IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 28, 2019, accepted January 2, 2020, date of publication January 14, 2020, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966520

Hybrid Feature Selection for Wafer Acceptance Test Parameters in Semiconductor Manufacturing

HONGWEI XU[1], JIE ZHANG[1], YOULONG LV[1], AND PENG ZHENG[2]
[1]College of Mechanical Engineering, Donghua University, Shanghai 201620, China
[2]School of Mechanical Engineering, Institute of Intelligent Manufacturing and Information Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
Corresponding author: Jie Zhang (mezhangjie@dhu.edu.cn)

This work was supported in part by the National Natural Science Foundation, China, under Grant 51435009, in part by the China Postdoctoral Science Foundation under Grant 2018M641890, in part by the Shanghai Sailing Program under Grant 18YF1400800, and in part by the National Engineering and Research Center for Commercial Aircraft Manufacturing, China, under Grant COMAC-SFGS-2018-36175.

ABSTRACT Wafer acceptance test (WAT) is a key process of semiconductor manufacturing. The collected testing parameters can be used in identification of wafer defects, improvement of product yield, and control of production costs. However, WAT parameters regularly have characteristics such as high dimensions and strong redundancy, which prevent the wafer yield from accurate prediction and effective improvement. To overcome these shortcomings, a hybrid feature selection method is proposed to identify key WAT parameters influencing wafer yields. This method is composed of two stages, i.e. filter selection and wrapper selection. In filter selection, the minimum Redundancy Maximum Relevance (mRMR) filtering parameter pre-screening criterion based on mutual information (MI) is proposed. The relevance between each parameter and the wafer yield value is calculated by MI. At the same time, the criterion of MI is used to measure the redundancy between each parameter to select the minimum redundancy parameters, and reduce feature size for further searches. In wrapper selection, a wrapped key parameter identification model based on genetic algorithm (GA) and deep belief network (DBN) is designed. The coding and optimization of candidate input parameters are realized by GA. The wafer yield prediction error value of the DBN and the weight of the selected features are solved as the fitness function to realize the selection process of the combined parameters. In experiment, both testing data sets and industrial data are used to demonstrate the efficiency of this proposed method.

INDEX TERMS Hybrid feature selection, wafer acceptance test parameters, semiconductor manufacturing, minimal redundancy maximal relevance, genetic algorithm, deep belief network.

I. INTRODUCTION

Semiconductor manufacturing is one of the most important industries in the world [1]. Among the processes of semiconductor manufacturing, quality control is significant for its cost saving and in-time delivery. [2]. FIGURE 1 illustrates the procedure of quality control in semiconductor manufacturing, which includes the defect detections during manufacturing, the WAT after all manufacturing processes and the circuit probing (CP) process for each grain on the wafer. Among these quality control steps, the CP process determines the wafer yield, but this process needs tremendous time spent on expensive and specialized equipment. The prediction of wafer yield based on WAT parameters is therefore used by engineers to reduce manufacturing time and production costs spent on CP process [3], [4].

As the size of integrated circuits continues decreasing and the processing technology becomes more complicated, the number of parameters that needs to be tested is gradually increased during the WAT process, the corresponding time consumption and test cost are increased at the same times [5]. In view of the large amount of WAT parameters, the relationship between parameters is complex and the redundancy problem is prominent. In addition, the key parameters are difficult to be obtained [6]. In which case, traditional methods based on statistical process control (SPC) are limited in large-scale parameter identification and automatic identification [7]. Therefore, it is significant for quickly discovering failure wafers and improving wafer yield to ensure the accuracy of wafer yield prediction during wafer manufacturing.

The associate editor coordinating the review of this manuscript and approving it for publication was Jerry Chun-Wei Lin.

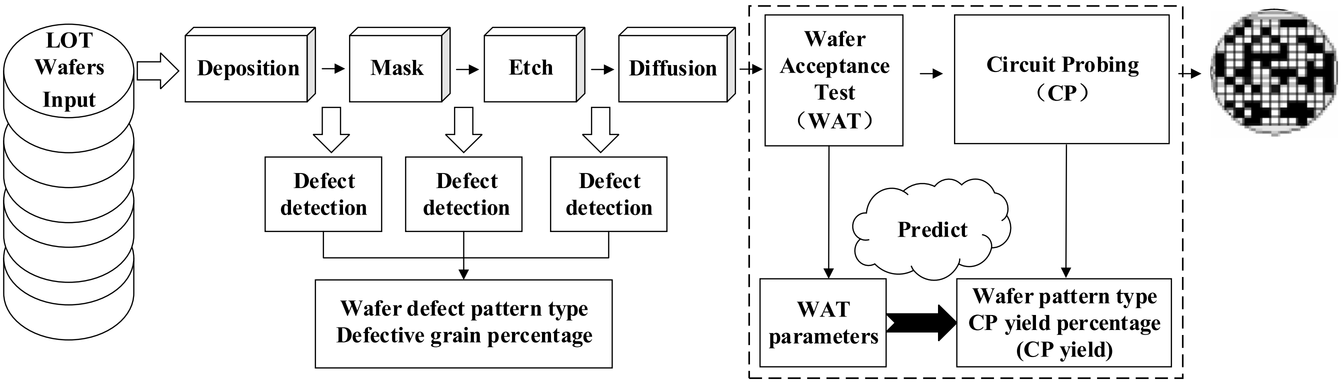


FIGURE 1. Production and quality control flowchart of the wafer.

In the existing literatures, expert experience methods, relevance analysis, principal component analysis (PCA), cluster analysis, information entropy, and heuristic-based analysis methods are used to identify key quality parameters of wafers. Chien *et al.* [8] screened 12 high-relevance WAT factors through expert experience, and then designed an improved analysis method based on modified Partial Least Squares (mPLS), finally, they used these high relevance factors as the model input parameters. However, this method requires expert experience to screen key parameters, and it is difficult for quality analysts with shallow experience to quickly master the skill. Zhang *et al.* [9] used the relevance between WAT parameters and CP yield value to screen out key WAT parameters. Meanwhile, the Backpropagation Neural Network (BPNN) and the General Regression Neural Network (GRNN) models were utilized to realize the establishment of the yield prediction model. However, this method mainly considered the relevance between the single variable and the yield value in the WAT parameter, without considering the relevance between the combined variable and the yield value. Tseng [10] used PCA to reduce the features and dimensions of high-dimensional data in the quality management database, and then used logistic regression analysis to perform data mining. But, the method converts high-dimensional quality parameters into low-dimensional uncorrelated linear comprehensive indicators through PCA, and lost the physical information of the original quality data, and it is difficult to analyze and regulate the quality reasons from the source; Chen *et al.* [11] clustered the high-quality process quality parameters by clustering method, and then inferred the clustering results by the Decision Tree Inference Rules (DTIR). Then, the DTIR was employed to infer the parameter clustering result, thereby constructing a complete wafer manufacturing yield analysis data mining architecture. Since the method uses the decision tree to make decision and analysis, it also needs to learn from the expert experience to set the decision nodes, which restricts the automatic mining ability of key parameters to a certain extent. Wang *et al.* [12] designed a key parameter selection method based on information entropy method, which comprehensively measures the relevance, redundancy and complementarity between parameters. Based on the relevance, redundancy and complementarity between the parameters, a filtering key parameter identification algorithm is proposed to filter out the key parameters that affect the fluctuation of the production cycle. This method usually measures the predictive ability of each feature separately with the high feature selection efficiency, but it cannot effectively measure the predictive ability of noise-sensitive combined variable. Hong [13] designed the Molecular-Inspired Particle Swarm Optimization (MI-PSO) algorithm to analyze the measured WAT parameters to find the best combination of parameters in accordance with the design specifications and circuit layout, which provides critical WAT parameters for wafer failure analysis. The key variable selection method based on MI-PSO algorithm performs better than other heuristic methods in most small datasets, but this method is failed for the dataset with high dimensional characteristics which requires much time cost in running the calculation process [14]. In general, the wafer key parameter identification methods are mainly divided into traditional statistical analysis method, filtering method represented by information entropy, wrapped method represented by heuristic algorithm [15]–[17], and machine learning method represented by PCA. Due to the defects of each method, these methods have great difficulties to automatically and efficiently identify the key WAT parameters of the wafer under the condition that the wafer parameter scale is continuously expanded and the constraint factors are gradually increased.

There are some problems with the existing key parameter identification method, for example, sacrificing the stability of selection model for the relationship between a single parameter and the target, and sacrificing time efficiency for the associated effects of combined parameters on the target [18]. Furthermore, considering the high-dimensional characteristics and the redundancy characteristics [19], this paper proposes a novel feature selection method for identifying key parameters of wafer acceptance test based on Hybrid Feature Selection (HFS).

This paper is organized as follows. Key WAT parameter identification framework based on HFS is presented in Section 2. Single WAT parameter filtering pre-screening

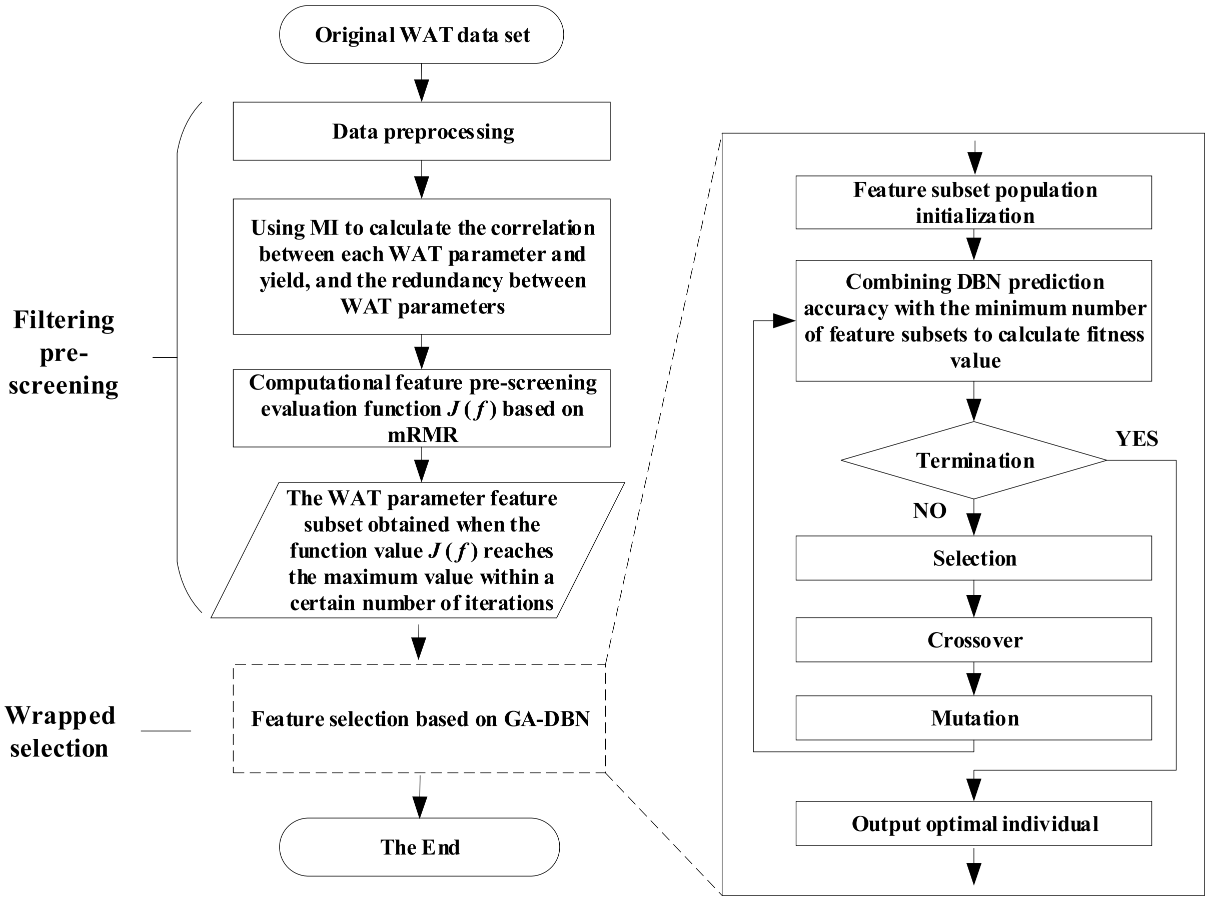


FIGURE 2. Key WAT parameter identification framework based on HFS.

method based on MI is described in Section 3. Combined WAT parameter wrapped selection method based on Genetic Algorithm and Deep Belief Network (GA-DBN) is outlined in Section 4. Experiments on HFS method and real data are discussed in Section 5. Finally, conclusions are given in Section 6.

II. HYBRID FEATURE SELECTION FRAMEWORK

The framework of key parameter identification of WAT based on HFS method can be seen in FIGURE 2. Firstly, data preprocessing is performed on the missing values, outliers, and dimensional differences to obtain the input parameters needed for further data analysis and modeling. Secondly, a filtering parameter pre-screening method based on MI is designed to obtain parameters with high relevance to the yield value. Meanwhile, the relevance between parameters is calculated by MI to reflect the redundancy characteristics between WAT parameters. In addition, the WAT parameters are filtered and pre-screened one by one in combination with mRMR characteristics. Then a wrapped feature selection model based on GA-DBN is established. Taking the prediction accuracy of the DBN model and the number of feature subset as the fitness function, multi-objective optimization is carried out to realize the complex relationship modeling between the combined WAT parameters and the yield, then output the key WAT parameters affecting the wafer yield.

III. FILTERING SELECTION BASED ON RELEVANCE AND REDUNDANCY

A. DATA PREPROCESSING

The WAT process is used to monitor the manufacturing conditions and the quality of products by applying current or voltage on the wafer [20]. Due to equipment shutdown, current surge, etc., the case of missing values, and outliers, etc. exist in the WAT parameters recorded by the enterprise quality management center. For such problems, considering the characteristics of the large amount of data set, the missing and abnormal values in the WAT parameters are not obvious, the process of statistics, screening, and culling are performed to them. For the difference in dimension of each WAT parameter, the data processing method of maximum and minimum normalization [21] is used to map the inconsistent data to the interval of 0-1. With the standardization of data processing and the processing of dimensional differences between parameters, input parameters that are more easily to be obtained for further data analysis and the establishment of prediction model.

B. RELEVANCE ANALYSIS

In the complex manufacturing process of semiconductors, the grain on the wafer may cause functional abnormalities due to certain manufacturing problems, which will directly affect the final wafer yield. Therefore, engineers believe that wafer yields are inextricably linked to some specific WAT parameters. Due to the complexity of the semiconductor manufacturing steps and the complex interaction between the parameters, the WAT parameters and the yield values show complex relevance characteristics. It is difficult for engineers to effectively identify the cause of the abnormality in a short period of time, resulting in the loss of yield. Therefore, a relevance analysis method based on MI is designed in this section, which analyzes all WAT parameters one by one with wafer yield, and then pre-screens WAT parameters having maximum relevance to yield.

MI is a metric method for describing the interdependence between two random variables [22]. For continuous random variables such as WAT parameters and yield values, the MI method as shown in equation (1) is implement to analysis the univariate relevance of each WAT parameter and wafer yield value.

$$I_c(X_i;Y) = \int_Y \int_{X_i} p(x_i, y) \log \left(\frac{p(x_i, y)}{p(x_i)p(y)} \right) dxdy \tag{1}$$

where $p(x_i, y)$ represents the joint probability density function of the current WAT parameter X_i and the yield value Y, $p(x_i)$ and $p(y)$ represents the edge probability density function of current WAT parameter X_i and yield value Y. The result Ic between each WAT parameter and the wafer yield is calculated by the MI criterion, and then the results of MI are reversely arranged to obtain the WAT parameter which are strongly correlated with the wafer yield.

C. REDUNDANCY ANALYSIS

In actual wafer manufacturing process, the integrated WAT parameters are stored in the form of the Mean, Maximum, Minimum, and Standard deviation values. For example, the leakage current values on a wafer will be finally stored in the form of the Mean, Maximum, Minimum, and Standard deviation into the wafer data management system. Therefore, there are duplicate attributes and strong relevance between WAT parameters, showing strong redundancy. However, data redundancy will not only occupy the amount of information storage, but also affect the stability of the establishment with the wafer yield prediction model. Therefore, the MI criterion is carried out to measure the redundancy of WAT parameters, and selects representative key data from many redundant data, in which case, the MI criterion can reduce the dimensionality of data and the data redundancy obviously.

The relevance between each WAT parameter is calculated by MI criterion as shown in equation (2). The higher of the MI value, the higher of the relevance and the more prominent of the redundancy between variables will be [23].

$$I_r(X_i;X_j) = \int_{X_j} \int_{X_i} p(x_i, x_j) \log \left(\frac{p(x_i, x_j)}{p(x_i)p(x_j)} \right) dx_i dx_j \tag{2}$$

where $p(x_\mathrm{i},\ x_\mathrm{j})$ represents the joint probability density function of the current WAT parameter X_i and the WAT parameter X_j, $p(x_\mathrm{i})$ and $p(x_\mathrm{j})$ represent the edge probability density function of current WAT parameter X_i and WAT parameter X_j.

D. MINIMUM REDUNDANCY AND MAXIMUM RELEVANCE

In the WAT test process, due to the large number of parameters required to be tested, the high-dimensional WAT parameter set is finally caused, moreover, it is difficult to effectively estimate the high-dimensional probability density. What's more, the high-dimensional feature selection will take a long time and is inefficient [24]. In order to effectively carry out feature pre-screening, this section starts from the perspective of mRMR [25], and the pre-screening evaluation index of the minimum redundant and maximum relevance filtering parameter pre-screening method based on mutual information (mRMR-MI) is further designed [28], [29], as shown in equation (3).

$$J(f) = I_c(X;Y) - \frac{1}{|S|} \sum_{X_i, X_j \in S} I_r(X_i;X_j) \tag{3}$$

where $J(f)$ is a pre-screening evaluation function based on MI. f is the selected WAT feature parameters, and $f \in X$, $I_c(X;Y)$ represents the MI value of each WAT parameter and wafer yield value, S is the selected subset of feature parameters, $|S|$ indicates the number of feature parameters currently selected, $I_r(X_i;\ X_j)$ represents the MI value between each WAT parameter. The termination condition of the designed mRMR algorithm is that the feature subset f_i corresponding to the evaluation function $J(f)$ reaching the maximum value on the basis of a certain number of iterations. The mRMR algorithm is used to combine the selection criteria of the minimum redundancy and the maximum relevance between the WAT feature parameters and the yield values. Therefore, the method of mRMR-MI can implement the pre-screening process of the WAT parameters. The pseudocode for filtering selection based on relevance and redundancy is shown in algorithm 1.

However, the WAT parameters screened at this time only consider the relevance and redundancy between the single WAT parameter variable and the wafer yield, without effectively reflecting the effect of the combined WAT parameters on the wafer yield. Therefore, it is significant for the next step to analyze the influence of the combined WAT parameter variables on the wafer yield, and obtain the less key WAT parameters affecting the wafer yield.

IV. WRAPPED SELECTION BASED ON GA AND DBN

The GA-DBN selection model is based on the wrapped feature selection model that the subsequent learning algorithm is embedded in the feature selection process. By testing the prediction performance of the combined feature subsets on the algorithm, the selected combination features are evaluated for their merits and demerits. Therefore, a combined WAT parameter selection model based on GA-DBN is

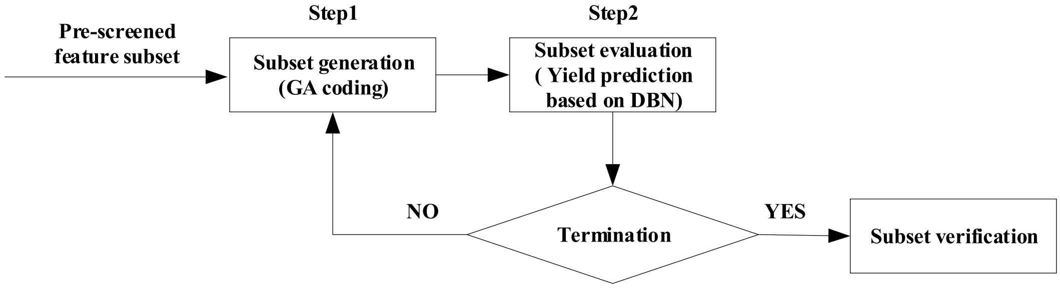


FIGURE 3. Combined WAT parameter selection flow chart based on GA-DBN.

Algorithm 1 Filtering Selection Based on Relevance and Redundancy

```
Input: X = [X_1, X_2,...,X_n] : Original WAT dataset
       Y : Wafer yield value
Output: Pre-screened WAT dataset
Begin
  i ←0 // i: Iterations number
  While (not termination condition) do
    relevance measure I_c(X_i; Y);
    redundancy measure I_r(X_i; X_j);
    fitness eval(X) by J(i); // J(i): mRMR fitness function designed in this paragraph
    if J(i) < J(i+1) then
      select the characteristic parameter corresponding to J(i+1) case;
    end
    i ← i+1;
  end
  output Pre-screened WAT dataset;
End
```

designed as shown in FIGURE 3, This GA-DBN criterion mainly consists of two parts. Step1, using GA algorithm to realize the process of encoding and updating process for the WAT parameters [30]. Step2, using the DBN deep learning model to establish a complex nonlinear mapping relationship between WAT parameters and wafer yield to predict the wafer yield [31], and accurate prediction of wafer yield can be obtained after a limited number of iterative processes.

A. SUBSET GENERATION

The subset generation process refers to generating candidate feature subsets according to a certain search strategy. In this section, the GA based feature subset initialization process is designed, and the initial selection of candidate WAT parameter features is performed by random selection. Moreover, the candidate WAT parameters are encoded in binary encoding to convert the feature variables into chromosomes in GA algorithm. Then, the feature subsets are scored by using the evaluation function, and then the process of global optimal solution is gradually approached according to the heuristic rules. Therefore, the optimal WAT parameter feature subset will be searched.

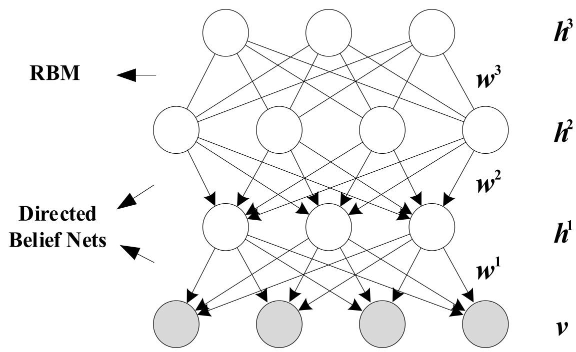


FIGURE 4. DBN structure.

B. SUBSET EVALUATION

Since the characteristics of random algorithm relies on random factors for parameter selection, it is difficult to reproduce when the experimental results of optimal solution were found. Therefore, it is necessary to design the fitness function for the randomness problem in GA algorithm. Combining the accuracy of DBN prediction model and the minimum number of feature subsets as the evaluation index, the multi-objective optimization is carried out. Furthermore, the evaluation of the selected features and the convergence of feature selection results are realized.

As shown in FIGURE 4, the DBN model is a deep learning model consisting of multiple Restricted Boltzmann Machine (RBM) models and BPNN models [32]. The unsupervised feature extraction of the input samples is realized by the structure of multi-layer RBM. Therefore, the information of features and weight values of the samples are obtained. Then, the BPNN is used to supervise the obtained features and the weight information, and the process of predicting the wafer yield by using various input WAT parameters is realized.

Calculating the accuracy of wafer yield prediction and the number of selected features as the calculation criterion for the key WAT parameter selection fitness function. Therefore, it is necessary to design an objective fitness function that can optimize two targets at the same time to solve the problem of multi-objective optimization, and realize the fitness function design with high accuracy and few numbers of features.

$$fitness = W_A \times \text{DBN\_accuracy} + W_F \times \left(\sum_{i=1}^{n} I_c(X_i; Y) \times F_i \right)^{-1} \quad (4)$$

$$F_i = \begin{cases} 0; & \text{Feature } i \text{ is not selected in pre} - \text{screening} \\ 1; & \text{Feature } i \text{ has been selected in pre} - \text{screening} \end{cases} \quad (5)$$

Designing a fitness function as shown in equation (4), the fitness function having two predetermined weights, W_A represents the weight of prediction accuracy in DBN model; W_F indicates the weight of the selected number of features. In the model optimization process, if the prediction accuracy of the DBN model is considered to be the most important, the W_A precision value can be adjusted to 100%, and usually the W_A precision value can be set between 75% and 100% according to the demand, the W_F precision value is usually set between 0% and 25%. The value of F_i is 0 or 1 respectively. When $F_i = 0$, the current feature i is a feature discarded in the mRMR-MI pre-screening; when $F_i = 1$, the current feature i is a feature retained in the mRMR-MI pre-screening. $I_c(X_i; Y)$ represents the MI value between the WAT feature parameter X_i and the wafer yield value Y to measure the degree of importance of the current feature. Through the design of the fitness function, chromosomes with high fitness values can be saved to the next generation as much as possible, so the parameters can also be set according to requirements.

The termination condition of GA-DBN model is that the fitness function fitness reaches the maximum value based on a certain number of iterations. Through the GA-DBN algorithm, the related WAT parameters are associated with the wafer yield in the form of combined parameter features. Therefore, the selection process of the combined WAT parameters is realized. The pseudocode for wrapped selection based on GA and DBN is shown in algorithm 2.

Algorithm 2 Wrapped Selection Based on GA and DBN

Input: $X' = [x_1, x_2, \ldots, x_n]$: Pre-screened WAT dataset
Y' : Yield value corresponding to the pre-screened WAT dataset
GA parameters
W_A: The weight of prediction accuracy in DBN model
W_F : The weight of the selected number of features in the model optimization process
Output: best WAT dataset solution
Begin
$t \leftarrow 0$ // t: generation number
initialize $P(t)$ by **encoding routine**; // $P(t)$: population of chromosomes
fitness $eval(P)$ by **decoding routine**; //$eval(P)$: fitness function designed in this paragraph
While (**not** termination condition) **do**
  crossover $P(t)$ to yield $C(t)$; // $C(t)$: offspring
  mutation $P(t)$ to yield $C(t)$;
  fitness $eval(C)$ by **decoding routine**;
  select $P(t+1)$ from $P(t)$ and $C(t)$;
  $t \leftarrow t+1$;
end
output best WAT dataset solution;
End

V. EXPERIMENTS AND DISCUSSION

This paper designs a key WAT parameter identification method based on HFS. Firstly, the UCI data set is used to verify the validity and reliability of the new model. Secondly, the prediction and selection test of WAT instance data is carried out. Finally, the key WAT parameters of the final selection are analyzed and compared, and the effectiveness of the HFS method is verified.

A. UCI DATASETS

Datasets with the same properties as wafer yield prediction were specifically selected for standard dataset validation. That is, the training and test data are suitable for the input and output of the regression analysis model. The input features have certain high-dimensional characteristics, and the output values are continuous features rather than discrete features. Therefore, after screening analysis, the following four sets of UCI standard data sets were finally chosen to verify the validity of the proposed HFS model. Including low-dimensional Abalone Dataset and Wine Quality Dataset, as well as high-dimensional Residential Building Dataset and UJI Indoor Loc Dataset. Then parameter selection tests were carried out for abalone age, wine quality grade, house price forecast, and residential floor location prediction. The mRMR filtering method and the GA-BPNN wrapped method are used to compare the HFS method. The parameters screened by the mRMR method are used as the input parameters of BPNN model for prediction experiments, therefore, the mRMR-BPNN filtering prediction model is composed.

In order to increase the redundancy characteristics and relevance characteristics in the standard dataset samples, this paragraph expands and enhances the data of these four standard datasets, and adds random noise characteristics to test the stability of the model under different data conditions. The number of original features and the number of random noise features added are as shown in Table 1 below. In the experiment process, the number of features selected in the case of Minimum Average Relative Error (MARE) is used as the index of model evaluation, and the selected features are analyzed to check whether the noise characteristics can be effectively eliminated. The test results are shown in Table 1 below.

It can be seen from Table 1 that all three parameter selection methods have the ability of filtering most noise parameters, and can reduce the input dimension of the prediction model to reduce the operation time. For low-dimensional

TABLE 1. Standard data set feature selection test.

Dimension information	Data set	Number of original features	Increased number of features	MARE			Number of features selected under MARE		
				mRMR-BPNN	GA-BPNN	HFS	mRMR-BPNN	GA-BPNN	HFS
Low dimensional data	Abalone	8	4	0.589	0.342	0.285	9	6	6
	Wine Quality	11	5	0.552	0.413	0.416	13	10	11
High dimensional data	Residential-Building	104	20	0.626	0.309	0.157	110	101	95
	UJI Indoor Loc	528	20	0.515	0.314	0.141	534	520	486

TABLE 2. WAT parameter information.

Serial number:	Parameters	Number of related parameters:
1	Breakdown Voltage with Drain and Source (BVDS）	24
2	Leakage current under device shutdown (IOFF）	24
3	Connection resistance (RCFV）	72
4	Insulation in current mode (SPAFI）	40
5	Saturation current in working condition (IDSAT）	40
6	MOSFET Capacitor (VBG）	32
7	FET turn-on voltage （VTFM1）	8
8	Turn-on voltage in constant current mode (VTI）	16
9	Surface resistance (RSFV）	36
10	Leakage current at the junction of the junction diode (JLEA）	8
11	Electrical thickness of the capacitor in accumulation mode (TXA）	16
12	Linear region threshold voltage (VTI）	48
13	Junction diode connection capacitor (CJ）	8
14	Polymetallic continuity resistance value (CONTI）	40
15	Capacitor leakage current (GLK）	4
16	Substrate current caused by hot electrons (ISUB）	8
17	FET turn-on voltage (VTFPO）	8
	Total number of parameters:	432

data, it is larger for mRMR-BPNN method in MARE to compare with GA-BPNN method and the HFS method, and the number of selected parameters is higher than the other two methods. The number of parameters selected by HFS method is consistent with the GA-BPNN method, and even slightly better than the GA-BPNN method. However, for high-dimensional data, the HFS method can ensure that fewer feature parameters are selected with high accuracy. Therefore, the proposed HFS method has more significant advantages in dealing with the key parameter extraction process of high-dimensional data.

B. CASE STUDY

This paper selects the actual data of a 300mm wafer production line of an enterprise in Shanghai for model analysis and verification. Usually in the wafer production process, one unit is equal to one Lot, and each Lot contains 25 wafers. During the WAT process, 432 parameters on each wafer are tested separately. The total number of wafers is more than 8,000, and the wafers produced in the same batch are divided into 8 groups, numbered from Lot_ID_A to Lot_ID_H. In the following process, 80% of the WAT test data and its corresponding yield value were selected as the training set for supervised regression training, and the remaining 20% of the data was used as the test set for test verification.

1) WAT PARAMETERS

The WAT parameters are electrical test parameters for detecting wafer's circuit device, and the WAT parameters mainly includes an open voltage leakage current, a saturation current, and a breakdown voltage, etc. related to the MOS transistor. Chip resistance, contact resistance, etc. related to resistors and capacitors. Gate oxygen breakdown voltage, gate oxide thickness, etc. related to gate oxide characteristics. The main WAT test objects and the number of related parameters of the test items are shown in Table 2 below.

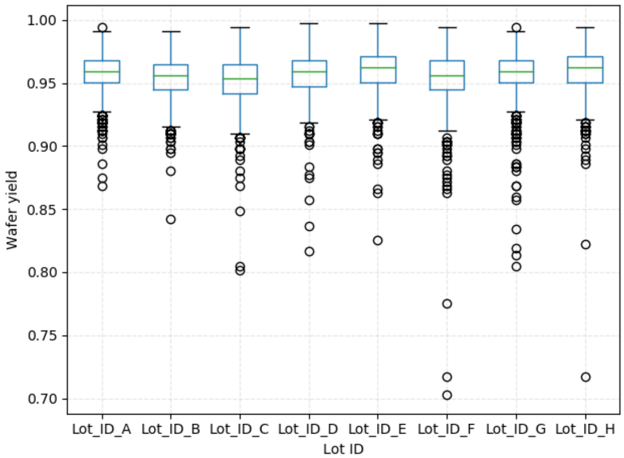


FIGURE 5. Wafer yield statistics in each set of lots.

2) WAFER YIELD VALUES

For semiconductor manufacturers, the later the discovery time of the failed semiconductor device, the greater the cost of the corresponding cost will be. After the CP test process, quality control engineers can determine whether the wafer needs to continue to the next step or discarded. Therefore, the grain that passed the CP test indicates good product, so as to obtain the yield information of the whole wafer, and the yield information of each wafer is stored in the quality management system.

In this paper, we use the box plot to analyze the eight sets of wafer yield values, as shown in Figure 5 below. There are nearly 1000 yield information in each set of data. The median of the yield values is above 0.95, and the 50% yield information value is between the upper and lower quartiles. Moreover, 90% of the data is concentrated between the upper and lower edges of the box plot, only a small amount of information belongs to the exception information. Therefore, for this part of the abnormal data, we have eliminated it before the yield prediction process.

3) RESULTS AND DISCUSSION

a: MODEL PARAMETER SETTING

The hyperparameters mainly involved in the HFS method includes the number of iterations k_1 in the filtering pre-screening process, initial population p, cross operation probability value r, mutation operation probability value m of GA, the number of layers l, the number of nodes in each layer n, and the number of iterations k_2 of the DBN model in wrapper selection process. However, the iteration number k_1 value is calculated by the mRMR fitness function, and the number of iterations when the algorithm convergence tends to be stable is 500 times, so the k_1 value is set to 500 in each test process. The initial population number p value of GA is adaptively set according to the feature pre-screening result, usually the number of pre-screening features is between 130 and 150. The crossover operation probability value r and the mutation operation probability value m are set to 0.8 and 0.01 respectively by referring to the setting methods in the existing literature [25]. For the DBN model, it can be seen from the orthogonal experiment that when the model layer number l is set to 3, the number of input layer nodes is adaptively installed as the number of features after the pre-screening process, the number of nodes in the hidden layer is decremented in the form of an arithmetic progression, and the prediction layer is set to 1, the DBN model can achieve the best prediction accuracy. Moreover, when the number of iterations k_2 reaches 3000 times, the convergence of DBN model tends to be stable, so the number of iterations k_2 set by the prediction process is set to 3000 times.

b: COMPARISON BETWEEN DIFFERENT SELECTION METHODS

The WAT training data and test data are substituted into the mRMR-BPNN model, the GA-BPNN model, and the HFS model. The absolute error comparison between the predicted values and the true values of the eight wafer sets is shown in FIGURE 6 below. Comparing the yield predictions of different Lot wafers, the error of three models in the Lot_ID_A data set are significantly higher than those of other groups, and the main reason is that the wafer of Lot_ID_A batch belongs to the trial production stage of wafer production, and the process of wafer production have not been stabilized. After the process technology are stable, the wafer yield error value is generally reduced. So, the wafer yield prediction error value produced in the middle and late stages is gradually reduced and tends to be stable. However, for each set of Lot wafers, the predicted absolute error values of the three yield prediction models are different obviously, the HFS method has the lowest prediction error value and is better than the other two comparison methods.

c: SCREENING RESULT COMPARATIVE TEST

The HFS method proposed in this paper needs to use the DBN model to predict the yield value. In order to effectively evaluate the prediction effect, we added the Mean Square Error (MSE) and the R^2 value evaluation index based on the MARE indicator. Furthermore, the number of features selected in the case of the minimum MARE is selected as the validity of the feature selection model. Therefore, from the Table 3 we can see that the MARE and MES values of the HFS model are smaller than the other two models, and higher R^2 values can be obtained under the same conditions, so the HFS model has a more stable prediction effect.

At the same time, we separately predict the wafer yields of the eight sets of wafers and record the time consuming on each training and test [33]. As can be seen from the table below, the time consumption of the HFS model is between the other two models. Compared to the filter model, the HFS model takes a relatively long time, but is much smaller than the wrapper model. Therefore, the HFS model has higher credibility from the perspective of time and prediction error.

Since the HFS method is combined by filtering pre-screening and wrapped selection process, when the HFS

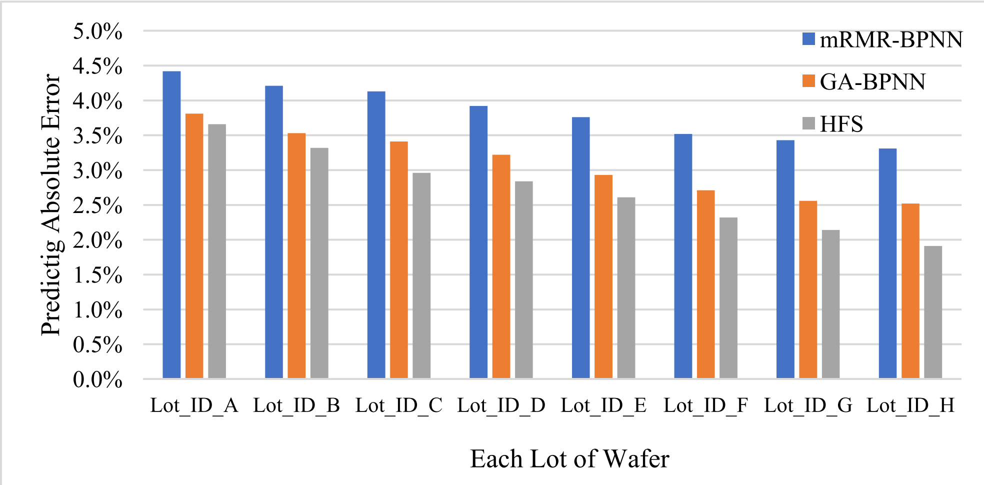


FIGURE 6. Comparison of absolute error prediction for each model of 8 sets of wafers.

TABLE 3. Predictive model accuracy and time complexity.

Lot ID	MARE			MSE			R^2			Time consuming/(s)		
	mRMR-BPNN	GA-BPNN	HFS	mRMR-BPNN	GA-BPNN	HFS	mRMR-BPNN	GA-BPNN	HFS	mRMR-BPNN	GA-BPNN	HFS
Lot_ID_A	0.0442	0.0381	0.0366	0.0512	0.0468	0.0387	0.87	0.91	0.92	160	320	172
Lot_ID_B	0.0421	0.0353	0.0332	0.0498	0.0457	0.0376	0.88	0.89	0.91	165	310	170
Lot_ID_C	0.0413	0.0341	0.0296	0.0468	0.0445	0.0354	0.88	0.9	0.92	161	315	168
Lot_ID_D	0.0392	0.0322	0.0284	0.0443	0.0451	0.0332	0.86	0.91	0.9	156	311	169
Lot_ID_E	0.0376	0.0293	0.0261	0.0421	0.0421	0.0321	0.89	0.92	0.93	158	308	165
Lot_ID_F	0.0352	0.0271	0.0232	0.0430	0.0413	0.0311	0.91	0.91	0.92	154	316	162
Lot_ID_G	0.0343	0.0256	0.0214	0.0426	0.0398	0.0302	0.92	0.92	0.93	153	312	161
Lot_ID_H	0.0331	0.0252	0.0191	0.0415	0.0387	0.0295	0.91	0.93	0.94	152	313	163
Average value	**0.0384**	**0.0309**	**0.0272**	**0.0452**	**0.0430**	**0.0335**	**0.89**	**0.91**	**0.92**	**157**	**313**	**166**

TABLE 4. Number of features of parameter screening in two stages.

Lot ID	Number of original features	Number of features remaining after Filtering pre-screening	Number of features left after Wrapped selection
Lot_ID_A	432	**158**	**62**
Lot_ID_B	432	155	56
Lot_ID_C	432	148	55
Lot_ID_D	432	143	52
Lot_ID_E	432	135	53
Lot_ID_F	432	132	52
Lot_ID_G	432	**131**	**51**
Lot_ID_H	432	**131**	52
Average value	**432**	**142**	**54**

is implemented, the number of features selected by the intermediate pre-screening process can be obtained. Therefore, the number of features pre-screened by the filter model and the number of features remaining through the wrapper selection process are as shown in Table 4 below. we can see that the maximum number of features remaining after pre-screening is 158 parameters, and the minimum is only 131 parameters. Compared with the original 432 WAT parameters, the unrelated noise features can be greatly filtered. However, the number of features remaining after the further wrapped process is 62 at most, and 51 at least. Therefore, from the perspective of the minimum parameter number, the HFS method can effectively ensure this point, and realize the effect of reflecting the actual wafer yield with fewer key parameters. What's more, it can be seen from the analysis of the experimental results that by the HFS method, of the 432 related WAT parameters, only less than one-third of the WAT parameters have an impact on wafer yield. Moreover, the HFS method can achieve smaller prediction error and higher prediction accuracy with the least number of input features, and thus has more significant advantages.

VI. CONCLUSION

In this paper, for the problems of the high dimension of WAT parameters, strong redundancy between data, and the key parameters are difficult to obtained, a WAT parameter identification method based on HFS method is proposed. Based on the design of the mRMR-MI filtering parameter pre-screening process, the GA-DBN model is designed. Relevance analysis of single WAT parameter variable and wafer yield is achieved by the mRMR-MI method, and the relevance effect of combined WAT parameters on wafer yield is realized by GA-DBN model. The experimental comparison and analysis showing the certain superiority.

The contributions of this paper are as follows:

- A filtering parameter pre-screening method based on MI is designed. The WAT parameters are filtered one by one according to mRMR characteristics, and some unrelated features are eliminated, in which case, the dimension of dataset and subsequent calculating time can be greatly reduced.
- A wrapped key parameter identification model based on GA-DBN is designed. The coding and optimization of combined candidate input parameters are realized by GA. Then, the DBN model is used to predict the wafer yield. In which case, the wrapped feature selection of key WAT parameters is implemented in closed loop form.
- The Filtering feature selection method and the wrapped feature selection method are combined to form the HFS method, and the HFS model considers the effect of single WAT parameters and the combined WAT parameters on wafer yield, which can not only improve the time efficiency of the algorithm, but also significantly improve the effect of key parameter recognition.
- The proposed HFS method can effectively filter the noise parameters, and achieve accurate prediction of wafer yield with less key WAT parameters input. Therefore, wafer manufacturers can use this method to predict wafer yields with less key WAT parameters to reduce test damage for wafers and equipment investment.

REFERENCES

[1] Y. Zheng, D. Ling, Y.-W. Wang, S.-S. Jang, and B. Tao, "Model quality evaluation in semiconductor manufacturing process with EWMA run-to-run control," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 8–16, Feb. 2017.

[2] J. Ruth and R. Berndt, "Quality control for ultrafiltration of ultrapure water production for high end semiconductor manufacturing," in *Proc. 27th Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, May 2016, pp. 16–22.

[3] F. Tan, T. Pan, Z. Li, and S. Chen, "Survey on Run-to-run control algorithms in high-mix semiconductor manufacturing processes," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1435–1444, Dec. 2015.

[4] J. Moyne, J. Samantaray, and M. Armacost, "Big data capabilities applied to semiconductor manufacturing advanced process control," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 4, pp. 283–291, Nov. 2016.

[5] Q. Zhu, N. Wu, Y. Qiao, and M. Zhou, "Optimal scheduling of complex multi-cluster tools based on timed resource-oriented petri nets," *IEEE Access*, vol. 4, pp. 2096–2109, 2016.

[6] F. Yang, N. Wu, Y. Qiao, M. Zhou, R. Su, and T. Qu, "Petri net-based efficient determination of optimal schedules for transport-dominant single-arm multi-cluster tools," *IEEE Access*, vol. 6, pp. 355–365, 2018.

[7] D. You, X. Wu, L. Shen, S. Deng, Z. Chen, C. Ma, and Q. Lian, "Online feature selection for streaming features using self-adaption sliding-window sampling," *IEEE Access*, vol. 7, pp. 16088–16100, 2019.

[8] C.-F. Chien, P.-C. Lee, R. Dou, Y.-J. Chen, and C.-C. Chen, "Modeling collinear WATs for parametric yield enhancement in semiconductor manufacturing," in *Proc. 13th IEEE Conf. Autom. Sci. Eng. (CASE)*, Aug. 2017, pp. 739–743.

[9] Z. Bingyu, C. Yutai, and W. Zhenyu, "Modeling of wafer die yield by WAT parameters," *J. Qual.*, vol. 18, no. 6, pp. 519–538, 2011.

[10] W. J. Tseng, "Hybird recursive statistical methods to improve features selection and data mining process," Dept. Ind. Eng. Eng. Manage., Tsinghua Univ., Beijing, China, 2011, pp. 1–48.

[11] C.-F. Chien, W.-C. Wang, and J.-C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 192–198, Jul. 2007.

[12] J. Wang, J. Zhang, and X. Wang, "A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 173–182, Feb. 2018.

[13] H. Chongwei, "Using molecular-inspired particle swarm optimization to solve the design of process parameters in semiconductor test vehicle," Dept. Ind. Eng. Eng. Manage., Tsinghua Univ., Beijing, China, 2011, pp. 1–62.

[14] G. Z. Li, J. Y. Yang, and Y. Q. Zhang, "Feature selection for ensemble learning and its application," *Mach. Learn. Bioinf.*, pp. 135–155, Nov. 2008.

[15] T.-S. Li, C.-L. Huang, and Z.-Y. Wu, "Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system," *J. Intell. Manuf.*, vol. 17, no. 3, pp. 355–361, Jun. 2006.

[16] K. Chen, P.-Y. Chang, and C.-H. Yeh, "Wafer die yield prediction by heuristic methods," in *Proc. 40th Int. Conf. Comput. Ind. Eng.*, Jul. 2010, pp. 1–4.

[17] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 231–240, 2006.

[18] W. Zhang, P. Duan, L. T. Yang, F. Xia, Z. Li, Q. Lu, W. Gong, and S. Yang, "Resource requests prediction in the cloud computing environment with a deep belief network," *Softw. Pract. Exper.*, vol. 47, no. 3, pp. 473–488, Mar. 2017.

[19] C.-F. Chien, Y.-J. Chen, and J.-Z. Wu, "Big data analytics for modeling WAT parameter variation induced by process tool in semiconductor manufacturing and empirical study," in *Proc. Winter Simulation Conf. (WSC)*, Dec. 2016, pp. 2512–2522.

[20] R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, "Supervised feature selection with a stratified feature weighting method," *IEEE Access*, vol. 6, pp. 15087–15098, 2018.

[21] W.-J. Shen and H.-X. Li, "A sensitivity-based group-wise parameter identification algorithm for the electric model of li-ion battery," *IEEE Access*, vol. 5, pp. 4377–4387, 2017.

[22] X. Su, L. Li, F. Shi, and H. Qian, "Research on the fusion of dependent evidence based on mutual information," *IEEE Access*, vol. 6, pp. 71839–71845, 2018.

[23] B. Zhang, T. Xi, X. Gong, and W. Wang, "Mutual information maximization-based collaborative data collection with calibration constraint," *IEEE Access*, vol. 7, pp. 21188–21200, 2019.

[24] J. Huang and X. Yan, "Quality relevant and independent two block monitoring based on mutual information and KPCA," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6518–6527, Aug. 2017.

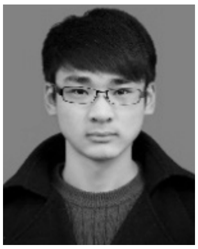
[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[26] L. Chen, X. Pan, Y.-H. Zhang, M. Liu, T. Huang, and Y.-D. Cai, "Classification of Widely and Rarely Expressed Genes with Recurrent Neural Network," *Comput. Struct. Biotechnol. J.*, vol. 17, pp. 49–60, 2019.

[27] J. Li and T. Huang, "Predicting and analyzing early wake-up associated gene expressions by integrating GWAS and eQTL studies," *Biochim. et Biophys. Acta (BBA)-Mol. Basis Disease*, vol. 1864, no. 6, pp. 2241–2246, Jun. 2018.

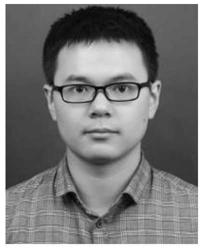
[28] N. Zhang, M. Wang, P. Zhang, and T. Huang, "Classification of cancers based on copy number variation landscapes," *Biochim. et Biophys. Acta (BBA)-Gen. Subjects*, vol. 1860, no. 11, pp. 2750–2755, Nov. 2016.

[29] L. Chen, Y.-H. Zhang, G. Huang, X. Pan, S. Wang, T. Huang, and Y.-D. Cai, "Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection," *Mol. Genet. Genomics*, vol. 293, no. 1, pp. 137–149, Feb. 2018.
[30] H. Takshi, G. Dogan, and H. Arslan, "Joint optimization of device to device resource and power allocation based on genetic algorithm," *IEEE Access*, vol. 6, pp. 21173–21183, 2018.
[31] D. Bibin, M. S. Nair, and P. Punitha, "Malaria parasite detection from peripheral blood smear images using deep belief networks," *IEEE Access*, vol. 5, pp. 9099–9108, 2017.
[32] T. Wen and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals," *IEEE Access*, vol. 6, pp. 25399–25410, 2018.
[33] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, Jun. 2016.

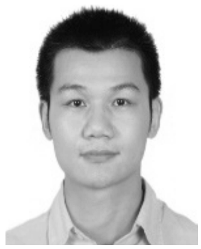


HONGWEI XU was born in Nantong, Jiangsu, China, in 1995. He received the B.S. degree in mechanical engineering from Jiangsu University, China, in 2017. He is currently pursuing the M.S. degree in mechanical engineering with Donghua University, China. From 2017 to 2020, he was a Research Assistant with the Intelligent Manufacturing and Information Engineering Institute, Shanghai, China. His research interests include the modeling techniques and big data analytics of complicated manufacturing system based on artificial intelligence algorithms.

JIE ZHANG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 1997. She was with the Institute of Intelligent Manufacturing and Information Engineering, Shanghai Jiao Tong University, China. She is currently the Dean of the College of Mechanical Engineering, Donghua University, China. Her research interests include industrial big data analysis, intelligent production scheduling, production control in intelligent manufacturing systems, and intelligent quality analysis.



YOULONG LV received the B.S. degree from the Huazhong University of Science and Technology, China, in 2011, and the Ph.D. degree from Shanghai Jiao Tong University, in 2017. She is currently a Lecturer with the College of Mechanical Engineering, Donghua University, China. Her current research interests include the modeling techniques, production control, and big data analytics of complicated manufacturing system based on artificial intelligence algorithms.



PENG ZHENG received the M.S. degree from the Wuhan University of Science and Technology, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Shanghai Jiao Tong University, China. His current research interests are big data analytics and operation in complex manufacturing systems.