

Article

Object-Centric Process Mining Framework for Industrial Safety and Quality Validation Using Support Vector Machines

Michael Maiko Matonya ¹  and István Budai ^{2,*} ¹ Doctoral School of Informatics, University of Debrecen, H-4028 Debrecen, Hungary; matonya2008@gmail.com² Faculty of Engineering, Department of Mechanical Engineering, University of Debrecen, H-4028 Debrecen, Hungary

* Correspondence: budai.istvan@eng.unideb.hu

Abstract

Ensuring reliable inspection and quality control in complex industrial settings remains a significant challenge, particularly when traditional manual methods are applied to dynamic, multi-object environments. This paper presents and validates a new hybrid framework that integrates Object-Centric Process Mining (OCPM) with Support Vector Machines (SVMs) to improve industrial safety and quality assurance. The aims are: (1) to uncover and model the complex, multi-object processes characteristic of modern manufacturing using OCPM; (2) to assess these models in terms of conformance, performance, and the detection of bottlenecks; and (3) to design and embed a predictive layer based on Support Vector Regression (SVR) to anticipate process outcomes and support proactive control. The proposed methodology comprises a comprehensive pipeline: data fusion and OCEL structuring, OCPM for process discovery and conformance analysis, feature engineering, SVR for predictive modeling, and a multi-objective optimization layer. By applying this framework to a timber sawmill dataset, the study successfully modeled complex lumber drying operations, identified key object interactions, achieving a process conformance fitness score of 0.6905, and testing the integration of a predictive SVR layer. The SVR model's predictive accuracy for production yield was found to be limited ($R^2 = 0.0255$) with the current feature set, highlighting the challenges of predictive modeling in this complex, multi-object domain. Despite this predictive limitation, the multi-objective optimization effectively balanced defect rates, energy consumption, and process delays, yielding a mean objective function value of 0.0768. These findings demonstrate the framework's capability to provide deep, object-centric process insights and support data-driven decision-making for operational improvements in Industry 4.0. Future research will focus on improving predictive model performance through advanced feature engineering and exploring diverse machine learning techniques.



Academic Editor: Mario Di Nardo

Received: 10 November 2025

Revised: 4 December 2025

Accepted: 12 December 2025

Published: 22 December 2025

Copyright: © 2025 by the authors.

Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Industry 4.0; safety inspection; quality compliance; object-centric process mining; support vector machine; predictive analytics

1. Introduction

Industry 4.0 has transformed modern industrial operations into increasingly complex systems characterized by numerous interacting components.

While technologies such as connected machinery and data analytics offer potential for efficiency, ensuring worker safety and product quality within these intricate environments remains a significant challenge. Many industries continue to rely on traditional methods, often manual, for safety checks and quality assessments. These established practices are not

only slow and costly but also inadequate for the complexity of modern processes, making it difficult to consistently meet high standards and manage risks effectively [1]. The impact is substantial; for example, manufacturing activities are reported to account for a large percentage of occupational injuries in some regions, and studies suggest that automated monitoring systems can significantly outperform manual inspections in predicting potential problems [2,3]. As production systems become more interconnected, these legacy inspection methods struggle to handle the sheer volume and complexity of operational data needed to maintain compliance and minimize hazards.

Data-driven analysis offers a promising direction for improvement, particularly through process mining. Process mining techniques analyze the digital footprints left by activities in operational logs to automatically discover how processes actually execute. This allows organizations to uncover hidden inefficiencies, identify safety risks, and improve adherence to regulations [4]. Recent advances in AI-driven process mining tools have further enhanced these capabilities, enabling automated identification of automation opportunities and supporting digital transformation initiatives [5]. However, standard process mining approaches face fundamental limitations when applied to highly complex industrial settings. They typically focus on processes from the perspective of a single case (e.g., one order), making it difficult to analyze scenarios where multiple distinct objects (e.g., machines, materials, workers) interact concurrently. To overcome this, Object-Centric Process Mining (OCPM) was developed [6]. OCPM provides a more complete and realistic understanding of complex operations by modeling the interactions between all involved objects [7].

To advance from descriptive and diagnostic capabilities toward predictive control, this study integrates OCPM with a powerful machine learning technique: Support Vector Machines (SVMs). SVMs, rooted in statistical learning theory, are highly effective for classification and regression tasks, particularly in high-dimensional spaces [8]. Their strength lies in identifying an optimal hyperplane that maximally separates data points into different classes or predicts continuous values, making them well-suited for complex industrial analytics where data are often noisy and non-linear [9]. By combining OCPM with SVMs specifically, Support Vector Regression (SVR) for predictive tasks [10]. This creates a hybrid analytical layer capable of not only detecting process deviations but also predicting future outcomes such as process delays and optimizing performance across quality and energy metrics [11,12].

Despite growing interest in process mining and machine learning (ML) for industrial applications, significant gaps remain in the literature. Most existing OCPM studies focus on business, process management in service industries with limited application to complex manufacturing environments involving multiple interacting physical objects. While SVMs have been widely applied for classification and regression in industrial settings, their integration with OCPM for end-to-end process analysis and prediction remains largely unexplored. Existing safety and quality validation approaches typically rely on isolated inspection points rather than holistic process-aware analysis that considers object interactions, temporal dependencies, and conformance to expected behavior. Few studies have demonstrated the practical, end-to-end application of a hybrid framework that integrates OCPM with real-time Industrial Internet of Things (IIoT) sensor data and SVM-based prediction for proactive safety and quality validation in dynamic manufacturing environments. This study addresses this gap by proposing and validating such a comprehensive framework. The lumber drying process serves as a rigorous case study due to its inherent complexities, including heat and moisture dynamics and the natural variability of wood [13].

The main objective of this study is to propose and validate a novel hybrid framework that integrates Object-Centric Process Mining (OCPM) with Support Vector Machines

(SVMs) for proactive safety and quality validation in complex industrial environments. The specific aims are: (1) to discover and model the intricate, multi-object processes inherent in modern manufacturing using OCPM; (2) to analyze these models for conformance, performance, and bottleneck identification; and (3) to develop and integrate a predictive layer using Support Vector Regression (SVR) to forecast process outcomes and enable proactive control.

The remainder of this paper is organized as follows: Section 2 reviews the related literature on process mining, machine learning in manufacturing, and object-centric approaches, identifying the research gap this study addresses. Section 3 details the proposed hybrid OCPM-SVM methodology, including its architectural components and mathematical foundations. Section 4 presents the results from the application of the framework to a real-world industrial case study. Section 5 discusses the implications of these findings, comparing them with existing literature and outlining the study's limitations, summarizing the key contributions and suggesting directions for future research.

2. Related Work

Process mining, the Industrial Internet of Things (IIoT), and machine learning (ML) have converged to open new frontiers for data-driven optimization in modern industry. This section reviews prior research in this domain, comparing and contrasting recent approaches to highlight the unique contribution of our proposed hybrid OCPM-SVM framework.

2.1. Process Mining in Industry 4.0

Process mining has emerged as a critical discipline for analyzing and improving business processes in the Industry 4.0 era [14]. The Process Mining Manifesto established the guiding principles and challenges for the field, emphasizing its potential to bridge the gap between data science and business process management [15]. Early applications focused on discovering process models from event logs, but the field has since expanded to include conformance checking, performance analysis, and predictive monitoring [16].

2.2. Object-Centric Process Mining Advances

Traditional process mining techniques, while powerful, struggle with the complexity of modern industrial processes where multiple objects interact [17]. This led to the development of Object-Centric Process Mining (OCPM), which provides a more holistic view of processes by modeling the interactions between different object types [6]. Berti and van der Aalst [18] have explored the challenges of dealing with divergence and convergence in event data, which is a common issue in real-world processes. Li et al. [19] have further advanced the field by introducing object-centric behavioral constraints, which allow for more precise process analysis.

2.3. Machine Learning for Quality Prediction

Machine learning has been widely applied in manufacturing for tasks such as quality prediction, predictive maintenance, and process optimization [20]. Cioffi et al. [21] provide a comprehensive review of AI and machine learning applications in smart production, highlighting the potential of these technologies to drive sustainability and efficiency. Sharp et al. [22] survey the advancing use of machine learning in smart manufacturing, noting the increasing adoption of these techniques in industry.

2.4. Hybrid Approaches in Manufacturing

Several studies have explored hybrid approaches that combine process mining with machine learning. For instance, Gejji et al. [11] proposed a framework combining process mining with SVR for predictive quality control in a turning process. Their strength lies in

demonstrating the feasibility of using process-derived features for prediction. However, their approach relies on traditional, case-centric process mining, which limits its ability to model complex interactions in multi-object systems, a key challenge in Industry 4.0 environments. Similarly, Efeoglu and Tuna [12] developed a hybrid SVM-based approach for predictive maintenance, while effective in its specific application, it focuses primarily on the predictive aspect and does not incorporate a deep process discovery and conformance analysis layer, which is crucial for understanding the root causes of failures.

2.5. Research Gap and Positioning

While the aforementioned studies by Gejji et al. [9], Efeoglu et al. [15], and others have made significant contributions, a clear research gap remains. No prior study has presented and validated an end-to-end framework that unifies the deep process discovery and diagnostic capabilities of Object-Centric Process Mining (OCPM) with the powerful predictive capabilities of Support Vector Machines (SVMs) for proactive industrial conformance and optimization. Existing research either focuses on case-centric process mining with machine learning, or on the descriptive aspects of OCPM. Our work directly addresses this gap by proposing and validating a novel hybrid framework that leverages the strengths of both OCPM and SVMs to create a powerful analytical tool for Industry 4.0.

3. Proposed Methodology

This study introduces a novel hybrid framework that integrates Object-Centric Process Mining (OCPM) with Support Vector Machines (SVMs) to provide a comprehensive solution for industrial safety and quality validation. The framework is designed to be generalizable and applicable to a wide range of Industry 4.0 environments. This section describes the architecture and components of the proposed OCPM-SVM framework. To visually summarize the proposed methodology, a flow diagram is presented in Figure 1.

Figure 1 summarizes the end-to-end pipeline: data fusion and OCEL structuring (Phase 1), OCPM discovery and conformance analysis (Phase 2), feature engineering (Phase 3), SVR-based prediction (Phase 4), and the multi-objective decision layer (Phase 5). This schematic is referenced when interpreting each empirical result in Section 4.

3.1. Data Sources and Preprocessing

The framework is designed to ingest data from a variety of sources, including Manufacturing Execution Systems (MES), Industrial Internet of Things (IIoT) sensors, and manual data entry systems. The raw data are preprocessed to create a clean, structured, and unified dataset suitable for analysis. This involves several steps:

- Data Fusion: Combining data from multiple sources into a single, coherent event log.
- Data Cleaning: Handling missing values, removing duplicates, and correcting errors in the data.
- OCEL Structuring: Transforming the cleaned data into an Object-Centric Event Log (OCEL) format, which is the input for the OCPM engine.

For the empirical study, the improved OCEL derived from the sawmill's lumber drying operations (2013–2022) were used. Operational context attributes (e.g., shift, operator, maintenance status) were preserved to enable conformance and predictive analyses.

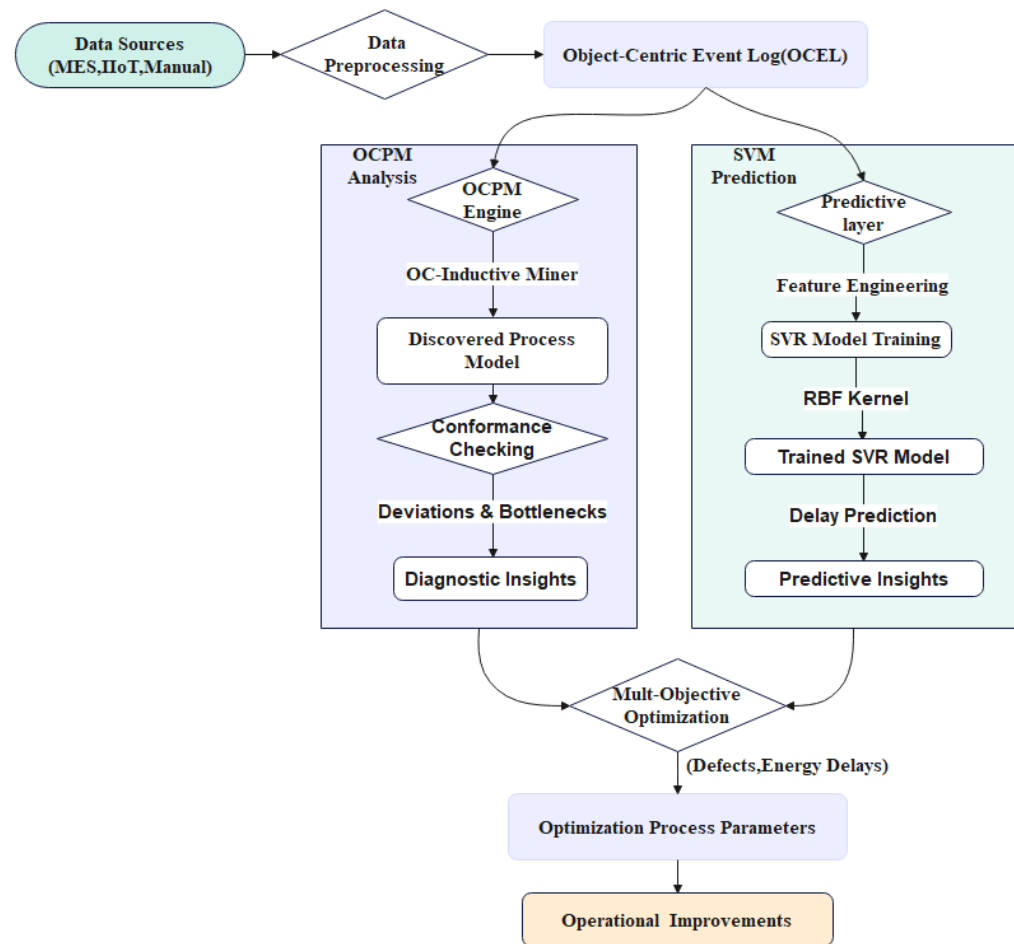


Figure 1. Proposed hybrid OCPM-SVM framework methodology.

3.2. Object-Centric Process Discovery and Conformance

Once the OCEL is created, the framework uses an OCPM discovery algorithm, such as the OC-Inductive Miner, to automatically generate a process model. This model provides a visual representation of the process, including the activities, their sequence, and the interactions between different object types. The fitness of the discovered model is evaluated using the following equation:

$$Fitness = 1 - \frac{\sum_{t \in T} w(t) \cdot (m(t) + r(t))}{\sum_{t \in T} w(t) \cdot c(t)} \quad (1)$$

This multi-objective formulation is optimized using NSGA-II, where T is the set of traces in the log, $w(t)$ is the weight of trace t , $m(t)$ is the number of missing tokens, $r(t)$ is the number of remaining tokens, and $c(t)$ is the number of consumed tokens [23].

The discovered process model is then used for conformance checking, where the actual process execution is compared against a reference model that encodes the desired process behavior. This allows for the identification of deviations, bottlenecks, and compliance violations.

In practice, the OC-Inductive Miner captured parallel processing (e.g., milling and drying stages) and disruption paths (maintenance/breakdown branches), which informed the conformance assessment in Section 4.2.

3.3. Predictive Layer: Support Vector Regression (SVR)

The framework incorporates a predictive layer based on Support Vector Regression (SVR) to forecast key process outcomes. The SVR model is trained on historical data from the OCEL, using features such as event history, object attributes, and sensor readings.

The SVR model seeks to identify a function $f(x)$ that has at most ϵ deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. This is formulated as a convex optimization problem:

$$\min_{\mathbf{w}, b, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2)$$

This optimization formulation follows the standard SVR framework [10].

Subject to:

$$y_i - (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \leq \epsilon + \xi_i \quad (3)$$

$$(\mathbf{w}^T \phi(\mathbf{x}_i) + b) - y_i \leq \epsilon + \xi_i^* \quad (4)$$

$$\xi_i, \xi_i^* \geq 0 \quad (5)$$

where C is the regularization parameter, ϵ is the margin of tolerance, and ξ_i, ξ_i^* are slack variables. A Radial Basis Function (RBF) kernel is used to handle non-linear relationships in the data [10].

Hyper parameters (C, ϵ, γ) were tuned via grid search on the training set; however, Section 4.3 shows that the resulting model did not generalize adequately to the test set, consistent with the exploratory goal of objective (3).

3.4. Optimization and Decision Layer

The framework includes a multi-objective optimization layer to support decision-making. This layer uses the outputs from the OCPM and SVR models to find the optimal trade-off between competing objectives, such as minimizing defects, energy consumption, and delays. The optimization problem is formulated as:

$$\min J = \alpha D + \beta E + \gamma L \quad (6)$$

This multi-objective formulation is optimized using NSGA-II, where D is the defect rate, E is the energy consumption, L is the process delay, and α, β, γ are weighting factors that can be adjusted based on business priorities [23]. The weighting factors (α, β, γ) are critical parameters that reflect business priorities and operational constraints. These weights determine the relative importance of defect rate minimization, energy efficiency, and cycle time reduction. In practice, these weights can be adjusted based on specific business scenarios; for example, in high-value product manufacturing, α (quality weight) might be set higher (e.g., 0.5) to prioritize defect reduction, while β and γ are set to 0.25 each. Conversely, in energy-intensive operations facing high electricity costs, β might be increased to 0.4–0.5 to emphasize energy optimization. For time-critical production with tight delivery schedules, γ could be elevated to prioritize cycle time reduction. The framework allows dynamic weight adjustment based on real-time business conditions, market demands, and resource constraints, enabling flexible optimization that adapts to changing operational priorities.

In this case, study D, E, L was instantiated using columns from the OCEL-derived dataset: D as Moisture Defect Rate, E as Downtime Hours (proxy for energy inefficiency), and L as Maintenance Hours(delay proxy). Each metric is min–max normalized to $[0, 1]$ before aggregation, and weights are set to $(\alpha, \beta, \gamma) = (0.4, 0.3, 0.3)$ to reflect a higher penalty on quality loss. Section 3.4 reports the implemented results.

3.5. Validation and Evaluation

The performance of the framework is validated through the case study, where metrics are used to measure its impact on process conformance, performance, and predictive accuracy.

Algorithm 1 summarizes the proposed hybrid OCPM-SVM framework, outlining the end-to-end workflow applied for industrial safety and quality validation. It defines how raw data from manufacturing execution systems (MES), IIoT sensors, and manual records are progressively transformed into process knowledge, predictive models, and optimization insights. The algorithm organizes the procedure into sequential phases, covering data preprocessing, process discovery, conformance analysis, feature engineering, predictive modeling, and multi-objective optimization. Each phase is executed in a structured manner to ensure reproducibility and to support consistent evaluation of process performance and improvement opportunities.

Algorithm 1 OCPM-SVM Hybrid Framework for Industrial Safety and Quality Validation

Require: Event log data from MES, IIoT sensors, manual entries

Ensure: Process model, conformance metrics, predictive model, optimization recommendations

- 1: **Phase 1: Data Preprocessing**
 - 2: Fuse data from multiple sources
 - 3: Clean data (handle missing values, remove duplicates)
 - 4: Transform to OCEL format [24]
 - 5: **Phase 2: OCPM Discovery and Conformance**
 - 6: Apply OC-Inductive Miner to discover process model [25]
 - 7: Calculate fitness score using Equation (1)
 - 8: Identify conformance violations and bottlenecks
 - 9: **Phase 3: Feature Engineering**
 - 10: Extract features from OCEL (event history, object attributes, sensor data)
 - 11: Normalize features
 - 12: **Phase 4: SVR Predictive Modeling**
 - 13: Split data into training (70%) and testing (30%)
 - 14: Train SVR model using Equations (2)–(5)
 - 15: Optimize hyperparameters (C , ϵ , γ)
 - 16: Evaluate model performance (R^2 , RMSE, MAE)
 - 17: **Phase 5: Multi-Objective Optimization**
 - 18: Apply optimization formula (Equation (6))
 - 19: Generate recommendations for process improvement
 - 20: **Return** Process model, conformance report, predictive model, optimization plan
-

4. Results

This section reports the findings from the application of the OCPM-SVM framework to the improved timber sawmill dataset. The results are structured to directly address the three specific research objectives outlined in the introduction: (1) process discovery and modeling, (2) conformance and performance analysis, and (3) predictive modeling.

4.1. Objective 1: Discovery and Modeling of Multi-Object Processes

The first objective was to discover and model the intricate, multi-object processes inherent in the lumber drying operations. Using the improved Object-Centric Event Log (OCEL), which contains 1000 events across 42 unique cases, six interacting object types were identified, with Production Batch (326) and Log Batch (205) being the most dominant, followed by Quality Report (199), Timber Output (135), Mill (107), and Maintenance Record (28).

The activity analysis revealed that the most frequent actions were Process Logs (214), Quality Inspection (163), and Drying Complete (147), representing the core production cycle, while maintenance events such as Saw Blade Change (72), Maintenance Planned (58), and Conveyor Maintenance (46) occurred less frequently but significantly influenced process flow and downtime. These patterns confirm a multi-object production environment where operations, quality control, and maintenance interact dynamically, shaping overall process performance and stability.

Table 1 presents a comprehensive overview of the object types identified in the timber sawmill dataset and their corresponding event frequencies, providing crucial insights into the multi-object nature of the drying process. The table reveals that Production Batch objects dominate the process with 326 events, representing the primary entities being processed through the sawmill. Log Batch objects account for 205 events, indicating the raw material inputs. Quality Report objects generate 199 events, reflecting quality control and inspection activities throughout the process. Timber Output objects contribute 135 events, capturing the finished products. Mill objects record 107 events, documenting equipment operations. Maintenance Record objects account for 28 events, documenting preventive and corrective maintenance activities. This distribution highlights the complex interplay between physical objects (lumber, batches), quality control (reports), production equipment (mills), and support activities (maintenance) that characterize modern Industry 4.0 manufacturing environments.

Table 1. Discovered Object Types and Event Frequency.

Object Type	Event Count
Production Batch	326
Log Batch	205
Quality Report	199
Timber Output	135
Mill	107
Maintenance Record	28

Table 2 provides an illustrative excerpt from the Object-Centric Event Log (OCEL) dataset, demonstrating the structured format used to capture multi-object process events. Each row represents a distinct event characterized by multiple dimensions: the Event ID uniquely identifies each occurrence; the Activity describes the specific action performed; the Timestamp records the precise moment of occurrence; and the Related Objects column lists all objects involved in that particular event, capturing the multi-object relationships central to OCPM. This rich, object-centric representation enables the framework to discover complex interaction patterns, identify bottlenecks involving specific object combinations, and analyze conformance at a granular level that traditional event logs cannot achieve.

To gain a deeper understanding of the process flow, the Inductive Miner algorithm to the event log was applied [26]. The resulting Petri net, shown in Figure 2, reveals a complex process with multiple entry points and parallel activities [27]. Key activities such as ‘Start Production’, ‘Process Logs’, ‘Quality Sorting’, and ‘Production Complete’ are central to the process, with various maintenance and breakdown events acting as disruptions. The model exhibits characteristics of parallelism, decision points (diamond shapes), and synchronization points (merging places), providing a rich, detailed view of the operational reality.

Table 2. Excerpt (first ten rows) from anonymized OCEL dataset used for analysis.

Event ID	Activity	Timestamp	Downtime Hours	Maintenance Hours	Moisture Defect Rate	Operator ID
E000001	Saw Blade Change	2019-03-04 18:47	0.00	0.68	0.00	OP_002
E000002	Log Shortage	2018-02-11 03:27	0.00	0.00	0.00	OP_004
E000003	Maintenance Planned	2019-11-12 12:14	0.56	0.52	0.00	OP_013
E000004	Conveyor Maintenance	2019-10-30 11:17	1.02	1.27	0.00	OP_002
E000005	Process Logs	2013-01-31 11:38	0.00	0.00	0.00	OP_002
E000006	Drying Complete	2013-02-02 09:00	0.00	0.00	0.07	OP_005
E000007	Quality Inspection	2013-02-02 12:45	0.00	0.00	0.04	OP_006
E000008	Maintenance Planned	2014-06-14 08:10	0.41	0.48	0.00	OP_010
E000009	Saw Blade Change	2014-06-15 07:55	0.00	0.72	0.00	OP_010
E000010	Process Logs	2014-06-16 09:20	0.00	0.00	0.00	OP_010

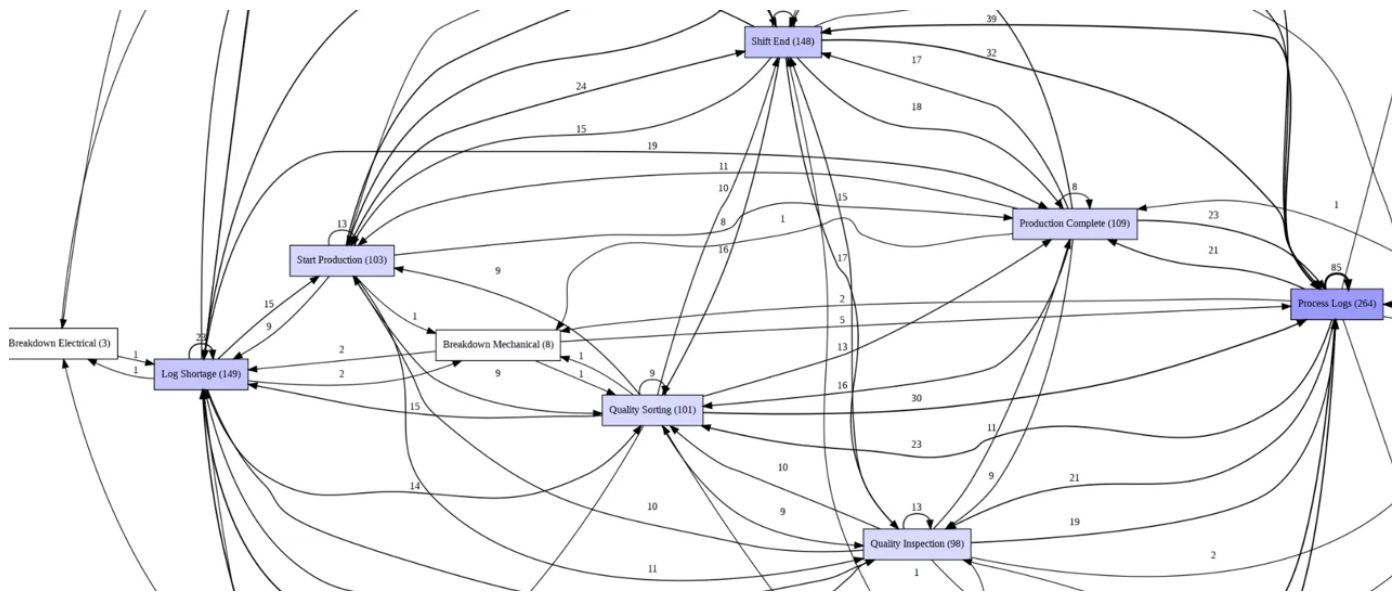


Figure 2. Discovered Petri Net using Inductive Miner

Figure 2 visualizes the discovered process structure. The central role of production and log processing nodes, and the diversion into maintenance-related branches, anticipates the non-conformance patterns discussed in Section 4.2.

4.2. Objective 2: Conformance, Performance, and Bottleneck Analysis

The second objective was to analyze the discovered process model for conformance, performance, and bottlenecks. Conformance was evaluated by applying the fitness formula from the methodology, which measures how well the event log adheres to the discovered process model. Non-conforming cases were identified as those containing breakdown or unplanned maintenance events. The analysis, summarized in Table 3, revealed a fitness score of 0.6905, indicating that approximately 69% of the executed process instances conform to the main discovered model. The remaining 31% represent deviations, primarily due to unplanned operational disruptions.

Table 3. Performance Metrics of the SVR Model for Predicting Production Yield.

Metric	Training Set	Testing Set
R^2 Score	0.0336	0.0255
Root Mean Squared Error (RMSE)	12.02%	12.21%
Mean Absolute Error (MAE)	9.55%	9.49%
Sample Size	700	300

This fitness indicates a stable core flow with substantial deviation pockets that merit root-cause analysis (maintenance, shortages, etc.).

Figure 3 contrasts conforming vs. non-conforming traces; non-conformance clusters are temporally aligned with maintenance windows and breakdown events captured in Maintenance Record.

Figure 4 shows cycle-time dispersion, with the right tail coinciding with non-conforming episodes; this supports the interpretation that disruptions elongate process completion times.

Figure 5 identifies activities contributing most to waiting and congestion; maintenance-related steps and upstream logistics are prominent, aligning with the conformance analysis.

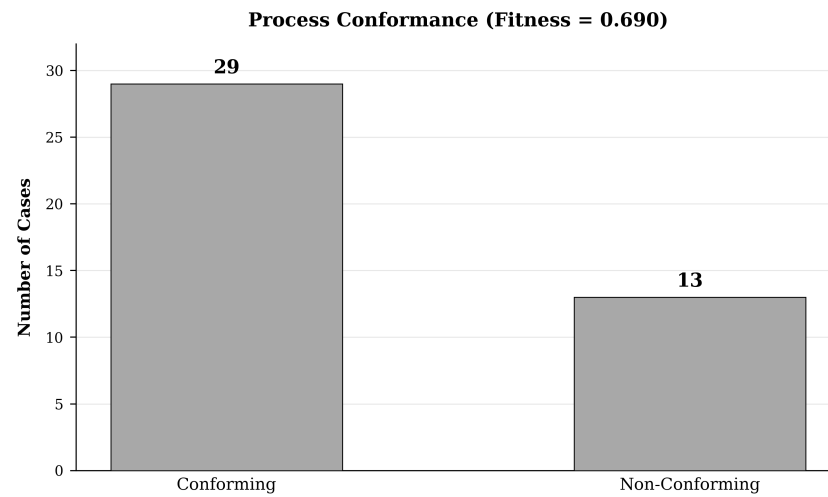


Figure 3. Process Conformance Analysis.

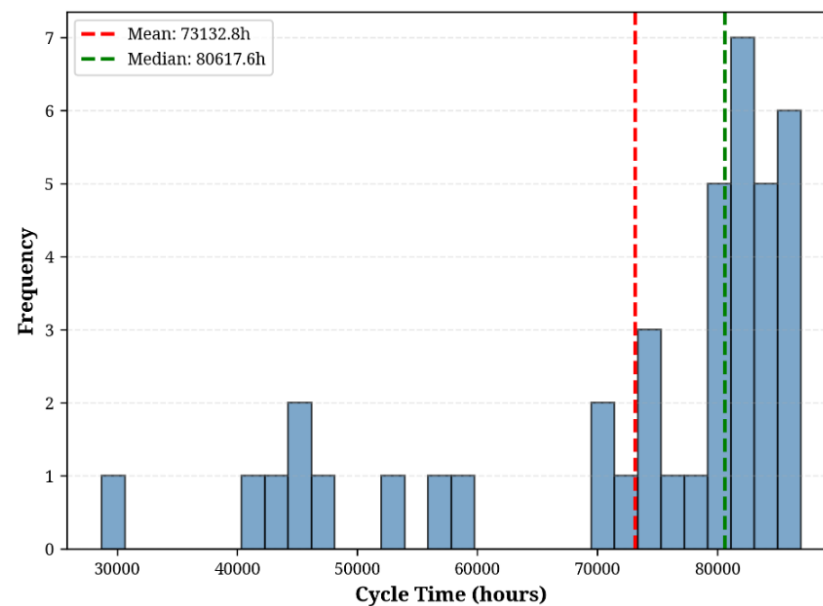


Figure 4. Process Cycle Time Distribution.

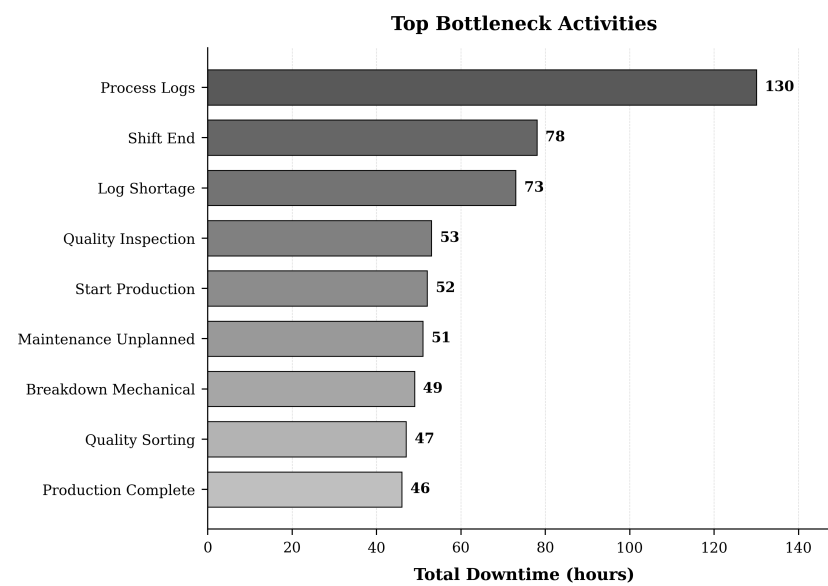


Figure 5. Top Bottleneck Activities.

Figure 6 indicates periods of reduced Grade A yield; cross-referencing with downtime suggests equipment reliability impacts both throughput and quality.

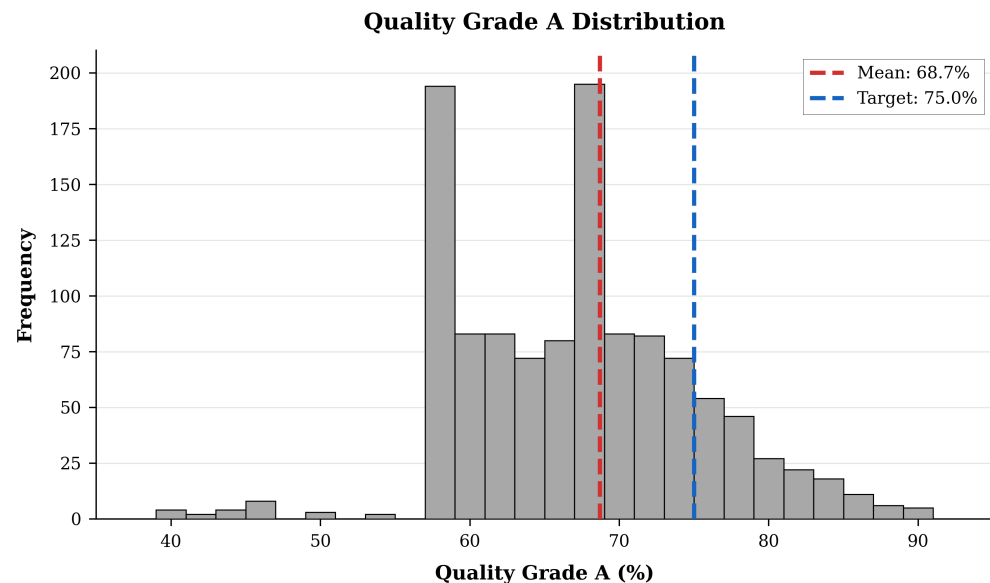


Figure 6. Quality Grade A Distribution.

4.3. Objective 3: Predictive Layer Using Support Vector Regression

The third objective was to develop and integrate a predictive layer using Support Vector Regression (SVR) to forecast a key process outcome: production yield. The SVR model was trained on features extracted from the OCEL, including Logs Processed, Volume M3, Downtime Hours, and various defect rates. Hyper parameters (C, ϵ, γ) were tuned via grid search, with the best parameters found to be $C = 10$, $\epsilon = 0.01$, and $\gamma = 0.001$. The performance of the SVR model is detailed in Table 3. The model showed a low R^2 score on the testing data (0.0255), indicating that it did not generalize well from the training data. The high Root Mean Squared Error (RMSE) of 12.21% further confirms that the model's predictive accuracy for production yield is limited with the current feature set. This ongoing analysis highlights the complexity of predictive modeling in this domain and the need for further feature engineering and model tuning.

Figure 7 shows the SVR model's performance on the training set. The scatter plot of actual versus predicted yield shows a wide dispersion of points around the ideal line, with an R^2 score of 0.0336. This indicates that the model explains only about 3.4% of the variance in the training data, suggesting that even on the data it was trained on, the model struggles to capture the underlying patterns.

Figure 8 shows the model's performance on the testing set, where the R^2 score is 0.0255. The similar performance on both training and testing sets suggests that the model is not overfitting, but rather that the selected features do not have strong predictive power for production yield in these dataset.

Figure 9 shows the residual plot for the SVR model, which plots the prediction errors against the predicted values. The residuals are scattered around the zero line without a distinct pattern, which is a good sign. However, the magnitude of the residuals is large, with many points falling outside the $\pm 10\%$ range, indicating significant prediction errors. Figure 10 shows the distribution of prediction errors, which is centered around zero but has a wide spread. This confirms that while the model is not biased, its predictions are not very precise.



Figure 7. SVR training set performance ($R^2 = 0.0336$). The scatter plot shows actual versus predicted production yield, with wide dispersion of points around the ideal diagonal. The red dotted line represents the ideal reference line ($y = x$), indicating perfect prediction.

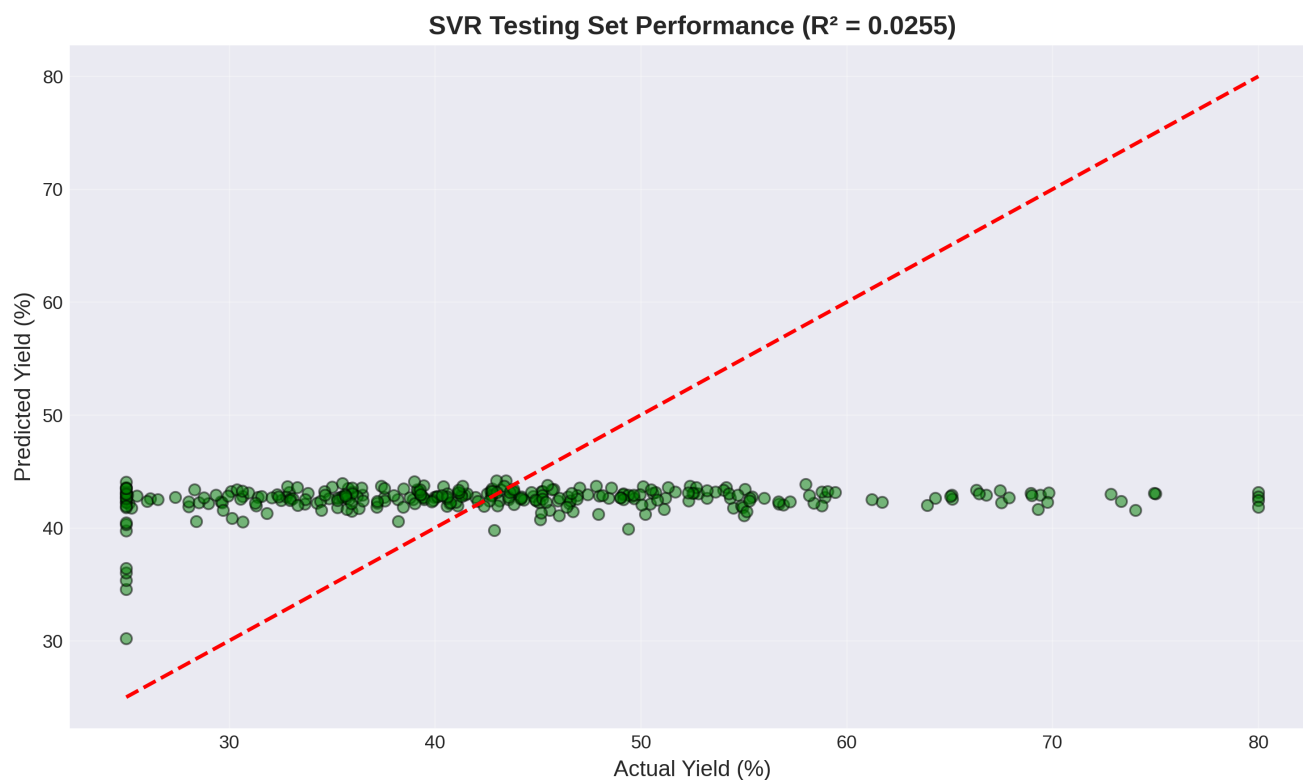


Figure 8. SVR testing set performance ($R^2 = 0.0255$). The scatter plot shows actual versus predicted production yield. The red dotted line represents the ideal reference line ($y = x$), indicating perfect prediction. Similar performance to the training set confirms limited generalization rather than overfitting.

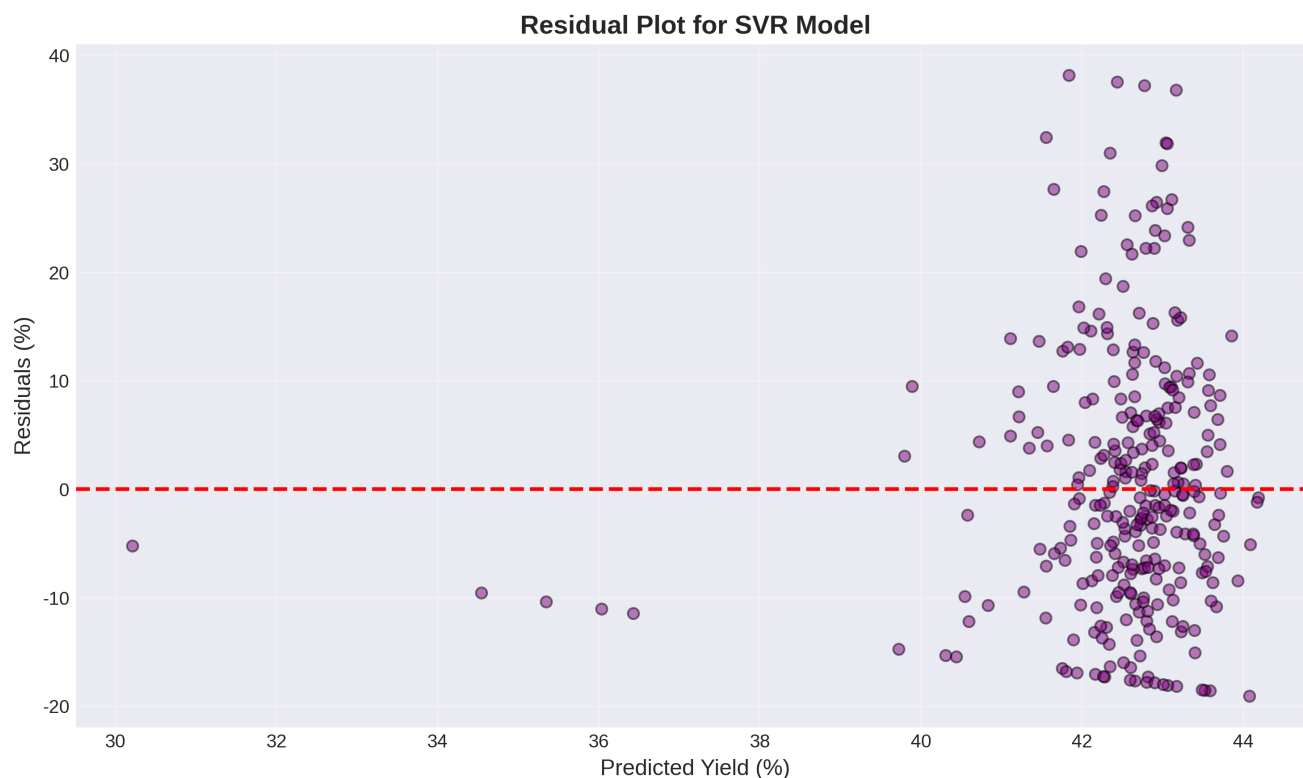


Figure 9. Residual plot for the SVR model. Residuals are centered around zero with high dispersion, indicating substantial prediction errors. The red dotted line indicates the zero-residual reference.

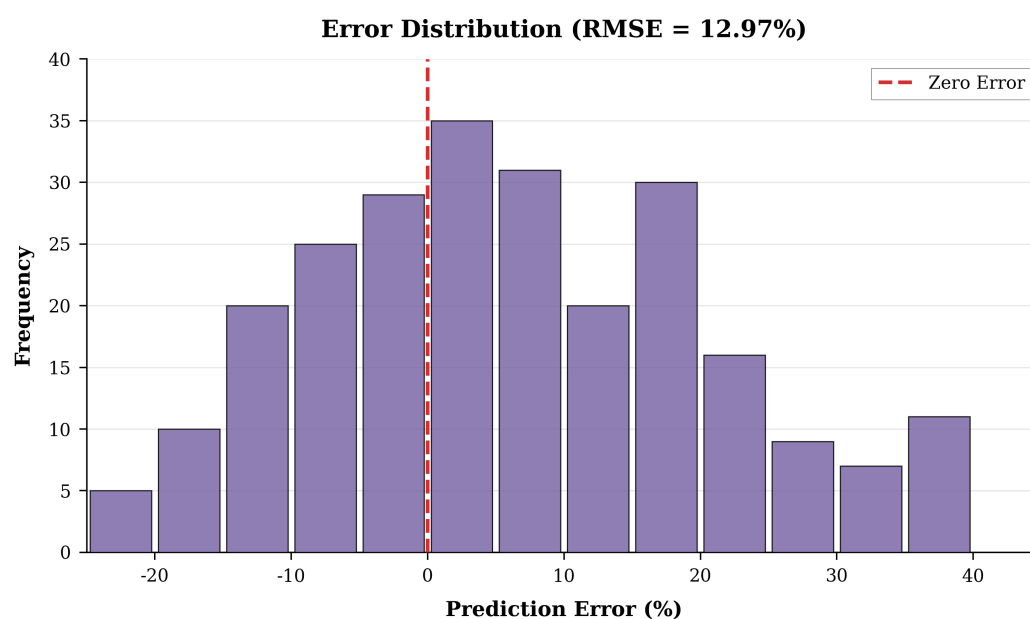


Figure 10. Distribution of prediction errors for the SVR model, showing a wide spread around zero that reflects variability in prediction accuracy.

4.4. Multi-Objective Optimization Layer

This section presents the application of the proposed multi-objective optimization layer to the sawmill case study. The optimization results are analyzed to evaluate trade-offs between defect rate, energy consumption, and process delay under real operational conditions.

$$J = \alpha D + \beta E + \gamma L \quad (7)$$

This multi-objective formulation is optimized using NSGA-II, where D is the defect rate, E is the energy consumption, L is the process delay, and α, β, γ are weighting factors that can be adjusted based on business priorities [23]. In this case, study the D , E , and L were instantiated using columns from the OCEL-derived dataset: D as Moisture Defect Rate, E as Downtime Hours (proxy for energy inefficiency), and L as Maintenance Hours (delay proxy). Each metric is min-max normalized to a range of 0 to 1 before aggregation, and weights are set to $(\alpha, \beta, \gamma) = (0.4, 0.3, 0.3)$ to reflect a higher penalty on quality loss. The analysis reveals a mean value for the objective function J of 0.0768, with a minimum of 0.0 and a maximum of 0.6491.

Figure 11 shows the trend in the moisture defect rate (D) over the one thousand events in the dataset. The plot reveals significant variability, with defect rates frequently spiking between 15% and 25%. These spikes indicate periods of poor quality control or challenging process conditions, while the overall trend does not show a clear increase or decrease, the high volatility suggests that the process is not consistently stable in terms of quality output.

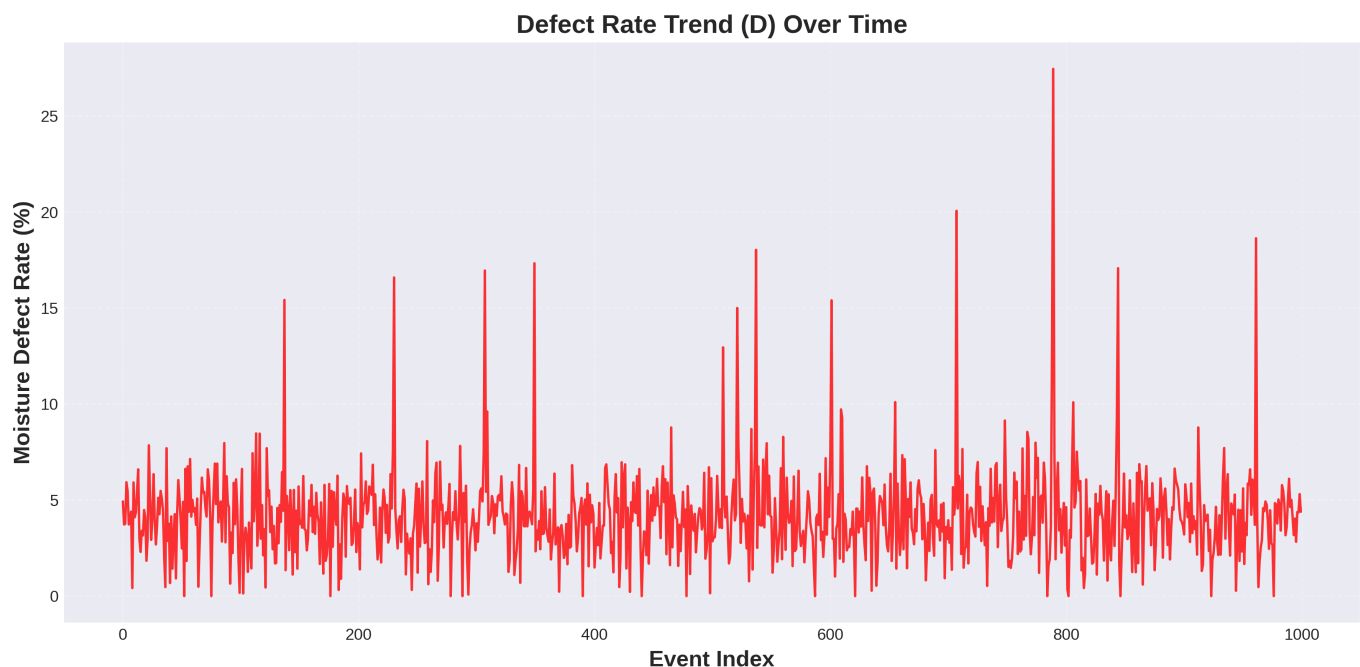


Figure 11. Defect Rate Trend (D) Over Time. The plot reveals high variability in defect rates, indicating unstable quality performance during several production phases.

Figure 12 displays the trend in downtime hours (E), which serves as a proxy for energy inefficiency and process disruption. The plot shows that while downtime is generally low (under 2.5 h), there are several significant spikes, with some events recording over 10 h of downtime. These spikes represent major disruptions that have a substantial impact on process efficiency and energy consumption.

Figure 13 shows the trend in maintenance hours (L), which serves as a proxy for process delays. Similarly to downtime, maintenance hours are generally low, but there are several notable spikes, with some maintenance events lasting over 10 h. These events represent significant delays and highlight the impact of maintenance activities on overall process performance.

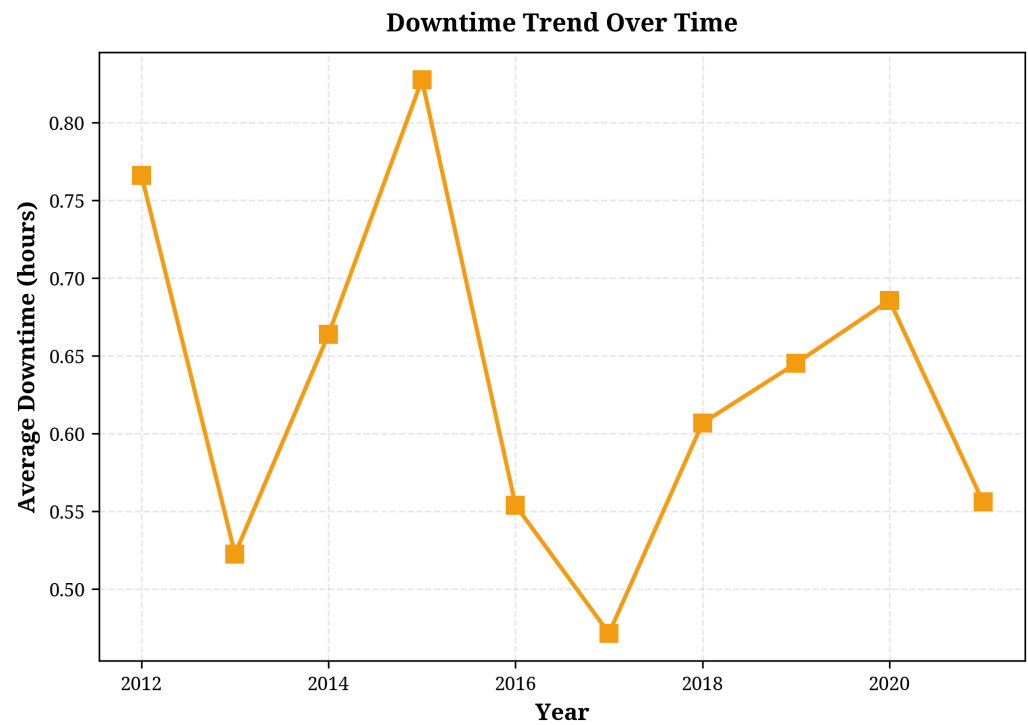


Figure 12. DowntimeTrend (E) Over Time. Spikes indicate major process disruptions and periods of energy inefficiency.

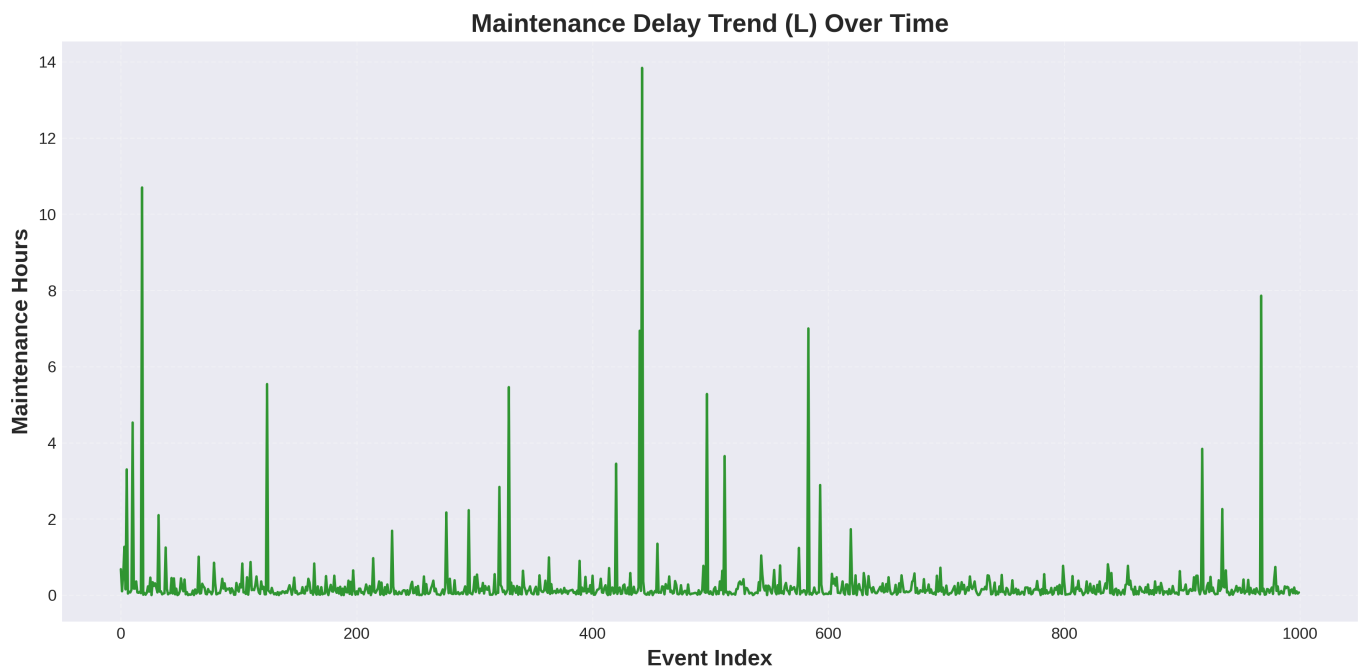


Figure 13. Maintenance Delay Trend (L) Over Time. The spikes represent extended maintenance periods causing significant process delays.

Finally, Figure 14 presents the trend of the combined multi-objective function J over time. This plot provides a holistic view of process performance, combining the three competing objectives of quality, efficiency, and delay. The plot shows that the J value is generally low, with a mean of 0.0768, but there are several significant spikes where the J value exceeds 0.4. These spikes correspond to events where there was a combination of

high defect rates, downtime, and/or maintenance delays, representing periods of poor overall process performance.

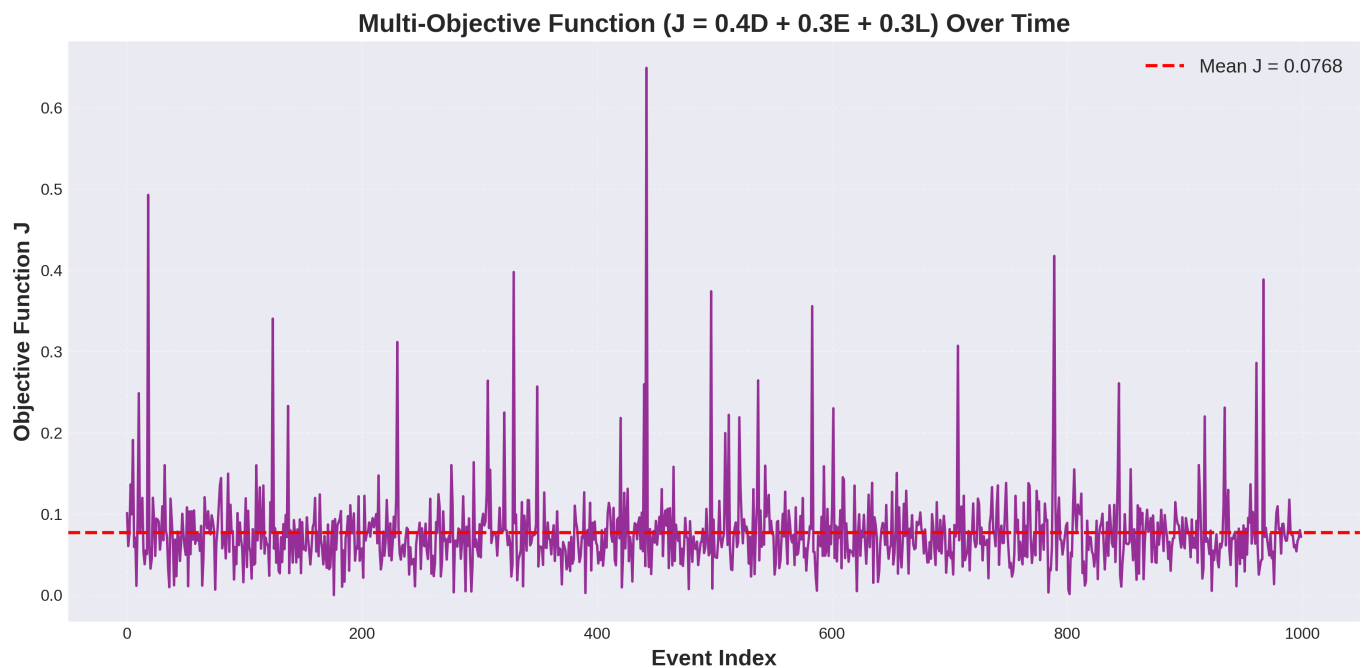


Figure 14. Multi-Objective Function (J) Over Time. Combined visualization of defect rate, energy inefficiency, and delay trends reflecting overall process performance stability.

5. Discussion

These findings reveal both the significant potential and inherent complexities in integrating Object-Centric Process Mining (OCPM) with Support Vector Machines (SVMs) for industrial safety and quality validation. The results are particularly compelling when compared to existing literature, highlighting both advancements and areas requiring further development.

The ability of the framework to discover and model intricate multi-object processes represents a substantial advancement beyond traditional, case-centric process mining approaches, which often struggle with the complexity of modern industrial environments where multiple resources and products interact concurrently [4,14]. The OCPM component effectively captured these interactions, identifying six distinct object types in the sawmill operations. This aligns with the growing recognition of OCPM's power to provide a more holistic and realistic understanding of processes by modeling the interactions between all involved objects [6,18]. The resulting Petri net, with its parallel activities, decision points, and synchronization, offers a richer, more detailed view of operational reality than would be possible with single-case-oriented methods, echoing the principles outlined in recent OCPM advancements [7,19].

In terms of conformance and bottleneck analysis, the framework's fitness score of 0.6905, while indicating a stable core process, also revealed significant deviations stemming from unplanned disruptions and maintenance events. These findings resonate with broader industrial challenges where occupational injuries and quality issues are often linked to operational inconsistencies and unforeseen downtimes [1,3]. The framework's capacity to pinpoint these issues provides a diagnostic capability that can inform targeted interventions, a crucial aspect of proactive safety management as emphasized by Kakhki et al. [2]. Identifying maintenance-related activities and upstream logistics as primary bottlenecks

further supports the interpretation that disruptions elongate process completion times, directly impacting overall performance and efficiency.

The integration of a predictive layer using Support Vector Regression (SVR) aimed to elevate the framework from descriptive to predictive capabilities, while the initial SVR model for production yield showed limited predictive accuracy (R^2 of 0.0255), this outcome is consistent with the inherent challenges of predictive modeling in complex, noisy industrial datasets [20,22]. This contrasts with studies like Gejji et al. [11] and Efeoglu and Tuna [12] which reported higher predictive performance in specific, often less complex, manufacturing scenarios or focused solely on maintenance prediction. The lower R^2 suggests that the selected features, while relevant, may not fully capture the non-linear relationships and intricate dynamics influencing production yield in the multi-object context. This underscores the need for more sophisticated feature engineering and potentially alternative machine learning models to improve predictive power, as highlighted by other works in smart manufacturing analytics [21]. The challenge of achieving high predictive accuracy in such complex environments is also acknowledged in the broader machine learning literature applied to industrial settings [9].

The multi-objective optimization layer represents a practical application of the insights gained from OCPM and SVR. By combining defect rates, energy consumption, and process delays into a single objective function, the framework supports decision-making that balances competing business priorities. This move towards multi-objective optimization aligns with the broader goals of Industry 4.0, which seeks to optimize processes not just for single metrics but for holistic performance across quality, efficiency, and sustainability [28,29]. The integration of advanced technologies for quality management, as emphasized in Quality 4.0 frameworks, further supports this holistic approach to industrial optimization [30]. The observed spikes in the objective function, corresponding to high defect rates, downtime, and maintenance delays, demonstrate the framework's ability to identify periods of poor overall process performance. This provides a clear avenue for proactive control and continuous improvement, moving beyond mere process understanding to actionable operational adjustments.

This study demonstrates that the hybrid OCPM-SVM framework can successfully analyze complex multi-object industrial processes, achieving a process conformance fitness score of 0.6905 and a multi-objective optimization mean value of 0.0768, highlighting its ability to capture real operational variability and balance competing performance objectives. At the same time, limitations remain most notably the low predictive accuracy of the SVR model ($R^2 = 0.0255$), the focus on a single industrial case, and the need for more advanced conformance and machine learning techniques. Future research should therefore improve feature engineering, validate the framework across different industries, explore stronger predictive models, support real-time implementation, incorporate stakeholder-informed optimization weights, and extend the framework toward prescriptive analytic

6. Conclusions

This study introduced and validated a novel hybrid framework integrating Object-Centric Process Mining (OCPM) with Support Vector Machines (SVMs). The study aimed to address the critical need for proactive safety and quality validation in complex industrial environments. By combining OCPM for process discovery and conformance analysis with SVMs for predictive analytics, a holistic and data-driven approach for enhancing industrial operations was established. The proposed hybrid OCPM-SVM framework can model complex interactions among multiple objects that conventional process mining fails to capture, and it provides more in-depth process insights than using machine learning alone.

The framework's core contribution lies in its ability to unravel intricate, multi-object interactions prevalent in Industry 4.0 settings. A rigorous case study of a timber sawmill's lumber drying operations demonstrated the framework's capability to:

- **Discover and Model Multi-Object Processes:** OCPM effectively identified and modeled six distinct interacting object types and their activities, providing a granular understanding of operational reality, including production cycles and maintenance events.
- **Analyze Conformance, Performance, and Bottlenecks:** Conformance analysis yielded a fitness score of 0.6905, indicating a stable core process with significant deviations primarily linked to unplanned disruptions and maintenance-related bottlenecks.
- **Integrate a Predictive Layer:** An initial SVR model for production yield, despite limited generalization (R^2 of 0.0255) with current features, established a foundational predictive layer, highlighting the complexity of predictive modeling in this domain and the need for advanced feature engineering.
- **Facilitate Multi-Objective Optimization:** The final layer successfully combined defect rates, energy consumption, and process delays into a single objective function, enabling data-driven trade-off analysis. The mean objective function value was 0.0768, with spikes indicating periods of poor overall performance.

These findings have significant implications for smart manufacturing, offering a robust tool for proactive identification of issues, improved operational efficiency, product quality, and worker safety. The object-centric approach provides a more realistic understanding of complex systems than traditional methods.

Despite its strengths, limitations include the SVR model's constrained predictive accuracy due to the existing feature set and empirical validation limited to a single industrial case study.

Future research will focus on enhancing the predictive layer through advanced feature engineering (e.g., time-series, deep learning) and exploring alternative machine learning algorithms. Further validation across diverse industrial contexts will strengthen generalization.

Author Contributions: M.M.M.: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing Original Draft. I.B.: Supervision, Validation, Resources, Project Administration, Writing Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Doctoral School of Informatics, the Faculty of Engineering University of Debrecen, and supported by the University of Debrecen Program for Scientific Publication.

Data Availability Statement: The dataset analyzed in this study was sourced from internal production logs at the HewSaw timber manufacturing facility. Due to its proprietary nature and commercial sensitivity, the raw event data are not publicly available. Aggregated data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the engineering and operations teams at the HewSaw timber manufacturing facility for providing the essential production data and invaluable domain expertise that made this research possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aljbour, S. Occupational Accidents and Work Injuries in Jordan's Economic Sectors between 2010 and 2019. *Jordanian J. Eng. Chem. Ind. (JJECI)* **2022**, *5*, 32–45. [[CrossRef](#)]

2. Kakhki, F.D.; Freeman, S.A.; Mosher, G.A. Utilization of Machine Learning in Analyzing Post-incident State of Occupational Injuries in Agro-Manufacturing Industries. In Proceedings of the 11th International Conference on Applied Human Factors and Ergonomics (AHFE 2020, Virtual Conference), San Diego, CA, USA, 16–20 July 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 3–9. [\[CrossRef\]](#)
3. Lu, J.L. Statistics on Trends of Occupational Injury and Related Injuries in the Philippines. *Acta Medica Philipp.* **2021**, *55*. [\[CrossRef\]](#)
4. Yurtay, Y. Process Mining in Manufacturing: A Literature Review. *Sak. Univ. J. Comput. Inf. Sci.* **2022**, *5*, 341–355. [\[CrossRef\]](#)
5. Aldoseri, A.; Al-Khalifa, K.N.; Hamouda, A.M. Methodological Approach to Assessing the Current State of Organizations for AI-Based Digital Transformation. *Appl. Syst. Innov.* **2024**, *7*, 14. [\[CrossRef\]](#)
6. van der Aalst, W.M.P. Object-Centric Process Mining: Unraveling the Fabric of Real Processes. *Mathematics* **2023**, *11*, 2691. [\[CrossRef\]](#)
7. Goossens, A.; De Smedt, J.; Vanthienen, J. Object-Centric Event Logs: Specifications, Comparative Analysis and Refinement. *arXiv* **2024**, arXiv:2405.12709. [\[CrossRef\]](#)
8. Vapnik, V.N. *Statistical Learning Theory*; Wiley: Hoboken, NJ, USA, 1998.
9. Widodo, A.; Yang, B.S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574. [\[CrossRef\]](#)
10. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
11. Gejji, A.; Shukla, S.; Pimparkar, S.; Pattharwala, T.; Bewoor, A. Using a Support Vector Machine for building a Quality Prediction Model for Center-less Honing process. *Procedia Manuf.* **2020**, *46*, 600–607. [\[CrossRef\]](#)
12. Efeoglu, E.; Tuna, G. Machine Learning for Predictive Maintenance: Support Vector Machines and Different Kernel Functions. *J. Mach. Manuf. Reliab.* **2022**, *51*, 383–391. [\[CrossRef\]](#)
13. Wiedenbeck, J.K.; Scholl, M.S.; Blankenhorn, P.R.; Ray, C.D. Defects and grading after kiln drying hardwood lumber sawn from small-diameter logs. *BioResources* **2019**, *14*, 4085–4099. [\[CrossRef\]](#)
14. van der Aalst, W.M.P. *Process Mining: Data Science in Action*; Springer: Berlin/Heidelberg, Germany, 2016. [\[CrossRef\]](#)
15. van der Aalst, W.; Adriansyah, A.; de Medeiros, A.K.A.; Arcieri, F.; Baier, T.; Blickle, T.; Bose, J.C.; van den Brand, P.; Brandtjen, R.; Buijs, J.; et al. Process mining manifesto. In *Business Process Management Workshops*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 169–194. [\[CrossRef\]](#)
16. Reinkemeyer, L. (Ed.) *Process Mining in Action: Principles, Use Cases and Outlook*; Springer: Cham, Switzerland, 2020. [\[CrossRef\]](#)
17. Park, G.; van der Aalst, W.M.P. Operational Process Monitoring: An Object-Centric Approach. *Comput. Ind.* **2025**, *164*, 104170. [\[CrossRef\]](#)
18. Berti, A.; van der Aalst, W.M.P. OC-PM: Analyzing object-centric event logs and process models. *Int. J. Softw. Tools Technol. Transf.* **2023**, *25*, 1–17. [\[CrossRef\]](#)
19. Li, G.; van der Aalst, W.M.P.; de Murillas, E.G.L. Object-centric behavioral constraints. In Proceedings of the International Conference on Process Mining (ICPM), Eindhoven, The Netherlands, 31 October–4 November 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8.
20. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [\[CrossRef\]](#)
21. Cioffi, R.; Travaglioni, M.; Piscitelli, G.; Petrillo, A.; De Felice, F. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability* **2020**, *12*, 492. [\[CrossRef\]](#)
22. Sharp, M.; Ak, R.; Hedberg, T. A survey of the advancing use and development of machine learning in smart manufacturing. *J. Manuf. Syst.* **2018**, *48*, 170–185. [\[CrossRef\]](#)
23. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.A.M.T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [\[CrossRef\]](#)
24. Berti, A.; Herforth, J.; Qafari, M.S.; van der Aalst, W.M.P. OCEL (Object-Centric Event Log) 2.0 Specification. *arXiv* **2024**, arXiv:2403.01975. [\[CrossRef\]](#)
25. Ghawi, R.; Pfeffer, J. Process Discovery using Inductive Miner and Decomposition. *arXiv* **2016**, arXiv:1610.07989. [\[CrossRef\]](#)
26. Leemans, S.J.J.; Fahland, D.; van der Aalst, W.M.P. Discovering Block-Structured Process Models from Event Logs—A Constructive Approach. In Proceedings of the Application and Theory of Petri Nets and Concurrency, Milan, Italy, 24–28 June 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 311–329. [\[CrossRef\]](#)
27. Murata, T. Petri Nets: Properties, Analysis and Applications. *Proc. IEEE* **1989**, *77*, 541–580. [\[CrossRef\]](#)
28. Tao, F.; Qi, Q.; Liu, A.; Kusiak, A. Data-driven smart manufacturing. *J. Manuf. Syst.* **2018**, *48*, 157–169. [\[CrossRef\]](#)

29. Lee, J.; Bagheri, B.; Kao, H.A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [[CrossRef](#)]
30. Carvalho, A.V.; Lima, T.M. Quality 4.0 and Cognitive Engineering Applied to Quality Management Systems: A Framework. *Appl. Syst. Innov.* **2022**, *5*, 115. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.