Data-driven Predictive Analysis for Smart Manufacturing Processes Based on a Decomposition Approach

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

23-07-2021 / 28-07-2021

CITATION

Ghahramani, Mohammadhossein (2021): Data-driven Predictive Analysis for Smart Manufacturing Processes Based on a Decomposition Approach. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.15045426.v1>

DOI

[10.36227/techrxiv.15045426.v1](https://doi.org/10.36227/techrxiv.15045426.v1)

Data-driven Predictive Analysis for Smart Manufacturing Processes Based on a Decomposition Approach

Mohammadhossein Ghahramani, Member, IEEE
University College Dublin
Dublin, Ireland
sepehr.ghahramani@ucd.ie

Mengchu Zhou, Fellow, IEEE
New Jersey Institute of Technology
New Jersey, USA
zhou@njit.edu

Abstract—Smart Manufacturing refers to leveraging advanced analytics approaches and optimization techniques that are implemented in production operations. Recently, considerable effort has been devoted to embedding Artificial Intelligence (AI) and state-of-the-art technologies across manufacturing systems. With the widespread increase in deploying various networked sensors in manufacturing processes, there is a progressive need for optimal and effective data management approaches. Embracing such modern technologies to take advantage of manufacturing data allows us to overcome associated challenges, including real-time manufacturing process control and maintenance optimization. In line with this goal, a hybrid decomposition-based method consisting of an evolutionary algorithm and an artificial neural network is proposed to make manufacturing smart. The proposed dynamic approach helps us obtain useful insights for controlling manufacturing processes and gain perspective on various dimensions that enable manufacturers to access effective predictive technologies.

Index Terms—Smart Manufacturing, Artificial Intelligence, Industrial Internet of Things.

I. INTRODUCTION

Adaptation and innovation are vitally important to better peacebuilding and success in the modern industrial environment in these changing times. New enabling technologies such as Internet of Things (IoT), Big Data, and Machine Learning are permeating different aspects of manufacturing industry and can endow associated processes with intelligence. The rapid development and implementation of the mentioned technologies have allowed for various possibilities in technological advancements in different aspects of manufacturing. IoT solutions and real-time data processing can empower the processing of a massive amount of data captured from interconnected machines and sensors. Big data analytics, together with AI-based solutions, can help tackle many concerns in order to achieve smart prediction, evaluation, optimization, and decision-making. Recent advances in technology-based solutions, e.g., IoT, cloud/fog

computing, and big data, can expedite and simplify the production process and make new development of manufacturing possible [1]–[6]. These advances should drive the evolution of manufacturing architectures into integrated networks of automation devices and enable the smart characteristics of being self-adaptive, self-sensing, and self-organizing. Providing such solutions includes addressing several challenges like interoperability, decentralization, distributed control, real-time manufacturing process control, service orientation, and maintenance optimization [1], [4].

The main focus of smart manufacturing studies is on product life-cycle management, manufacturing process management, industry-specific communication protocols, and manufacturing strategies. Traditional fault detection and diagnosis systems interpret sensory signals as single values [7]. Then, these values are fed into a model to verify product status. The main drawback of this approach is that it fails to determine the most important features/operations involved in production processes and may result in sensory data loss. Moreover, sensory data are characterized by heterogeneous structures and might consist of noise, outliers, and missing values. Hence, relying on traditional methods can lead to inaccurate modeling of manufacturing processes. To address the concerns discussed above, we propose an intelligent and dynamic algorithm consisting of a feature extraction phase. We conduct a case study using a semiconductor manufacturing dataset that is publicly available to illustrate the proposed method's application. This dataset is imbalanced (like most manufacturing datasets) because the manufacturing processes' defective rate is quite low in practice. To address this potential issue, we implement an imbalanced classification technique to improve the model performance. We propose an integrated algorithm to solve a multiobjective problem based on an Artificial Neural Network (ANN) and Genetic Algorithm (GA). The generated model is then used

to establish a fault diagnosis solution by extracting the most relevant features. These features are used as an input for classifiers. To that end, a decomposition-based approach including a weighted sum technique is considered, and a comparison between the proposed solution and traditional methods is presented. This work offers twofold contributions:

- 1) It proposes a hybrid model based on a decomposition approach to model manufacturing processes; and
- 2) It integrates the capabilities of AI-based techniques to implement a highly flexible and personalized smart manufacturing environment;

The remainder of this paper is organized as follows: some related work about manufacturing processes and application of AI is described in Section II; preprocessing procedures are discussed in Section III; the proposed approach with its associated discussions are given in Section IV; the experimental settings and the classification results are shown in Section V; and conclusions are presented in Section VI.

II. RELATED WORK

The key to leveraging manufacturing data lies in constant monitoring of processes, which can be associated with different issues, e.g., noisy signals. Dimensionality reduction and feature selection/extraction methods play a critical role in dealing with noise and redundant features and must be considered as a preprocessing stage of manufacturing data analysis, which leads to better insights and robust decisions [8]. Some previous manufacturing fault detection studies have focused on utilizing the mentioned techniques for extracting the most relevant features and classification. A support vector machine (SVM) is used to detect semiconductor failures in [9]. The authors have developed their approach based on an RBF kernel to address the high dimensionality issue. In [10], an incremental clustering method is adopted for fault detection. A Bayesian model has been proposed to infer manufacturing processes. The authors have considered the root causes of manufacturing problems. However, their approach heavily relies on an expert's knowledge regarding the related field. Zheng et al. have proposed a convolution neural network [11]. They have decomposed multivariate time-series datasets into univariate ones. Then, features have been extracted and an MLP-based method has been implemented for data classification. Lee et al. have compared the performance of different fault detection models, including feature extraction algorithms and classification approaches [12]. They have revealed that developing an algorithm based on features that are not suitable for a specific model can significantly deteriorate classifiers' performance. Therefore, it is desirable to consider both feature extraction and classification stages simultaneously to maximize a model's performance.

Most studies in the literature have focused on using PCA and KNN algorithms for manufacturing data classification. However, PCA-based approaches project features to another space based on a linear combination of original features. Therefore they cannot be interpreted in the original feature space [13], [17]. Moreover, most PCA-related work has considered linear PCA, which is not efficient in exploring non-linear patterns [14]–[16]. Although these techniques try to cover maximum variance among manufacturing variables, inappropriate selection of parameters, e.g., principal components, may result in great data loss [18]. KNN is a memory-based classifier. Hence, in cases of high dimensional datasets, its performance degrades dramatically with data size. To overcome the mentioned concerns, we propose an efficient global search method to model manufacturing processes.

III. DATA PREPROCESSING

The dataset used in this work is obtained from a semiconductor factory, SECOM (Semiconductor Manufacturing) dataset. It consists of various operation observations, i.e., wafer fabrication production data, including 590 features (operation measurements). The target feature is binomial (Failure and Success), referring to the production status and encoded as 0 and 1. The first step in data analysis is data cleansing to address various data quality issues, e.g., noise, outliers, inconsistency, and missing values. We have dealt with missing values and noise resulting from inexact data collection. These can negatively affect later processes. Outlier labeling methods and T-squared statistics (T^2) have been utilized. Any observation beyond the interval has been eliminated.

Class Imbalance Issue

The observations that have been labeled as *Failure* are relatively rare (104 cases) as compared to the *Success* class. Hence, we face an imbalanced classification issue [19]. In other words, *Success class* (the majority) outnumbered *Failure class* (the minority), and both classes do not make up an equal portion of our dataset. We have implemented a density-based SMOTE [20] technique, and by synthetically adding *Failure class* instances, the distribution has been made more balanced. The implemented technique is an over-sampling method in which the Failure class is over-sampled by generating its synthetic instances.

IV. PROPOSED MODEL

As stated, the dataset consists of nearly 600 features. Datasets with high dimensions can cause serious challenges such as overfitting in learning processes, known as the curse of dimensionality. To address these challenges, dimensionality needs to be reduced. An integrated feature selection approach consisting of a *metaheuristic* algorithm (i.e., GA) and

an Artificial Neural Network has been proposed in this work. GA is a heuristic search method and inspired by Charles Darwin's theory of natural evolution. Since selecting features can be considered as a binary problem, we have developed our model based on binary GA that treated candidate features (chromosomes in GA terminology) as bit-strings. Selecting an appropriate pressure measurement (β in this work) can maintain a balance between exploration and exploitation. Parameter β has been used in the parent selection stage, and candidate individuals have been taken into account in the generation production. This operation, iteratively, has been repeated until the termination criteria (number of iterations or number of function evaluations (NFE)) are met. The best individual (the one with the minimum cost) is selected, and in this way, optimal features are then identified.

Our objective is to modify the output of each iteration (a subset of features) by searching feature space and finding proper values for the input features such that the measured cost is minimized. Our proposed feature selection model consists of different phases. It starts with defining an initial population, i.e., individuals including m -dimensional chromosomes.

$$\zeta = (v_1, v_1, \dots, v_m)$$

where v_i is either 1 or 0, and corresponds to the status of the i^{th} variable (feature), selected or not. While some individuals are admitted to the new generation unchanged, others may be subject to some genetic operators (crossover and mutation). The cost related to each individual is evaluated by the ANN. We have also employed the Boltzmann Selection method which is inspired by Simulated Annealing [21]. The probability of an individual being selected is calculated according to the below Boltzmann probability:

$$p(i) = \frac{e^{-\beta J_i}}{\sum_{k=1}^{\eta_p} e^{-\beta J_k}}$$

where η_p is the size of the initial population and J is the defined cost function. β is the selection pressure. It is clear that parents are selected based on probabilities that are proportional to the costs measured in the initial phase. This means that individuals with a lower cost are more likely to be chosen than ones with a greater cost. It should be mentioned that we have selected the β parameter such that $\sum_{i \in H} p(i) = 0.7$, where H is the set of half of the best individuals (population is sorted according to their cost values, and $\eta_p/2$ of them are selected). Consequently, the Roulette Wheel method is used to sample (selecting parents using stochastic sampling with replacement based on the Boltzmann probability function). A circular wheel is considered and divided into η_p pies, each of which is proportional to

the cost values. The wheel is spun, and the individual related to the pie on which it stops is then selected. We have repeated this procedure until our predefined number of parents is selected. In this way, individuals with the largest cost value have a minimal chance to be selected. Parents are selected according to the weighted slots, and crossover operations are then applied to them. On this basis, the chromosomes of selected parents are combined to create new offspring. A random portion of the first individual is swapped with a random portion of the second one. In this process, the chromosome combination can be carried out in different ways, e.g., single-point, double-point, or uniform crossover. In the single-point crossover, one random position in the array of bits is selected, and exchanging then takes place, while in the double-point method, two positions are chosen, and chromosomes are swapped. In the uniform crossover, parents' chromosomes are selected for random exchange. Parents contribute to creating new offspring based on a bit string known as the crossover mask. Let ξ be the predefined cross-over mask, e.g., $\xi = \{1, 1, 0, 0, 0, 1, \dots, 0, 0, 1\}$. As discussed earlier, after the initial population is created, the parent selection operation should be conducted in the reproduction phase. Our goal is to select individuals from those with minimum costs in the population. Consequently, parents are selected to create offspring for the next generation. The procedures are presented in Algorithm 2. The cost function and the way we have integrated ANN to calculate this measurement are next described.

A. Cost Function and MLP

Our objective in the feature selection phase is to explore a hypothesis space, find the optimal number of features, and reduce dimensionality. In other words, we are looking for a subset of the original dataset, $J : X' \subseteq X \rightarrow \mathbf{R}$, such that two criteria are satisfied. The cost function obtains different subsets of features, and target values as the input and the corresponding costs are calculated. Given the conventions adopted earlier, let F be the original feature set with the cardinality of $|X| = m$. Now, let $J(X')$ be an evaluation measure to be optimized given the below criteria:

- Set $|X'| = k < m$. Find $X' \subset X$, such that $J(X')$ is maximized. It is equal to minimizing the Mean Squared Error:

$$\epsilon = \frac{1}{n} \sum_{i=1}^n (Y_i - J_i)^2$$

- Find optimal features (in the case of both the number of features and discrimination) while minimizing $|X'| = n_j$.

It should be mentioned that we are facing a multi-objective optimization problem [22], [23]. We define our objective function with weights as:

$$J = \epsilon \times (1 + (\Omega \times |X'|)) \quad (1)$$

where $|X'|$ is the dimension of selected features in each iteration, and Ω can be considered as a cost parameter for choosing new features. If $\Omega = 0$, all features are selected, while a large number to Ω results in no feature being selected. This parameter is a trade-off between relevancy and redundancy and must be designated carefully. As stated, our objective is to minimize objective function J . In doing so, we have integrated Artificial Neural Network (ANN) and GA. GA gets the defined cost function (i.e., Feature-Selection-Cost, J) as the input and employs the ANN to calculate cost values. Iteratively, different individuals (bit strings) consist of 0 and 1 (where 1 refers to a feature being selected and 0 refers to is not being selected) are generated and evaluated by GA's operations. Multilayer Perceptron (MLP) is utilized to calculate ϵ (in each iteration). Multi-layer perceptrons with a Levenberg-Marquardt training algorithm are used (since it converges faster and more accurately towards our problem) and consists of two layers (15 neurons in the hidden layer) adaptive weights with full connectivity among neurons in the input and hidden layers. The procedures for defining the algorithm's operations are presented in Algorithm 1. All costs are calculated, and the best features are selected such that the corresponding cost is minimized.

V. RESULTS

The proposed algorithm for feature selection is based on an adaptive and dynamic GA combined with a neural network. Our meta-heuristic method evaluates various subsets of features to optimize our defined cost function, whose calculation has been given to a multilayer perceptron. We consider the volume of our data and the number of features and samples for defining the initial population rate. We choose the number of neurons based on a trial and error method. It should be mentioned that we have used the neural network as a cost function, and in this context, the main objective is to decrease the cost function's values. The algorithm gets the initial solutions (manufacturing operations) and obtains the optimal features after a series of iterative computations (given the termination criteria, e.g., Number of Function Evaluation).

Finally, we have examined various classification techniques, and the most appropriate one is selected. To do so, different classification models, e.g., Gaussian Support Vector Machine, Random Forest, Linear Discriminant, K-NN, and SVM with RBF kernel, have been tested. All classifiers' performances have been evaluated based on their classification accuracies. Given the results, SVM has been chosen.

TABLE I
ALGORITHMS COMPARISONS

Method	Selected Features	Accuracy(%)
Lasso	54	74.2
PCA	48	79.2
FWE	15	47.8
FDR	12	41.5
SP	71	81.5
SFS	88	72.9
SBS	40	72.4
FFS	111	43.6
CFS	92	61.0
Ensemble	76	70.9
Proposed method	36	90.2

For more information we refer to the journal version of this paper [1]. The ability of each method to accurately predict the correct class has been measured and expressed as a percentage. ROC curves are used to determine the predictive performance of the examined classification algorithms. The area under a ROC curve can be considered as an evaluation criterion to select the best classification algorithm. When the area under the curve is approaching 1, it indicates that the classification has been carried out correctly. Fig. 1 shows AUC - ROC curve resulted from the proposed model.

We have also tested most popular feature selection algorithms, e.g., Family-Wise Error Rate (FWE), False Discovery Rate (FDR), Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Filtration Feature Selection (FFS), Correlation-based Feature Selection (CFS), Lasso Regression and Ensemble methods [24]. These traditional methods have been used to reduce the dimensionality of our dataset and the obtained results are presented in Table I. To do so, the extracted features are used as the input for the chosen classifier.

GA has different parameters and the performance of a GA-based model depends on these parameters. We have discussed how they have been selected throughout this work. Table II reveals the impacts of different parameter setting.

VI. CONCLUSIONS AND FUTURE WORK

The goal of manufacturing enterprises is to develop cost-effective and competitive products. Manufacturing Intelligence can significantly improve effectiveness by bridging business and manufacturing models with the help of low-cost sensor data. It aims to achieve a high level of intelligence with the latest appropriate technology-based computing, advanced analytics, and new levels of Internet connectivity. The landscape of Industry 4.0 includes achieving visibility on real-time processes, mutual recognition, and establishing an effective relationship among the workforce, equipment,

TABLE II
COMPARING THE RESULTS OF OUR HYBRID MODEL GIVEN DIFFERENT PARAMETER SETTING

Crossover Rate	Mutation Rate	Population Size	Neurons	Corresponding cost for selected solution in different iterations								
0.6	0.3	200	10	1.8311	1.8311	1.8311	1.8266	1.8266	...	0.5446	0.5446	0.5446
		250	15	1.5182	1.3109	1.3109	1.1102	1.1102	...	0.4902	0.4852	0.4852
		300	20	1.5330	1.5330	1.5330	1.4016	1.4001	...	0.4767	0.4767	0.4561
0.7	0.3	450	10	0.6013	0.5182	0.5182	0.5468	0.5012	...	0.1926	0.1859	0.1859
		500	15	0.4165	0.4002	0.4002	0.3922	0.3774	...	0.0991	0.0991	0.0982
		550	20	0.3911	0.3891	0.3891	0.3496	0.3347	...	0.0932	0.0932	0.0932
0.8	0.3	650	10	0.1542	0.1523	0.1523	0.1478	0.1478	...	0.0652	0.0652	0.0652
		700	15	0.1227	0.1214	0.1214	0.1205	0.1205	...	0.0481	0.0472	0.0415
		750	20	0.1304	0.1302	0.1246	0.1246	0.1228	...	0.0478	0.0478	0.0478

Fig. 1. ROC result from the proposed method

and products. Most studies in the area of manufacturing data analysis are based on PCA-based approaches. They are not able to recognize nonlinear relationships among features and extract complex patterns. To address this concern, we have proposed a dynamic feature selection method based on GA and ANN. We have compared the results achieved in this work with traditional approaches to prove the effectiveness of our proposed solution. As part of our future work, we plan to consider other MOEAs, e.g., dominance-based algorithms, to solve our optimization problem so that both feature selections objective functions are optimized simultaneously [25]–[28].

REFERENCES

- [1] M. Ghahramani, Y. Qiao, M. C. Zhou, A. O'Hagan and J. Sweeney, "AI-based modeling and data-driven evaluation for smart manufacturing processes," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 1026-1037, July 2020, doi: 10.1109/JAS.2020.1003114.
- [2] M. Ghahramani, M. C. Zhou, and C. T. Hon, "Extracting Significant Mobile Phone Interaction Patterns Based on Community Structures,"

Algorithm 1: Pseudo-code for the feature selection model

```

1 Function  $[O1, O2] = CrossoverFn(P1, P2)$ 
2    $CrossoverMethod \leftarrow$ 
   {Single-Point, Multi-Point, Uniform Crossover}
3   randomly select one method given probabilities
   defined for each of them;
4   return two offsprings;
5 end
6 Function  $M = MutationFn(P)$ 
7   Apply mutation operator;
8   return mutant;
9 end
10 Function  $J = CostFn(dataset)$ 
11   Employ ANN;
12   return  $\epsilon \times (1 + (\Omega \times |X'|))$ ;
13 end

```

IEEE Transactions on Intelligent Transportation Systems, vol. 20, pp. 1031-1041, 2019.

- [3] M. Ghahramani, M. C. Zhou, and C. T. Hon, "Mobile Phone Data Analysis: A Spatial Exploration toward Hotspot Detection," *IEEE Transactions on Automation Science and Engineering*, vol. 16, pp. 351-362, 2019.
- [4] M. H. Ghahramani, M. C. Zhou, and C. T. Hon, "Toward cloud computing QoS architecture: analysis of cloud systems and cloud services," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, pp. 6-18, 2017.
- [5] D. P. Bertsekas, "Feature-based aggregation and deep reinforcement learning: a survey and some new implementations," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 1, pp. 1-31, Jan. 2019.
- [6] M. Ghahramani, M. Zhou and G. Wang, "Urban sensing based on mobile phone data: approaches, applications, and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 627-637, May 2020, doi: 10.1109/JAS.2020.1003120.
- [7] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 3768-3774, 2015.
- [8] S. M. Meerkov and M. T. Ravichandran, "Combating Curse of Dimensionality in Resilient Monitoring Systems: Conditions for Lossless Decomposition," *IEEE Transactions on Cybernetics*, vol. 47, pp. 1263-1272, 2017.
- [9] R. Baly and H. Hajj, "Wafer Classification Using Support Vector

- Machines,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 3, pp. 373-383, 2012.
- [10] J. Kwak, T. Lee, and C. O. Kim, ”An Incremental Clustering-based Fault Detection Algorithm for Class-imbalanced Process Data,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 28, no. 3, pp. 318-328, 2015.
- [11] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. Zhao, ”Time series classification using multi-channels deep convolutional neural networks,” in *Proc. WAIM*, Macau, pp. 298-310, 2014.
- [12] K. Lee, S. Cheon, and C. Kim, ”A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, pp. 135-142, May 2017.
- [13] J. Zhang, H. Chen, S. Chen, and X. Hong, ”An Improved Mixture of Probabilistic PCA for Nonlinear Data-Driven Process Monitoring,” *IEEE Transactions on Cybernetics*, vol. 49, 2019.
- [14] M. Ghahramani, N. J. Galle, F. Duarte, C. Ratti, F. Pilla, ”Leveraging AI and machine learning to analyse citizens’ opinions on Urban Green Space”, *City and Environment Interactions*. 2021.
- [15] M. Ghahramani and F. Pilla, ”Leveraging Artificial Intelligence to Analyze the COVID-19 Distribution Pattern based on Socio-economic Determinants”, *Sustainable Cities and Society*, 2021.
- [16] S. Hu, A. O’Hagan, J. Sweeney, and M. Ghahramani ”A Spatial Machine Learning Model for Analyzing Customers’ Lapse Behaviour in Life Insurance”, *Annals of Actuarial Science*, 2020.
- [17] W. Xiong et al., ”Wafer Reflectance Prediction for Complex Etching Process Based on K-Means Clustering and Neural Network,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 34, no. 2, pp. 207-216, May 2021, doi: 10.1109/TSM.2021.3068974.
- [18] M. Ghahramani, A. O’Hagan, M. Zhou and J. Sweeney, ”Intelligent Geodemographic Clustering Based on Neural Network and Particle Swarm Optimization,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, doi: 10.1109/TSMC.2021.3072357.
- [19] H. Liu, M. Zhou and Q. Liu, ”An Embedded Feature Selection Method for Imbalanced Data Classification,” *IEEE/CAA Journal of Automatica Sinica*, 6(3), pp. 703-715, May 2019.
- [20] X. Zhang, Y. Zhuang, W. Wang, and W. Pedrycz, ”Transfer Boosting with Synthetic Instances for Class Imbalanced Object Recognition,” *IEEE Transactions on Cybernetics*, vol. 48, 2018.
- [21] S. Lee and S. B. Kim, ”Parallel Simulated Annealing with a Greedy Algorithm for Bayesian Network Structure Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1157-1166, 1 June 2020, doi: 10.1109/TKDE.2019.2899096.
- [22] Q. Peng, M. Zhou, Q. He, Y. Xia, C. Wu and S. Deng, ”Multi-Objective Optimization for Location Prediction of Mobile Devices in Sensor-Based Applications,” *IEEE Access*, vol. 6, pp. 77123-77132, 2018, doi: 10.1109/ACCESS.2018.2869897.
- [23] A. Gupta, Y. Ong, L. Feng, and K. Tan, ”Multiobjective Multifactorial Optimization in Evolutionary Multitasking,” *IEEE Transactions on Cybernetics*, vol. 49, pp. 1652-1665, 2017.
- [24] C. Hou, F. Nie, X. Li, D. Yi, and Y. Wu, ”A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, Issue 1, pp. 16-28, 2014.
- [25] S. Han, K. Zhu and M. Zhou, ”Utilization Method-assisted Multimodal Multiobjective Optimization and Application to Credit Card Fraud Detection,” to appear in *IEEE Trans. on Computational Social Systems*, Feb. 2021.
- [26] WenQing Xiong, Yan Qiao, LiPing Bai, Mohammadhossein Ghahramani, NaiQi Wu, and PinHui Hsieh, ”Wafer Reflectance Prediction for Complex Etching Process Based on K-Means Clustering and Neural Network”, *IEEE Transactions on Semiconductor Manufacturing*, 2021.
- [27] H. Zhu, et al. ”Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to credit card fraud detection,” *Neurocomputing*, 407, pp. 50-62, May 2020.
- [28] X. Guo, M. Zhou, S. Liu, and L. Qi, ”Lexicographic Multiobjective Scatter Search for the Optimization of Sequence-Dependent Selective Disassembly Subject to Multiresource Constraints,” *IEEE Transactions on Cybernetics*, 50(7), pp. 3307-3317, July 2020.

Algorithm 2: Pseudo-code for the feature selection model

Input : $GA(|F|, CostFn)$
Output: *Individual* $\leftarrow$ [Selected Features (a binary vector)]

```

1  $I^{Max} \leftarrow$  Maximum number of Iterations
2  $\theta \leftarrow$  Crossover rate,  $\mu \leftarrow$  Mutation rate
3  $\eta_p \leftarrow$  Size of population
4 # Initialise Population
5 for  $i \leftarrow 1, \dots, \eta_p$  do
6    $Pop.position \leftarrow$  randomly-generated chromosomes
7    $Pop.cost \leftarrow$  calculated costs given each chromosome
8 end
9  $Pop \leftarrow$  Sort population ( $Pop$ );
10 # Main Loop
11  $\eta_c \leftarrow$  Size of crossover (based on  $\theta$ )
12  $\eta_m \leftarrow$  Size of mutation (based on  $\mu$ )
13 for  $i \leftarrow 1, \dots, I^{Max}$  do
14   # Crossover operation
15   Calculate probabilities based on

$$Pr(s \in Pop) = \frac{exp^{(-\beta) * \frac{J_s}{LargestCost}}}{\sum_{k=1}^{\eta_p} exp^{(-\beta) * \frac{J_k}{LargestCost}}}$$

16   for  $i \leftarrow 1, \dots, \eta_c$  do
17     Select two parents ( $P1$  and  $P2$ ) based on a
      Roulette Wheel method given probabilities
      measured above;
18      $[I(i, 1).position, I(i, 2).position] \leftarrow$ 
      CrossoverFn( $P1.position, P2.position$ )
19      $[I(i, 1).cost, I(i, 2).cost] \leftarrow$ 
      CostFn( $P1.position, P2.position$ )
20   end
21   # Mutation operation
22   for  $i \leftarrow 1, \dots, \eta_m$  do
23     Select one parents ( $P$ ) based on Roulette
      Wheel method
24      $[Mutant(i).position] \leftarrow$ 
      MutationFn( $P.position$ )
25      $[Mutant(i).cost] \leftarrow$  CostFn( $P.position$ )
26   end
27    $Pop \leftarrow$  [ $Pop$ , Offsprings, and Mutants];
28    $Pop \leftarrow$ 
      Sort population and select first  $\eta_p$  individuals;
29    $BestSolution \leftarrow$ 
      Select first chromosome’s position,  $Pop(1).position$ ;
30    $BestCost(i) \leftarrow$ 
      Select first chromosome’s cost,  $Pop(1).cost$ ;
31 end
32 end
33 return Individual (Selected Features);
```