Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

INTERPRETABLE ANOMALY DETECTION FOR KNOWLEDGE DISCOVERY IN SEMICONDUCTOR MANUFACTURING

Mattia Carletti
Marco Maggipinto
Alessandro Beghi
Gian Antonio Susto

Department of Information Engineering
University of Padova
Via Gradenigo 6/B
Padova, 35131, ITALY

Natalie Gentner
Yao Yang
Andreas Kyek

Infineon Technologies AG
Am Campeon 1-15
Neubiberg, 85579, GERMANY

ABSTRACT

Machine Learning-based Anomaly Detection approaches are efficient tools to monitor complex processes. One of the advantages of such approaches is that they provide a unique anomaly indicator, a quantitative index that captures the degree of 'outlierness' of the process at hand considering possibly hundreds or more variables at the same time, the typical scenario in semiconductor manufacturing. One of the drawbacks of such approaches is that Root Cause Analysis is not guided by the system itself. In this work, we show the effectiveness of a method, called DIFFI, to equip Isolation Forest, one of the most popular Anomaly Detection algorithms, with interpretability traits that can help corrective actions and knowledge understanding. Such approach is validated on real world semiconductor manufacturing data related to a Chemical Vapor Deposition process.

1 INTRODUCTION AND LITERATURE REVIEW

In complex manufacturing environments like semiconductor manufacturing with hundreds of machines involved in the day-by-day production of a single fab, monitoring of processes is a crucial activity to perform corrective actions in time and ensure elevated production quality and reduced costs associated with scrap products.

Typically monitoring is achieved by means of univariate control charts over quantities like key process indicators or statistics/features of the sensor data collected from the equipment. This approach, widespread in many industries, is effective for its simplicity, but has several drawbacks, namely:

- univariate approaches in many cases are not able to capture multi-dimensional anomalies (see Figure 1);
- univariate control charts rely on the fact that monitored quantities follow Gaussian distributions, that is difficultly the case in real-world applications;
- when dealing with dozens of processes (Pampuri et al. 2012), this generates hundreds of control charts to be monitored, making users overwhelmed by this quantity of information and leading such users to not act systematically or even ignore the suggestions coming from these tools.

Machine Learning (ML)-based Anomaly Detection (AD) approaches allow to overcome the above-mentioned issues, providing a unique module that is able to jointly consider all the data related to a system

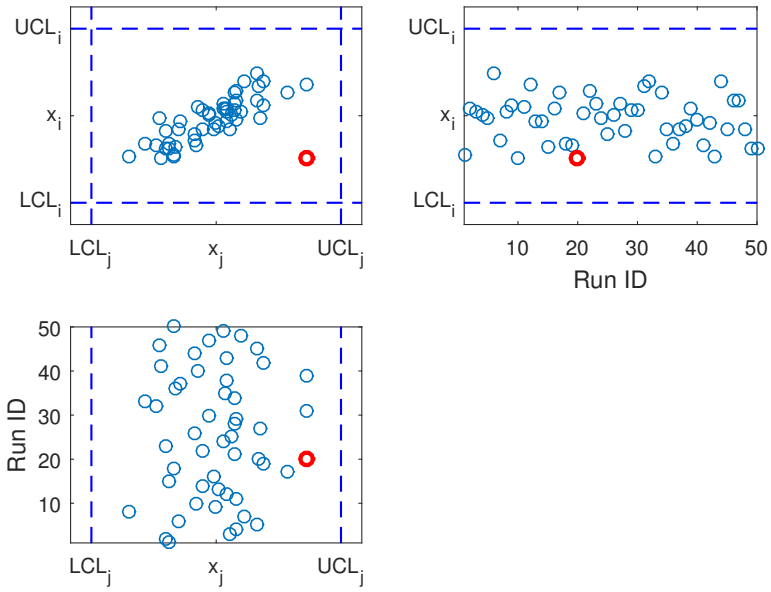


Figure 1: Univariate Control Charts; in this 2-dimensional example, an outlier, represented by the red circle (Top-left panel), cannot be detected by monitoring the control charts related to x_i (Top-right panel) and x_j (Bottom-left panel).

and to provide a degree of 'outlierness' for the process iteration under exam; AD solutions provide an *anomaly score* - also called with various other names like *deviance index* or *health factor* (Susto et al. 2018) - that summarizes with a unique index how much the system in its current state differs from its past, typical behaviour.

In semiconductor manufacturing literature, several works have proven the effectiveness of AD techniques when dealing with hundreds of monitored quantities (Susto et al. 2017), (Puggini and McLoone 2018). While some works in the literature have used supervised approaches to deal with anomaly detection (Nakazawa and Kulkarni 2019), (Rostami et al. 2016), typically when speaking of AD tasks we are considering forecast approaches when dealing with time-series data or unsupervised approaches for 'static', feature-based monitoring problems. In the past recent years, several Deep Learning-based solutions have been presented for their capabilities of handling time-variant, complex data structures (Chen et al. 2018), (Kim et al. 2018), (Liao et al. 2018), (Maggipinto et al. 2019); nevertheless, feature-based approaches are more widespread and versatile, since they can be applied to any monitoring problem (Puggini and McLoone 2018), (Puggini and McLoone 2016), (Susto et al. 2017). In this context, we will consider in this work the Isolation Forest (Liu et al. 2008), (Liu et al. 2012), a powerful model both in terms of outlier detection accuracy and computational efficiency, that has already been employed in the context of semiconductor manufacturing (Puggini and McLoone 2018).

Beside their main advantages, AD methods have a major drawback that limits their diffusion in many real scenarios, that is the lack of interpretability and, consequently, the lack of a natural approach to guide the operators in performing Root Cause Analysis (RCA) procedures. In this work we employ for the first time in semiconductor manufacturing an interpretability method for Isolation Forest called *Depth-based Isolation Forest Feature Importance* (DIFFI) (Carletti et al. 2019), that provides the user with a sorted list of the variables monitored in the AD task according to their importance; the list provided by DIFFI also comes with scores quantifying the importance of each variable in the AD task, enabling both quick and informed RCA and enhanced process understanding.

We demonstrate the effectiveness of the proposed approach compared with other interpretability methods like Permutation Importance (Altmann et al. 2010), also this applied for the first time, to the best of our

knowledge, in the context of semiconductor manufacturing. As case study, we investigate a Chemical Vapor Deposition (CVD) (Susto et al. 2011) monitoring problem.

The rest of the paper is organized as follows: in Section 2 the Isolation Forest is reviewed and the DIFFI method is illustrated. Section 3 is devoted to provide the experimental settings and results on a CVD case study. Finally, conclusions and future works are highlighted in Section 4.

2 METHODOLOGY

2.1 Isolation Forest

The Isolation Forest (Liu et al. 2008; Liu et al. 2012) is among the most used AD algorithms due to its good performance on a wide variety of application domains, the limited number of hyperparameters to be tuned and the low computational cost. It is an ensemble method which combines the predictions of base estimators, called *isolation trees*, characterized by a tree-like structure. The core idea behind the functioning of the Isolation Forest is based on the fact that anomalies are "*few and different*" (Liu et al. 2008), i.e. they are lower in number and they have very different features values with respect to normal instances. Relying on these properties, each isolation tree performs an isolation procedure aimed at distinguishing between normal and abnormal instances. This is achieved by means of split tests performed at each internal node of the tree, starting from the root node: instances associated to the generic node v are routed to either the left child or the right child, depending on whether the value of the splitting feature f associated to node v is smaller than or greater than a given splitting threshold. Both the splitting feature and the splitting threshold associated to a specific internal node are chosen randomly. The procedure is repeated until the node only contains one instance or the tree reaches a predetermined maximum height. As a result, all instances end up in a leaf node, i.e. a node with no children. In light of the properties listed above, we should expect anomalies to be associated with leaf nodes closer to the root since the isolation of data points far away from the others should require less split tests with respect to normal instances. As mentioned above, the Isolation Forest consists of a collection (an *ensemble*) of isolation trees and typically each tree is trained on a subset (with fixed size ψ) of instances sampled from the original training set, allowing for low computational cost, low memory consumption and reduced swamping and masking effects. We refer the curious reader to (Liu et al. 2008; Liu et al. 2012) for more details on the implications of sub-sampling.

At the evaluation stage, for the generic instance x, we need a systematic way to put together the information provided by each tree and define a unique measure of outlierness, the so-called *anomaly score*. This is done as follows:

$$s(x,\psi) = 2^{-\frac{E(h(x))}{c(\psi)}} \tag{1}$$

where $E(h(x))$ is the average path length of x (i.e. the average number of edges x passes through in its path from the root to the leaf node), $c(\psi)$ is a normalization term given by

$$c(\psi) = \begin{cases} 2H(\psi-1) - 2(\psi-1)/\psi & \text{if } \psi > 2, \\ 1 & \text{if } \psi = 2, \\ 0 & \text{otherwise,} \end{cases}$$

with the harmonic number $H(i)$ estimated as $H(i) \approx \ln(i) + 0.5772156649$. The last step consists in a thresholding operation on the values of anomaly score in order to associate a predicted label $\hat{y}$ with any data point of interest: $\hat{y} = 1$ represents a predicted anomaly, while $\hat{y} = 0$ represents a predicted inlier. Based on the value of the predicted labels, the set $\mathscr{D}$ of data points of interest is partitioned into:

- the subset of predicted inliers $\mathscr{P}_I = \{x \in \mathscr{D} \,|\, \hat{y} = 0\}$;
- the subset of predicted outliers $\mathscr{P}_O = \{x \in \mathscr{D} \,|\, \hat{y} = 1\}$.

2.2 Feature Importance for Isolation Forest

Despite many appealing properties, the Isolation Forest exhibits a major drawback: the lack of interpretability. As a matter of fact, the output at the evaluation stage is just a prediction, in the form of either an anomaly score or a predicted binary label, and no explanation about the logic behind the prediction is provided. While this is not necessary a problem in many applications, additional information about the process under examination may be the key for the development of actionable insights in industrial scenarios. In this work, we rely on the model-specific method proposed in (Carletti et al. 2019), called DIFFI, for the interpretation of Isolation Forest. The method is completely unsupervised and has been proven to have good performance, especially in presence of correlated features. The core of the DIFFI method is represented by an iterative procedure carried out separately over the subset of predicted inliers $\mathscr{P}_I$ and the subset of predicted outliers $\mathscr{P}_O$. Such a procedure is based on the evaluation of splits performed in each internal node (of each isolation tree) in terms of the capability of isolating data points and outputs a vector of importance scores, whose i-th component indicates the score associated with the i-th feature. As pointed out in (Carletti et al. 2019), in the context of AD a feature is deemed 'important' if it facilitates the isolation of anomalies and, at the same time, does not contribute to the isolation of normal data points. These are the reasons why the final feature importance scores are computed as the ratio between the scores obtained on the subset of predicted outliers and subset of predicted inliers. Indeed, for the generic feature f a high importance score on the subset of predicted inliers denotes a significant relevance of feature f in the task of isolating inliers, which is clearly the opposite objective with respect to the AD task. The DIFFI method also takes into account the randomness inherently present in the Isolation Forest model (i.e. random selection of splitting features and splitting thresholds); the adopted approach performs a normalization procedure to filter out the effect of randomness on the final feature importance scores.

2.3 Selection of the best predictions

As stated in (Carletti et al. 2019), the DIFFI method as outlined above has a major drawback: since we iterate over $\mathscr{P}_I$ and $\mathscr{P}_O$, the computational cost depends on the total number of data points n. The simplest solution to this problem would be to randomly select a fraction of predicted inliers and outliers, but we can easily define a smarter criterion which turns out to be beneficial also for the quality of the produced feature importance scores. The strategy, suggested in (Carletti et al. 2019), consists in selecting the fraction of predicted inliers with the lowest anomaly scores, and the fraction of predicted outliers with the highest anomaly scores. We will define the resulting subsets as the set of *selected inliers* $\mathscr{S}_I$ and the set of *selected outliers* $\mathscr{S}_O$, respectively. Obviously, it holds that $\mathscr{S}_I \subset \mathscr{P}_I$ and $\mathscr{S}_O \subset \mathscr{P}_O$. Notice that $\mathscr{S}_I$ and $\mathscr{S}_O$ can be specified according to two different and alternative criteria:

- by defining the fraction of predicted inliers α_I and predicted outliers α_O to be considered, or
- by defining a fixed number of predicted inliers m_I and predicted outliers m_O to be considered.

The DIFFI computational cost when adopting the first selection strategy still depends, although indirectly, on the total number of data points n (more precisely, it depends on the cardinality of $\mathscr{P}_I$ and $\mathscr{P}_O$, which in turn are influenced by the total number of data points n). The second selection strategy, instead, implies a constant computational cost since the cardinality of $\mathscr{S}_I$ and $\mathscr{S}_O$ in this case are fixed and independent of the cardinality of $\mathscr{P}_I$ and $\mathscr{P}_O$, respectively.

Notice that through the selection process we focus the attention on data points whose associated predictions are more confident. This reveals to a greater extent the inner logic of the Isolation Forest since less confident predictions may inject noise into the feature importance scores computed by the DIFFI method.

In Figure 2 a pictorial overview of the approach adopted in our experiments is displayed. Notice that DIFFI relies on the inner structure of the Isolation Forest (big blue arrow). The selection process discards less confident predictions (characterized by a lighter shade).

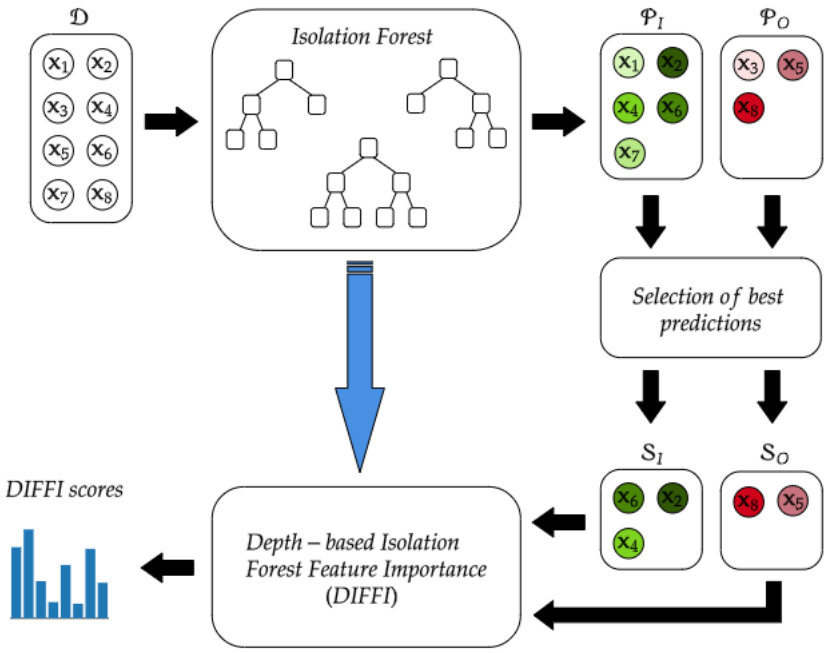


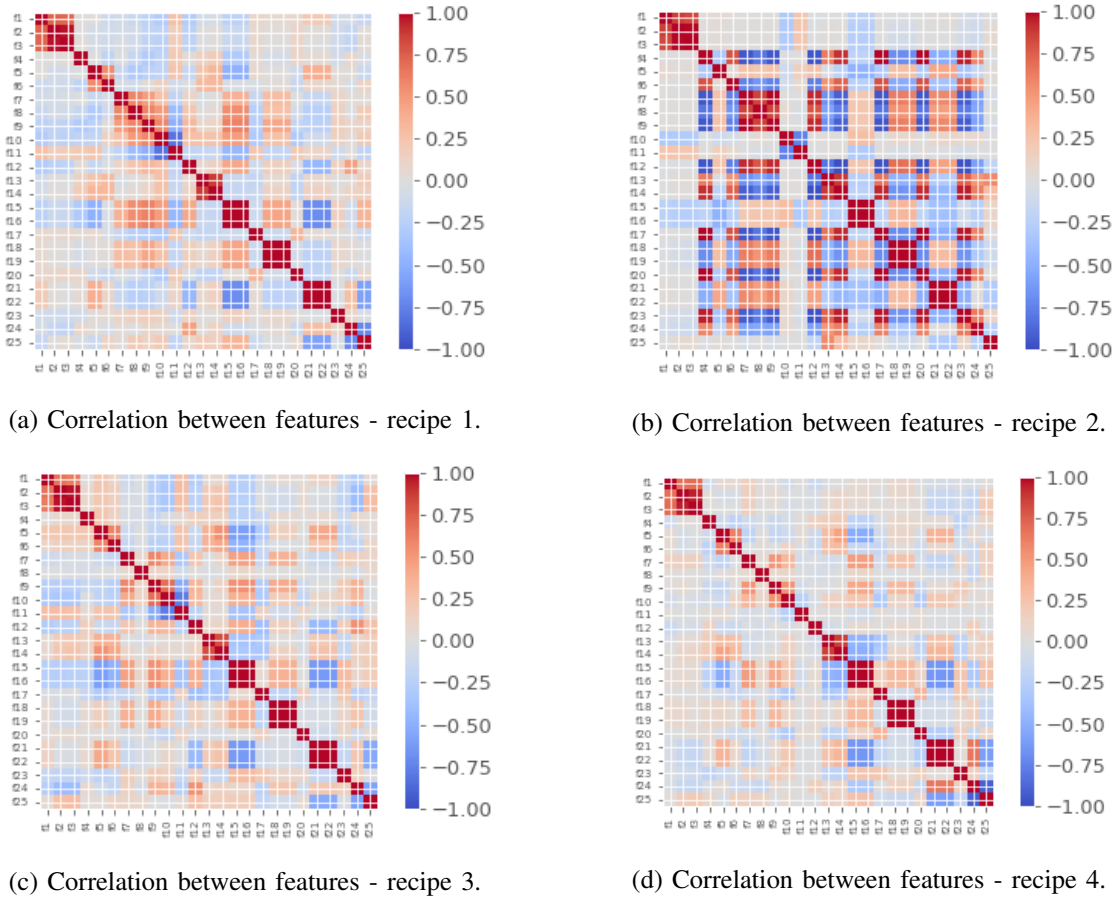
Figure 2: Overview of the approach adopted in our experiments. Circles represent data points, where green indicates predicted inliers and red predicted outliers. The darker the shade, the more confident the prediction.

3 EXPERIMENTAL RESULTS

As anticipated in Section 1, our experiments are focused on a CVD process monitoring problem. Specifically, we consider 4 datasets representing 4 different recipes, with different characteristic in terms of cardinality (see Table 1). Data points are represented as 25-dimensional feature vectors, where each feature is derived from sensor measurements installed on the equipment. It is worth highlighting that none of the considered datasets is equipped with ground truth binary labels indicating whether a specific data point is anomalous or not. In light of this, the exploitation of the collected data in the context of an AD problem might seems senseless since the detection performance cannot be assessed. Instead, we claim that the formulation of an AD task is reasonable and useful in many respects:

- in industrial scenarios labels are rarely available, thus the formulation of supervised problems is practically almost never possible;
- AD algorithms (such as the Isolation Forest) which associate an anomaly score to each data point are particularly useful in flagging potentially problematic situations, even in cases where the data point under examination does not clearly represent an anomaly;
- unsupervised algorithms offer a multivariate alternative to simpler univariate control charts for monitoring purposes.

In addition to the above considerations, in the specific case study considered in this work we can also rely on prior knowledge due to the collaboration with domain experts. More in detail, thanks to the physical interpretation that can be associated with most of the features, we have additional information about which are the most relevant features for the CVD process at hand. This can be considered as a (noisy) ground truth for the feature importance ranking and can be exploited to assess how well the feature importance ranking



(a) Correlation between features - recipe 1.

(b) Correlation between features - recipe 2.

(c) Correlation between features - recipe 3.

(d) Correlation between features - recipe 4.

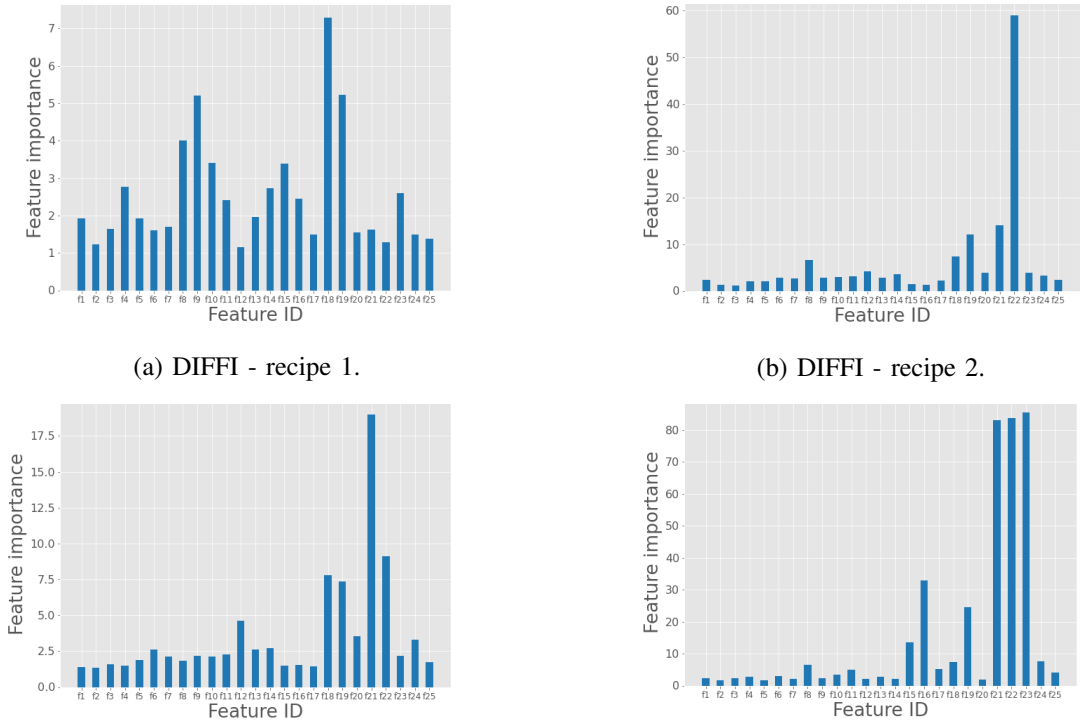
Figure 3: Correlation between features: recipe 1 dataset (a); recipe 2 dataset (b); recipe 3 dataset (c); recipe 4 dataset (d).

based on the DIFFI scores is aligned with domain knowledge. A good alignment between true and predicted feature importance could help in increasing the user's trust towards the predictions produced by the Isolation Forest. If we denote the first feature with f_1, the second one with f_2 and so on, the prior knowledge is represented by the claim that the most important features are $\{f_{15}, f_{16}, f_{17}, f_{18}, f_{19}, f_{20}, f_{21}, f_{22}, f_{23}\}$.

For our experiments, we train an instance of the Isolation Forest with the default hyperparameters setting suggested in the original paper, i.e. 100 isolation trees and sub-sampling size $\psi = 256$. For the DIFFI method, we define a fixed number of predicted inliers and predicted outliers $m = m_I = m_O = 50$, so that we have constant computational time, while ensuring satisfactory statistical significance of the obtained results.

For the sake of simple visualizations, we report experimental results referred to a single experiment, which is representative of a set of experiments performed with different random seeds, all leading to similar results.

We compare the DIFFI scores with the feature importance scores obtained with the Permutation Importance (PIMP) method, adapted to unsupervised settings by evaluating the deviation from original predictions rather than the drop in detection performance under permutations of feature values. This choice is motivated by the fact that PIMP method has been proven to achieve performance comparable with DIFFI when features are not highly correlated (Carletti et al. 2019) and this is actually the case for recipe 1, recipe 3 and recipe 4 datasets, as shown in Figure 3.



(a) DIFFI - recipe 1.

(b) DIFFI - recipe 2.

(c) DIFFI - recipe 3.

(d) DIFFI - recipe 4.

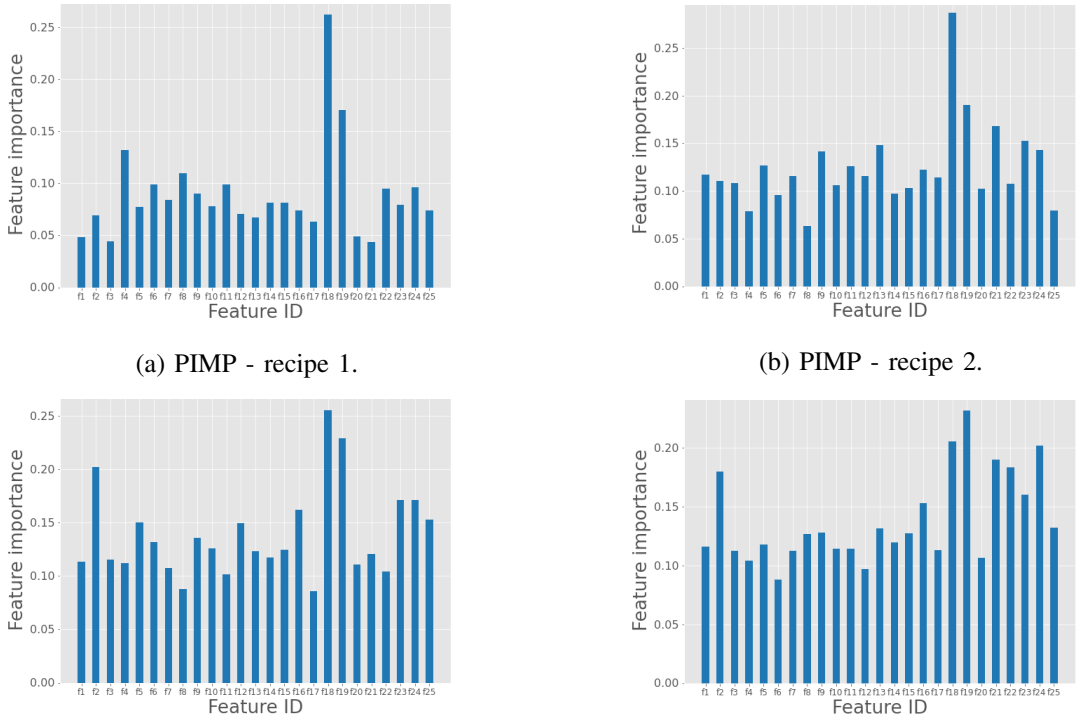
Figure 4: DIFFI scores: recipe 1 dataset (a); recipe 2 dataset (b); recipe 3 dataset (c); recipe 4 dataset (d).

In Figure 4 are reported the DIFFI scores for recipe 1 dataset (top left), recipe 2 dataset (top right), recipe 3 dataset (bottom left) and recipe 4 dataset (bottom right), respectively. It is evident that the DIFFI method successfully identifies relevant features for the AD task since the important features according to DIFFI align well with those indicated by domain experts.

Similar considerations hold for the feature importance scores obtained exploiting the PIMP method, as can be seen in Figure 5. Despite the fact that the relative importance between most important and least important features is not as pronounced as in the DIFFI scores (especially for recipe 4 dataset), it is fair to say that PIMP method exhibits satisfactory performance as well. In light of this consideration, a fundamental question arises: why do we need a new model-specific interpretability method for the Isolation Forest if well-established model-agnostic methods such as PIMP perform comparably? The answer can be found in Table 1, where the execution times of the two approaches are compared. Thanks to the fact that its computational cost does not depend on the cardinality of the dataset, the DIFFI method is arguably the best choice as the execution time is negligible if compared to the huge amount of time necessary for the computation of PIMP scores.

It is worth noticing that interpretability in general (being it achieved thanks to DIFFI or other methods) is highly valuable when dealing with semiconductor manufacturing processes as a tool to tackle many other interesting questions, such as the following:

- Should we train different ML models for different machines? Or should we better train a single ML model for all machines?
- Should we train different models for different ML chambers within the same machine? Or should we better train a single ML model for all chambers?
- Should we train different ML models for different recipes? Or should we better train a single ML model for all recipes?



(a) PIMP - recipe 1.

(b) PIMP - recipe 2.

(c) PIMP - recipe 3.

(d) PIMP - recipe 4.

Figure 5: PIMP scores: recipe 1 dataset (a); recipe 2 dataset (b); recipe 3 dataset (c); recipe 4 dataset (d).

For the case study presented in this paper, a quick visual inspection of the diagrams in Figure 4 and Figure 5 suggests that different recipes should be kept separated w.r.t. AD modules. Indeed, feature importance scores profiles are not consistent across the recipes. Put differently, features that are important for a specific recipe might not be relevant to others and for this reason a model trained on data related to several different recipes might struggle in achieving satisfactory detection performance.

Table 1: Execution times comparison: DIFFI vs PIMP.

Dataset	**Num samples**	**DIFFI exec time [s]**	**PIMP exec time [s]**
Recipe 1	956	0.37	9.12
Recipe 2	34716	0.41	272.38
Recipe 3	3580	0.38	18.75
Recipe 4	13459	0.38	55.73

4 CONCLUSIONS AND FUTURE WORKS

In this paper we have shown in the context of semiconductor manufacturing how unsupervised Anomaly Detection can achieve interpretability; more specifically, we employed feature importance approaches to unsupervised outlier detection for the first time in this application context. In the proposed comparison we focused on the Isolation Forest algorithm, one of the most popular approaches in AD, and we considered the model-specific approach, DIFFI, for feature importance: we have shown the effectiveness of such approach in a real CVD case study.

Future works in the area of interpretability for AD in semiconductor manufacturing will consider more complex data scenarios, like time-series data or Optical Emission Spectroscopy data, upon which effective AD approaches have already been implemented in the literature.

ACKNOWLEDGMENTS

Infineon Technologies AG is gratefully acknowledged for financing this research activity. Part of this work was supported by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016).

REFERENCES

Altmann, A., L. Toloşi, O. Sander, and T. Lengauer. 2010. "Permutation importance: a corrected feature importance measure". *Bioinformatics* 26(10):1340–1347.

Carletti, M., C. Masiero, A. Beghi, and G. A. Susto. 2019. "Explainable Machine Learning in Industry 4.0: Evaluating Feature Importance in Anomaly Detection to Enable Root Cause Analysis". In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 21–26. IEEE.

Chen, T. et al. 2018. *Anomaly detection in semiconductor manufacturing through time series forecasting using neural networks*. Ph. D. thesis, Massachusetts Institute of Technology.

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

Liao, D.-Y., C.-Y. Chen, W.-P. Tsai, H.-T. Chen, Y.-T. Wu, and S.-C. Chang. 2018. "Anomaly Detection for Semiconductor Tools Using Stacked Autoencoder Learning". In *2018 International Symposium on Semiconductor Manufacturing (ISSM)*, 1–4. IEEE.

Liu, F. T., K. M. Ting, and Z.-H. Zhou. 2008. "Isolation forest". In *2008 Eighth IEEE International Conference on Data Mining*, 413–422. IEEE.

Liu, F. T., K. M. Ting, and Z.-H. Zhou. 2012. "Isolation-based anomaly detection". *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6(1):1–39.

Maggipinto, M., A. Beghi, and G. A. Susto. 2019. "A Deep Learning-based Approach to Anomaly Detection with 2-Dimensional Data in Manufacturing". In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, Volume 1, 187–192. IEEE.

Nakazawa, T., and D. V. Kulkarni. 2019. "Anomaly Detection and Segmentation for Wafer Defect Patterns Using Deep Convolutional Encoder–Decoder Neural Network Architectures in Semiconductor Manufacturing". *IEEE Transactions on Semiconductor Manufacturing* 32(2):250–256.

Pampuri, S., A. Schirru, G. A. Susto, C. De Luca, A. Beghi, and G. De Nicolao. 2012. "Multistep virtual metrology approaches for semiconductor manufacturing processes". In *2012 IEEE International Conference on Automation Science and Engineering (CASE)*, 91–96. IEEE.

Puggini, L., and S. McLoone. 2016. "Feature selection for anomaly detection using optical emission spectroscopy". *IFAC-PapersOnLine* 49(5):132–137.

Puggini, L., and S. McLoone. 2018. "An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data". *Engineering Applications of Artificial Intelligence* 67:126–135.

Rostami, H., J. Blue, and C. Yugma. 2016. "Equipment anomaly detection and automatic fault fingerprint extraction in semiconductor manufacturing". In *Procedding of International Symposium on Semiconductor Manufacturing Intelligence*, 8–10.

Susto, G. A., A. Beghi, and C. De Luca. 2011. "A virtual metrology system for predicting cvd thickness with equipment variables and qualitative clustering". In *ETFA2011*, 1–4. IEEE.

Susto, G. A., A. Schirru, S. Pampuri, A. Beghi, and G. De Nicolao. 2018. "A hidden-gamma model-based filtering and prediction approach for monotonic health factors in manufacturing". *Control Engineering Practice* 74:84–94.

Susto, G. A., M. Terzi, and A. Beghi. 2017. "Anomaly detection approaches for semiconductor manufacturing". *Procedia Manufacturing* 11:2018–2024.

AUTHOR BIOGRAPHIES

MATTIA CARLETTI is currently a Ph.D. student in the Department of Information Engineering and Human Inspired Technology Research Centre at University of Padova. He earned his M.Sc. degree in Automation Engineering from the University of Padova. His research interests include Explainable Machine Learning and Anomaly Detection, with special focus on industry 4.0 and healthcare applications. His e-mail address is mattia.carletti@unipd.it.

NATALIE GENTNER is currently an industrial Ph.D. student with the University of Padova and Infineon Technologies AG (Neubiberg, Germany). She earned her M.Sc in Mathematics from the Technical University in Munich. Her research interests include machine and deep learning, automatization in semiconductor manufacturing, system theory and applied analysis. Her email address is natalie.gentner@studenti.unipd.it / natalie.gentner@infineon.com.

YAO YANG is Data Scientist at Infineon Technologies AG (Neubiberg, Germany). She earned her Ph.D. from the University of Mannheim with OR (operations research) related topics. She started her career as a scientist solving supply chain optimization problems in BASF SE (Ludwigshafen, Germany). After joining Infineon in 2017, she was involved in a large variety of semiconductor supply chain projects and is now focusing on data science topics, applying Big Data and Artificial Intelligence methods in manufacturing. Her email address is yao.yang@infineon.com.

ANDREAS KYEK is Data Scientist at Infineon Technologies AG (Neubiberg, Germany). He started his career in industry as a unit process development engineer at Infineon in 2001. From the beginning he was involved in Advanced Process Control and took over the development department for Advanced Process Control in 2007. During this time he also co-initiated the ENIAC funded project IMPROVE and was speaker of the executive board. In the years 2009 and 2010 he was member of the steering committee of the AEC/APC conference Europe. Later he joined a Manufacturing Excellence group, where he got involved in Factory Physics. Today he is heading projects where the usage of Big Data and Artificial Intelligence methods on manufacturing data is investigated. He holds a Ph.D. in Physics from the Technical University in Munich. His email address is andreas.kyek@infineon.com.

MARCO MAGGIPINTO received the M.S. degree (summa cum laude) in telecommunications engineering from the University of Padova, Padova, Italy, in 2016, where he is currently pursuing the Ph.D. degree. He has been an Erasmus Student with the Universitat Politécnica de Catalunya (UPC), Barcelona, Spain, and a Visiting Researcher with Queen's University Belfast (QUB), Belfast, U.K., Infineon Technology AG, Munich, Germany and University of Pennsylvania, Philadelphia, USA. His current research interests include deep learning, machine learning, and computer vision. His email address is marco.maggipinto@phd.unipd.it.

ALESSANDRO BEGHI received the Laurea degree (cum laude) in electrical engineering and the Ph.D. degree in control systems engineering from the University of Padova, Italy, in 1993 and 1989, respectively. In 1994, he joined the Department of Information Engineering, University of Padova, where he is currently Full Professor of control systems engineering. He held visiting positions with universities and international research centers both in Europe and in the USA. He has researched on a wide range of control applications, from control of fusion devices, guidance algorithms for virtual vehicles, to control of HVAC&R systems, adaptive optics systems, and semiconductor manufacturing processes. He has been responsible of research projects funded by the European Union, the Italian Ministry for Instruction and Research, the University of Padova, and of research activities with national and international industrial partners. He is a Co-Inventor of

six patents on the use of advanced control techniques in different applications fields. He has authored over 200 publications in journals, books, and conference proceedings. His research interests include modeling and control, filtering and identification, model reduction, fault detection, and isolation and their applications. His email address is beghi@dei.unipd.it. His website is http://automatica.dei.unipd.it/people/beghi.html.

GIAN ANTONIO SUSTO is currently an Assistant Professor with the University of Padova. He earned his Ph.D. from University of Padova and he held visiting positions at University of California at San Diego, Maynooth University and Infineon Technologies Austria AG. He is co-founder of Statwolf LtD. His current research interests include machine and deep learning, industry 4.0 and natural language processing. Prof. Susto received the IEEE-CASE Best Student Conference Paper Award in 2011, the IEEE/SEMI-ASMC Best Student Paper Award in 2012, and the IEEE-MSC Best Student Paper Award in 2012. He is an Associate Editor of the IEEE Transactions on Semiconductor Manufacturing for the area of Process Modeling. His email address is gianantonio.susto@unipd.it. His website is http://automatica.dei.unipd.it/susto.html.