Ensemble Machine Learning Algorithms for Anomaly Detection in Multivariate Time-Series

Youssef Trardi*
Aix Marseille Université
LIS UMR 7020 CNRS
Marseille, France
youssef.trardi@univ-amu.fr

Bouchra Ananou
Aix Marseille Université
LIS UMR 7020 CNRS
Marseille, France
bouchra.ananou@univ-amu.fr

Philip Tchatchoua
Aix Marseille Université
LIS UMR 7020 CNRS
Marseille, France
philip.tchatchoua@lis-lab.fr

Mustapha Ouladsine
Aix Marseille Université
LIS UMR 7020 CNRS
Marseille, France
mustapha.ouladsine@univ-amu.fr

2022 International Conference on Control, Automation and Diagnosis (ICCAD) | 978-1-6654-9794-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCAD55197.2022.9853995

***Abstract*—This paper proposes a multivariate time-series anomaly detection approach using multiple transform techniques and ensemble machine learning (EML) algorithms. The objective is to detect the presence of abnormal wafers during the semiconductor manufacturing process. Therefore, we evaluate a set of eleven features derived from an intermediate manufacturing chain to characterize the wafer status. Data from each feature is recorded over a 150-second time frame. To address the computational complexity of large-scale data processing, a dimensionality reduction step is highly desirable. Indeed, independent component analysis (ICA), principal component analysis (PCA), and factor analysis (FA) are used for comparison purposes. As well, to extract the most significant components from each feature sequence and build a thoroughly combined subset of characteristics. In the sequel, decision trees, bootstrap aggregating, boosting, one of the prevalent evolutions of EML algorithms, are fitted to the obtained characteristics to define the best anomaly detection ranking. The selected model is validated using 7000 samples (*i.e.* wafers) divided into 5000 normal samples and 2000 abnormal samples. The results highlight the strengths of the proposed approach, which could serve as a valuable decision-making support for abnormal wafer detection in the semiconductor manufacturing process.**

***Index Terms*—Anomaly detection, Data-driven methods, Ensemble machine learning, Multivariate time series analysis, Semiconductor manufacturing.**

I. Introduction

Anomaly detection (AD) is an important class of time series analysis problems. It has been widely discussed in statistics and machine learning. Synonymously, it is also called "outlier detection", "novelty detection", "deviation detection" and "exception mining" [1]. The goal is to identify abnormal occurrence patterns that do not conform to the expected behavior [2]. The complexity of such tasks stems from the nature of the data, the availability of their labeling, and their application framework [3, 4]. Over the past decade, time series anomaly detection has received significant interest and emerged in many real-world data analysis and prediction applications such as medical and public health [5–7], finance, fraud detection, intrusion detection [8, 9], manufacturing [10], and many other subjects [11, 12]. Technological advances and the evolution of sophisticated machine learning and data mining techniques have contributed significantly towards the emergence of effective modeling for AD. Although most of the existing research studies are devoted to AD in univariate time series, relatively few studies have addressed AD problems in Multivariate Time Series (MTS) [13]. Indeed, several factors can complicate the detection of anomalies in MTS—first, the undefined nature of the anomaly paradigm in a multivariate context. Second, the presence of atypical data (i.e., too high or too low values) and random subsequences (i.e., changes in shape) that are typically considered anomalies in a univariate time series [14]. However, multivariate techniques focus beyond the anomalous data or random subsequences to examine the statistical relationships between features.

Many studies have been proposed to solve AD in MTS. Examples include AD using statistical features. Feature-based AD algorithms rely heavily on features extracted from time-series data. Thus, the collected data are segmented and converted into a set of features using descriptive statistics. Descriptive statistics techniques can be applied to single or multiple measurement sequences. Two common classes of descriptive statistics are used: (i) measures of central tendency (*e.g.* mean, median, and mode) and (ii) variability measures (*e.g.* interquartile range, variance, standard deviation, coefficient of variation, and asymmetry coefficient). In addition, non-linear time-domain features have also been used to describe the non-linear dynamics of MTS data [7]. However, feature-based AD algorithms are a costly process since the features embedded in time series data are difficult to highlight and cannot accommodate all the essential properties of the data. To summarise, a given dataset cannot always be simplified into a reduced statistical representation. For this reason, approaches based on dimensionality reduction are more suitable for categorizing MTS data and emphasizing abnormal patterns. Indeed, Baydogan, M. G. et al. [15] uses Learned Pattern Similarity (LPS) to extract segments from MTS and train regression trees to detect dependencies and classify unknown samples. Buvé, C. et al. [16] discusses a number of MTS methods: Principal Component Analysis (PCA), Partial Least Squares regression (PLS), Parallel Factor Analysis (PFA), and ANOVA Simultaneous Component Analysis (ANOVA-SCA), applicable to food quality surveys. Shyu, M.-L. et al. [17] have also adopted PCA, in which a predictive model is constructed based on major and minor PCs of regular instances. Further, Kwitt, R. et al. [18] suggested an alternative variant of this technique, in which the Minimum Covariance

Determinant (MCD) is applied to compute the covariance and correlation matrix instead of the standard estimators. Wong, J., et al. [19] have developed an AD function called Robust Anomaly Detection (RAD). The function uses robust-PCA to detect anomalies. Sorzano, C., et al. [20] categorize dimensionality reduction techniques, along with the underlying mathematical concepts. PCA with AdaBoost and decision tree algorithms are used in [21] to classify defective wafers. The main objective of this paper is to propose an EML-based anomaly detector for MTS. In this context, we investigate various dimensionality reduction techniques and study their performance in retaining relevant insights in a large-scale data representation paradigm. Therefore, a pipeline of transformers and classifiers is conducted to evaluate the performance of each combination and designate the best AD model. In effect, ICA, PCA, and FA are applied to each sequence of samples' feature data to depict the main embedded components. The data represent the evolution of eleven features derived from an intermediate production process, recorded over a 150-second time frame. The decision-making task is performed using an ensemble of six classification algorithms, namely, Bagging, ETC, RFC, AdaBoost, GB, and XGB. Each is fitted to the obtained subsets of components to define the best anomaly detection ranking. All methods have been applied to the same experimental data to compare the consistency of the selected subset and conduct a comparative study of the results.

II. Exploratory Data Analysis

A. Data specification:

Data is collected every second for a recipe cycle of about 150 seconds. The monitoring consists of 11 simulated variables, similar to fundamental production dynamics, such as gas flows, pressures, temperatures, and etching tool capacity. In this paper, we used 5000 samples of normal data and 2000 samples of abnormal data, i. e. a ratio of 28.6% of abnormal data, satisfying the low frequency of abnormal data in the semiconductor industry. The abnormal dataset describes 5 different defects types, equally distributed in the dataset (400 samples per defect type). The 5 types of faults in the data stand for common faults produced during wafer processing. These faults can be either atomic or aggregated anomalies and affect various variables. Atomic anomalies are instances of deviant values for one variable. In contrast, aggregate anomalies result from groups of deviant variables as a collective. These include contextual anomalies (faults 1, 2 & 5), and collective anomalies (faults 3 & 4). Fault 1 is a breakage point, creating a deviant cycle. Fault 2 is a temporary value modification. Faults 3 and 4 are similar to additive noise and sinusoidal disturbances, respectively, innovative outliers creating a trend change. Fault 5 is a peripheral point, a literal independent data point outlier resulting from a sudden increase in value. Figure 1 illustrates the 5 types of faults.

B. Default specification:

Time-series data in semiconductor manufacturing consist of 3-dimensional information, including wafer, variable status identification (SVID), and recorded time. The SVID represents the status of equipment or machine such as temperature, pressure, and gas flow ([23]). Below, we demonstrate an example of SVID data collected by production equipment. Fig. 2 use Andrews graphs to visualize the multivariate data represented by the normal pattern and the five defects. An Andrews curve

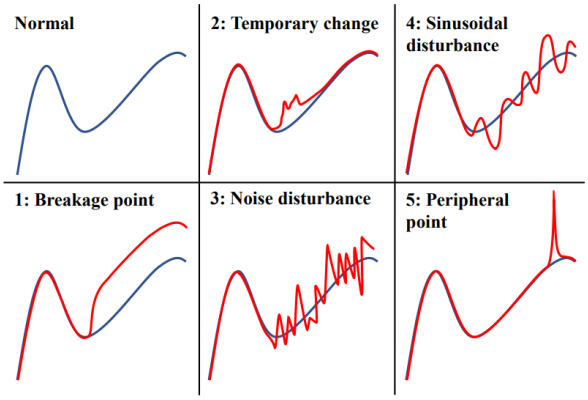

Fig. 1: Description of fault types ([22])

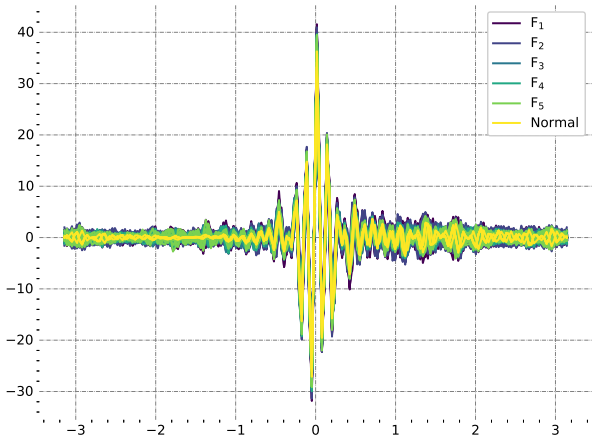

Fig. 2: The 1st variable plot using Andrews curves

is a graphical data analysis technique for mapping multivariate data. This consists in applying the following transformation to the data:

$$\frac{x_1}{\sqrt{2}} + x_2 sin(t) + x_3 cos(t) + \cdots + x_i sin(t) + x_{i+1} cos(t) + \cdots \quad (1)$$

where t varies from $-\pi$ to π and $\{x_1, x_2, \cdots\}$ are the variables (i.e., columns) of data. An Andrews curve is generated for each row of data. The diagrams highlight the abnormal and normal wafer signatures over a 150-second window for the first 3 variables. This study considers that this signature applies to the product status and can be exploited to distinguish a faulty wafer. The remaining variables are approximately similar to the three cases illustrated. The main distinction is the impact of each defect on the variable concerned.

III. Multivariate Data Analysis: terminology

Assuming that there are p sensors installed in the equipment unit. Thus, p SVIDs are collected for each wafer processing. Each SVID is collected as a temporal signal during n instants. Hence, the data of wafer k is given by a $(n \times p)$ matrix as follows:

$$\mathbf{S^{(k)}} = \begin{array}{c} \begin{matrix} \text{SVID}_1 & \text{SVID}_2 & \dots & \text{SVID}_p \end{matrix} \\ \begin{bmatrix} s_{1,1,k} & s_{1,2,k} & \dots & s_{1,p,k} \\ s_{2,1,k} & s_{2,2,k} & \dots & s_{2,p,k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,1,k} & s_{n,2,k} & \dots & s_{n,p,k} \end{bmatrix} \end{array} \begin{matrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{matrix} \tag{2}$$

where $s_{n,p,k}$ represents the sensor reading value for k^{th} wafer of the p^{th} sensor (SVID_p) at time t, for $k = 1, \dots, N$. Here n denote the total number of recorded time points for k^{th} wafer.

To process the N-wafer time series matrix at once, we stack all sparse characteristics $S^{(k)}$ for p SVID :

$$W = \left(S^{(1)}, S^{(2)}, \cdots, S^{(N)}\right) \in \mathbb{R}^{(n \times p) \times N} \tag{3}$$

Certain studies perform a dimensionality reduction step on the W matrix.

$$W \in \mathbb{R}^{(n \times p) \times N} \quad \mapsto \quad W_r \in \mathbb{R}^{m \times N}, \quad \text{where} \quad m \leq (n \times p) \tag{4}$$

The reduced dataset W_r suppose to summarize all the essential information of each variable in n instances. This can be performed using any dimensionality reduction technique, such as PCA, FA, ICA, or others. In our case, each variable consists of 150 measures $(n = 150)$, yielding $150 \times 11 = 1650$ features for 11 variables $(p = 11)$ (i.e., SVID).

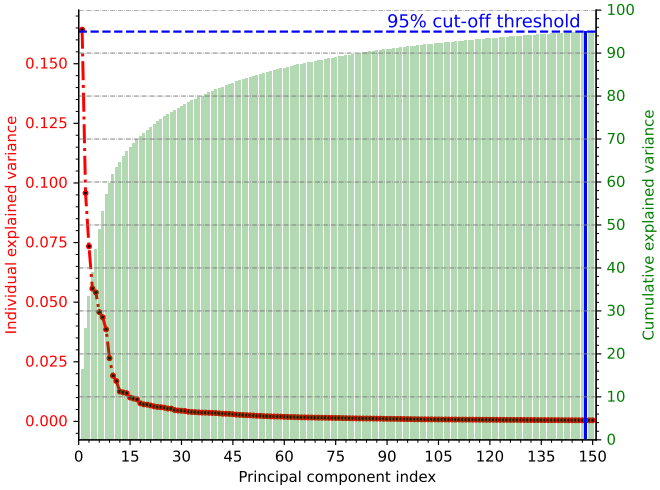

Fig. 3: Scree plot – PCA

For instance, we explore PCA as a dimensionality reduction method applied to the W matrix. The obtained results are illustrated in Fig. 3, which describes the individual explained variance per variable and the cumulative explained variance. As shown in Fig. 3, we focused primarily on the statistics of the first 150 of the 1650 components. The cumulative explained variance reaches 95% at the 148 components (the blue line) and gets the 100% cumulative variation at the 1232 components. The average rate of progression past the first 150 components is approximately 0.00325% for each increment.

Typically, we are interested in an explained variance between 95–99%. In this case, to ensure 95% of the explained variance, we need 148 principal components. The challenge remains to define an adjustment strategy to select the most suitable number of features to meet the desired performance. Therefore, we need to test all subsets from 148 to 360 variables, i.e. to cover 99% of the explained variance. It's a very computationally and time-intensive optimization process.

As a potential improvement, we propose to reshape the wafer matrix data into SVID matrices. This amounts to evaluating the SVID information independently of each other. Accordingly, the dimensionality reduction step is performed separately for each SVID matrix. Then all the resulting reduced subsets are stacked. Each SVID matrix can be explained as follows:

$$\text{SVID}_i = \begin{array}{c} \begin{matrix} t_1 & t_2 & \dots & t_n \end{matrix} \\ \begin{bmatrix} s_{1,i,1} & s_{2,i,1} & \dots & s_{n,i,1} \\ s_{1,i,2} & s_{2,i,2} & \dots & s_{n,i,2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1,i,N} & s_{2,i,N} & \dots & s_{n,i,N} \end{bmatrix} \end{array} \in \mathbb{R}^{N \times n} \tag{5}$$

where $s_{n,p,k}$ represents the sensor reading value for k^{th} wafer of the i^{th} sensor (SVID_i) at time t, for $k = 1, \dots, N$, and $i = 1, \dots, p$.

The Eq.6 determines the following metrics: the $N = 7000$ wafer matrix of dimension 150×11 are transformed into $p = 11$ SVID matrix of dimension 7000×150.

$$\begin{pmatrix} S^{(1)} \\ S^{(2)} \\ \vdots \\ S^{(N)} \end{pmatrix} \rightarrow \text{reshaping} \rightarrow \begin{pmatrix} \text{SVID}_1 \\ \text{SVID}_2 \\ \vdots \\ \text{SVID}_p \end{pmatrix} \tag{6}$$

To process the p-SVID time series matrix, we applie a dimentionality reduction separately for each one:

$$\text{SVID}_i \in \mathbb{R}^{N \times n} \quad \mapsto \quad M_i \in \mathbb{R}^{N \times d}, \quad \text{where} \quad d \leq n$$

$$\begin{pmatrix} \text{SVID}_1 \\ \text{SVID}_2 \\ \vdots \\ \text{SVID}_p \end{pmatrix} \rightarrow \text{data reduction} \rightarrow \begin{pmatrix} \text{M}_1 \\ \text{M}_2 \\ \vdots \\ \text{M}_p \end{pmatrix} \tag{7}$$

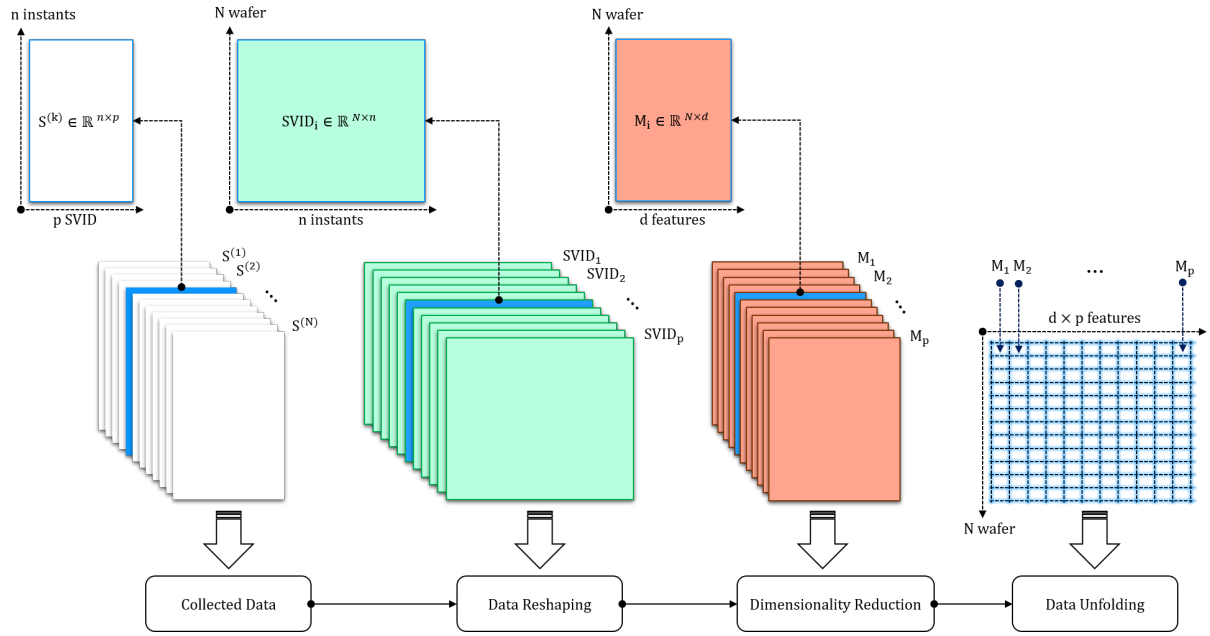

Fig. 4: Synoptic diagram of the entire data preprocessing

Thus, different dimensionality reduction techniques can scale the SVID_i matrices. In the current framework, we are interested in PCA, FA, and ICA. These are designed to infer a reduced subset of features M_i (Fig.4). The subsequent dimension of the resulting matrices M_i is fixed iteratively by setting the number of components between 10 and 20 for each technique on each SVID_i matrix.

IV. System Design

Thus, different dimensionality reduction techniques can scale the SVID_i matrices. In the current framework, we are interested in PCA, FA, and ICA. These are designed to infer a reduced subset of features M_i. The subsequent dimension of the resulting matrices M_i is fixed iteratively by setting the number of components between 10 and 20 for each technique on each SVID_i matrix. The resulting matrix is assumed to have $p \times d$ characteristic, and describes the essential information extracted from each SVID_i matrix, unfolded into an X_{DRT}-matrix, which can be used to fit the classifiers in a subsequent hyper-parameter tuning operation. The matrix X_{DRT} can be expressed as follows:

$$\text{X}_{\text{DRT}} \in \mathbb{R}^{N\times(d\times p)}, \quad \text{where } \text{p} = 11, \ \& \ d = \{10, \cdots, 20\} \tag{8}$$

As mentioned previously, ensemble machine learning (EML) algorithms are used in the current study to classify normal and abnormal wafers. Typically, EML attempts to build a robust classifier by combining multiple weak models. Naturally, such algorithms create and connect numerous learners to achieve better results. Generally, depending on the approach used to develop and train the models, the EML techniques can be grouped into two categories: (1) Bagging methods such as Bootstrap Aggregating (Bagging), Extra Trees Classifier (ETC), and Random Forest Classifier (RFC), and (2) Boosting methods like Adaptive Boosting (AdaBoost), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB).

Sure hyper-parameters must be carefully determined to build an appropriate learning strategy and provide an adequate AD model. For instance, standard hyper-parameter tunings for the boosting and bagging methods concern the maximum depth, the number of estimators, and the learning rate. We searched over a wide range of values from 1 to 10 for the maximum depth, and over a number of estimators ranging from 50 to 600 and a learning rate between 0.1 and 1. Overall, we evaluated $(3 \times 11 \times 6 = 198)$ configurations, consisting of 3 dimensionality reduction techniques, 11 component combinations between 10 and 20, and 6 EML algorithms. Furthermore, each configuration is evaluated using all of the hyper-parameters mentioned above to define the most suitable configuration for the problem. Each model is trained and tested using the same database to ensure comparability and consistency. The dataset is partitioned on train and test datasets before standardization according to the ratio $8:2$ and by respecting the fault types stratifying. Thus, 80% of the dataset (*i.e.,* 4000 normal samples and 1600 abnormal samples) is used to fit the EML algorithms, and the remaining 20% (*i.e.,* 1000 normal samples and 400 abnormal samples presenting 80 samples from each of the 5 faults) is used to test and evaluate the fitted algorithms performance.

V. Experiments and Discussion

This section describes the experimental results of the proposed method. We conducted experiments applying three anomaly detection systems based on PCA, FA, and ICA. Each strategy involves six of the most common EML algorithm up-

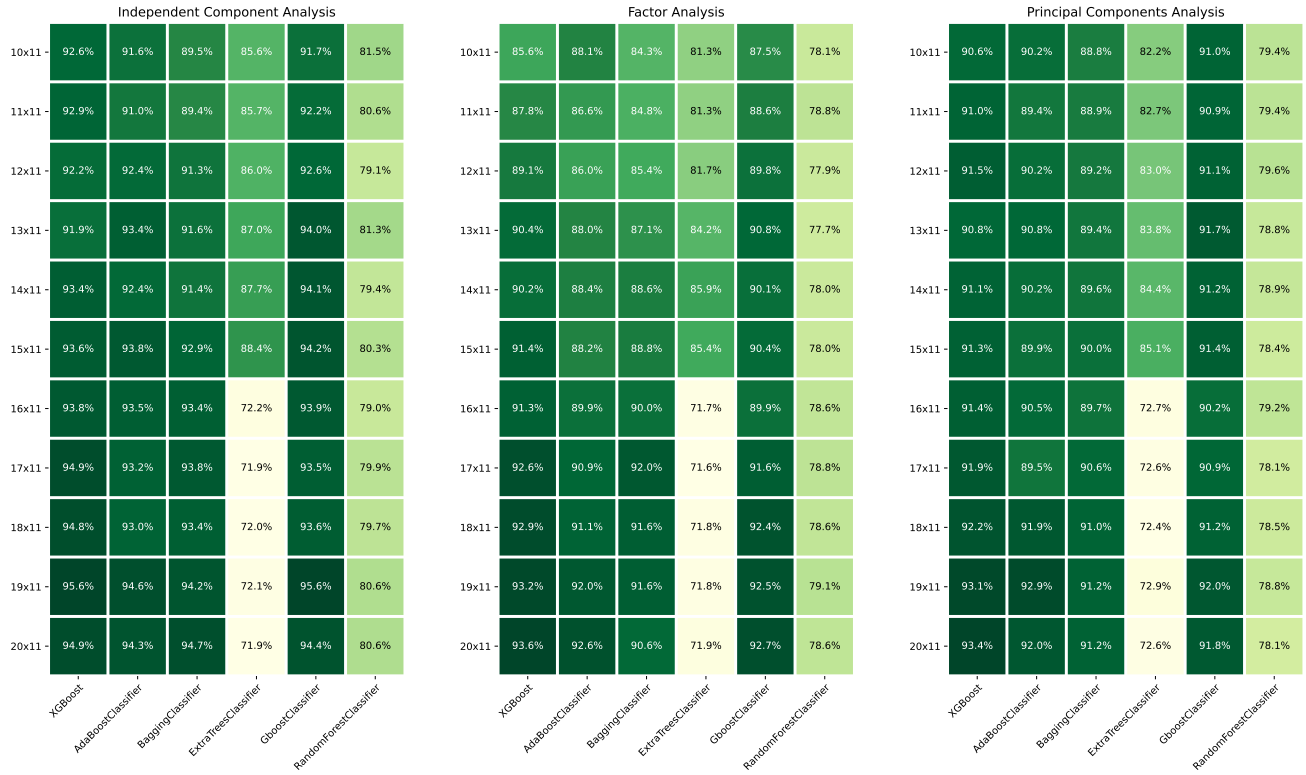

Fig. 5: Detection model performances. Herein, we report the best accuracy-test values obtained for each classifier. The performance is computed in percentages. The study includes subsets of 10 to 20 components built by PCA, FA, and ICA.

grades, such as Bootstrap Aggregating (Bagging), Extra Trees Classifier (ETC), Random Forest Classifier (RFC), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB). All terms are the best results for the anomaly detection models obtained with optimized parameters.

In Fig. 5, we present the three anomaly detection models based on PCA, FA, and ICA. For each model, we have indicated: the best recognition rate obtained for each classifier, considering different combinations of features between 10 and 20 expressed as a multiple of 11, equal to the number of reduced $SVID_i$ matrices. A global overview of performance as a function of classifiers, the dimensionality reduction technique used, and the number of retained components are summarized in the figure below.

Out of the presented results, the best model is obtained using ICA and XGBoost (i.e., XGB) with 19×11 components. The model achieves the best accuracy-test of 95.64%. The model has been built based on the following hyper-parameters: a number of estimators equal to 298, a maximum depth equal to 3, and a learning rate equal to 0.3. For a more in-depth insight into the model's performance, we use the Table below, which indicates the scores of the different metrics across the entire test database and the performance against each type of defect. The XGB model classifies a total of 1339 samples correctly (i.e. 987 out of 1000 normal samples and 352 out of 400 abnormal samples). The 61 incorrectly classified samples are distributed as follows: 13 samples from normals, 5 samples from fault 1, 4 from fault 2, 13 from fault 3, 14 from fault 4, and 12 from fault 5 (metric scores are provided in Tab.I). Let's mention that faults 3, 4, and 5 are the most difficult to detect compared to the first two because of their nature (noise and sinusoidal disturbances, and peripheral point) which makes them very difficult to identify, see Fig. 1. We need to clarify that the precision metric in Tab.I refers to the positive class of normals. The findings demonstrate the interest in EML algorithms for anomaly detection and wafer classification during semiconductor manufacturing. In addition to being highly effective for anomaly detection, EML algorithms are designed based on rules generation that sentences can describe.

TABLE I: Detection Performance

	Best model evaluation metrics for each fault				
	Pre (%)	Sen (%)	Spe (%)	F1 (%)	Acc (%)
Overall	95.36	98.7	88.0	97.0	95.64
Fault 1	99.5	98.7	93.75	99.1	98.33
Fault 2	99.6	98.7	95.0	99.15	98.43
Fault 3	98.7	98.7	83.75	98.7	97.59
Fault 4	98.6	98.7	82.5	98.65	97.5
Fault 5	98.8	98.7	85.0	98.75	97.69

Such rules can improve the production process by locating the most common causes of anomalies.

VI. Conclusion

In this paper, we have studied the problem of anomaly detection in multivariate time series using different transformation techniques and numerous state-of-the-art algorithms. This study is conducted on a set of eleven features derived from intermediate manufacturing equipment that characterize the wafer status. We proposed a preprocessing scheme that restructures the data distribution and reshapes the wafer matrix into SVID_i data matrices to exploit the maximum Machine-derived informational features. Overall, we evaluated $(3 \times 11 \times 6 = 198)$ anomaly detection designs. In effect, we opted for three dimensionality reduction techniques, each of which was used to infer several subsets of data using a multitude of components between 10 and 20 per SVID_i matrix, i.e., 11 structures to be evaluated. Six EML algorithms are used to score each given data subset. The best classifiers are determined according to the recognition rate of a normal and abnormal wafer. This work aims to provide a reliable contribution to the research community on the performance of novel approaches that are still not implemented in this area of investigation and present a precious basis for comparison between robust algorithms and dimensionality reduction techniques to detect anomalies.

References

1. Ahmed, M., Naser Mahmood, A. & Hu, J. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications* (2016).
2. Chen, S.-M. & Chen, S.-W. Fuzzy Forecasting Based on Two-Factors Second-Order Fuzzy-Trend Logical Relationship Groups and the Probabilities of Trends of Fuzzy Logical Relationships. *IEEE Transactions on Cybernetics* (2015).
3. Li, Z., Zhao, Y., Botta, N., Ionescu, C. & Hu, X. *COPOD: Copula-Based Outlier Detection* 2020.
4. Zhao, Y., Rossi, R. A. & Akoglu, L. *Automating Outlier Detection via Meta-Learning* 2021.
5. Kang, H. & Choi, S. Bayesian common spatial patterns for multi-subject EEG classification. *Neural Networks* (2014).
6. Trardi, Y., Ananou, B., Haddi, Z. & Ouladsine, M. *A Novel Method to Identify Relevant Features for Automatic Detection of Atrial Fibrillation* in *26th MED* (2018).
7. Trardi, Y., Ananou, B. & Ouladsine, M. *An Advanced Arrhythmia Recognition Methodology Based on R-waves Time-Series Derivatives and Benchmarking Machine-Learning Algorithms* in *ECC* (2020).
8. Lahmiri, S. A Variational Mode Decompoisition Approach for Analysis and Forecasting of Economic and Financial Time Series. *Expert Systems with Applications* (2016).
9. Rosas-Romero, R., Dıéaz-Torres, A. & Etcheverry, G. Forecasting of Stock Return Prices with Sparse Representation of Financial Time Series over Redundant Dictionaries. *Expert Syst. Appl.* (2016).
10. Hsu, C.-Y., Chen, W.-J. & Chien, J.-C. Similarity matching of wafer bin maps for manufacturing intelligence to empower Industry 3.5 for semiconductor manufacturing. *Computers & Industrial Engineering* (2020).
11. Wang, L., Wang, Z. & Liu, S. An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. *Expert Systems with Applications* (2016).
12. Nayak, R., Pati, U. C. & Das, S. K. A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing* (2021).
13. Braei, M. & Wagner, S. *Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art* 2020. eprint: 2004.00433.
14. Cheng, H., Tan, P.-N., Potter, C. & Klooster, S. in *Proceedings of the 2009 SIAM International Conference on Data Mining (SDM)* (2009).
15. Baydogan, M. G. & Runger, G. Time series representation and similarity based on local autopatterns. *Data Mining and Knowledge Discovery* (2016).
16. Buvé, C. *et al.* Application of multivariate data analysis for food quality investigations: An example-based review. *Food Research International* (2022).
17. Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K. & Chang, L. *A novel anomaly detection scheme based on principal component classifier* tech. rep. (2003).
18. Kwitt, R. & Hofmann, U. Robust methods for unsupervised PCA-based anomaly detection. *Proc. of IEEE/IST WorNshop on Monitoring, AttacN Detection and Mitigation* (2006).
19. Wong, J., Colburn, C., Meeks, E. & Vedaraman, S. Rad outlier detection on big data. *Web blog post. The Netflix Tech Blog. Netflix* (2015).
20. Sorzano, C., Vargas, J. & Montano, A. A survey of dimensionality reduction techniques (2014).
21. Fan, S.-K. S., Lin, S.-C. & Tsai, P.-F. Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, AdaBoost and decision tree. *Journal of Industrial and Production Engineering* (2016).
22. Tchatchoua, P., Graton, G., Ouladsine, M. & Juge, M. *A Comparative Evaluation of Deep Learning Anomaly Detection Techniques on Semiconductor Multivariate Time Series Data* in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)* (2021).
23. Hsu, C.-Y. & Liu, W.-C. Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. *Journal of Intelligent Manufacturing* (2021).