Proceedings of Machine Learning Research 343:363–378, 2026 Probabilistic Graphical Models (PGM)

Using P&IDs and Process Data To Construct Bayesian Networks For Anomaly Detection

Jakob Ø. Hansen JAKOB@HUGIN.COM
Department of Computer Science, Aalborg University, HUGIN EXPERT, Aalborg, Denmark

Anders L. Madsen ANDERS@HUGIN.COM
Department of Computer Science, Aalborg University, HUGIN EXPERT, Aalborg, Denmark

Thomas Dyhre Nielsen TDN@CS.AAU.DK
Department of Computer Science, Aalborg University, Aalborg, Denmark

Joseph Mietkiewicz JOTOMI@CS.AAU.DK
Department of Computer Science, Aalborg University, Aalborg, Denmark
School of Food Science & Environmental Health, TU Dublin, Dublin, Ireland

Gabriele Baldissone GABRIELE.BALDISSONE@POLITO.IT
Dipartimento Scienza Applicata e Tecnologia (DISAT), Politecnico di Torino, Torino, Italy

Micaela Demichela MICAELA.DEMICHELA@POLITO.IT
Dipartimento Scienza Applicata e Tecnologia (DISAT), Politecnico di Torino, Torino, Italy

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Industrial processes are often safety-critical and consist of multiple interacting components. The complexity of such systems, combined with the scarcity of faulty-operation data, makes the construction of reliable anomaly detection models challenging. To address this issue, we investigate how qualitative information in technical process documentation - specifically Piping and Instrumentation Diagrams (P&IDs) - can guide the construction of Bayesian networks for anomaly detection. P&IDs describe the instrumentation and piping of industrial systems, including the placement of sensors and actuators and their interconnections. We show how this structural information can be exploited to decompose an industrial process into smaller subsystems for which local models can be specified or learned independently. For anomaly detection, we employ a conflict-based measure that quantifies inconsistencies in observed process data. We adapt existing threshold calibration and smoothing concepts to the conflict-measure setting, introducing three sequential filters that balance false-alarm frequency against detection latency for true anomalies.

The proposed methodology has been integrated into PredictIT, a predictive maintenance platform for industrial processes and systems. Experiments on a simulated formaldehyde production plant demonstrate that P&ID-informed models outperform corresponding full-system models, yielding lower false positive rates and reduced detection latency.

Keywords: Bayesian Networks, Anomaly Detection, P&ID

© 2026 J.Ø. Hansen, A. L. Madsen, T.D. Nielsen, J. Mietkiewicz, G. Baldissone & M. Demichela.

1. Introduction

The increasing adoption of Industry 4.0 technologies has intensified interest in predictive maintenance and data-driven industrial monitoring (Vaidya et al., 2018). Rather than reacting to failures after they occur, predictive maintenance aims to detect abnormal system behaviour early using sensor data and condition monitoring. However, reliable anomaly detection in industrial systems remains challenging due to noisy measurements, complex process dependencies, and scarcity of fault data (Naqvi et al., 2026). Purely data-driven approaches may struggle to generalise to rare fault conditions, while expert-driven approaches often require extensive manual modeling effort.

Piping and Instrumentation Diagrams (P&IDs) (King, 1998) are standard technical documentation of industrial plants, capturing equipment and their interconnections. This structural information may be utilized in combination with probabilistic graphical models (PGMs) (Pearl, 1991): in particular, Bayesian Networks (BNs) (Jensen and Nielsen, 2007), can combine qualitative structural information with quantitative sensor measurements, making them well suited for enabling industrial monitoring tasks in safety-critical environments.

This work investigates the use of P&ID-guided decomposition of a system for anomaly detection in industrial processes. Given available process variables, the objective is to detect deviations from normal operation as early as possible while minimising false positives and supporting fault localization. The proposed approach decomposes the plant into subsystems based on the structural information in the P&ID, for which local models can then be specified or learned independently. Furthermore, the modular structure of the approach may support reuse of model components across plants containing similar subsystems.

Anomaly detection for industrial systems has been explored extensively. A recent review notes advances in image processing (Alzarooni et al., 2025), with anomalies detected using student-teacher approaches (Batzner et al., 2024) and foundation models (Gu et al., 2024). There has also been work on developing increasingly challenging benchmarks to reflect real-world conditions (Wang et al., 2024). For time series anomaly detection, Graph Neural Networks, and in particular Graph Attention Networks, have been utilised to learn inter-sensor relationships (Deng and Hooi, 2021) and to capture spatial and temporal dependencies between variables (Tian et al., 2023). Recently, Li et al. (2026) have integrated domain knowledge with a graph attention network to reduce spurious correlations.

Probabilistic approaches using BNs have also been explored for anomaly detection in sensor data. Nielsen and Jensen (2007) detect faults using a conflict-based measure that quantifies inconsistencies in observed process data. Conflict-based anomaly detection has also been explored in tabular data (Dufraisse et al., 2020) and to detect anomalies in electrical grids (Madsen et al., 2025). Unlike Rong et al. (2024), whose classification model is trained on clusters of previously observed abnormal patterns, the conflict measure requires only a BN of normal operation and no prior examples of faults. Onnes et al. (2024) use a similar conflict measure with a calibrated threshold for AI monitoring, but not for industrial fault detection. The novelty of our work is the integration of technical documentation (P&IDs) and probabilistic anomaly detection, which, to our knowledge, has not been explored previously. We also introduce a conflict threshold calibration method and three smoothing filters to reduce false positives.

The remainder of this paper is organised as follows. Section 2 introduces the formaldehyde plant case study and problem setting. Section 3 presents the mathematical preliminaries and notation used throughout the paper. Section 4 describes a conflict threshold calibration method, three smoothing filters and an algorithm for constructing BN models from the P&ID. Section 5 presents the experimental design and results. Section 6 describes the integration of the approach into the PredictIT application[1]. Finally, Section 7 concludes the paper and outlines directions for future work.

2. Case Study: Formaldehyde Plant

Industrial processes are complex, safety-critical systems for which real fault data is rarely available and ground truth is difficult to establish. To evaluate our approach in a controlled setting, we use a simulated formaldehyde production plant, which allows faults to be injected with known onset, magnitude, and location while capturing the multivariate sensor interactions typical of real industrial processes. The system under study is an experimental formaldehyde production plant (Demichela et al., 2017), developed within the Collaborative Intelligence for Safety-Critical Systems (CISC) project[2] and previously used for industrial decision support and operator assistance research (Demichela et al., 2017; Amazu et al., 2025; Abbas et al., 2025; Mietkiewicz et al., 2023, 2024; Amazu et al.; Briwa et al., 2026). Figure 1 shows the P&ID of the plant with an indication of its 6 different subsystems. The

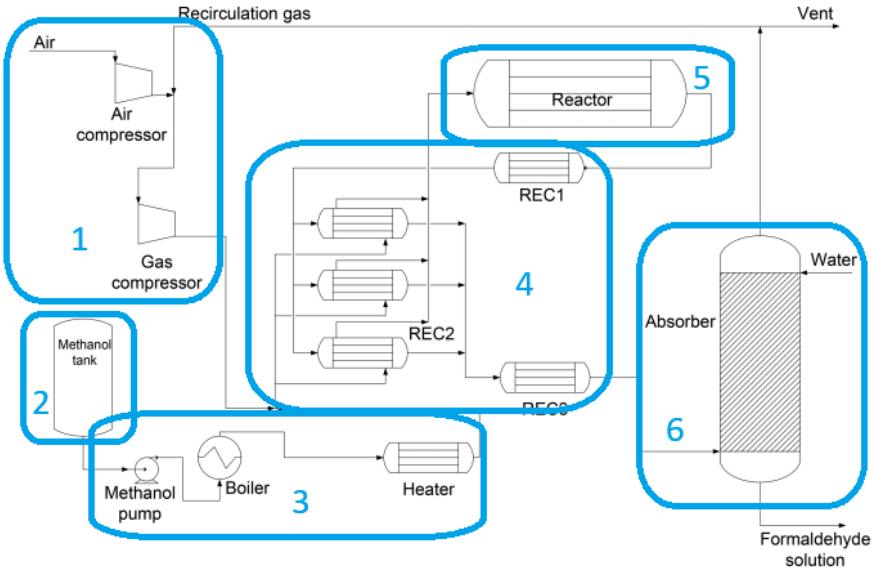


Figure 1: Overview of the formaldehyde plant under study adapted from Demichela et al. (2017).

plant produces formaldehyde through partial oxidation of methanol with air, and consists of six main subsystems: Compressor (1), Tank (2), Methanol Heater (3), Heat Exchangers (4),

1. https://idocpredictit.dk/
2. https://www.ciscproject.eu/

Reactor (5), and Absorber (6). The Tank stores liquid methanol and the Methanol Heater vaporizes the feed stream before reaction, while the Compressor supplies the airflow required for oxidation; the Reactor then converts methanol into formaldehyde over a catalyst, with the Heat Exchangers and Absorber recovering energy and capturing the final product. Currently, the plant relies primarily on threshold-based sensor alarms, and in complex operating conditions a large number of simultaneous alarms may increase operator workload and delay response times (Briwa et al., 2026), making it essential that any anomaly detection system maintains a low false positive rate and is robust to sensor noise and transient deviations. For further inspection of the system, consider the Tank subsystem, shown in Figure 2: sensors are depicted by circles with tag numbers indicating the measured variable and modifiers, pipes by lines, and equipment by predefined (sometimes plant-specific, custom) symbols[3].

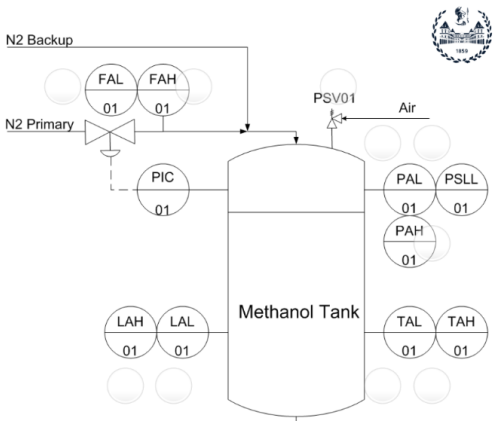


Figure 2: Tank subsystem of the formaldehyde plant.

3. Preliminaries

3.1. P&ID

Piping and Instrumentation Diagrams (P&IDs) are standardised engineering diagrams describing process components, instrumentation, and their interconnections according to the ISA-5.1 standard (International Society of Automation (ISA), 2024; King, 1998). These diagrams provide a structural representation of industrial plants that is widely used in engineering practice. Automatic extraction of structured information from P&IDs has been studied, primarily using deep learning approaches (Toral et al., 2021; Yoochan et al., 2021; Kim et al., 2021; Mani et al., 2020). However, challenges such as limited annotated data, symbol variability, and class imbalance remain unresolved in general settings (Jamieson et al., 2024). In this work, we focus on the use of P&IDs for probabilistic modeling rather than data extraction. We represent a P&ID as $\mathcal{D} = (\mathcal{V}, \mathcal{U})$, where $\mathcal{V} = \{V_1, \ldots, V_n\}$ is the set of process variables, each corresponding to a measurable or observable quantity like equipment, settings, derived variables or other process elements, and $\mathcal{U} = \{U_1, \ldots, U_k\}$ is a partition of $\mathcal{V}$ into subsystems, where $U_i \subseteq \mathcal{V}$, $\bigcup_{i=1}^{k} U_i = \mathcal{V}$, semantically groups the variables belonging to a physically coherent section of the plant, as defined by the P&ID.

Example: Formaldehyde Plant Consider again the formaldehyde plant introduced in Section 2 and illustrated in Figure 1. Each plant section corresponds to a subsystem $U_i \in \mathcal{U}$, and the variables in each group measure physical quantities such as pressure, temperature, and flow rate within that section. Figure 2 shows the Tank subsystem U_{Tank}, whose measurable sensor variables are shown as circles. These include nitrogen flow (FAL, FAH) measured in Nm3/h, pressure (PAL, PSLL, PAH, PIC) measured in Pascal, tank level (LAH, LAL) measured in meters and temperature (TAL, TAH) measured in °C.

3. https://kimray.com/training/most-common-control-valve-symbols-pid

3.2. Bayesian Networks

A BN (Jensen and Nielsen, 2007) is a probabilistic graphical model represented by the tuple $\mathcal{B} = (G, \mathcal{P})$, where $G = (X, E)$ is a directed acyclic graph (DAG) over random variables $X = \{X_1, \ldots, X_n\}$ with edges E, and $\mathcal{P}$ is the corresponding set of conditional probability distributions. For discrete variables, each distribution is represented as a conditional probability table, encoding the distribution of a variable X_i given its parents $\mathbf{Pa}(X_i)$, while for continuous variables the distribution is a conditional linear Gaussian distribution (Lauritzen and Jensen, 2001). The joint probability distribution over all variables in the network factorizes as $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathbf{Pa}(X_i))$.

Our approach to anomaly detection is based on learning a BN representation of the system under study from normal operation data. Normal operation data is data collected over a period of where the system is expected to run normally.

3.3. Conflict Measure

An anomaly can be viewed as a combination of observations that are unlikely to occur according to a probabilistic model. This is quantified through the conflict measure introduced in Jensen et al. (1990).

$$\operatorname{conf}(e) = \log\left(\frac{\prod_{i=1}^{n} P(e_i)}{P(e_1, \ldots, e_n)}\right). \tag{1}$$

Here, $e = \{e_1, \ldots, e_n\}$ denotes the evidence, where each e_i is an observation of the i-th variable, and P denotes a model of the system learned under normal operation. The conflict measure assumes that the observations are positively correlated: $\operatorname{conf}(e) > 0$ means that the product of the marginal probabilities of each e_i is larger than the joint probability of the evidence, which indicates deviation from normal operation. For continuous variables, the probabilities are replaced by the value of the conditional probability density function evaluated at the observed value.

The global conflict can be decomposed into local and inter-component contributions (Nielsen and Jensen (2007)). For $e = e^x \cup e^y$, then:

$$\operatorname{conf}(e) = \operatorname{conf}(e^x) + \operatorname{conf}(e^y) + \operatorname{conf}(e^x, e^y). \tag{2}$$

This decomposition is particularly relevant for subsystem-based monitoring, as the global conflict can be exactly recovered from local conflicts and inter-conflict terms, ensuring that no information is lost under decomposition.

4. Methods

In this section, we first introduce methods for reducing false positives by using conflict threshold calibration and applying smoothing filters. These methods are independent from the P&ID. Next, we describe the decomposition procedure that constructs subsystem-level BNs from the P&ID representation and assigns the calibrated threshold to each subsystem.

4.1. Conflict Threshold Calibration

In practice, high conflict values may occasionally arise even under normal operating conditions, particularly in rare but valid operating regimes (Jensen et al., 1990). To reduce

such false alarms, we determine a subsystem-specific threshold τ based on the distribution of conflict scores under normal operation. This per-subsystem threshold τ is based on the conflict scores under normal operation over a window of T observations. Only positive conflict scores are considered, and τ is set at the p-th percentile of the observed scores,

$$\tau = \text{Percentile}_p\Big(\Big\{\text{conf}(e^{(t)}) \mid \text{conf}(e^{(t)}) > 0\Big\}\Big). \tag{3}$$

4.2. Conflict Score Smoothing

Threshold calibration addresses cases of positive conflict scores under normal operation, but it does not address outliers. We add a sequential threshold requirement by smoothing the conflict scores over a window of time. Stronger smoothing reduces spurious detections but increases detection delay, while weaker smoothing improves responsiveness at the cost of higher false alarm rates. The optimal setting balances the costs associated with false alarms and detection delay.

We evaluate three smoothing strategies spanning two design paradigms: one that aggregates raw conflict scores before comparing against a threshold, and one that compares each observation against the threshold first and then aggregates the resulting binary outcomes.

4.2.1. Sliding Window Average

The sliding-window filter computes the mean conflict score over the most recent k observations:

$$\bar{c}^{(t)} = \frac{1}{k}\sum_{i=0}^{k-1} \text{conf}(e^{(t-i)}). \tag{4}$$

An alarm is triggered if $\bar{c}^{(t)} > \tau$. This approach smoothens short-term fluctuations, reducing sensitivity to isolated spikes. A larger k improves robustness to noise but slows detection, while smaller k increases responsiveness at the cost of potential false alarms.

4.2.2. Exponentially Weighted Moving Average (EWMA)

The EWMA filter computes a recursively smoothed estimate:

$$\sum_{i=0}^{k-1} \mathbf{1}\Big[\text{conf}(e^{(t-i)}) > \tau\Big] \geq m, \quad k \geq m. \tag{5}$$

with $\hat{c}^{(0)} = 0$ and $\alpha \in (0, 1)$. An alarm is raised when $\hat{c}^{(t)} > \tau$. The parameter α controls responsiveness versus smoothing. Small values of α produce heavy smoothing with slow adaptation, and large values prioritize recent observations at the cost of noise sensitivity.

4.2.3. Count-Based Filter

The count-based filter operates on exceedance events rather than raw scores. Let $\mathbf{1}[\cdot]$ denote the indicator function. An alarm is triggered when:

$$\sum_{i=0}^{k-1} \mathbf{1}\Big[\text{conf}(e^{(t-i)}) > \tau\Big] \geq m, \quad k \geq m. \tag{6}$$

The ratio m/k governs the required persistence of exceedances, trading off latency against robustness to noise. By ignoring score magnitude, this filter is insensitive to outliers and provides robustness to extreme but isolated deviations. The parameter m/k is chosen to balance false alarms and detection delay.

4.3. Decomposition

The proposed approach decomposes the full system into a set of subsystems defined by the P&ID. The P&ID provides a domain-informed partitioning of the system that determines which variables are grouped together during model construction. This partitioning allows models to capture local dependency structures and avoid spurious correlations in other parts of the system. The decomposition procedure is summarised in Algorithm 1 and takes three inputs: a dataset $\mathbf{D} \in \mathbb{R}^{N\times T}$ representing N process variables observed over T time steps under normal operating conditions, a P&ID $\mathcal{D} = (\mathcal{V},\mathcal{U})$, where the partitioning of the P&ID into subsystems $U \in \mathcal{U}$ is based on the pre-defined sections of the plant described in Section 2, and finally a percentile setting for the threshold calibration described in Section 4.1.

For each subsystem $U \in \mathcal{U}$, the corresponding columns of $\mathbf{D}$ are extracted to form $\mathbf{D}_U \in \mathbb{R}^{|U|\times T}$ (line 3), containing only the observations relevant to subsystem U. The dependency structure $G_U = (X_U, E_U)$ is then learned from $\mathbf{D}_U$ via *learn_structure* (line 4), and the conditional probability distributions $\mathcal{P}_U$ are estimated via *learn_parameters* (line 5). Together, G_U and $\mathcal{P}_U$ define a BN $\mathcal{B}_U = (G_U, \mathcal{P}_U)$ for subsystem U (line 6). The per-subsystem threshold τ_{B_U} is determined by the *conflict_threshold_calibration* procedure at line 7 (discussed in Section 4.1). The result of Algorithm 1 is a set of BNs with calibrated thresholds $\{(\mathcal{B}_U, \tau_{B_U})\}_{U\in\mathcal{U}}$, each representing a probabilistic model of normal behaviour for its corresponding subsystem.

Algorithm 1: Construct BNs from P&ID and identify thresholds

Input: $\mathbf{D} \in \mathbb{R}^{N\times T}$, process variable observations over T time steps
Input: $\mathcal{D} = (\mathcal{V},\mathcal{U})$, structured P&ID descriptor
Input: $p \in (0, 100)$, percentile for conflict threshold calibration
Output: $\mathcal{M}$, set of BNs with calibrated thresholds

1 $\mathcal{M} \leftarrow \emptyset$
2 **foreach** $U \in \mathcal{U}$ **do**
3 $\quad \mathbf{D}_U \leftarrow \mathbf{D}[U]$
4 $\quad G_U \leftarrow$ *learn_structure*$(\mathbf{D}_U)$
5 $\quad \mathcal{P}_U \leftarrow$ *learn_parameters*$(G_U, \mathbf{D}_U)$
6 $\quad \mathcal{B}_U \leftarrow (G_U, \mathcal{P}_U)$
7 $\quad \tau_{\mathcal{B}_U} \leftarrow$ *conflict_threshold_calibration*$(\mathcal{B}_U, \mathbf{D}_U, p)$
8 $\quad \mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathcal{B}_U, \tau_{\mathcal{B}_U})\}$
9 **end**
10 **return** $\mathcal{M}$

Figure 3 shows the BNs generated by Algorithm 1 for the formaldehyde plant subsystems in HUGIN Grapihcal User Interface, where each graph represents the learned dependency

structure over the variables associated with that subsystem. It shows that the formaldehyde plant has subsystems that produce models of varying complexity and size, and that all variables are continuous, as evident by the inner circle of each node. However, our algorithm also supports discrete variables.

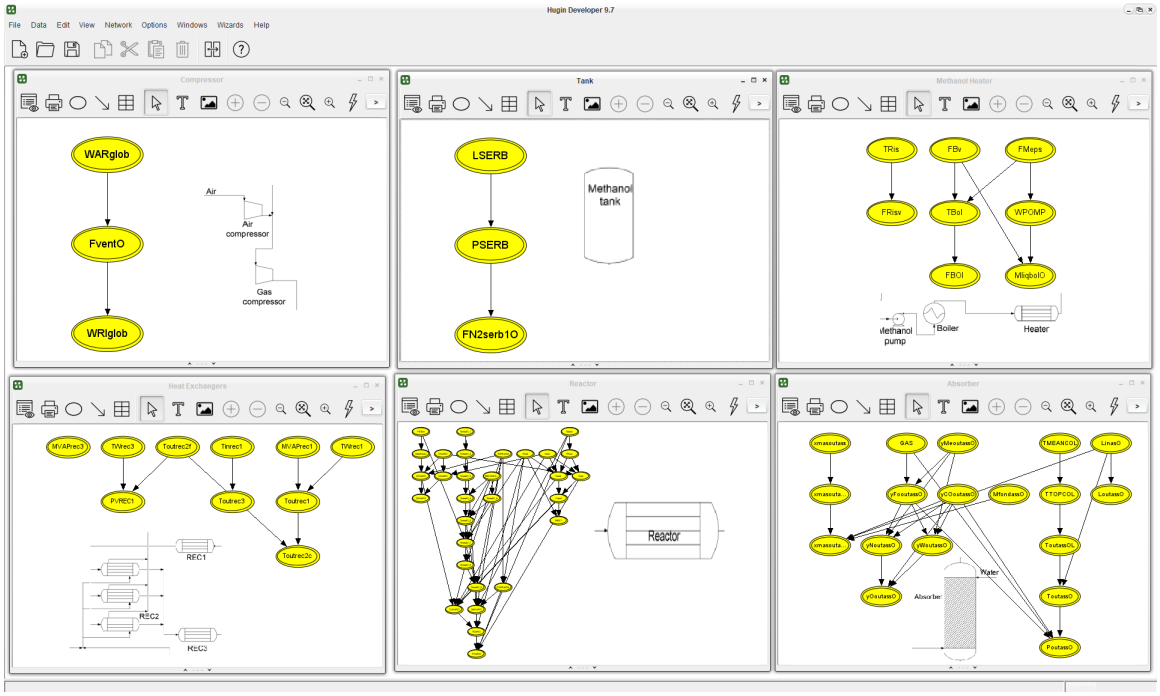


Figure 3: BNs learned for each subsystem of the formaldehyde plant.

5. Experiments

The experiments serve two purposes: first, to assess whether P&ID-informed subsystem decomposition improves anomaly detection relative to a full-system baseline, and second, to evaluate whether threshold calibration and sequential smoothing filters can further reduce false alarms without incurring prohibitive detection latency.

5.1. Experimental Setup

Training data consists of approximately six hours of normal operation of the simulated formaldehyde plant (Demichela et al., 2017), with a sampling interval of one second per time step. We evaluate two fault scenarios from Amazu et al. (2025). In both scenarios, the fault is introduced at $t = 60$ and persists for the remaining 140 seconds of the test sequence. In the first scenario, nitrogen flow into the tank is interrupted, causing a pressure drop that may lead to tank implosion if not mitigated. In the second scenario, reduced water flow to the heat recovery units causes a temperature imbalance, which can lead to reactor overheating. We focus on the tank and heat recovery subsystems as an initial, simple testbed, since their fault dynamics are relatively simple, before extending the approach to more complex

subsystems. For evaluation, the P&ID-informed models $\mathcal{B}_{\text{Tank}}$ and $\mathcal{B}_{\text{Heat_Recovery}}$ generated by Algorithm 1, are used for the first scenario and second scenario, respectively. A full-system baseline $\mathcal{B}_{\text{Full}}$ is trained using all process variables for comparison. For all models, structure learning is performed using the PC algorithm (Spirtes et al., 2001; Madsen et al., 2017) with significance level 0.05, and parameters are estimated using EM (Lauritzen, 1995) to handle missing data.

The simulator is deterministic and based on differential equations and does not model sensor noise. Hence, to reflect real world environments, noise is injected into training and test data at 1%, by taking each observation and randomly adding or subtracting 1% of the value, and all results are averaged over $n = 50$ iterations.

Performance is evaluated using false positives per minute (FP/min), defined as the proportion of normal observations incorrectly flagged as anomalous per minute, and Recall@N, defined as the fraction of iterations in which the fault is detected within N time steps of its onset, for $N \in \{1, \ldots, 20\}$. We compare P&ID-informed models against the full-system baseline under both uncalibrated and calibrated thresholds. The calibrated threshold is set at the 95th percentile of normal conflict scores, selected empirically to balance false alarm reduction and detection latency. Additionally, the three smoothing filters are applied to the uncalibrated conflict scores. Their performance is characterised using Pareto fronts relating FP/min to mean detection latency. Hyperparameter optimisation is performed over $k \in \{2, \ldots, 10\}$ for the sliding-window filter, $\alpha \in \{0.1, \ldots, 0.9\}$ for the EWMA filter, and $k \in \{5, \ldots, 20\}$ with exceedance fraction $m/k \in \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ for the count-based filter.

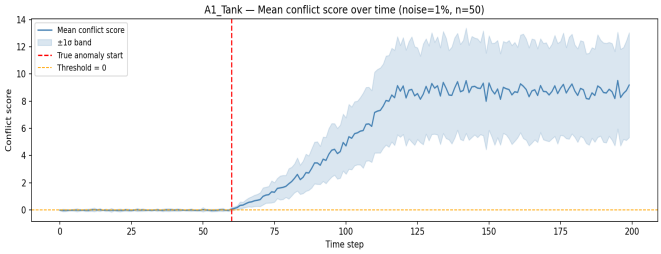
5.2. Detection Latency and FPR

Figures 4 and 5 report mean conflict scores, Recall@N curves, and FP/min for fault scenarios 1 and 2, respectively. For fault scenario 1 (Figure 4), $\mathcal{B}_{\text{Tank}}$ produces a near-zero conflict score during normal operation and a sharp, monotonically increasing signal after fault onset at $t = 60$. On the other hand, $\mathcal{B}_{\text{Full}}$ exhibits positive scores under normal operation with wide variance. The Recall@N curves (Figure 4c) show a clear difference between calibrated and uncalibrated thresholds. Without calibration, both models detect the fault within approximately $N \approx 5$, but at the cost of high FP/min (Figure 4d): 20.6 for $\mathcal{B}_{\text{Tank}}$ and 53.9 for $\mathcal{B}_{\text{Full}}$. These numbers indicate that it is necessary to perform the threshold calibration which results in a reduction to 0.04 FP/min for $\mathcal{B}_{\text{Tank}}$ and 3.9 FP/min for $\mathcal{B}_{\text{Full}}$. As a consequence of the threshold calibration the detection latency increases modestly for $\mathcal{B}_{\text{Tank}}$, but significantly for $\mathcal{B}_{\text{Full}}$.

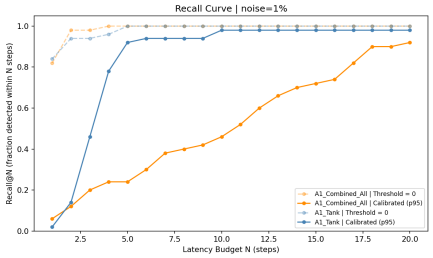
Fault scenario 2 (Figure 5) has similar results, in that threshold calibration is necessary to reduce FP/min (Figure 5d) from 35.1 to 3.5 for $\mathcal{B}_{\text{Heat_Recovery}}$, and from 53.6 to 4.0 for $\mathcal{B}_{\text{Full}}$, at the cost of increased latency.

Overall, the results show that uncalibrated thresholds lead to an unacceptable number of false alarms, while calibration substantially reduces false positives at the cost of increased detection latency. Furthermore, the P&ID-informed models consistently achieve lower FP/min and earlier detection than the full-system baseline. The high number of FP/min may be an artifact of the simulator as such high FP/min are not usually seen in practice, according to our experience.

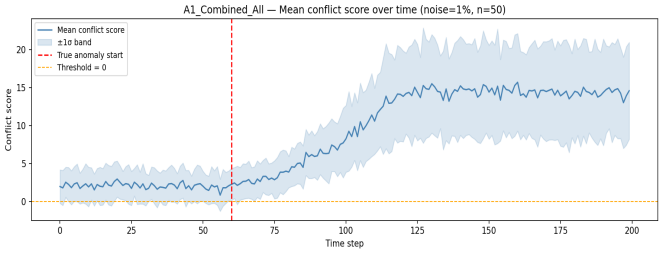
(a) $\mathcal{B}_{\text{Tank}}$ conflict score



(c) Recall@N curve



(b) $\mathcal{B}_{\text{Full}}$ conflict score



(d) FP/min

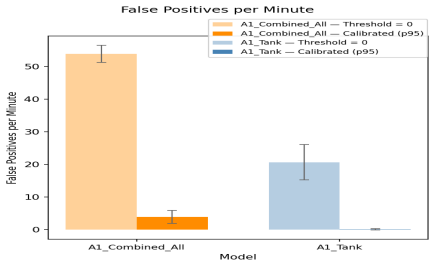


Figure 4: Mean conflict scores, Recall@N curves and FP/min for fault scenario 1.

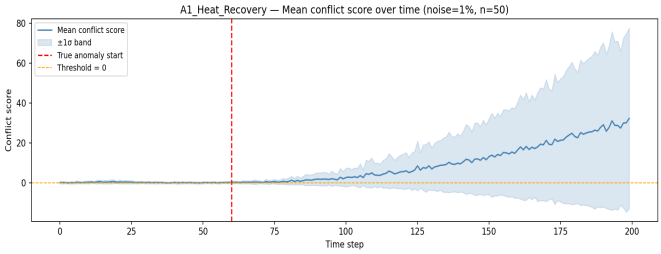
5.3. Smoothing Filter Evaluation via Pareto Front

To suppress isolated outliers we evaluate the three smoothing filters and study how each trades off FPR against detection latency. Across both submodels, the sliding window and EWMA filters broadly confirm that increasing the window size or the weight on previous observations lowers FPR at the cost of higher latency, but produce narrow Pareto fronts with limited room to tune the tradeoff. The count-based filter, with two hyperparameters k and m, yields a substantially wider and better-structured front. For $\mathcal{B}_{\text{Tank}}$ (Figure 6a), the knee lies around k=8, m=7, yielding a mean FPR below 0.05 (3 FP/min) at a latency of approximately 5 steps. $\mathcal{B}_{\text{Heat_Recovery}}$ (Figure 6b) exhibits a less favourable tradeoff, with the knee near k=9, m=8 reaching comparable FPR reduction only at a latency of around 25–30 steps.

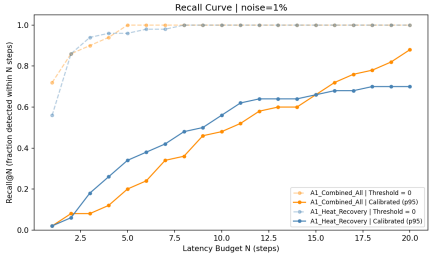
6. Software Implementation

The approach has been integrated into PredictIT (idoc PredictIT, 2026), a predictive maintenance platform designed for continuous monitoring of industrial processes. Figure 7a shows an overview of the dashboard for the formaldehyde plant. Each subsystem is represented as a widget displaying its current operational status and a live conflict score trend. When the conflict score of a subsystem exceeds its calibrated threshold, an alarm is raised and the corresponding widget is flagged, directing operator attention to the affected process component. Figure 7b shows the detailed view for $\mathcal{B}_{\text{Tank}}$, displaying the conflict score at the top alongside individual trend plots for each monitored variable.

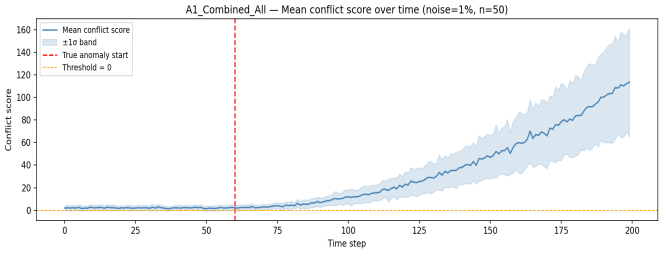
(a) $\mathcal{B}_{\text{Heat_Recovery}}$ conflict score



(c) Recall@N curve



(b) $\mathcal{B}_{\text{Full}}$ conflict score



(d) FP/min

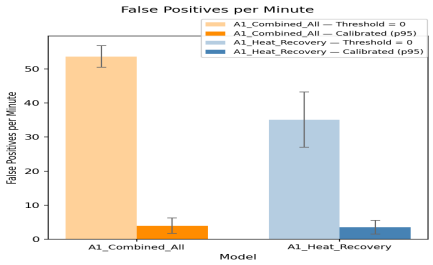
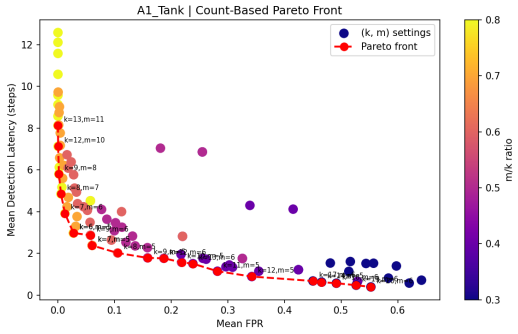


Figure 5: Mean conflict scores, Recall@N curves and FP/min for fault scenario 2.

(a) $\mathcal{B}_{\text{Tank}}$ Count-Based Filter Pareto Front



(b) $\mathcal{B}_{\text{Heat_Recovery}}$ Count-Based Filter Pareto Front

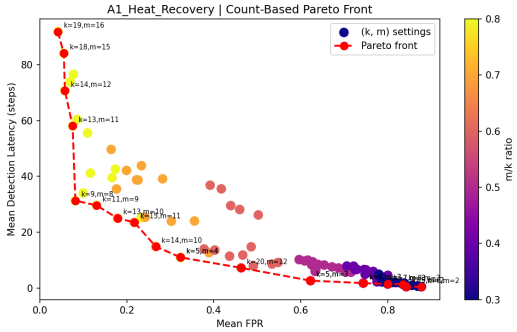
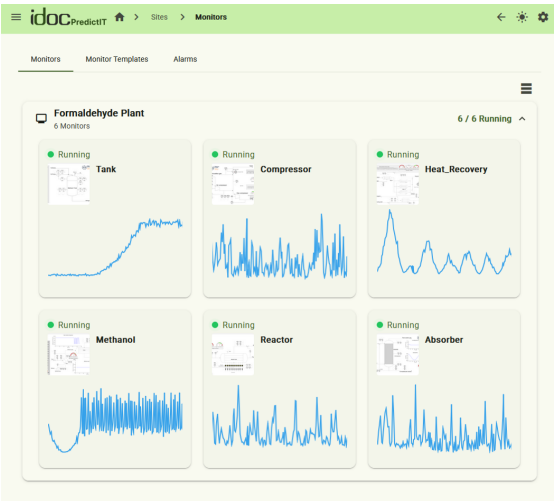


Figure 6: Pareto Fronts for Count-Based Filters for Both Fault Scenarios

7. Conclusion and Future Work

We have presented a P&ID-informed approach to anomaly detection in a simulated industrial process, in which the P&ID guides the decomposition of the plant into subsystems, each modelled as a BN learned from process data under normal operation. Anomalies are detected via the conflict score, a probabilistic measure quantifying inconsistency between observations and the learned model. To reduce false positives, we introduced a per-subsystem

(a) PredictIT dashboard overview



(b) Detailed view for $\mathcal{B}_{\text{Tank}}$ and U_{Tank}. Bottom plots show raw values while top plot shows conflict score.



Figure 7: Screenshots of the PredictIT platform for the formaldehyde plant.

threshold calibration method based on the conflict scores under normal operation, and three smoothing filters each trading detection latency against false alarm rate in distinct ways. These methods are independent of the P&ID decomposition and can be applied to any conflict-based anomaly detector.

Experiments on a simulated formaldehyde plant show that P&ID-informed subsystem models can reduce false positives per minute compared to full-system baselines, though results were mixed across the two scenarios: the subsystem models outperformed the full-system model in the first scenario and under most latency budgets in the second, while performance was comparable once the full-system model was calibrated in the second scenario.

The results also demonstrate that conflict threshold calibration and smoothing filters further reduce false positives at a modest cost in detection latency. Future work should investigate the impact of the missing inter-conflict term, and if significant, explore connecting submodels through interface nodes, for example using object-oriented BNs. Lastly, the approach should be validated on real industrial process data, beyond the controlled simulated setting.

8. Acknowledgements

This work is part of an Industrial PhD which is partly funded by the Innovation Fund Denmark (IFD) under grant 5376-00071B.

References

Ammar N. Abbas, Chidera W. Amazu, Joseph Mietkiewicz, Houda Briwa, Andres Alonso Perez, Gabriele Baldissone, Micaela Demichela, Georgios C. Chasparis, John D. Kelleher, and Maria Chiara Leva. Analyzing Operator States and the Impact of AI-Enhanced Decision Support in Control Rooms: A Human-in-the-Loop Specialized Reinforcement Learning Framework for Intervention Strategies. *International Journal of Human-Computer Interaction*, 41(12):7218–7252, 2025. ISSN 1044-7318. doi: 10.1080/10447318.2024.2391605. URL https://doi.org/10.1080/10447318.2024.2391605.

Abdelrahman Alzarooni, Ehtesham Iqbal, Samee Ullah Khan, Sajid Javed, Brain Moyo, and Yusra Abdulrahman. Anomaly Detection for Industrial Applications, Its Challenges, Solutions, and Future Directions: A Review. https://arxiv.org/abs/2501.11310v1, January 2025.

Chidera W. Amazu, Joseph Mietkiewicz, Ammar N. Abbas, Gabriele Baldissone, Davide Fissore, Micaela Demichela, Anders L. Madsen, and Maria Chiara Leva. Decision support impact and error prediction in control room: A behavioural data analysis. *Reliability Engineering & System Safety*, 272:112616. ISSN 0951-8320. doi: 10.1016/j.ress.2026.112616. URL https://www.sciencedirect.com/science/article/pii/S095183202600428X.

Chidera W. Amazu, Joseph Mietkiewicz, Ammar N. Abbas, Houda Briwa, Andres Alonso-Perez, Gabriele Baldissone, Davide Fissore, Micaela Demichela, and Maria Chiara Leva. Exploring the Influence of Human System Interfaces: Introducing Support Tools and an Experimental Study. *International Journal of Human–Computer Interaction*, 41(10): 6300–6317, 2025. ISSN 1044-7318. doi: 10.1080/10447318.2024.2376354. URL https://doi.org/10.1080/10447318.2024.2376354.

Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 127–137, 2024. doi: 10.1109/WACV57701.2024.00020.

Houda Briwa, Anders L. Madsen, and Maria Chiara Leva. Classifying Control Room Operators' Performance Using Bayesian Networks. In Kai Sauerwald and Matthias Thimm, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 17–30. Springer Nature Switzerland, 2026. ISBN 978-3-032-05134-9. doi: 10.1007/978-3-032-05134-9_2.

Micaela Demichela, Gabriele Baldissone, and Gianfranco Camuncoli. Risk-Based Decision Making for the Management of Change in Process Plants: Benefits of Integrating Probabilistic and Phenomenological Analysis. *Industrial & Engineering Chemistry Research*, 56(50):14873–14887, December 2017. ISSN 0888-5885. doi: 10.1021/acs.iecr.7b03059.

Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:4027–4035, 05 2021. doi: 10.1609/aaai.v35i5.16523.

Evan Dufraisse, Philippe Leray, Raphaël Nedellec, and Tarek Benkhelif. Interactive anomaly detection in mixed tabular data using bayesian networks. In Manfred Jaeger and Thomas Dyhre Nielsen, editors, *Proceedings of the 10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, pages 185–196. PMLR, 23–25 Sep 2020. URL https://proceedings.mlr.press/v138/dufraisse20a.html.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. Anomalygpt: Detecting industrial anomalies using large vision-language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38:1932–1940, 03 2024. doi: 10.1609/aaai.v38i3.27963.

idoc PredictIT. idoc PredictIT - predictive maintenance & AI software. https://idocpredictit.dk/, 2026. Accessed: 2026-05-11.

International Society of Automation (ISA). ISA 5.1 standard – instrumentation symbols and identification, 2024. URL https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa5-1. Accessed: 2026-05-17.

Laura Jamieson, Carlos Francisco Moreno-García, and Eyad Elyan. A review of deep learning methods for digitisation of complex documents and engineering diagrams. *Artificial Intelligence Review*, 57(6):136, May 2024. ISSN 1573-7462. doi: 10.1007/s10462-024-10779-2.

Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs: February 8, 2007*. Information Science and Statistics. Springer New York, New York, NY, second edition edition, 2007. ISBN 9780387682815.

Finn Verner Jensen, Bo Chamberlain, Torsten Nordahl, and Frank Jensen. Analysis in hugin of data conflict. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '90, page 519–528, USA, 1990. Elsevier Science Inc. ISBN 0444892648.

Hyungki Kim, Wonyong Lee, Mijoo Kim, Yoochan Moon, Taekyong Lee, Mincheol Cho, and Duhwan Mun. Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams. *Expert Systems with Applications*, 183:115337, 2021. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.115337. URL https://www.sciencedirect.com/science/article/pii/S0957417421007661.

C. F. King. Analysis, synthesis, and design of chemical processes. richard turton, richard bailie, wallace whiting, joseph shaeiwitz prentice hall, 1998. *Chemie Ingenieur Technik*, 71(11):1319–1320, 1998. doi: https://doi.org/10.1002/cite.330711124. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.330711124.

Steffen L. Lauritzen. The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2):191–201, February 1995. ISSN 0167-9473. doi: 10.1016/0167-9473(93)E0056-A.

Steffen L. Lauritzen and Frank Jensen. Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, 11(2):191–203, April 2001. ISSN 1573-1375. doi: 10.1023/A:1008935617754.

Qixuan Li, Yangjian Ji, Linjin Sun, Nian Zhang, and Tiannuo Yang. Anomaly detection for industrial time series in process industry using informed machine learning with graph attention networks. *Computers in Industry*, 176:104445, March 2026. ISSN 0166-3615. doi: 10.1016/j.compind.2026.104445.

Anders L. Madsen, Frank Jensen, Antonio Salmerón, Helge Langseth, and Thomas D. Nielsen. A parallel algorithm for bayesian network structure learning from large data sets. *Knowledge-Based Systems*, 117:46–55, 2017. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2016.07.031. URL https://www.sciencedirect.com/science/article/pii/S0950705116302465. Volume, Variety and Velocity in Data Science.

Anders L. Madsen, Somesh Bhattacharya, Christian D. Jensen, Rasmus L. Olsen, and Hans-Peter Schwefel. Wrong data detection in electricity distribution grids using bayesian networks. In Kai Sauerwald and Matthias Thimm, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 31–45, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-032-05134-9.

Shouvik Mani, Michael A. Haddad, Dan Constantini, Willy Douhard, Qiwei Li, and Louis Poirier. Automatic digitization of engineering diagrams using deep learning and graph search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 673–679, 2020. doi: 10.1109/CVPRW50498.2020.00096.

Joseph Mietkiewicz, Ammar N. Abbas, Chidera Winifred Amazu, Anders L. Madsen, and Gabriele Baldissone. Dynamic influence diagram-based deep reinforcement learning framework and application for decision support for operators in control rooms. In {Mário P.} Brito, {Terje } Aven, {Piero } Baraldi, {Marko } Čepin, and {Enrico } Zio, editors, *Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)*, pages 2806–2813. Research Publishing, 2023. doi: 10.3850/978-981-18-8071-1_P531-cd. The 33rd European Safety and Reliability Conference (ESREL 2023) ; Conference date: 03-09-2023 Through 08-09-2023.

Joseph Mietkiewicz, Ammar N. Abbas, Chidera W. Amazu, Gabriele Baldissone, Anders L. Madsen, Micaela Demichela, and Maria Chiara Leva. Enhancing Control Room Operator Decision Making. *Processes*, 12(2):328, 2024. ISSN 2227-9717. doi: 10.3390/pr12020328. URL https://www.mdpi.com/2227-9717/12/2/328.

Syed Meesam Raza Naqvi, Mohammad Ghufran, Christophe Varnier, Jean-Marc Nicod, and Noureddine Zerhouni. Anomaly detection using case-based reasoning and representation learning: Application to industrial maintenance. *Journal of Process Control*, 160:103677, April 2026. ISSN 0959-1524. doi: 10.1016/j.jprocont.2026.103677.

Thomas Dyhre Nielsen and Finn V. Jensen. On-line Alert Systems for Production Plants: A Conflict Based Approach. *International Journal of Approximate Reasoning*, 45(2): 255–270, 2007. ISSN 0888-613X. doi: 10.1016/j.ijar.2006.06.010.

Annet Onnes, Mehdi Dastani, and Silja Renooij. Normative monitoring using Bayesian networks: Defining a threshold for conflict detection. In Zied Bouraoui and Srdjan Vesic, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2023)*, volume 14294 of *Lecture Notes in Computer Science*, pages 149–159. Springer, 2024.

Judea Pearl. Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *Artif. Intell.*, 48:117–124, February 1991. doi: 10.1016/0004-3702(91)90084-W.

H. Rong, A. P. Teixeira, and C. Guedes Soares. A framework for ship abnormal behaviour detection and classification using AIS data. *Reliability Engineering & System Safety*, 247, 2024.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. January 2001. ISBN 978-0-262-28415-8. doi: 10.7551/mitpress/1754.001.0001.

Zhiwen Tian, Ming Zhuo, Leyuan Liu, Junyi Chen, and Shijie Zhou. Anomaly detection using spatial and temporal information in multivariate time series. *Scientific Reports*, 13, March 2023. doi: 10.1038/s41598-023-31193-8.

Luis Toral, Carlos Francisco Moreno-García, Eyad Elyan, and Shahram Memon. A deep learning digitisation framework to mark up corrosion circuits in piping and instrumentation diagrams. In Elisa H. Barney Smith and Umapada Pal, editors, *Document Analysis and Recognition – ICDAR 2021 Workshops*, pages 268–276, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86159-9.

Saurabh Vaidya, Prashant Ambad, and Santosh Bhosle. Industry 4.0 – A Glimpse. *Procedia Manufacturing*, 20:233–238, January 2018. ISSN 2351-9789. doi: 10.1016/j.promfg.2018.02.034.

Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jianning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection. https://arxiv.org/abs/2403.12580v1, March 2024.

Moon Yoochan, Jinwon Lee, Duhwan Mun, and Seungeun Lim. Deep learning-based method to recognize line objects and flow arrows from image-format piping and instrumentation diagrams for digitization. *Applied Sciences*, 11:10054, 10 2021. doi: 10.3390/app112110054.