.** Comparison of LACC of RF and DPRF at different missing rates based on hard voting.

Figure 6. Comparison of LACC of RF and DPRF at different missing rates based on soft voting.

Table 5. LACC values of RF and DPRF based on soft voting rules (missing rate = 0.1).

Fault Type	RF$_{soft}$	DPRF$_{soft}$
1	0.9738	0.9775
2	0.9763	0.9813
3	0.1550	0.1525
4	0.8863	0.8675
5	0.5475	0.5775
6	0.9950	0.9963
7	0.8988	0.8988
8	0.5963	0.6263
9	0.1675	0.1538
10	0.3888	0.4388
11	0.6075	0.6313
12	0.7200	0.7275
13	0.2200	0.2238
14	0.8350	0.8800
15	0.2275	0.2063
16	0.4788	0.4813
17	0.8400	0.8513
18	0.8225	0.8425
19	0.7275	0.7850
20	0.6263	0.6400
21	0.3313	0.3750

In order to quantify the comparison of the LACC of each category on the two algorithms more clearly, Table 6 lists the summary of the LACC of RF and DPRF under soft voting rules at different missing rates. Taking the first line of the table as an example, at the missing rate 0.1, there are 16 categories where the LACC value of DPRF is greater than or equal to that of RF among 21 fault categories, and the LACC value of DPRF is less than that of RF in the remaining 5 categories.

Table 6. Summary of LACC of RF and DPRF based on soft voting rules.

Missing Rate	RF$_{soft}$	DPRF$_{soft}$
0.1	5	15
0.2	4	17
0.3	5	16
0.4	6	15

It can be concluded from Table 6 that for about 70% of the categories under the soft voting rule, the LACC of DPRF is higher than that of RF, which indicates that the proposed DPRF algorithm improves the diagnosis effects of most of the fault categories, and, finally, promotes the OACC. It means that DPRF has better adaptability to different types of fault diagnoses. Figure 7 shows the LACC improvement of 21 types of faults using DPRF. Figure 7a depicts the absolute change value of LACC, while Figure 7b depicts the relative change value of LACC. When the LACC value is not improved by DPRF, the calculation result is a negative number filled with red. The more negative the result, the darker the red. Otherwise, the calculation result is a positive number filled with blue. The more positive the result, the darker the blue.

fault index	mr=0.1	mr=0.2	mr=0.3	mr=0.4
1	0.38	0.25	-1.13	-3.00
2	0.50	1.13	3.00	3.88
3	-0.25	-0.50	-0.63	-1.25
4	-1.87	-3.25	-3.75	-4.75
5	3.00	2.63	4.00	2.75
6	0.12	0.25	0.75	1.88
7	0.00	0.00	0.12	0.25
8	3.00	4.00	4.88	6.63
9	-1.38	-1.75	-2.88	-2.50
10	5.00	3.75	5.63	4.88
11	2.37	3.38	4.50	2.88
12	0.75	2.25	0.75	-0.12
13	0.38	0.63	1.13	1.38
14	4.50	5.12	5.25	5.25
15	-2.13	-4.38	-5.00	-4.88
16	0.25	3.13	2.50	1.00
17	1.13	1.00	1.63	1.37
18	2.00	3.00	4.25	5.12
19	5.75	8.12	5.88	3.13
20	1.38	2.75	2.88	0.25
21	4.38	6.13	7.00	6.88

(**a**)

fault index	mr=0.1	mr=0.2	mr=0.3	mr=0.4
1	0.39	0.27	-1.24	-3.50
2	0.51	1.16	3.16	4.13
3	-1.61	-5.26	-11.11	-34.48
4	-2.12	-4.07	-5.10	-7.18
5	5.48	5.47	9.82	7.59
6	0.13	0.25	0.76	1.94
7	0.00	0.00	0.17	0.38
8	5.03	7.02	8.78	13.05
9	-8.21	-12.84	-26.44	-29.85
10	12.86	10.38	17.65	17.33
11	3.91	6.05	8.78	5.97
12	1.04	3.20	1.07	-0.18
13	1.70	2.89	5.62	7.14
14	5.39	7.00	8.17	9.52
15	-9.34	-23.97	-34.48	-44.32
16	0.52	6.79	5.60	2.20
17	1.34	1.31	2.32	2.24
18	2.43	3.70	5.34	6.58
19	7.90	12.72	10.09	6.22
20	2.20	4.44	4.75	0.41
21	13.21	20.16	23.33	24.44

(**b**)

Figure 7. Details of LACC improvement by DRPF under the soft voting rule. (**a**) Absolute change value of LACC, (**b**) relative change value of LACC (%).

Most of the areas in Figure 7 are blue. However, the diagnosis effect of the 3, 4, 9, and 15 types of faults have not been improved under all the missing rates calculated by DPRF, and when the missing rates reach 0.3 and 0.4, the diagnostic effect of the first type of faults is also unable to be improved, despite the introduction of the new algorithm. Previous studies [46,67] have concluded that the detection of faults 3, 9, and 15 is very difficult, as there are not any observable changes in the means, variance, or the peak time. Therefore, these three faults cannot always be detected by any statistical technique. Random forest is a statistical algorithm based on information entropy, so its low accuracy is reasonable in the diagnosis of these three types of faults, and it is difficult for DPRF to improve the situation. Chiang et al. [68] proposed that the fourth type of fault will only cause the mean and standard deviation of each variable to differ less than 2% between the faulty status and normal operation. This phenomenon makes the observed signal belonging to fault four similar to the signal under the normal state, resulting in the distinction of the fourth type of fault difficult. Therefore, it is difficult to improve the diagnostic accuracy of the fourth type of fault [69]. It is detected that fault one fails to be improved by DPRF if the missing rate is over 0.3; however, the real reason behind the phenomenon is not clear. Further studies should be carried out.

In conclusion, the test on the TE process confirms the DPRF's effectiveness in accuracy and stability in FDD. DPRF can be easily applied to other chemical processes of different amounts of samples and features, and it will outperform other traditional FDD algorithms in cases of incomplete data. This is not surprising, since the proposed method utilizes the correction procedure, which helps to alleviate the adverse impacts of missing data. DPRF is also adaptable to harsh working conditions where large amounts of data are hard to detect, and thus numerous vacancies occur in records, since the experiment shows that DPRF's diagnostic accuracy maintains a high value at the missing rate 0.4. This is definitely useful in the fault diagnosis of extreme cases, such as large-scope instrument failure.

5. Conclusions

This paper proposes an improved RF algorithm based on the decision path named DPRF. Firstly, an ensemble classifier with a fixed internal tree structure is constructed on the basis of the traditional RF algorithm. Next, for each test sample input to RF, the classifier returns a specific decision path, which contains features for splitting to reach the final classification result. Then, according to the prediction results, the local importance for the nodes and reliability scores is calculated, and the original prediction results are revised. Finally, two different voting rules are used for obtaining the final classification results.

In addition to the DPRF algorithm under the two voting rules, five other FDD strategies are applied, including the RF algorithm based on the hard voting rule, the RF algorithm based on the soft voting rule, BP, DBN, and RBF, to show the effectiveness of the DPRF algorithm. Compared with other methods, DPRF shows strong stability when the dataset is incomplete. When the data missing rate is set from 0.1 to 0.4, DPRF always performs better than the traditional RF, especially under the soft voting rule, with the OACC value increasing by more than 1%. The diagnostic accuracy rate of DPRF is far ahead of that of BP, DBN, and RBF. With the increase in the missing rate, the proposed DPRF algorithm also maintains a high robustness score, and its average robustness score, reaching 3.25, is the best among the compared algorithms. Additionally, DPRF has excellent performance in terms of the diagnostic accuracy rate of local categories, when comparing RF and DPRF under the same voting rule and missing rate. Therefore, DPRF has improved the overall accuracy, robustness, and label accuracy compared to the traditional RF algorithm. When applied in actual chemical processes, where missing data are ubiquitous, DPRF has advantages in accurately detecting and diagnosing faults, and maintains its applicability in the case of large amounts of missing data.

The correction coefficient is a common approximation strategy that is applied in the engineering field. This article draws on this idea to introduce the reliability coefficient, consequently improving the diagnostic accuracy. In the future, when encountering other abnormal data, except for missing data, other existing machine learning algorithms can also be modified with the idea of correction. However, it should be mentioned that the proposed DPRF relies on the choice of modified strategy, i.e., different definitions of correction coefficients will generate different results. Besides, the computational time of DPRF is longer compared to the other four FDD algorithms, due to the embedding of the iterative correction procedure. These shortcomings will limit the practical application of this method in complex chemical processes. In future studies, we will discuss how to combine DPRF with feature extraction strategies to reduce the computation cost, which will help to effectively diagnose in industrial chemical processes.

Author Contributions: Conceptualization, Y.Z., L.L. and X.J.; methodology, Y.Z. and Y.D.; software, Y.Z. and L.L.; validation, X.J., L.L. and Y.D.; formal analysis, Y.Z.; investigation, Y.Z.; resources, Y.D.; data curation, L.L.; writing—original draft preparation, Y.Z.; writing—review and editing, L.L. and Y.D.; visualization, Y.Z. and L.L.; supervision, X.J.; project administration, Y.D.; funding acquisition, X.J. and Y.D. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions, e.g., privacy or ethical.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclatures

T	the total number of decision trees in the random forest
$\mathcal{Y}$	a collection of all target labels y
t	decision tree t
x	single sample
y	the predicted label of x in the random forest, $y \in \mathcal{Y}$
$r_k^t(y)$	the ratio of samples belonging to label y on node k in decision tree t
$LI_k^t(y)$	the local importance of node k in decision tree t with predicted label y the prediction of sample x in decision tree t
$f_t(x)$	
$v_t(x,y)$	the binary voting variable of decision tree t on whether the sample prediction equals to y
$P_t(x,y)$	the possibility that the sample prediction equals to y in decision tree t
$RS_k(x,y)$	the reliability score calculated by the decision path sample x passes in decision tree t
$RS_{min}(x,y)$, $RS_{max}(x,y)$	the minimum and maximum values of RS in all decision trees
$H_{RF,hard}$, $H_{RF,soft}$, $H_{DPRF,hard}$, $H_{DPRF,soft}$	the voting results of *RF* and *DPRF* under the hard and soft voting rule, respectively
$\mathcal{DP}(t,x)$, $\mathcal{DP}_{miss}(t,x)$	the collection of all the nodes of the decision path in decision tree t, and the collection of all the corresponding missing nodes, respectively.

References

1. Zhang, J.X.; Luo, W.J.; Dai, Y.Y. Integrated Diagnostic Framework for Process and Sensor Faults in Chemical Industry. *Sensors* **2021**, *21*, 822. [CrossRef]
2. Zhu, H.B.; He, Z.M.; Wei, J.H.; Wang, J.Q.; Zhou, H.Y. Bearing Fault Feature Extraction and Fault Diagnosis Method Based on Feature Fusion. *Sensors* **2021**, *21*, 2524. [CrossRef]
3. Alauddin, M.; Khan, F.; Imtiaz, S.; Ahmed, S. A Bibliometric Review and Analysis of Data-Driven Fault Detection and Diagnosis Methods for Process Systems. *Ind. Eng. Chem. Res.* **2018**, *57*, 10719–10735. [CrossRef]
4. Liu, S.Y.; McDermid, J.A.; Chen, Y.T. A Rigorous Method for Inspection of Model-Based Formal Specifications. *IEEE Trans. Reliab.* **2010**, *59*, 667–684. [CrossRef]
5. Nor, N.M.; Hassan, C.R.C.; Hussain, M.A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems. *Reviews in Chemical Engineering* **2020**, *36*, 513–553. [CrossRef]
6. Nor, N.M.; Hussain, M.A.; Hassan, C.R.C. Multi-scale kernel Fisher discriminant analysis with adaptive neuro-fuzzy inference system (ANFIS) in fault detection and diagnosis framework for chemical process systems. *Neural Comput. Appl.* **2020**, *32*, 9283–9297. [CrossRef]
7. Shu, Y.D.; Ming, L.; Cheng, F.F.; Zhang, Z.P.; Zhao, J.S. Abnormal situation management: Challenges and opportunities in the big data era. *Comput. Chem. Eng.* **2016**, *91*, 104–113. [CrossRef]
8. Zhang, H.; Chen, H.; Guo, Y.; Wang, J.; Li, G.; Shen, L. Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on PCA and preprocessed by combined clustering. *Appl. Therm. Eng.* **2019**, *160*. [CrossRef]
9. Musleh, A.S.; Debouza, M.; Khalid, H.M.; Al-Durra, A. IEEE. Detection of False Data Injection Attacks in Smart Grids: A Real-Time Principle Component Analysis. In Proceedings of the 45th Annual Conference of the IEEE Industrial Electronics Society (IECON), Lisbon, Portugal, 14–17 October 2019; pp. 2958–2963.
10. Cai, L.F.; Tian, X.M. A new fault detection method for non-Gaussian process based on robust independent component analysis. *Process. Saf. Environ.* **2014**, *92*, 645–658. [CrossRef]
11. Rad, M.A.A.; Yazdanpanah, M.J. Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee Eastman process. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 149–157. [CrossRef]
12. Shi, X.; Yang, H.; Xu, Z.; Zhang, X.; Farahani, M.R. An Independent Component Analysis Classification for Complex Power Quality Disturbances With Sparse Auto Encoder Features. *IEEE Access* **2019**, *7*, 20961–20966. [CrossRef]
13. Hong, Y.; Kim, M.; Lee, H.; Park, J.J.; Lee, D. Early Fault Diagnosis and Classification of Ball Bearing Using Enhanced Kurtogram and Gaussian Mixture Model. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4746–4755. [CrossRef]
14. Yoo, Y.-J. Fault Detection Method Using Multi-mode Principal Component Analysis Based on Gaussian Mixture Model for Sewage Source Heat Pump System. *Int. J. Control Autom. Syst.* **2019**, *17*, 2125–2134. [CrossRef]
15. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [CrossRef]
16. Hoang, D.-T.; Kang, H.-J. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cogn. Syst. Res.* **2019**, *53*, 42–50. [CrossRef]

17. Wang, H.; Li, S.; Song, L.; Cui, L. A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. *Comput. Ind.* **2019**, *105*, 182–190. [CrossRef]

18. Dai, Y.Y.; Zhao, J.S. Fault Diagnosis of Batch Chemical Processes Using a Dynamic Time Warping (DTW)-Based Artificial Immune System. *Ind. Eng. Chem. Res.* **2011**, *50*, 4534–4544. [CrossRef]

19. Zhao, J.S.; Shu, Y.D.; Zhu, J.F.; Dai, Y.Y. An Online Fault Diagnosis Strategy for Full Operating Cycles of Chemical Processes. *Ind. Eng. Chem. Res.* **2014**, *53*, 5015–5027. [CrossRef]

20. Shu, Y.D.; Zhao, J.S. Fault Diagnosis of Chemical Processes Using Artificial Immune System with Vaccine Transplant. *Ind. Eng. Chem. Res.* **2016**, *55*, 3360–3371. [CrossRef]

21. Holst, C.A.; Lohweg, V. A Redundancy Metric Set within Possibility Theory for Multi-Sensor Systems. *Sensors* **2021**, *21*, 2508. [CrossRef]

22. Chen, D.; Yang, S.; Zhou, F. Transfer Learning Based Fault Diagnosis with Missing Data Due to Multi-Rate Sampling. *Sensors* **2019**, *19*, 1826. [CrossRef] [PubMed]

23. Zhang, Y.; Liu, Y.; Chao, H.C.; Zhang, Z.J.; Zhang, Z.Y. Classification of Incomplete Data Based on Evidence Theory and an Extreme Learning Machine in Wireless Sensor Networks. *Sensors* **2018**, *18*, 1406. [CrossRef] [PubMed]

24. Askarian, M.; Escudero, G.; Graells, M.; Zarghami, R.; Jalali-Farahani, F.; Mostoufi, N. Fault diagnosis of chemical processes with incomplete observations: A comparative study. *Comput. Chem. Eng.* **2016**, *84*, 104–116. [CrossRef]

25. Dong, Y.R.; Peng, C.Y.J. Principled missing data methods for researchers. *SpringerPlus* **2013**, *2*, 1–17. [CrossRef] [PubMed]

26. Sharpe, P.K.; Solly, R.J. Dealing with Missing Values in Neural-Network-Based Diagnostic Systems. *Neural Comput. Appl.* **1995**, *3*, 73–77. [CrossRef]

27. Gabrys, B. Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems. *Int. J. Approx. Reason.* **2002**, *30*, 149–179. [CrossRef]

28. Llanes-Santiago, O.; Rivero-Benedico, B.C.; Galvez-Viera, S.C.; Rodriguez-Morant, E.F.; Torres-Cabeza, R.; Silva-Neto, A.J. A Fault Diagnosis Proposal with Online Imputation to Incomplete Observations in Industrial Plants. *Rev. Mex. De Ing. Quim.* **2019**, *18*, 83–98. [CrossRef]

29. Wang, Z.; Zhang, X. Fuzzy Set-Valued Information Systems and the Algorithm of Filling Missing Values for Incomplete Information Systems. *Complexity* **2019**, *2019*. [CrossRef]

30. Usman, K.; Kamaruddin, M.; Chamidah, D.; Saleh, K.; Eliskar, Y.; Marzuki, I. Modified Possibilistic Fuzzy C-Means Algorithm for Clustering Incomplete Data Sets. *Acta Polytech.* **2021**, *61*, 364–377. [CrossRef]

31. Beaulac, C.; Rosenthal, J.S. BEST: A decision tree algorithm that handles missing values. *Comput. Stat.* **2020**, *35*, 1001–1026. [CrossRef]

32. Shah, A.D.; Bartlett, J.W.; Carpenter, J.; Nicholas, O.; Hemingway, H. Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *Am. J. Epidemiol.* **2014**, *179*, 764–774. [CrossRef]

33. Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost. *IEEE Access* **2018**, *6*, 21020–21031. [CrossRef]

34. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front. Aging Neurosci.* **2017**, *9*. [CrossRef] [PubMed]

35. Soltaninejad, M.; Zhang, L.; Lambrou, T.; Yang, G.; Allinson, N.; Ye, X. MRI Brain Tumor Segmentation and Patient Survival Prediction Using Random Forests and Fully Convolutional Networks. In Proceedings of the 3rd International Workshop on Brain-Lesion (BrainLes) held jointly at the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI), Quebec City, QC, Canada, 14 September 2017; pp. 204–215.

36. Nussbaum, M.; Spiess, K.; Baltensweiler, A.; Grob, U.; Keller, A.; Greiner, L.; Schaepman, M.E.; Papritz, A. Evaluation of digital soil mapping approaches with large sets of environmental covariates. *Soil* **2018**, *4*, 1–22. [CrossRef]

37. Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci. Rep.* **2018**, *8*, 663. [CrossRef] [PubMed]

38. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

39. Talekar, B.; Agrawal, S. A Detailed Review on Decision Tree and Random Forest. *Biosci. Biotechnol. Res. Commun.* **2020**, *13*, 245–248. [CrossRef]

40. Fang, W.; Zhou, J.; Li, X.; Zhu, K.Q. Unpack Local Model Interpretation for GBDT. In Proceedings of the 23rd International Conference on Database Systems for Advanced Applications, Gold Coast, Australia, 21–24 May 2018.

41. Hatwell, J.; Gaber, M.M.; Azad, R.M.A. CHIRPS: Explaining random forest classification. *Artif. Intell. Rev.* **2020**, *53*, 5747–5788. [CrossRef]

42. Palczewska, A.; Palczewski, J.; Robinson, R.M.; Neagu, D. Interpreting random forest models using a feature contribution method. In Proceedings of the 2013 IEEE 14th International Conference on Information Reuse & Integration (IRI), San Francisco, CA, USA, 14–16 August 2013.

43. Park, P.; Di Marco, P.; Shin, H.; Bang, J. Fault Detection and Diagnosis Using Combined Autoencoder and Long Short-Term Memory Network. *Sensors* **2019**, *19*, 4612. [CrossRef]

44. Gao, X.; Hou, J. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing* **2016**, *174*, 906–911. [CrossRef]

45. Wang, Y.L.; Yang, H.B.; Yuan, X.F.; Shardt, Y.A.W.; Yang, C.H.; Gui, W.H. Deep learning for fault-relevant feature extraction and fault classification with stacked supervised auto-encoder. *J. Process. Control* **2020**, *92*, 79–89. [CrossRef]
46. Yin, S.; Gao, X.; Karimi, H.R.; Zhu, X.P. Study on Support Vector Machine-Based Fault Detection in Tennessee Eastman Process. *Abstr. Appl. Anal.* **2014**. [CrossRef]
47. Downs, J.J.; Vogel, E.F. A Plant-Wide Industrial-Process Control Problem. *Computers & Chemical Engineering* **1993**, *17*, 245–255. [CrossRef]
48. Krishnannair, S. Fault Detection of Tennessee Eastman Process using Kernel Dissimilarity Scale Based Singular Spectrum Analysis. In Proceedings of the 13th International-Federation-of-Automatic-Control (IFAC) Workshop on Adaptive and Learning Control Systems (ALCOS), Winchester, UK, 4–6 December 2019; pp. 204–209.
49. Onel, M.; Kieslich, C.A.; Pistikopoulos, E.N. A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process. *Aiche J.* **2019**, *65*, 992–1005. [CrossRef]
50. Mingxuan, L.; Yuanxun, S. Deep Compression of Neural Networks for Fault Detection on Tennessee Eastman Chemical Processes. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 March 2021; pp. 476–481.
51. Fazai, R.; Mansouri, M.; Abodayeh, K.; Nounou, H.; Nounou, M. Online reduced kernel PLS combined with GLRT for fault detection in chemical systems. *Process Saf. Environ.* **2019**, *128*, 228–243. [CrossRef]
52. Yin, S.; Ding, S.X.; Haghani, A.; Hao, H.Y.; Zhang, P. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *J. Process Control* **2012**, *22*, 1567–1581. [CrossRef]
53. Wang, Y.L.; Wu, D.Z.; Yuan, X.F. LDA-based deep transfer learning for fault diagnosis in industrial chemical processes. *Comput. Chem. Eng.* **2020**, *140*, 106964. [CrossRef]
54. Lee, K.B.; Cheon, S.; Kim, C.O. A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes. *IEEE Trans. Semicond. Manuf.* **2017**, *30*, 135–142. [CrossRef]
55. Guan, Z.Y.; Liao, Z.Q.; Li, K.; Chen, P. A Precise Diagnosis Method of Structural Faults of Rotating Machinery based on Combination of Empirical Mode Decomposition, Sample Entropy, and Deep Belief Network. *Sensors* **2019**, *19*, 591. [CrossRef]
56. Li, A.Y.; Yang, X.H.; Dong, H.Y.; Xie, Z.H.; Yang, C.S. Machine Learning-Based Sensor Data Modeling Methods for Power Transformer PHM. *Sensors* **2018**, *18*, 4430. [CrossRef]
57. Li, G.Q.; Deng, C.; Wu, J.; Xu, X.B.; Shao, X.Y.; Wang, Y.H. Sensor Data-Driven Bearing Fault Diagnosis Based on Deep Convolutional Neural Networks and S-Transform. *Sensors* **2019**, *19*, 2750. [CrossRef]
58. Chen, R.; Yuan, Y.; Zhang, Z.; Chen, X.; He, F.; IOP. Fault Diagnosis for Transformers Based on FRVM and DBN. In Proceedings of the 4th International Conference on Advances in Energy Resources and Environment Engineering (ICAESEE), Chengdu, China, 7–9 December 2019.
59. Chen, X.-M.; Wu, C.-X.; Wu, Y.; Xiong, N.-x.; Han, R.; Ju, B.-B.; Zhang, S. Design and Analysis for Early Warning of Rotor UAV Based on Data-Driven DBN. *Electronics* **2019**, *8*, 1350. [CrossRef]
60. Su, X.; Cao, C.; Zeng, X.; Feng, Z.; Shen, J.; Yan, X.; Wu, Z. Application of DBN and GWO-SVM in analog circuit fault diagnosis. *Sci. Rep.* **2021**, *11*, 7969. [CrossRef] [PubMed]
61. Tran, D.A.T.; Chen, Y.M.; Chau, M.Q.; Ning, B.S. A robust online fault detection and diagnosis strategy of centrifugal chiller systems for building energy efficiency. *Energy Build.* **2015**, *108*, 441–453. [CrossRef]
62. Xiaoqin, L.; Haijun, S.; Mengchan, W.; Jiayao, Y. Based on Rough Set and RBF Neural Network Power Grid Fault Diagnosis. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *300*, 042113. [CrossRef]
63. Yang, L.; Su, L.; Wang, Y.; Jiang, H.; Yang, X.; Li, Y.; Shen, D.; Wang, N. Metal Roof Fault Diagnosis Method Based on RBF-SVM. *Complexity* **2020**, *2020*, 1–12. [CrossRef]
64. Wang, J.; Zhang, S.; Hu, X. A Fault Diagnosis Method for Lithium-Ion Battery Packs Using Improved RBF Neural Network. *Front. Energy Res.* **2021**, *9*, 418. [CrossRef]
65. Little, T.D.; Jorgensen, T.D.; Lang, K.M.; Moore, E.W.G. On the Joys of Missing Data. *J. Pediatric Psychol.* **2014**, *39*, 151–162. [CrossRef]
66. Zhu, J.; Ge, Z.; Song, Z.; Gao, F. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annu. Rev. Control* **2018**, *46*, 107–133. [CrossRef]
67. Chiang, L.H.; Russell, E.L.; Braatz, R.D. *Fault Detection and Diagnosis in Industrial Systems*; Springer: London, UK, 2001.
68. Russell, E.L.; Chiang, L.H.; Braatz, R.D. *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*; Springer: London, UK, 2000.
69. Li, H.; Xiao, D.-Y. Fault diagnosis of Tennessee Eastman process using signal geometry matching technique. *EURASIP J. Adv. Signal Process.* **2011**, *2011*, 1–19.