Advances in Reliability, Safety and Security, Part 9 – Kołowrocki, Dąbrowska (eds.)
 ISBN 978-83-68136-21-0 (printed), ISBN 978-83-68136-08-1 (electronic)

Advances in Reliability, Safety and Security

ESREL 2024
Monograph Book Series

Causal Failure Root Cause Detection In Sensor Production Image And Machine Data

Mohammad Qalaji[a], Jawad Tayyub[b], Ivo Häring[c], Thomas Brox[d]

[a]*Work done at Endress+Hauser, Maulburg, Germany, Master Student Embedded Systems, Technical Faculty, University of Freiburg*
[b]*Endress+Hauser, Maulburg, Germany*
[c]*Fraunhofer EMI, Freiburg, Germany*
[d]*Department of Computer Science, University of Freiburg, Germany*

Abstract

This work explores the application of causality in production data, with a specific focus on a dataset that consists of both images and machine parameters. Traditional data analysis methods often rely on correlation-based approaches, that cannot effectively distinguish between causation and mere association. This research aims at causal relationships within the variables, to provide a more profound understanding of the underlying mechanisms that drive production efficiency and quality. The primary objective is to identify production failure root causes through causal discovery in terms of causal graph identification and inference methodologies in terms of causal graph quantification. For the present work, an evaluation scheme of the approach based on failure prediction with anomaly detection is presented that uses only the identified main failure root cause production parameters and measurements extracted from images. In addition, the causal approach is compared to the most influencing features of state-of-the-art feature selection methods. The work shows that when comparing the algorithms NOTEARS, DAS, LinGAM and SCORE for causal graph identification, as well as graph structure preserving algorithms using multiple noise propagation or outlier effect generation of noise at single node for causal quantification, that the NOTEARS algorithnm achieves the highest balanced accuracy compared to state of the art feature selection methods such as SHAP, mutual information regression, ANOVA and RFE when using XGBoost for anomaly detection as overall evaluation metric of the failure root cause detection algorithm pipeline. The additional insights generated within the causal approach comprise the effect of the number of selected features on the outcome, identification of indirect dependencies using causal graphs and radar plots for visualization of the causal influence of the most important subsets of the heterogeneous input data. It is indicated how the approach can be further improved and be applied for iteratively improving electronic hardware production lines.

Keywords: sensor or hardware production data, failure root cause identification, heterogeneous data, feature extractoin from images, causal graph identification, causal graph quantification, set of potential root caues, set of main features, validation through anomaly prediction, main feature extraction, metrics for prediction assessment

1. Introduction

Sensor hardware and chip production by now is digitalized in the sense that big heterogenous data are collected and stored on the level of single production items, i.e. it can be traced on which machine and with which machine parameters production steps were conducted as well as related production measurements and quality assessment test measurements or images. Many of this big data is already used to adjust the process and to decide after each of the up to ca. 100 production steps if quality of item allows for further production, lets expect item with somewhat reduced quality or even is production loss. Similar assessments are done at final quality assessment step. If production loss is high and/or not understood engineers use stored data to identify and mitigate failure root causes, which involves the inspection and statistical assessment of stored data, typically by even manually assessing hypotheses or even only educated guesses by production engineers. The latter process also takes place in case of major numbers of failure events of components in the field.

As of now and due the long and successful production history of established production lines most standard and not too seldom failure root causes can be expected to be known but to be accepted because of trade-off of

cost considerations regarding improvement costs and of costs of current production. However, options for improvement include: (i) to faster identify new types of failure root causes, e.g. due to change of materials, processes or components, etc.; (ii) to better identify more seldom events such as double failures, tribble failures; (iii) to better identify transient failures, e.g. temporary temperature, light condition changes, change of machine operating policies or even human inspectors as well as ramp-up or ramp-down effects; (iv) to increase explainability of already known failure root causes as well as to quantify their influence to determine additional root causes that might be hidden behind main effects; (v) and finally to support existing human manual root cause identification processes.

The present approach follows the conceptual process of Pearl and Mackenzie that causal discovery can be structured in nine steps and that the degree or quality of causal discovery can be ranked using three rungs of a causal ladder (see (Pearl and Mackenzie 2020), Figure I, Figure 1.2). Key properties of the former are that hypothesis testing should be based on causal graphs representing causal relationships, i.e. either observation (or measurement, quantity, observable, etc.) depends on another one or not, and vice versa or both as expressed in directed graphs as well as causal graph quantification (causal inference) expressing the actual level of dependencies, including second order dependencies, etc. The causal ladder concept distinguishes between associative, interventional and counterfactual causal relations, which are linked to increasing levels of insights (or rungs) from correlation-like causal relations, to causal reasoning regarding system behavior in case of defined changes that can be observed, and finally to counterfactual changes that cannot be observed. For instance, for a gluing process it can be observed that production failure is strongly correlated to lower temperatures of glue (rung 1), intervention analysis could determine effect of temperature (rung 2), and counterfactual analysis could assume that temperature variations are completely controlled in production hall environments (rung 3) to determine any other effects.

Regarding mathematical representation, hence the aim of the paper is to find approaches to construct credible causal graphs (directed graph) that point to production item failure node from heterogenous production data nodes as well as to quantify the causal graphs in terms of dependency strengths or weights of graph arrows, e.g. as determined by a corresponding Bayesian network quantifying the causal graph. Note that the latter is only one option.

For validation of the approach, the assumption is made that the causal graph topology, i.e. if a measurement node is connected to the item failure node or not, as well as the relative weights of the connection (in case of direct connection) or connections (in case of indirect connections) can be used to select production measurements that are relevant for prediction of production failure. The corresponding set of measurements is identified as set of likely root causes. Goodness of prediction of failure of item from selected root cause production data point set is then used as validation of the overall approach using selected prediction evaluation metrics. The assumption is that a valid root cause detection approach selects all relevant main root cause measurements which then also suffice for anomaly prediction algorithm for good predictions.

The remaining paper is structured as follows. Section 2 names recent approaches to determine failure root causes in chip or sensor production data. Section 3 gives an overview of the present method chain using causal approaches including the validation idea in terms of failure prediction with reduced data set. Section 4 gives sample results for the implemented approaches and discusses them, also comparing with state-of-the-art feature identification approaches. Section 5 summarizes and concludes.

2. State of the art, gaps and own approach focus

Failure root cause detection has been applied to chip production data using statistical assessment of at sub-batch level (Chien and Chuang 2014). The idea of clustering approaches for the identification of failure root causes is to cluster production data such that the label production step failed and not failed as well reveals additional distinguishing features that are only present in the clustering results in the case of failure, see e.g. (Liu and Chien 2013). Also, support vector machine (SVM) (He et al. 2016) or random forest approach (Kim et al. 2023) are used to classify production data measurements.

Multi-time series analysis uses as input the heterogeneous production data translated into time series and predicts the time history of production step quality with the expectation that changes in e.g. machine parameters are signatures of subsequent production failure (Wang et al. 2023b), for instance using convolutional neural network (CNN) approaches (Jain et al. 2020). Somewhat similar to multi-time series analysis anomaly identification approaches can be used to predict production failure steps, i.e. the expectation is that an anomaly is predicted for each production failure step. In this case the set of input data can be reduced to narrow the possible root cause data to a subset containing potential root causes.

Variations of artificial neural networks (ANNs) used in published literature include autoencoder approach (Cha et al. 2022), long-term short-term memory (LSTM) (Kao and Chien 2023), CNN (Nagamura et al. 2021), and deep neural networks (Zhang et al. 2022). In these cases, the ANN variation predicts the link between input parameters and production step success and failure. The challenge is to extract knowledge regarding causal relations from the constructed networks, e.g. in terms of strengths of weights between nodes or by using explainable intermediate layers.

Most of the approaches use a restricted set of data types, e.g. only machine parameters, time series data, test measurement data, or only images. In all cases data spotting, preprocessing, and preparation, e.g., normalization is conducted. Within this step for instance also principal component analysis (PCA) (He et al. 2016) (Ong et al. 2015) or feature extraction (Lundberg and Lee 2017) (Beraha et al. 2019) (Shakeela et al. 2020) (Ogunleye and Wang 2018) are conducted to reduce the dimensionality of the input data.

Closest to the present approach is (Wang et al. 2023a), which is however considering image data only and only NOTEARS algorithm for causal graph construction as detailed in section 3. Causal root cause detection has to the authors' knowledge not yet been applied to chip or sensor production data, even if the potential has been identified and first conceptual solutions have been developed (Fujiwara et al. 2023).

3. Methodology and implementation

The present method chain uses as input heterogeneous data including machine parameters, measurement data, time-series data, and selected extracted image features, see Figure 1. It follows the idea to first construct a causal graph and then to quantify it (causal inference). Based on the main contributions of data features considering several graph layers, the main contributing features are extracted (set of key features) that are expected to be sufficient within an anomaly approach to predict failure or success of the production step. The anomaly prediction with reduced data set is compared with the anomaly prediction with the full data set as well as current critical feature extraction algorithms. Note that all of them generate a subset of data features that are believed to be relevant for predicting failure or success of the production step.

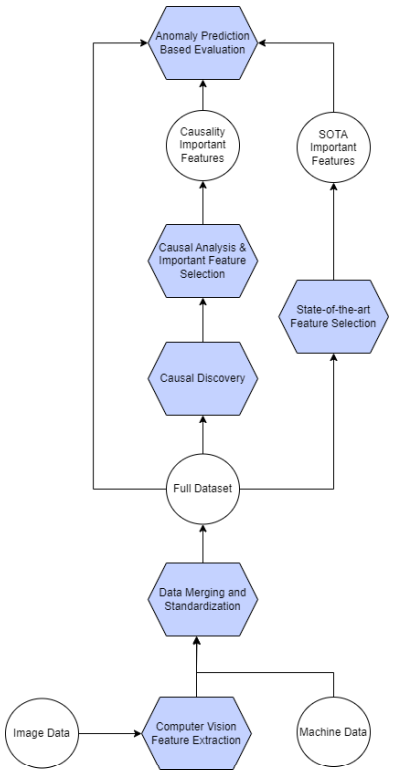


Fig. 1. Causal method chain overview for production failure root cause feature identification and evaluation approaches.

For each step, up to serval distinct algorithms are used for comparison as listed in Table 1, which gives the acronym of the approach, short description, references used as well as code sources employed.

Table 1. Methods used for causal production root cause failure detection and validation.

Step	Acronym, Short description	References, Code source used
Hand-crafted feature extraction from images	E.g. Roundness computation; Open Computer Vision (CV) Python library OpenCV;	e.g. (Sui and Zhang 2012); e.g. (Howse and Minichino 2020)
Data pre-processing and normalization	The dataset is standardized by subtracting the mean value of each feature and scaled by dividing by its standard deviation using the Python library Scikit-learn.	(Pedregosa and et al. 2011)
Causal graph construction	Discovery at Scale, DAS, is an extension to SCORE; Dodiscover Python library;	(Montagna et al. 2023) (Li et al. 2023)
	Linear Non-Gaussian Acyclic Model, LiNGAM; Independent component analysis (ICA) based LiNGAM algorithn in Causal-learn python library; iterations are limited;	(Shimizu et al. 2006) (Zheng et al. 2023)
	SCORE, Score matching of non-linear additive noise models; Dodiscover Python library	(Rolland et al. 2022) (Li et al. 2023)
	NOTEARES, Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning	(Zheng et al. 2018b); (Zheng et al. 2018a)
Causal graph quantification (Causal inference)	Intrinsic causal cause quantification. A Gaussian noise term is introduced to each node and its contribution to the output node is quantified.	(Janzing et al. 2020);
	Causal structure-based root cause analysis of outliers assessing outlier generation effect when adding noise to single node.	(Budhathoki et al. 2022);
Root cause feature set selection using feature extraction methods	SHAP (SHapley Additive exPlanations) assigns each feature an importance value for a particular prediction, SHARPLY.	(Lundberg and Lee 2017);
	Mutual information regression.	(Beraha et al. 2019)
	Analysis of Variance (ANOVA) F statistic.	e.g. (Shakeela et al. 2020);
	Recursive Feature Elimination (RFE) or logistic regression; implemented in Skylearn (Scikit learn).	e.g. (Ogunleye and Wang 2018); (Hao and Ho 2019)
Root cause feature set selection using quantified causal graph or feature extraction methods	Selection of top 5 contributors for each method chain; Visualization using radar plots; Radar plot implementation and metrics over data feature plots with Python packages.	(Draper et al. 2009); (Stancin and Jovic 2019)
Anomaly prediction using selected candidate feature set	ONEClass support vector machine (SVM) for time series prediction based on feature set only; One Class SVM; implemented using Phython Scikit-learn.	(Amer et al. 2013); (Hao and Ho 2019)
	XGBoost (eXtreme Gradient Boosting) is a tree boosting system that uses a sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning.	(Chen and Guestrin 2016); (Pedregosa and et al. 2011)
Metrics for evaluation of prediction based on selected key data features	Example accuracy metrics are: balanced-accuracy, Area Under the Receiver Operating Characteristic Curve (ROC-AUC), F1 score; implemented in Python Scikit-learn.	(Pedregosa and et al. 2011); (Hao and Ho 2019)

4. Results and discussion

This section sows example results for each step of the root cause identification pipeline described in Section 3. Section 4.1 discusses the extracted image feature results and interprets differences between images of normal and failed production steps. In section 4.2, causal graphs are given based on the 4 different algorithms listed in Table 1 and compared regarding their structural credibility. Section 4.3 conducts causal graph quantification using 2 different approaches. Section 4.4 lists the failure anomaly prediction metrics using full feature data set, causal inference-based feature subsets for 4 different algorithms and importance feature extraction-based set of parameters using again 4 algorithms.

Table 2. Example measurement comparisons between bad and good images.

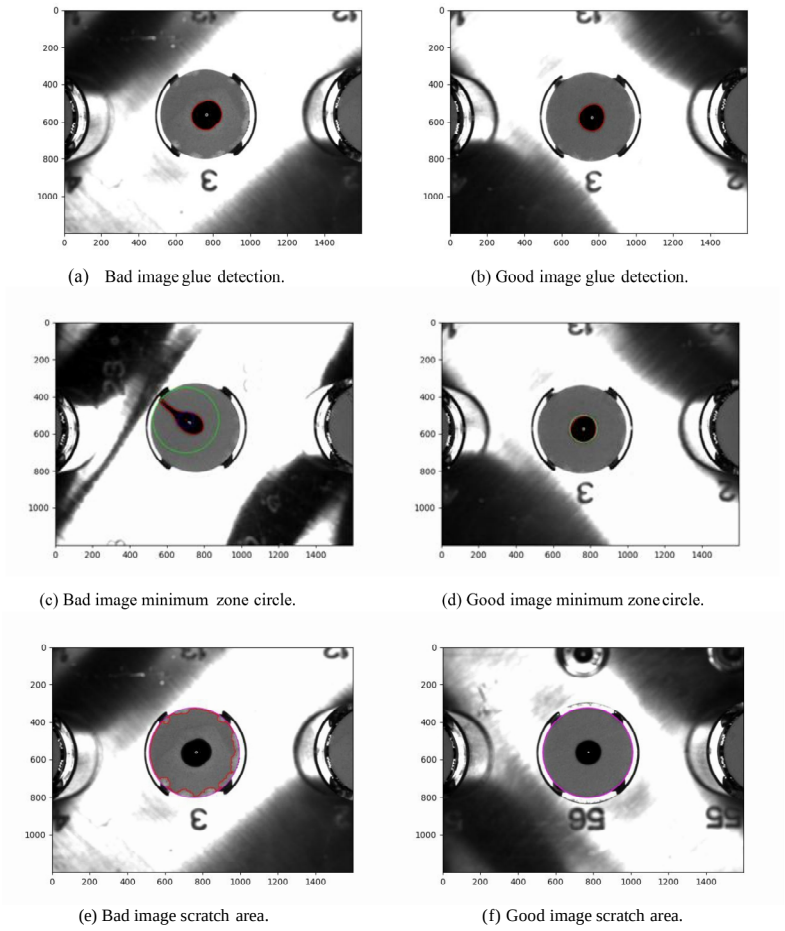
Type	G-MGV	G-Circ	G-Area	G-MZC	Offset	C-MGV	S-Area
Explana-tion	Glue mean grey value	Glue Circularity	Glue area	Glue minimum zone circle	Offset between glue and component center	Component mean grey value	Scratch area
Bad	5.51	0.88	18808.5	135.9	10	72.24	22175
Good	5.33	0.88	15025.5	10.63	2.33	86.22	2000

4.1 Extracted image features

Figure 2 illustrates contours detected for the image dataset to extract useful features (see Table 2) that will be beneficial in the later stages. Figure 2.a, 2.c, 2.e are images classified as failure while Figure 2.b, 2.d, 2.f are good images. Images 2.a, 2.b represent the contour for detecting the deposited glue and the measurements we

learn from these images are glue area, glue circularity and glue mean grey value. Figure 2.c, 2.d represent two measurements: the minimum zone circle and the offset between the center of the glue contour and the Component center. Figure 2.e, 2.f represent the area of the scratch for the component drive which subtracts the total area with the area detected and measures the component mean grey value. See further examples in (Qalajia 2024).

Table 3 illustrates the value difference between good and bad images. Figure 2.a, 2.b have similar glue mean grey value and circularity value. However, the glue area is much larger for the bad image. Figure 2.c, 2.d have big difference in the mean grey value and the offset between good and bad images. Images 2.e, 2.f have a small difference in the component mean grey value, while the scratched area of the component is much greater in the bad image. Please note that in many cases the measurements between the good and bad images are very similar and cannot be separated.



(a) Bad image glue detection. (b) Good image glue detection.

(c) Bad image minimum zone circle. (d) Good image minimum zone circle.

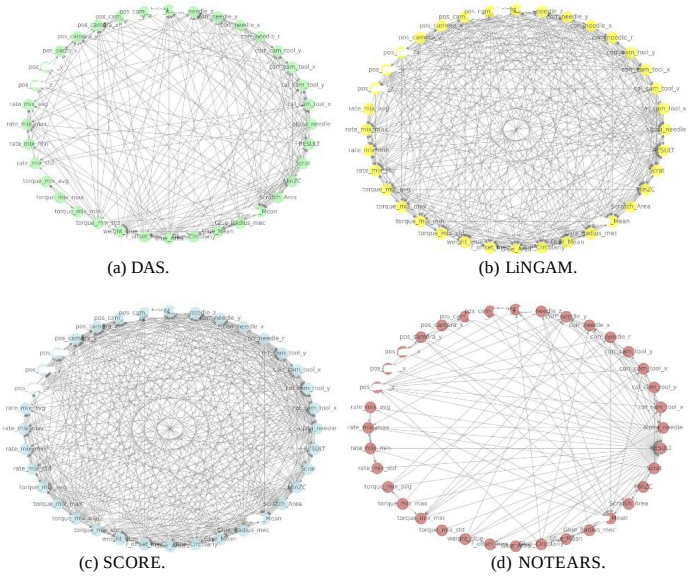
(e) Bad image scratch area. (f) Good image scratch area.

Fig. 2. Extracted features from images.

4.2 Causal graph construction using extracted features and machine parameters

This section considers production data consisting of machine measurements and image extracted features as explained in section 4.1. Each data point of the production step contains 26 machine parameters and 9 image-based measurements, in total 35, which are explained in detail in (Qalajia, 2024)]. We compare the structure of the causal graphs and the subset of the most important features. Our aim is to determine which of the extracted features from the images and the machine parameters influence the outcome of the gluing process. In total 2800 normal data points and almost 400 failed data-points are used to construct the graphs and in the feature

quantification. The result of the causal discovery for the 4 approaches of Table 1 can be seen in Figure 3.



(a) DAS. (b) LiNGAM.

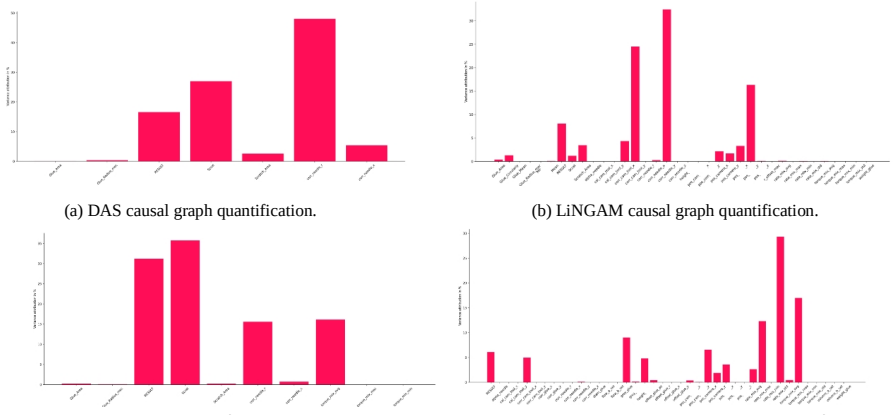
(c) SCORE. (d) NOTEARS.

Fig. 3. Causal graphs based on machine- and image-extracted feature measurements.

Figure 3 illustrates the causal graphs for joint assessment of machine measurements and image extracted features. Figure 3.a represents the output using DAS algorithm and we can see that it has the least connections to the outcome RESULT node and the second least total connections. Figure 3.b represents the LiNGAM algorithm and has a lot of interconnections and connections directly to the outcome node. Figure 3.c represents the SCORE algorithm and it has also multiple interconnections and connections directly to the outcome node. Figure 3.d represents NoTEARS algorithm and it has the least interconnections but multiple connections to the outcome node.

4.3 Causal graph quantification

Next, we estimate the contribution of each variable to the outcome. The results can be seen in the bar char plots of Figure 4. The quantification of the percentage of contribution of each node considers direct and indirect contributions using the algorithms listed in Table 1.



(a) DAS causal graph quantification. (b) LiNGAM causal graph quantification.

(c) SCORE causal graph quantification. (d) NOTEARS causal graph quantification.

Fig. 4. Key features from machine and image data from causal graph quantification.

Figure 4 represents the contribution of each node to the variance of the outcome node, which sum up to 100%. Figure 4.a represents the quantification of DAS algorithm and it can be seen that only two machine features and three image extracted feature have an influence on the outcome node. In addition, the result node has small influence due to missing confounders, i.e. a node that is connected to the result node as well as a node that is also connected to the result node. Note that contributions are also plotted if they are not direct contributions e.g. parents of direct contributions, and their parents. Figure 4.b represents the quantification of the LiNGAM algorithm and we can observe that more variables are influencing the outcome. The machine parameters Corr(elation)-needle-y, corr-cam(era)-tool-x and pos(ition)-component-y have the strongest influence on the outcome while some of the image extracted features have some influence but much lower. Figure 4.c represents the quantification of the SCORE algorithm where Scrat(ch) feature has the highest influence on the outcome and we can also observe that torque-mix(ed)-avg(average) and corr-needle-r have also some influence. Note that $r = \sqrt{x^2 + y^2}$. Moreover, the outcome node has the highest influence compared to the other algorithms which accounts to missing confounders. Figure 4.d represents the quantification of NOTEARS algorithm and we can see that corr-cam-tool-x variable has almost a 70% influence on the outcome while the other variables have lower influence rate.

4.4 Production failure prediction using anomaly detection with main root cause candidate set

Now the machine parameters and image parameters with main contributions are selected to do anomaly prediction of production step failure. For each algorithm the 5 most contributing factors are considered:

- The subset using DAS is {corr-needle-r, corr-needle-x, Scrat, Scratch-Area, Glue-Radius-mec};
- for SCORE {corr- needle-r, corr-needle-x, Scratch-Area, Scrat, torque-mix(ed)-avg};
- for LiNGAM {corr-needle-y, corr-cam-tool-y, pos-component-y, corr-cam-tool-x, Scratch-area};
- and for NOTEARS {corr-cam-tool-y, torque-mix-max, corr-needle-r, Scrat, cal- cam-tool-y}.

Observe that the subset for each algorithm is different and there is not one variable that is in all subsets.

Now the aim is to evaluate the subsets using anomaly detection. In Table 3 several accuracy metrics are calculated for each of the 4 causal algorithms and compared to the full data subset and to 4 state-of-the-art feature selection algorithms that are listed in more detail already in Table 1.

Table 3. Metrics for production step failure anomaly prediction with ONEClass SVM using data feature subset. The metrics are used to evaluate the quality of the root cause data subset identification.

Method for causal discovery	Balanced accuracy	F1 score	Precision recall curve
DAS	0.67	0.91	0.95
SCORE	0.73	0.93	0.95
LiNGAM	0.67	0.91	0.95
NOTEARS	0.74	0.93	0.96
Full dataset	0.72	0.93	0.96
Shapely	0.72	0.93	0.96
Mutual info regression	0.85	0.96	0.98
ANOVA-F statistic	0.71	0.92	0.95
RFE-logistic regression	0.62	0.90	0.94

Table 3 represents the evaluation for the most influential features using ONEClass support vector machine (SVM) for anomaly prediction, see method details in Table 1. For each subset we compute the balanced accuracy, f1 and precision recall curve scores. This is a first example of anomaly detection with most influential measurements as identified with feature extraction methods, see section 4.1. The 2nd to 5th row represents the scores for the causality feature selected as illustrated in blue. We can observe that NOTEARS has the highest balanced accuracy and SCORE is one percent lower. The F1 and precision recall curve are similar for all algorithms. The balanced score for the full dataset is 0.72 which is lower than NOTEARS and SCORE but higher than DAS and LiNGAM. Moreover, the F1 and precision recall curve scores for the full dataset are the same for NOTEARS.

The causality-based feature selection algorithms can compete with start of the art feature selection algorithms and in most cases have higher scoring metrics. The NOTEARS and SCORE algorithms have higher balanced accuracy than Shapley, ANOVA-F and RFE-logisitc regression, while maintaining similar F1 and precision recall. Mutual Info classification has highest balanced accuracy, and higher f1 and precision recall values by a margin.

Table 5 represents the evaluation of the subsets using XGBoost for anomaly detection. We can observe that DAS algorithm has the highest balanced accuracy compared to other causality-based algorithms. The F1 and precision recall curve score is almost similar for all the causal graph based algorithms. NOTEARS has the highest balanced accuracy for OneClassSVM, however, surprisingly has the lowest for XGBoost. Causality based feature selection achieves similar results in F1 and precision recall when compared to full dataset and state of the art feature extraction algorithms. DAS algorithm has the highest balanced accuracy score compared to all algorithms and other causality-based algorithm differ with only with a small margin to state-of-the-art methods.

Table 5. Metrics for production step failure anomaly prediction with XGBoost using data feature subset. Again, the metrics are used to evaluate the quality of the root cause data subset identification.

Method for causal discovery	Balanced accuracy	F1 score	Precision recall curve
DAS	0.93	0.98	0.99
SCORE	0.90	0.99	0.99
LiNGAM	0.89	0.98	0.98
NOTEARS	0.86	0.98	0.98
Full dataset	0.90	0.98	0.98
Shapely	0.91	0.99	0.99
Mutual info regression	0.91	0.99	0.99
ANVOA-F statistic	0.88	0.98	0.98
RFE-logistic regression	0.88	0.98	0.98

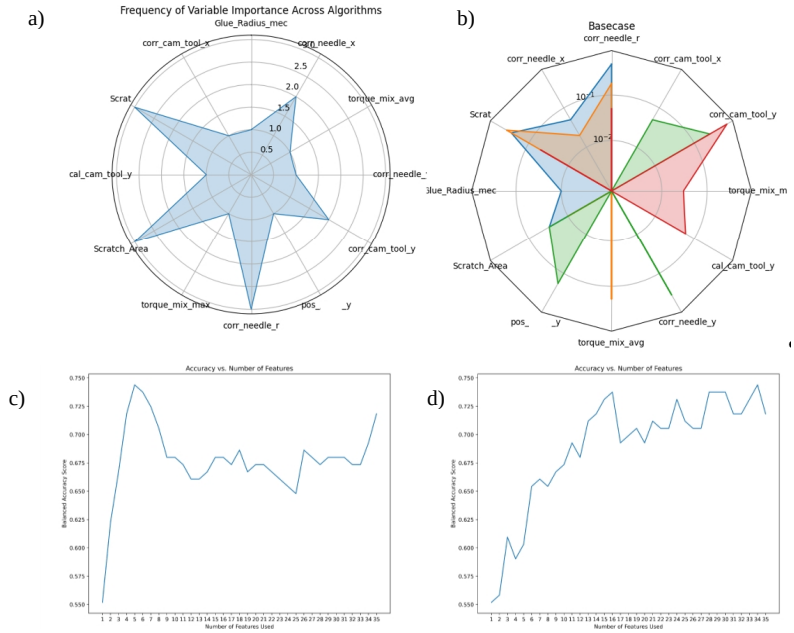


Fig. 5. Star plot analysis of features contributing to causal subsets of 5 most important data features. Balanced accuracy comparison for different number of features using Oneclass SVM.
(a) Star plot counting how often data feature is selected by all 4 causal algorithms; (b) Star plot subset comparison using logarithmic scale; (c) NOTEARS accuracy comparison;(d) LiNGAM accuracy comparison.

In Figure 5.c, 5.d and Figure 6.a, 6.b we can see the balanced accuracy comparison for different numbers of features considered for the 4 causal algorithms. The aim is to observe how the number of features affect the balanced accuracy. Figure 5.c, 5.d represents the balanced accuracy comparison for the NOTEARS and LiNGAM causal quantification using Oneclass SVM model. In Figure 5.c we can see that the balanced accuracy is initially increasing proportionally with the number of features. However, when the number of features is 5 we see the highest recorded accuracy and after that, as the number of features increases, we see no clear relationship to the accuracy. Figure 5.d illustrates LiNGAM algorithm where the accuracy is increasing but not in a

proportional manner. When the number of features is almost 16 we see one of the highest accuracy but then it keeps alternating until the number of features is 34, which marks the highest accuracy.

Figure 5.a counts how often a data feature was within the 5 most important ones as identified by the 4 causal algorithms. It ranges from 1 to 3 times. Figure 5.b gives the absolute values of the same data features using a logarithmic plot to basis 10 revealing that in most cases also the highest numbers of counts also correlate to the highest number of absolute values, see e.g. correlation needle radius.

Figure 6 represents the balanced accuracy comparison for the DAS and SCORE causal quantification using XGBOOST model. In Figure 6.a the highest accuracy is for one feature which is corr-needle-r, then it decreases and alternates by a small margin. Note that the difference between the highest and lowest accuracy is less than 0.2 % in this case only. For Figure 6.b the accuracy is low only for the initial two features and then it remains almost constant as the number of features is increased. The third variable that caused the high increase is also corr-needle-r. In Figure 6 we can clearly see that the number of features affecting the outcome are much less than in Figure 5.

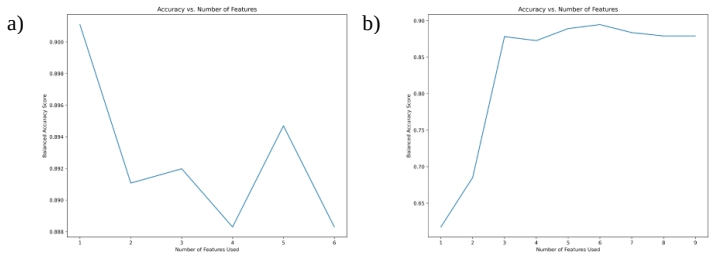


Fig. 6. Balanced accuracy comparison for different number of features using XGBOOST. (a) DAS-accuracy comparison; (b) SCORE-accuracy comparsion.

5. Conclusions and outlook

In summary, a method chain has been presented that digests heterogeneous hardware production step data, in particular image data and machine data, constructs a causal failure root cause graphs and quantifies them sufficient to determine most contributing data features. These were used to predict production step failure. The number of data features was selected from 35 to ca. 3 to 5 key features that generated similar anomaly prediction results than the full data set. It was found that the most promising 4 causal inference approaches compare very well and are often even better than state-of-the-art feature extraction methods.

Using the causal ladder terminology of the introduction section 1, the causal graph construction itself allows rung 2 and rung 3 analysis, i.e. the effect of variations of data feature values on the production step result can be estimated in terms of if there is an influence at all and how strong it is as well as counterfactual assessments could be attempted in terms of hypothesizing how the result would be affected if some arrow connections would be added or removed. Rung 2 considerations have been used in the present approach to assess the strengths of the effect of data features nodes on the production step results.

However, it remains somewhat unclear what can be considered as true root cause in the present case as several data features are analytically related to each other, e.g., x, y and r, or it has not yet been clarified in as far data features can be changed in production process at all or even in addition independent from each other. Nevertheless, the radar plot approach is expected to be very helpful for analysts of production failures at it gives strong hints where to search and eliminate failures in the production process. In addition, it allows to superpose the results of several causal inference approaches thus increasing likely robustness of root cause identification.

Future work could further explore rung 2 and rung 3 analysis and improvement options, in particular hypothesis ranking of best production process optimization options.

References

Amer, M., Goldstein, M., Abdennadher, S. 2013. Enhancing one-class support vector machines for unsupervised anomaly detection. In: ACM SIGKDD (Ed.), Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description (ODD'13), 8–15.

Beraha, M., Metelli, A.M., Papini, M., Tirinzoni, A., Restelli, M. 2019. Feature Selection via Mutual Information: New Theoretical Insights. In: 2019 International Joint Conference on Neural Networks (IJCNN), 1–9.

Budhathoki, K., Minorics, L., Bloebaum, P., Janzing, D. 2022. Causal structure-based root cause analysis of outliers. In: Proceedings of the 39th International Conference on Machine Learning, 2357–2369.

Cha, J., Park, J., Jeong. J. 2022. A Novel Defect Classification Scheme Based on Convolutional Autoencoder with Skip Connection in Semiconductor Manufacturing. In: 2022 24th International Conference on Advanced Communication Technology (ICACT), 347–352.
Chen, T., Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, 785–794.
Chien, C.-F., Chuang, S.-C. 2014. A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics. IEEE Trans. Semicond. Manufact. 27, 475–488.
Draper, G.M., Livnat, Y., Riesenfeld, R.F. 2009. A survey of radial methods for information visualization. IEEE Trans Vis Comput Graph 15, 759–776.
Fujiwara, D., Koyama, K., Kiritoshi, K., Okawachi, T., Izumitani, T., Shimizu, S. 2023. Prospects of continual causality for industrial applications. In: Proceedings of The First AAAI Bridge Program on Continual Causality 208, Proceedings of Machine Learning Research, PMLR, 18–24.
Hao, J., Ho, T.K. 2019. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. Journal of Educational and Behavioral Statistics 44, 348–361.
He, Z., He, Y., Wie, Y. 2016. Big data oriented root cause identification approach based on PCA and SVM for product infant failure. In: Proceedings of 2016 Prognostics and System Health Management Conference (PHM-Chengdu), 1–5.
Howse, J., Minichino, J. 2020. Learning OpenCV 4 computer vision with Python 3: Get to grips with tools, techniques, and algorithms for computer vision and machine learning. Packt Publishing, Birmingham, UK.
Jain, A.K., Grumber, C., Gelhausen, P., Häring, I., Stolz, A. 2020. A Toy Model Study for Long-Term Terror Event Time Series Prediction with CNN. European Journal for Security Research 5, 289–309.
Janzing, D., Blöbaum, P., Minorics, L., Faller, P., Mastakouri, A. 2020. Quantifying intrinsic causal contributions via structure preserving interventions, arXiv, 18 pp.
Kao, S.-X., Chien, C.-F. 2023. Deep Learning-Based Positioning Error Fault Diagnosis of Wire Bonding Equipment and an Empirical Study for IC Packaging. IEEE Trans. Semicond. Manufact. 36, 619–628.
Kim, E., An, J., Cho, H.-C., Cho, S., Lee, B. 2023. A sensor data mining process for identifying root causes associated with low yield in semiconductor manufacturing. DTA 57, 397–417.
Li, A., Lee, J., Montagna, F., Trevino, C., Ness, R. 2023. Dodiscover: Causal discovery algorithms in Python. https://github.com/py-why/dodiscover.
Liu, C.-W., Chien, C.-F. 2013. An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing. Engineering Applications of Artificial Intelligence 26, 1479–1486.
Lundberg, S., Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. arXiv, 10
Montagna, F., Noceti, N., Rosasco, L., Zhang, K., Locatello, F. 2023. Scalable Causal Discovery with Score Matching. In: Proceedings of the Second Conference on Causal Learning and Reasoning (CleaR), PMLR 213, 752–771.
Nagamura, Y., Arima, K., Arai, M., Fukumoto, S. 2021. Layout Feature Extraction Using CNN Classification in Root Cause Analysis of LSI Defects. IEEE Trans. Semicond. Manufact. 34, 153–160.
Ogunleye, A., Wang, Q.-G. 2018. Enhanced XGBoost-Based Automatic Diagnosis System for Chronic Kidney Disease. In: 2018 IEEE 14th International Conference on Control and Automation (ICCA), 805–810.
Ong, P.-L., Choo, Y.-H., Kamilah Muda, A. 2015. A manufacturing failure root cause analysis in imbalanced data set using PCA weighted association rule mining. Jurnal Teknologi 77, 103-111.
Pearl, J., Mackenzie, D. 2020. The book of why: The new science of cause and effect. Basic Books, New York.
Pedregosa, F., et al. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.
Qalajia, M. 2024. Root cause detection of failure in Production data using Causal Graphs. Master Thesis, University of Freiburg.
Rolland, P., Cevher, V., Kleindessner, M., Russell, C., Janzing, D., Schölkopf, B., Locatello, F. 2022. Score Matching Enables Causal Discovery of Nonlinear Additive Noise Models. In: Proceedings of the 39th International Conference on Machine Learning, 18741–18753.
Shakeela, S., Shankar, N.S., Reddy, P.M., Tulasi, T.K., Koneru, M.M. 2020. Optimal Ensemble Learning Based on Distinctive Feature Selection by Univariate ANOVA-F Statistics for IDS. International Journal of Electronics and Telecommunications 67, 267–275.
Shimizu, S., Hoyer, P.O., Hyvarinen, A., Kerminen, A. 2006. A Linear Non-Gaussian Acyclic Model for Causal Discovery. Journal of Machine Learning Research 7, 2003–2030.
Stancin, I., Jovic, A. 2019. An overview and comparison of free Python libraries for data mining and big data analysis. In: Proceedings of 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 977–982.
Sui, W., Zhang, D. 2012. Four Methods for Roundness Evaluation. Physics Procedia 24, 2159–2164.
Wang, H., Li, J., Zhu, G. 2023a. A Data Feature Extraction Method Based on the NOTEARS Causal Inference Algorithm. Applied Sciences 13, 8438, 22
Wang, S., Zhao, Q., Han, Y., Wang, J. 2023b. Root cause diagnosis for complex industrial process faults via spatiotemporal coalescent based time series prediction and optimized Granger causality. Chemometrics and Intelligent Laboratory Systems 233, 104728, 10
Zhang, C., Wang, D., Jia, J., Wang, L., Chen, K., Guan, L., Liu, Z., Zhang, Z., Chen, X., Zhang, M. 2022. Potential failure cause identification for optical networks using deep learning with an attention mechanism. J. Opt. Commun. Netw. 14, 122-133.
Zheng, X., Aragam, B., Ravikumar, P., Xing, E.P. 2018a. DAGs with NO TEARS. https://github.com/xunzheng/notears.
Zheng, X, Aragam, B., Ravikumar, P., Xing, E.P. 2018b. DAGs with NO TEARS: Continuous Optimization for Structure Learning. In: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), 12 pp.
Zheng, Y., Huang, B., Chen, W., Ramsey, J., Gong, M., Cai, R., Shimizu, S., Spirtes, P., Zhang, K. 2023. Causal-learn: Causal Discovery in Python, arXiv, 10 pp.