3rd International Conference on Industry 4.0 and Smart Manufacturing

Predictive discarding of wafers based on power leakage predictions from single layer misalignment data

Geert H. van Kollenburg [a,*], Mike Holenderski [a], Patrizia Vasquez [b], Nirvana Meratnia [a]

[a]*Eindhoven University of Technology, Interconnected Resource-aware Intelligent Systems, Den Dolech 2, 5612AZ, Eindhoven, The Netherlands*
[b]*STMicroelectronics, Zona Industriale Stradale Primosole 50, 95121 Catania, Italy*

Abstract

Photolithography is a process used in the manufacturing of dies, which are at the core of complex integrated circuits. During this process several layers of semi-conducting material are stacked on top of each other. Precise alignment of the layers is crucial to the performance of a die. Upon completion, each die is subjected to several electrical tests. If many dies of a wafer fail the test, the whole wafer is considered faulty and has to be discarded, or reworked). This paper proposes the use of machine learning models to predict the outcome of a crucial test for MOS power leakage, from misalignments of a single layer. Wafers which are predicted to be faulty when finished can be predictively discarded, saving costs and resources otherwise spent on finishing the faulty wafer.

1. Introduction

Integrated circuits are an essential aspect of modern life. At the core of these circuits lie dies made from precisely stacked layers of semi-conducting material. Dies are manufactured hundreds at a time on a single wafer.

After the manufacturing stage, each die undergoes a number of electrical tests to ensure that it functions correctly. If many dies from a wafer fail these tests, the entire wafer is discarded because the dies that did pass the tests are not deemed reliable enough for future use [1]. One of these test is the evaluation power leakage of the dies. Power leakage is a critical performance measurement of dies [17].

The electrical tests are only available after completion of a wafer. As a result, completed wafers may have to be discarded resulting in wasted resources. During production, however, metrology systems monitor the alignment of each layer of the wafer [2]. Critical misalignments are directly dealt with by redoing parts of the production, or

---

* Corresponding author. Tel.: +31 40 247 2971
E-mail address: g.h.v.kollenburg@tue.nl

aborting production. Yet, combinations of non-critical misalignments of single layers may also lead to faulty wafers [9].

Because misalignments play a crucial role in the functioning of dies, measurements from unfinished wafers may be predictive of the outcome of the electrical tests. When the lithographic process continues on faulty wafers, resources are wasted. Early predictive discarding of a wafer may therefore improve sustainability of the lithography process, by reducing unnecessary resource consumption otherwise required to finish a product which will not reach the required end-product standards.

In this paper, we report the application of predictive discarding on wafers based on predictions for the outcome of the high-voltage MOS power leakage test. The predictions were made with readily-available classification models, using data from the most critical layer, according to expert knowledge. The results indicate that the outcome of the power leakage test be predicted in a very early stage of the lithography process. Predictively discarding the faulty wafers can save on costs, but the benefits depend on the revenue that would be missed due to falsely discarding good wafers. The studied models can be easily implemented for real-time decision-making and only require data already collected as part of standard quality control procedures. Predictive discarding is therefore a resource-friendly method to help reduce costs, raw material usage and waste.

2. Related work

Monitoring the status of an unfinished product is standard practice in virtually all manufacturing processes. Some mishaps that directly influence the quality of a product may be easily detected and appropriate action (like reworking or discarding the unfinished product) can be easily decided upon. anomaly detection methods can also identify mishaps not seen before [13]. The issue remains that also normal variations during the production process may influence the end-product quality. Models are therefore needed to evaluate the effects of each production step on the end-product quality.

In process industry, data-driven prediction models are a standard addition to first-principles physical models. Product quality is often continuously predicted from measurements of process variables. Relationships between production units and the effect of those units on product quality can be studied using Process PLS [10]. The challenge of using such regression-type models in wafer production settings is that variables in the wafer manufacturing process are measured only once when a layer is finished, instead of continuously as in process industry

One could regard an unfinished wafer as an intermediate product which is used to make a next product (i.e., a wafer with one more layer) and follow the methodology proposed by [19]. In their work, the authors used sensory production data to train a neural network to predict the end-product quality from such intermediate products. While the goal of the method is similar to what we aim for in the current paper, the discontinuous nature of the photolithography phase in wafer manufacturing means that this existing type of modelling is not applicable. Hence a different approach is needed to model the effects of each distinct layer on the end-product quality.

Previous work specifically on quality predictions of wafers took into account the discrete nature of the manufacturing steps [12]. However, that approach predicted the quality after each production step was finished (for example after the lithography step). Evaluating the effects of separate layers was not considered. The work presented in the current paper can therefore be seen as an addition to the work by [12] by providing predictions on more detailed level (i.e. using data from distinct layers during the lithography phase).

3. Data

For the current application, data of 480 wafers were used, where the alignment of the critical layer was measured at 36 reference points. Misalignments were measured, in micron, in the horizontal (x) and vertical (y) directions with respect to the metrology device above the wafer (See Figure 1 for an example), resulting in $36 * 2 = 72$ data points per wafer. The data were originally collected as part of the standard quality control procedure and all wafers belonged to the same product/type.

The misalignment measurements can be used to calculate offsets, like rotations or skewness with respect to the ideal position [6]. Offsets are generally used to calibrate the machinery [1], but may hold predictive value as they combine the information from the alignment measurements. Twelve offsets were available in the data.
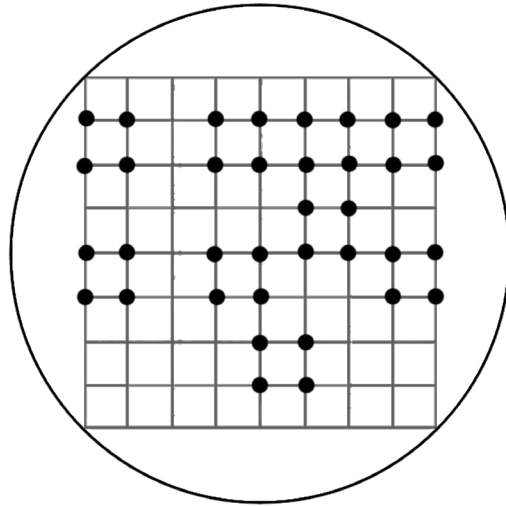
Fig. 1. Visual representation of alignment measurement locations. The circle represents a wafer at a particular layer, the virtual reference lines are represented by grey lines and the black dots represent the set of coordinates measured at this particular layer.

Variables which had missing values for more than half of the wafers were removed from the data. All variables were then standardized to have a mean of zero and unit variance. Any remaining missing values were imputed as zero, as this is the expected value of every misalignment.

The data contained information on how many dies on each wafer failed the power leakage test. This number serve as a measure of overall wafer quality (the lower the values the higher quality the wafer), and was calculated as the average of 25 (independent) tests, assumed to be replicate measurements of each wafer. In practice, if more than a certain number of dies fail the test, the entire wafer is discarded (or is reworked). The formal thresholds cannot be disclosed, so an artificial threshold was specified for this report. Wafers with higher than average number of faulty dies were considered 'faulty' (coded 1) and those with lower than average faults were considered 'within-spec' (coded 0).

4. Methods

4.1. Classification

Four readily available classification models were used, namely k-nearest neighbors (KNN) [5] Guassian Naive Bayes (NB) [4], random forest (RF) [3], and support vector classification (SVC) [15]. KNN groups similar observations and classifies new observation to the group most similar to it. NB classifies wafers based on which class (good/faulty) best fits the assumption of conditional independence between pairs of misalignment. RF fits multiple decision trees, each using a sub-sample of predictors, and a final classification is made based on the average result from all decision trees. SVC organizes observations in a (transformed) space in such a way that the distance between the groups is maximized.

For every classification model, data was split into a training set (67 %) and test set (33%) using stratified shuffle splits. Within each split, each classification model was trained twice. Once using the set of 72 alignments as predictors and once with the set of twelve offsets. Because the offsets are determined by raw misalignment measurements, the sets of features were not used in the prediction models at the same time.

The training data were used to perform a grid search for optimizing the model hyper-parameters. For RF, class weights were either equal for each class or balanced according to class sizes. The number of trees was either 100, 500,

or 1000. For KNN, the options for number of neighbors was either 3 or 5 and leaf sizes were either 1, or 5. The SVC used either a linear or a radial basis function ('rbf') kernel.

Using the hyper-parameters for each model which led to the highest true positive rate (i.e. highest precision) in the grid search [16], the models were refit to the training data, again optimizing the model precision.

The resulting model estimates were used to classify the previously held out test set. Wafers were classified as faulty if the model had a certain level of 'confidence' in its class predictions, quantified in the form of classification probabilities. Next to the default thresholds of .5, also a stricter value of .7 was used. A higher threshold may reduce the number of false positives, but at the same time lower detection rate of true positives.

The entire procedure described above was repeated 10 times for different random training/test splits. The analyses were done using the Python libraries Pandas [11], Numpy[7] and Sci-kit Learn (v0.24) [14].

4.2. Benefits of Predictive Discarding

In an optimal situation, a wafer of sufficient quality would never be discarded and all wafers of insufficient quality would be discarded as soon as possible. Compared to the current practice of finishing all wafers before discarding, the benefits of predictive discarding depend three things: the number of true positives (discarded faulty wafers), false positives (discarded good wafers) and the profit margin missed out on due to discarding good wafers.

Let $C = 1 = \sum_{i=1}^{L} C_i$ be the total costs of the lithography process with L layers. Each layer has a marginal cost, C_i. after a given layer l, the production costs can be split into sunk costs (already made up until that layer) and projected costs (to be made when continuing the production) [8]:

$$C = \sum_{i=1}^{l} C_i + \sum_{i=1+1}^{L} C_i \tag{1}$$

Sunk costs were disregarded in the following calculations, as they are made regardless of whether a wafer is discarded, faulty, or good. The projected costs to finish a wafer after layer l are therefore

$$C_{proj} = \sum_{i=1+1}^{L} C_i = 1 - \sum_{i=1}^{l} C_i = \tag{2}$$

The wafers used in this study consist of $L = 23$ layers, and the data used for predictions was related to the third layer, that is $l = 3$. According to expert knowledge, the differences in costs for producing each layer in the wafer are negligible, simplifying the marginal costs for producing a layer to $1/L$.

The projected costs for finishing a wafer after the critical layer equal to $1 - l/L = 1 - 3/23$, which means that we can save up to $1 - l/L = 20/23$ of the costs from the lithography process for each wafer we correctly discard predictively. On the other hand, if we reject a wafer which would not be faulty, we risk missing out on revenue for that wafer.

The savings when predictively discarding a wafer w based on predictions from a model m can be calculated as

$$\text{Savings}_m(w) = \begin{cases} 0, & \text{if continue with } w \\ 20/23 & \text{if correct discard of } w \\ -P, & \text{if wrong discard of } w \end{cases} \tag{3}$$

where $P = \text{Cost} - \text{Added Value}$ is the profit margin of the finished product. If $P = 0$, the added value of the lithography process is just enough to cover the production costs. The savings were calculated separately for the different profit margins $P \in [0, .05, .10, .25, .5, 1, 1.5]$

Summing up the savings for each discarded wafer w gives an indication of how much a particular model could have saved if it were applied to the wafers in the test set:

$$\text{Savings}_m = \sum_{w \in W} \text{Savings}_m(w) \tag{4}$$

5. Results

The test set classification results can be found in Figure 2. A wafer was classified as faulty if the model provides a classification probability of at least .5 that the wafer will turn out faulty.

With respect to the model precision (i.e. the true positive rate), NB (using misalignments data or the offsets as predictors) and RF (only when using misalignment values) perform best. On average, around 70-75% of the wafers classified as faulty were indeed faulty. However, NB only detects around 25% of all faulty wafers (i.e, recall of around .25). RF with misalignments on average has a recall of over .5, indicating that it detects over half of the faulty wafers. KNN had highest recall, but around half of those classfied as faulty were false positives. SVC performed very poorly as it did not provide any consistent results.
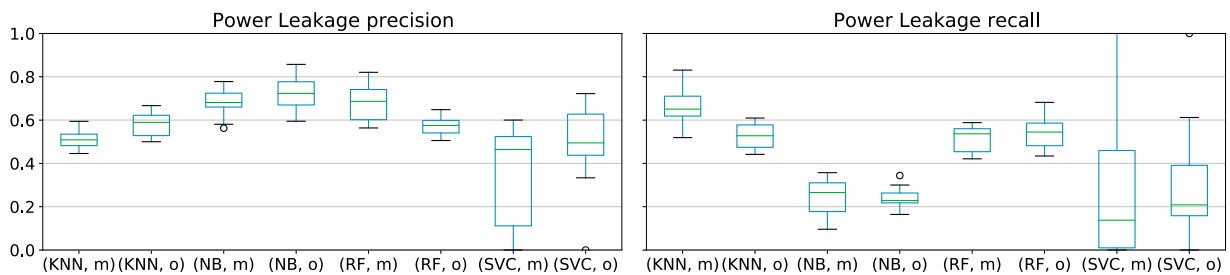

Fig. 2. Test set precision and recall on the PowerMOS test for the different models trained on misalignment measurements (m) or offsets (o). Confidence of 0.5 was used as threshold for classification.

To accurately evaluate the best Recall/precision trade-off, we need to consider the savings of the predictive discarding approach. Figure 3 compares the savings for all the models trained on data. The results show that the Random Forest trained on misalignment measurements can result in positive savings for a wide range of profit margins. The savings are high for low profit margins and decrease with increasing profit margins because of the missed out revenue for false positives.
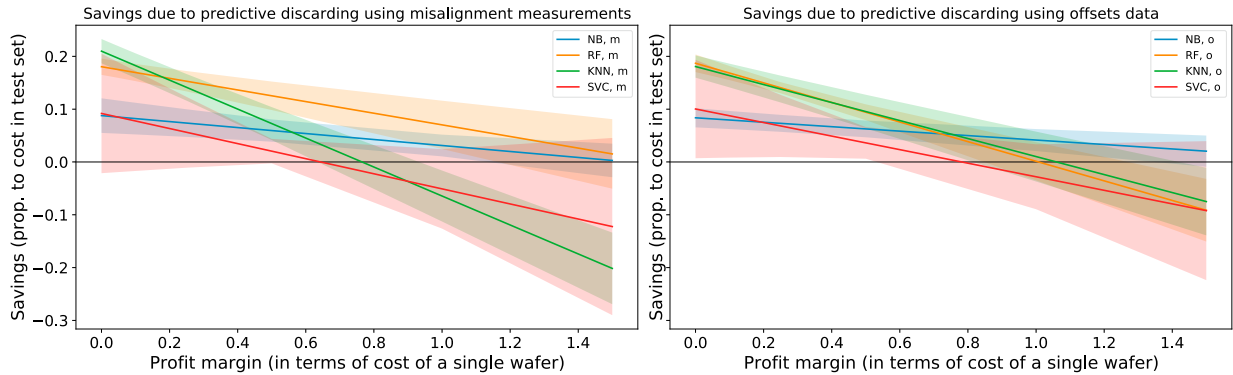
Fig. 3. Mean cost savings due to predictive discarding vs. profit margins, for different models trained on misalignment measurements (m) or offsets (o). Confidence of 0.5 was used as threshold. The shaded areas are the standard deviation over the 10 training/test splits.

To restrict the number of false positives, wafers were also classified using a threshold of .7 for the classification probability. The model performances for this threshold can be found in Figure 4. As expected, the the precision indeed increased for all the models (except for SVC, which apparently was almost never more than 70% confident that a wafer belonged to a certain class).
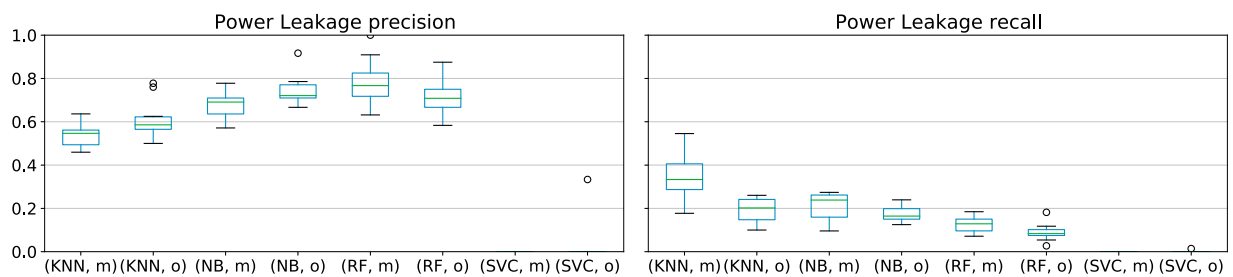

Fig. 4. Test set precision and recall on the PowerMOS test for the different models trained on misalignment measurements (m) or offsets (o). Confidence of 0.7 was used as threshold for classification.

Different classifications lead to different results for the savings. For the .7 threshold, the mean savings are shown in Figure 5. Again, as expected, with fewer false positives, the range of the profit margin in which predictive discarding still produces savings becomes wider, as less revenue is missed.
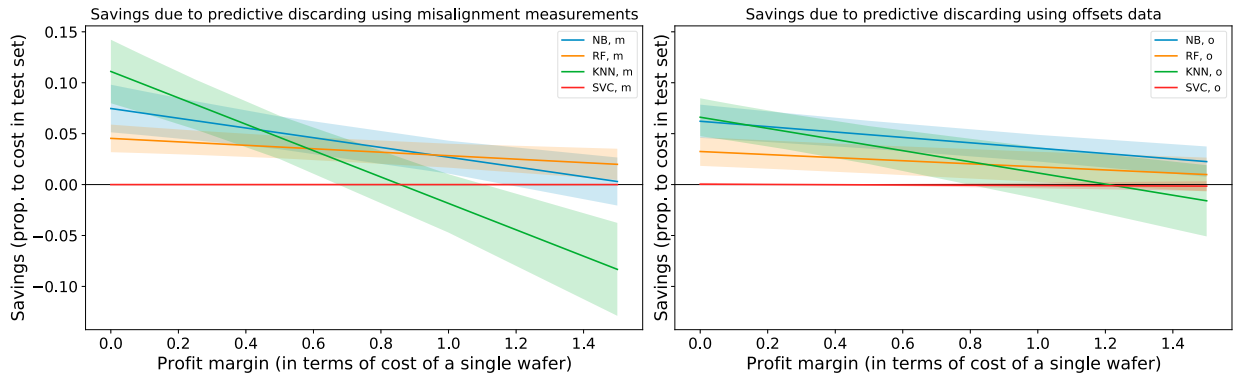
Fig. 5. Mean cost savings due to predictive discarding vs. profit margins, for different models trained on misalignment measurements (m) or offsets (o). Confidence of 0.7 was used as threshold. The shaded areas are the standard deviation over the 10 training/test splits.

6. Discussion

With the data available, it was possible to calculate the saved resources and missed profits/gains, because the actual outcomes (below/above average) were known. A more accurate picture of the benefits would be obtained by considering the probability that a wafer becomes faulty during production of the next layers. Future work may develop a more economically sound equation for the actual benefits.

Calculation of costs and profits was kept very simple. In practice there will be many more factors to consider. For now, we leave it to the reader to decide on what is a feasible added value of the lithography step and evaluate the benefits of predictive discarding in that light.

Models were optimized based on precision (i.e., low false positive rates). But as was shown in this paper, the benefits of predictive discarding depend on the profit margins. If the profit margin is known, other optimization schemes may be more beneficial for certain applications.

When interpreting the classification results presented in this paper, one should realise that the label 'faulty' was artificially created and all the wafers in the data were actually within specification. Wafers which are around the artificial cutoff may be very similar, which can explain why none of the models achieve perfect precision in the classification of 'faulty' wafers. In practice, when faulty wafers are clearly different from the good wafers, predictive discarding can definitely be an asset.

7. Conclusions

This work makes an important contribution towards prescriptive analytics in lithographic processes by showing how readily-available algorithms may be used to predict, at a very early stage of production, whether an end-product will be faulty. Discarding these wafers at such an early stage results in less waste and lower resource consumption.

It was shown that the end-product quality of certain wafer could be predicted from misalignment measurements of a single critical layer, using only within-spec misalignment data. The approach is therefore a promising basis for future work into end-product quality predictions and predictive discarding of faulty wafers in situations where faults may occur.

The models discussed above only focus on predicting single outcomes and used data from single layers. While such models already provide valuable information on effects of misalignments, future work may incorporate multiple layers into the predictive models.

In the current study only a single layer was used for predicting the end-product quality. During production, the data for different layers come in sequentially. Because each subsequent layer can add to the (prediction of) end-product quality, online predictive models that make latent representations of each layer could be a viable option for future research [18].

The benefits of predictive discarding extend much further into manufacturing industry than just wafer production. Especially in situations where production costs are high and there is a high risk of faulty end-products, predictive discarding may be a crucial asset for ensuring resource-optimized production.

Acknowledgements

References

[1] di Bella, R., Carrera, D., Rossi, B., Fragneto, P., Boracchi, G., 2019. Wafer defect map classification using sparse convolutional networks, in: International Conference on Image Analysis and Processing, Springer. pp. 125–136.

[2] den Boef, A.J., 2016. Optical wafer metrology sensors for process-robust cd and overlay control in semiconductor device manufacturing. Surface Topography: Metrology and Properties 4, 023001.

[3] Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

[4] Chan, T.F., Golub, G.H., LeVeque, R.J., 1982. Updating formulae and a pairwise algorithm for computing sample variances, in: COMPSTAT 1982 5th Symposium held at Toulouse 1982, Springer. pp. 30–41.

[5] Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory 13, 21–27.

[6] Flack, W.W., Avlakeotes, S., Chen, D., Flores, G.E., 1996. Efficient overlay optimization of stepper correctables, in: Metrology, Inspection, and Process Control for Microlithography X, International Society for Optics and Photonics. pp. 400–413.

[7] Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. doi:https://doi.org/10.1038/s41586-020-2649-2.

[8] Hussain, T., 2010. Engineering economics. Laxmi Publications, Ltd., New Delhi.

[9] Kandel, D., Levinski, V., Sapiens, N., Cohen, G., Amit, E., Klein, D., Vakshtein, I., 2012. Overlay accuracy fundamentals, in: Metrology, Inspection, and Process Control for Microlithography XXVI, International Society for Optics and Photonics. p. 832417.

[10] van Kollenburg, G., Bouman, R., Offermans, T., Gerretzen, J., Buydens, L., van Manen, H.J., Jansen, J., 2021. Process pls: Incorporating substantive knowledge into the predictive modelling of multiblock, multistep, multidimensional and multicollinear process data. Computers & Chemical Engineering , 107466doi:https://doi.org/10.1016/j.compchemeng.2021.107466.

[11] Wes McKinney, 2010. Data Structures for Statistical Computing in Python, in: Stéfan van der Walt, Jarrod Millman (Eds.), Proceedings of the 9th Python in Science Conference, pp. 56 – 61. doi:https://doi.org/10.25080/Majora-92bf1922-00a.

[12] Melhem, M., Ananou, B., Djeziri, M., Ouladsine, M., Pinaton, J., 2015. Prediction of the wafer quality with respect to the production equipments data. IFAC-PapersOnLine 48, 78–84. doi:https://doi.org/10.1016/j.ifacol.2015.09.508. 9th IFAC Symposium on Fault Detection, Supervision andSafety for Technical Processes SAFEPROCESS 2015.

[13] Papananias, M., McLeay, T.E., Obajemu, O., Mahfouf, M., Kadirkamanathan, V., 2020. Inspection by exception: A new machine learning-based approach for multistage manufacturing. Applied Soft Computing 97, 106787. doi:https://doi.org/10.1016/j.asoc.2020.106787.

[14] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[15] Platt, J., et al., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers 10, 61–74.

[16] Powers, D.M., 2020. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061 .

[17] Sze, S.M., 2008. Semiconductor devices: physics and technology. John wiley & sons.

[18] Vitale, R., Zhyrova, A., Fortuna, J.F., De Noord, O.E., Ferrer, A., Martens, H., 2017. On-the-fly processing of continuous high-dimensional data streams. Chemometrics and Intelligent Laboratory Systems 161, 118–129.

[19] Wang, G., Ledwoch, A., Hasani, R.M., Grosu, R., Brintrup, A., 2019. A generative neural network model for the quality prediction of work in progress products. Applied Soft Computing 85, 105683. doi:https://doi.org/10.1016/j.asoc.2019.105683.