Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-6 (2022) 476–481

Semiconductor Multivariate Time-Series Anomaly Classification based on Machine Learning Ensemble Techniques ⋆

Samia MELLAH * **Youssef TRARDI** *
Guillaume GRATON *,** **Bouchra ANANOU** *
El Mostafa EL ADEL * **Mustapha OULADSINE** *

* *Aix Marseille Univ., Université de Toulon, CNRS, LIS (UMR 7020), Avenue Escadrille Normandie-Niemen, F-13397 Marseille Cedex 20, France, (e-mail: samia.mellah@lis-lab.fr, youssef.trardi@lis-lab.fr)*
** *Ecole Centrale Marseille, Technopôle de Château-Gombert, 38 rue Frédéric Joliot-Curie, F-13451 Marseille Cedex 13, France.*

Abstract: This paper proposes an efficient multivariate time-series fault detection and classification approach aiming to detect faulty wafers (*i.e.* pieces of silicon) during semiconductor manufacturing process. This approach is based on using Independent Component Analysis (ICA) and several Machine Learning Ensemble Techniques. The main objective is to extract the most useful information from each time-series and combine them to build a set of fully concatenated features. Thereafter, Extra Trees, Random Forest, Gradient Boosting and Extreme Gradient Boosting, one of the prevalent evolutions of tree-based algorithms, are fitted to the extracted features subset to design and implement an efficient anomaly detection strategy. The obtained results show that the proposed technique is very efficient and very promising.

Keywords: Anomaly detection, Data-driven diagnosis methods, Machine learning, Multivariate time-series data.

1. INTRODUCTION

The semiconductor industry is specialized in the design and the fabrication of semiconductor devices such as electronic integrated circuits and silicon-based photovoltaic cells. This is done by applying different processing techniques on wafers like doping, implantation, etching...

Semiconductor manufacturing is a highly complex and long fabrication process with 300–500 steps Chien et al. (2013). In order to improve the production cost, delivery time and quality, the technology used in the semiconductor manufacturing processes is in a continuous evolution.

In the semiconductor factory, modifications in equipment physical status (*e.g.* pressure, temperature, chemical gas flow, humidity, ...) during the manufacturing processes may negatively impact the process. The occurrence of faults or anomalies during the manufacturing process is unwanted because it impacts the product quality or even it can lead to the process deterioration. Thereby, this implies the loss of hundreds of thousands or millions of dollars depending on the size and scale of the operations. Hence, early detection of faulty wafers is very important to ensure the process operations control and to reduce yield loss Hsu et al. (2020).

The fault detection and classification (FDC) requires developing models for supervising wafer manufacturing results. A good FDC model detects wafer faults at an early manufacturing stage and prevents the release of faulty wafers to the posterior steps. Thereby, it helps the production yield improvement by reducing the cost.

FDC is a widely studied problem in the literature. In the semiconductor field, numerous FDC approaches have been proposed, varying from statistical process control and virtual metrology Kim et al. (2012); Susto et al. (2012) to machine learning and deep learning algorithms.

For example, a multiblock Principal Component Analysis based on a combined index is proposed in Cherry and Qin (2006), whereas authors in He and Wang (2007) propose to reduce the data dimension using PCA before applying the K-Nearest Neighbours algorithm to detect faulty wafers. In Mahadevan and Shah (2009), One-class Support vector machine is introduced to detect wafer anomalies, and in Chien et al. (2013), the association between faulty wafers and sensor variables is analysed using decision trees. To establish the cause-effect relationships between root causes, equipment, and process parameters, authors in Nawaz et al. (2014) have introduced an expert knowledge-based approach, while AdaBoost and decision tree algorithms are used in Fan et al. (2016) to classify defective wafers. Regarding the Deep Learning methods, a method based on a Deep Neural Network is proposed in Lee et al. (2016),

⋆ This paper is conducted in the framework of the project MADEin4, which has received funding from the ECSEL JU (Electronic Components and Systems for European Leadership Joint Undertaking) under grant agreement No 826589. The JU receives support from the European Union's Horizon 2020 research and innovation program and France, Germany, Austria, Italy, Sweden, Netherlands, Belgium, Hungary, Romania and Israel.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2022.07.174

while a multiple time-series convolution neural network is considered in Hsu and Liu (2021).

In this paper, a new strategy is proposed to detect wafer faults at an early manufacturing stage using Independent Component Analysis (ICA) to extract relevant information from each time-series. After that, Machine Learning Ensemble Techniques are fitted to the extracted features subset to generate an efficient anomaly detection model.

The remainder of this paper is organized as follows: Section 2 presents the industrial context. Section 3 discusses the data preprocessing technique and the machine learning considered algorithms, followed by the data description. After that, machine learning algorithms configuration and evaluation metrics are presented in Section 4. This is followed by the obtained results and discussion in Section 5. Finally, Section 6 concludes the paper.

2. CONTEXT

Within the framework of industry4.0 and intelligent manufacturing, large amount of sensors are installed and make possible to record and collect automatically the data generated by the production and testing equipment during complex semiconductor manufacturing processes Oztemel and Gursev (2020). Since the sensors generate signals sequentially at some time intervals, the collected data which reveal some temporal dependencies can be considered as multivariate time-series.

Time-series data in semiconductor manufacturing are also called Fault detection and classification (FDC) data. They consist on three dimensional information including wafer, status variable identification (SVID), and recorded time. The SVID represents the status of equipment or machine such as temperature, pressure, and gas flow Rostami et al. (2018); Hsu and Liu (2021).

Multivariate time-series signals are one of the most complicated signals to analyze for detecting anomalies Kim et al. (2018). Fig. 1 summarizes the difference between univariate and multivariate anomaly detection. An univariate anomaly can be detected by analyzing the individual variable (blue signal in the second sub-figure), but a multivariate anomaly cannot be found without watching the multiple variables at once (green and blue signals in the last sub-figure).

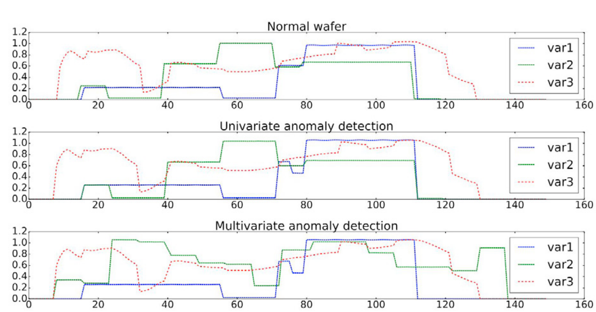

Fig. 1. Univariate and multivariate anomaly detection

Generally, the collected data is segmented and reduced into a set of statistical data such as the mean, the median, the minimum, the maximum, the standard deviation, or the percentiles Park et al. (2014). FDC methods are then elaborated using this set of statistical data. Nevertheless, this reduction presents a risk. Actually, it leads to the loss of information. Hence, a set of data cannot always be summarized into its reduced statistical representation Anscombe (1973). Thus, exploring the "raw" data can give better performance.

In this paper, an effective approach is proposed to detect faulty wafers in an early manufacturing stage using the "raw" collected data. To proceed, ICA is firstly used to extract useful information from each time-series. Then, the generated independent components are concatenated to build a new data subset containing significant information. Finally, four Machine Learning Ensemble Techniques, namely Extra Trees (ET), Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB), are fitted to the obtained subset to generate an efficient anomaly detection model.

3. METHODS

3.1 Batch process raw data

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 (time-series) during n instants. Hence, the data of wafer j is given by a $(n \times p)$ matrix as follows:

$$W(j) = \begin{bmatrix} s^j_{1,1} & s^j_{2,1} & \cdots & s^j_{p-1,1} & s^j_{p,1} \\ s^j_{1,2} & s^j_{2,2} & \cdots & s^j_{p-1,2} & s^j_{p,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s^j_{1,n-1} & s^j_{2,n-1} & \cdots & s^j_{p-1,n-1} & s^j_{p,n-1} \\ s^j_{1,n} & s^j_{2,n} & \cdots & s^j_{p-1,n} & s^j_{p,n} \end{bmatrix}$$

3.2 Data preprocessing

The preprocessing proposed methodology is illustrated in Fig. 2. The data is saved in a *csv* file under two-dimensional form where each column represents one SVID and each n rows represent one wafer data (each SVID is measured for n time steps for each wafer).

First of all, data reshaping is achieved to build a three-dimensional structured dataset. The reshaped data is composed of p two-dimensional matrices. Each two-dimensional matrix represents a time-series SVID measurements for all the database wafers, where the k rows represent the k wafers, and the n columns represent time instants.

After that, ICA is applied on each time-series (*i.e.* each SVID) to extract its relevant and significant information that consist of the Independent Components (ICs) of the time-series signals. This allows not only to extract the most useful information, but also to reduce the data dimensionality. All the similar time-series signals are grouped in the same IC.

Finally, data unfolding (*i.e.* three-dimensional data representation transforming into a two-dimensional matrix) is processed to get a two-dimensional representative data to fit the Machine Learning models.

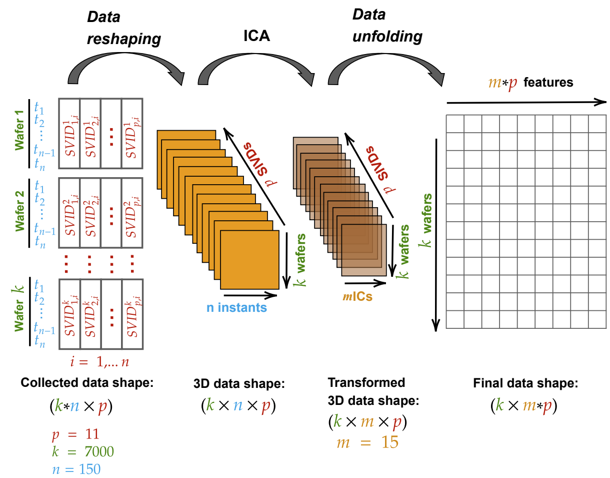

Fig. 2. Data preprocessing technique

3.3 Machine Learning Ensemble Techniques

As mentioned before, Ensemble Techniques are used in this work to get a classification model allowing to classify good and faulty wafers. Ensemble Techniques are Machine Learning algorithms that attempt to build a strong classifiers from a number of weak models. They are based on creating and combining multiple weak learners (*i.e* models) to produce improved results. Often, these techniques provide better performance than a single model. Generally, based on how they are created and trained, Ensemble Techniques can be grouped into two different categories, namely: 1) Bagging methods such as ET and RF classifiers, and 2) Boosting methods like GB and XGB classifiers, see Fig. 3.

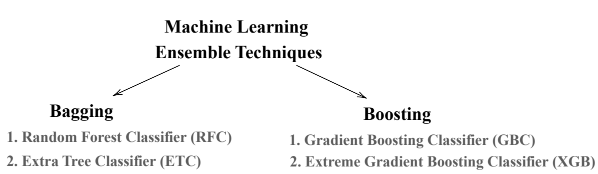

Fig. 3. Ensemble Techniques categories

The difference between these two categories is illustrated in Fig. 4. Bagging models are built independently and learn in a parallel way, while each new model in boosting techniques is influenced by the performance of previously built models. Another major difference is that in bagging techniques, different training data subsets are randomly selected with replacement from the entire training dataset, whereas in boosting techniques, every new subset contains the elements that are misclassified by previous models.

Besides of their high performances (see Section 5), Ensemble Techniques are based on generating rules, and the rules can be described by sentences, which are understandable by human beings. This allows to facilitate the engineer's work and can help to enhance the production process.

3.4 Data description

The raw data is measured and saved every second, for a recipe enduring approximately 150 seconds. A total of

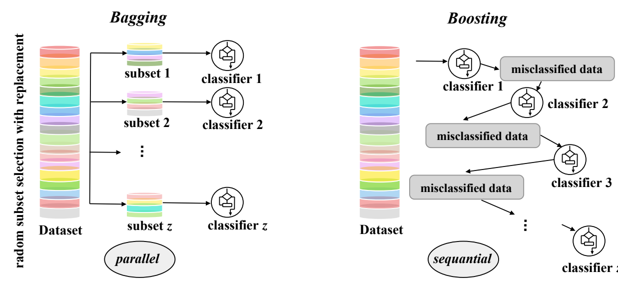

Fig. 4. Bagging and boosting methods building concept

11 SVIDs are monitored, which are simulations with a nature very close to real production variables like pressure, temperature, gas flows, and capacitance of an etch tool. For this study, the dataset is constituted of a total of 7000 samples (*i.e.* wafers) distributed as 5000 normal samples and 2000 faulty samples. The faulty data constitutes a ratio of 28.6% which is a good one regarding the rarity of faulty data in the semiconductor industry. The faulty dataset presents five different types of faults, equally distributed in the dataset (400 samples per fault type). These fault types consists on common faults occurring during wafer processing. They appear on different variables and are either atomic or aggregate anomalies. The atomic anomalies are cases with deviant values for one variable (univariate anomaly) while the aggregate ones result from groups of variables deviating as a collective (multivariate anomaly), see Fig. 1. These anomalies are present on some process steps and not during the whole processing.

Fig. 5 illustrates the five fault types. Fault 1 is like a bias. Fault 2 is a temporary change in value with a return to a regular level after some time steps. Faults 3 and 4 are similar to additive noise and sinusoidal disturbances respectively. Finally, Fault 5 is a peak resulting from a sudden rise in value.

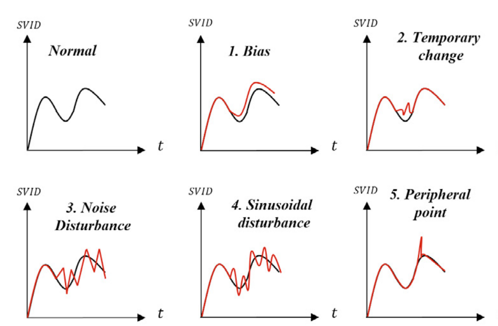

Fig. 5. Fault types

This work aims to detect faults, whatever their type. The classification and the diagnosis of the fault types is not the main focus here. However, the generated models will be evaluated on each fault type individually to study the performance of detecting each fault type.

4. MODEL CONFIGURATION AND EVALUATION METRICS

Since the measurements of the different SVIDs (*e.g.* pressure, temperature, ...) during the manufacturing process are not on the same numerical scales, the raw data are standardized to range between 0 and 1 before the preprocessing (*i.e.* before applying ICA) to avoid scale errors.

Regarding the number of the kept ICs for each SVID time-series, an optimization has been done and the optimal obtained number is in the order of 15. Hence, each SVID time-series with a length of n instants is represented by 15 ICs (see Fig. 2).

4.1 Models configuration

As mentioned before, four Ensemble Techniques are used in this work. Two bagging algorithms: 1) ETC, 2) RFC, and two boosting algorithms: 3) GBC, and 4) XGB.

The methods are implemented in Python 3.8.8 with the scikit-learn 0.24.1 library for RFC, ETC, and GBC, and the xgboost 1.5.0 library for XGB. Experiments are carried out on a personal computer with Windows 10 CPU @1.80 GHz, 16 Go RAM.

To get the optimal models, an optimization has been processed for all the considered techniques in order to get the hyper-parameters giving the best performances.

The optimal number of estimators are given for RFC, ETC, GBC, and XGB receptively by 142, 60, 482, and 316. The best criterion for both RFC and ETC is "gini". Consisting on a more regularized form of GBC, XGB requires more hyper-parameters to optimize. The optimal value obtained for the "max depth" parameter is 3, while the optimal obtained value of the "learning rate" parameter is 0.3.

The dataset is partitioned on train and test data sets before standardization according to the ratio 8 : 2 and by respecting the fault types stratification. Thus, 80% of the dataset (*i.e.* 4000 normal samples and 1600 faulty samples) is used to fit the Machine Learning algorithms, and the remaining 20% (*i.e.* 1000 normal samples and 400 faulty samples presenting 80 samples from each of the 5 faults) is used to test and evaluate the fitted algorithms performance. Regarding the dataset classes, the class of normal wafers has been assigned as the positive class with a positive label value (1), and the faulty wafers form the negative class with (-1) as label value.

4.2 Evaluation metrics

In a supervised binary classification where normal samples are assigned to the positive class and abnormal samples to the negative class, model evaluation metrics are based on terminologies extracted basically from a (2×2) matrix known as "confusion matrix" based on real and predictive classes (see Fig. 6).

These terminologies are defined as follows:

(a) True Positive (TP), designs the number of positive samples correctly predicted as positive.
(b) True Negative (TN), designs the number of negative samples correctly predicted as negative.
(c) False Negative (FN), designs the number of positive samples predicted as negative.
(d) False Positive (FP), designs the number of negative samples predicted as positive.

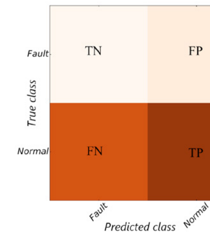

Fig. 6. Confusion matrix

Following metrics are calculated using these terminologies:

- Precision (*Pre*) refers to the classifier ability not to confuse an actual positive sample with a true negative.
$$\text{Pre} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$
- Sensitivity (*Sen*) refers to the classifier ability to detect all positive (*i.e.* abnormal) samples.
$$\text{Sen} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$
- Specificity (*Spe*) refers to the classifier ability to detect all negative (*i.e.* normal) samples.
$$\text{Spe} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$
- F1 score can be interpreted as a weighted average of the precision and recall.
$$\text{F1} = 2 \times \frac{\text{Pre} \times \text{Sen}}{\text{Pre} + \text{Sen}}$$
- Accuracy (*Acc*) refers to the total number of correct predictions performed by the model.
$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}$$

5. RESULTS AND DISCUSSION

The five aforementioned metrics are used in this work to evaluate the generated models performance for each of the four considered Ensemble Techniques. All the classifiers obtained performances for the overall test dataset, then for each fault type test data subset, are summarized in Table 1.

Considering the positive and the negative classes related to normal and faulty wafers respectively, precision, sensitivity, and F-measure metrics are linked to the normal class recognition, while specificity metric is linked to the faulty wafers detection performance. Finally, accuracy considers both normal and faulty wafers and studies their correct classification.

Bold terms in Table 1 are the best comparison results for the anomaly detection considered models. In the following, the results will be discussed firstly for bagging models, then for boosting models.

- Bagging models:
  (1) ETC classifies correctly a total of 1362 samples (*i.e.* 997 over 1000 normal samples and 365 over 400 faulty samples), which is a total of 97.28% of the test data. The 35 faulty misclassified samples are distributed as follows: 6 samples of Fault 1, 1 of Fault 2, 12 of Fault 3, 9 of Fault 4, and 7

Table 1. Classification results

Model		Evaluation metrics					Confusion matrix			
		Precision (%)	Sensitivity (%)	Specificity (%)	F-measure (%)	Accuracy (%)	TN	FP	TP	FN
ETC	Overall	96.61	99.70	91.25	98.13	97.29	365	35	997	3
	Fault 1	99.40	**99.70**	92.50	99.55	99.17	74	6	997	3
	Fault 2	**99.90**	**99.70**	**98.75**	**99.80**	**99.63**	79	1	997	3
	Fault 3	98.81	**99.70**	85.00	99.25	98.61	68	12	997	3
	Fault 4	99.11	**99.70**	88.75	99.40	98.89	71	9	997	3
	Fault 5	99.30	**99.70**	91.25	99.50	99.07	73	7	997	3
RFC	Overall	98.61	99.20	**96.50**	98.90	98.43	386	14	992	8
	Fault 1	**99.90**	99.20	**98.75**	99.55	99.17	79	1	992	8
	Fault 2	**99.90**	99.20	**98.75**	99.55	99.17	79	1	992	8
	Fault 3	99.30	99.20	91.25	99.25	98.61	73	7	992	8
	Fault 4	99.50	99.20	93.75	99.35	98.80	75	5	992	8
	Fault 5	**100**	99.20	**100**	99.60	99.26	80	0	992	8
GBC	Overall	96.70	99.70	91.50	98.18	97.36	366	34	997	3
	Fault 1	99.50	**99.70**	93.75	99.60	99.26	75	5	997	3
	Fault 2	99.70	**99.70**	96.25	99.70	99.44	77	3	997	3
	Fault 3	99.20	**99.70**	90.00	99.45	98.98	72	8	997	3
	Fault 4	98.42	**99.70**	80.00	99.06	98.24	64	16	997	3
	Fault 5	99.80	**99.70**	97.50	99.75	99.54	78	2	997	3
XGB	Overall	**98.62**	**99.70**	**96.50**	**99.15**	**98.79**	386	14	997	3
	Fault 1	**99.90**	**99.70**	**98.75**	**99.80**	**99.63**	79	1	997	3
	Fault 2	**99.90**	**99.70**	**98.75**	**99.80**	**99.63**	79	1	997	3
	Fault 3	**99.50**	**99.70**	**93.75**	**99.60**	**99.26**	75	5	997	3
	Fault 4	**99.40**	**99.70**	**92.50**	**99.55**	**99.17**	74	6	997	3
	Fault 5	99.90	**99.70**	98.75	**99.80**	**99.63**	79	1	997	3

of Fault 5. By analyzing these results, the ETC model recognises well normal class and Fault 2. However, it is not well performing in detecting the other faults, mostly Fault 3 and 4.

(2) RFC classes correctly a total of 1378 samples (*i.e.* 992 over 1000 normal samples and 386 over 400 faulty samples). This gives a total of 98.42% well classified data. Compared to ETC, RFC performs well on detecting the faults (*i.e.* it misclassifies a total of only of 14 faulty samples compared to 35 samples for ETC), but it is less good in classifying normal samples (*i.e.* it misclassifies 8 normal samples compared to only 3 samples for ETC). Considering the different types of faults, RFC classifies perfectly Fault 5, followed by Fault 1 and 2, while it is not very good on detecting Fault 3 and Fault 4.

To compare these two bagging models, ETC is the best one if we give more importance to the normal samples classification, when RFC is the best one if detecting the faulty wafers is the priority.

- Boosting models:

(3) GBC classes correctly a total of 1363 samples (*i.e.* 997 over 1000 normal samples and 366 over 400 faulty samples), which presents a total of 97.35% of the test data. GBC performs in the same way as RTC in classifying normal samples (*i.e.* it misclassifies only 3 samples), while it is better in classifying Fault 3 and 5.

(4) XBC classifies correctly a total of 1383 samples (*i.e.* 997 over 1000 normal samples and 386 over 400 faulty samples). This presents a total of 98.78% of the test data. Like ETC and GBC, XBG classifies perfectly the normal samples. With respect to faults, XGB like RFC, misclassifies only 14 over 400 samples. Compared to RFC, the results of detecting each fault type is quasi-similar.

Finally, since XBG model is the one that classifies better both of normal and faulty wafers, it is the best performing model with the following overall metrics: Precision: 98.62%, Sensitivity: 99.70%, Specificity: 96.50%, F-measure: 99.15%, and Accuracy: 98.79.

Note that Fault 3 and 4 are the least detected faults compared to the three others, because of their nature (*i.e.* Noise and Sinusoidal disturbances) which makes them very difficult to be detected, see Fig. 5.

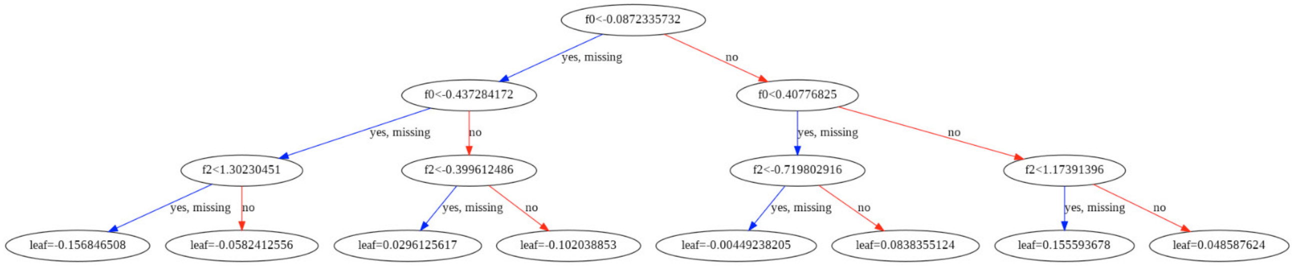

Fig. 7. One of the XGB weak learners

The results show that Ensemble Techniques are a very promising Machine Learning algorithms for wafers anomaly detection and classification during semiconductor manufacturing process. Depending on the equipment processing where they are implemented, they allow to detect anomalies at a early stage, and consequently, to reduce the production yield loss.

Besides of being very effective for the detection of anomalies, Ensemble Techniques are based on generating rules which can be described by sentences. This makes them understandable by human beings and allows to enhance the production process by localizing the most current fault causes. For example, Fig. 7 illustrates one weak model of the XGB generated algorithm. To classify a new test data, the output of each weak learner is taken into account. By following all the weak learners output, the classification result can be explained.

6. CONCLUSION

In this paper, wafer anomaly detection during manufacturing process issues is considered and a new approach based on data preprocessing and Ensemble Machine Learning Algorithms is proposed. The aim is to detect wafer anomalies in an early stage . Firstly, significant information is extracted from wafer time-series data using Independent Component Analysis (ICA). Then, the extracted features are concatenated to fit four Machine Learning Ensemble Techniques, namely Extra Trees, Random Forest, Gradient Boosting, and Extreme Gradient Boosting classifiers. The very optimist results obtained in classifying normal and faulty wafers show that the proposed approach is very promising.

REFERENCES

Anscombe, F.J. (1973). Graphs in statistical analysis. *The american statistician*, 27(1), 17–21.

Cherry, G.A. and Qin, S.J. (2006). Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Transactions on semiconductor manufacturing*, 19(2), 159–172.

Chien, C.F., Hsu, C.Y., and Chen, P.N. (2013). Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence. *Flexible Services and Manufacturing Journal*, 25(3), 367–388.

Fan, S.K.S., Lin, S.C., and Tsai, P.F. (2016). Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, adaboost and decision tree. *Journal of Industrial and Production Engineering*, 33(3), 151–168.

He, Q.P. and Wang, J. (2007). Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE transactions on semiconductor manufacturing*, 20(4), 345–354.

Hsu, C.Y., Chen, W.J., and Chien, J.C. (2020). Similarity matching of wafer bin maps for manufacturing intelligence to empower industry 3.5 for semiconductor manufacturing. *Computers & Industrial Engineering*, 142, 106358.

Hsu, C.Y. and Liu, W.C. (2021). Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. *Journal of Intelligent Manufacturing*, 32, 823–836.

Kim, C., Lee, J., Kim, R., Park, Y., and Kang, J. (2018). Deepnap: Deep neural anomaly pre-detection in a semiconductor fab. *Information Sciences*, 457, 1–11.

Kim, D., Kang, P., Cho, S., Lee, H.j., and Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications*, 39(4), 4075–4083.

Lee, H., Kim, Y., and Kim, C.O. (2016). A deep learning model for robust wafer fault monitoring with sensor measurement noise. *IEEE Transactions on Semiconductor Manufacturing*, 30(1), 23–31.

Mahadevan, S. and Shah, S.L. (2009). Fault detection and diagnosis in process data using one-class support vector machines. *Journal of process control*, 19(10), 1627–1639.

Nawaz, J.M., Arshad, M.Z., and Hong, S.J. (2014). Fault diagnosis in semiconductor etch equipment using bayesian networks. *JSTS: Journal of Semiconductor Technology and Science*, 14(2), 252–261.

Oztemel, E. and Gursev, S. (2020). Literature review of industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31(1), 127–182.

Park, E.L., Park, J., Yang, J., Cho, S., Lee, Y.H., and Park, H.S. (2014). Data based segmentation and summarization for sensor data in semiconductor manufacturing. *Expert systems with applications*, 41(6), 2619–2629.

Rostami, H., Blue, J., and Yugma, C. (2018). Automatic equipment fault fingerprint extraction for the fault diagnostic on the batch process data. *Applied Soft Computing*, 68, 972–989.

Susto, G.A., Pampuri, S., Schirru, A., De Nicolao, G., McLoone, S.F., and Beghi, A. (2012). Automatic control and machine learning for semiconductor manufacturing: Review and challenges. In *Proceedings of the 10th European Workshop on Advanced Control and Diagnosis (ACD 2012)*.