Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Closed-Loop Data-Driven Framework for Value Stream Mapping and Future-State Design

Edwin Montes Orozco

edwin.montes@udlap.mx

Universidad de las Américas Puebla: Universidad de las Americas Puebla https://orcid.org/0000-0001-8594-7006

Mariam Dopslaf

FH Bielefeld: Hochschule Bielefeld

Roman Anselmo Mora-Gutiérrez

Universidad Autónoma Metropolitana: Universidad Autonoma Metropolitana

Sergio Gerardo de-los-Cobos-Silva

Universidad Autónoma Metropolitana: Universidad Autonoma Metropolitana

Eric Alfredo Rincón-García

Universidad Autónoma Metropolitana: Universidad Autonoma Metropolitana

Miguel Ángel Gutiérrez-Andrade

Universidad Autónoma Metropolitana: Universidad Autonoma Metropolitana

Pedro Lara-Velázquez

Universidad Autónoma Metropolitana: Universidad Autonoma Metropolitana

Research Article

Keywords: Value Stream Mapping, Value Stream Design, Process Mining, Digital Twin, Explainable Machine Learning, Complex Network Analysis

Posted Date: September 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10924132/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

A Closed-Loop Data-Driven Framework for Value Stream Mapping and Future-State Design

Edwin Montes-Orozco[1*], Mariam Dopslaf[2], Roman Anselmo Mora-Gutiérrez[3], Sergio Gerardo de-los-Cobos-Silva[4], Eric Alfredo Rincón-García[4], Pedro Lara-Velázquez[4], Miguel Ángel Gutiérrez-Andrade[4]

[1*] Department of Computing, Electronics and Mechatronics, Universidad de las Américas Puebla, Puebla, Puebla, Mexico .

[2] Faculty of Engineering and Mathematics, Hochschule Bielefeld, Bielefeld, Germany .

[3] Department of Systems, Universidad Autónoma Metropolitana Azcapotzalco, Mexico City, Mexico .

[4] Department of Electrical Engineering, Universidad Autónoma Metropolitana Iztapalapa, Mexico City, Mexico .

*Corresponding author(s). E-mail(s): edwin.montes@udlap.mx;

Abstract

Value Stream Mapping (VSM) is widely used in Lean Manufacturing to identify waste and support continuous improvement, but conventional VSM relies on manual data collection, static process representations, and expert interpretation. This paper proposes a data-driven VSM/VSD decision-support framework that establishes a traceable path from operational event data to current-state reconstruction, process diagnosis, simulation-based evaluation, and future-state redesign. Process Mining reconstructs observed process behavior and supports quantitative current-state VSM generation; Complex Network Analysis characterizes structural dependencies; Explainable Machine Learning integrates operational and network-derived information to identify and interpret process criticalities; and Digital Twin simulation evaluates candidate improvement interventions before they are translated into Future-State Value Stream Designs. The framework was evaluated using a controlled synthetic manufacturing event log and three heterogeneous real-world event logs under a leakage-controlled temporal protocol. Network-derived features provided complementary predictive information beyond conventional operational variables, with the strongest gains observed for rework prediction: Random Forest F1-score increased from 0.2243 to 1.0000 for BPI2017 and from 0.4658 to 0.7451 for BPI2019. In the cumulative simulation-based evaluation, the complete framework reduced Lead Time by 32.40%, Work-In-Process by 45.09%, Defect Rate by 27.63%, and Rework Rate by 45.52%, while increasing Throughput by 71.73% relative to the baseline. The results support a reproducible transition from descriptive VSM toward evidence-based, quantitatively evaluated Future-State Value Stream Design.

Keywords: Value Stream Mapping, Value Stream Design, Process Mining, Digital Twin, Explainable Machine Learning, Complex Network Analysis

1 Introduction

The increasing digitalization of industrial operations has substantially expanded the availability of process execution data generated by Enterprise Resource Planning (ERP) systems, Manufacturing Execution Systems (MES), Supervisory Control and Data Acquisition (SCADA) platforms, Industrial Internet of Things (IIoT) devices, and cyber-physical production systems. This growing data availability creates new opportunities for continuous improvement methodologies that have traditionally relied on manual observation and expert-driven process analysis. In particular, Lean Manufacturing methods can increasingly be supported by computational approaches capable of reconstructing actual process behavior, identifying operational inefficiencies, and evaluating improvement alternatives from recorded operational data.

Value Stream Mapping (VSM) is one of the most established Lean methods for representing material and information flows, identifying waste, and supporting the transition from a current-state process toward an improved future-state configuration. Its main strength lies in providing an end-to-end representation of the value stream that connects process activities with operational indicators such as cycle time, waiting time, inventory, and information flows. However, conventional VSM is commonly constructed through shop-floor observations, interviews, time studies, and expert interpretation. Consequently, the resulting maps generally provide static representations of production systems whose actual behavior may exhibit substantial variability, rework, resource interactions, alternative execution paths, and time-dependent bottlenecks. These limitations become increasingly relevant in data-rich manufacturing environments, where large volumes of execution data are available but are not systematically incorporated into conventional value-stream analysis.

The digitalization of VSM has therefore emerged as an important research direction. Process Mining (PM), in particular, provides mechanisms for reconstructing actual process behavior from event logs and can reduce the dependence of value-stream analysis on manually collected observations. Recent research has demonstrated growing interest in combining Process Mining and VSM for production analysis, including automated or semi-automated generation of value-stream information from operational data [18]. Nevertheless, recent systematic evidence indicates that this research stream remains relatively young and that many existing approaches are conceptual, case-specific, or provide only partial integration between process-mining outputs and the information required for end-to-end value-stream analysis [18]. Thus, an important challenge is no longer merely whether VSM can be digitalized, but how operational event data can be transformed into actionable evidence for systematic process redesign.

Related developments further extend the analytical capabilities available for data-driven process improvement. Machine Learning (ML) can identify nonlinear relationships between operational conditions and undesirable outcomes such as delays, defects, rework, and bottlenecks, while Explainable Artificial Intelligence (XAI) can reveal the variables that drive those predictions. Graph-based and Complex Network representations provide a complementary structural perspective by quantifying activity and resource dependencies that may not be evident from conventional performance indicators alone. In parallel, Digital Twins (DTs) and simulation-based models provide mechanisms for evaluating alternative production configurations before physical implementation. Recent studies have already demonstrated the potential of integrating Digital Twins with VSM to support dynamic monitoring, simulation, and improvement of manufacturing processes [16]. These developments indicate a transition from static value-stream visualization toward computationally supported value-stream analysis.

Despite this progress, a methodological gap remains between *data-driven current-state analysis* and *evidence-based future-state design*. Existing approaches provide important capabilities for process discovery, digital VSM, predictive analysis, structural modeling, or simulation-based improvement, but these capabilities do not by themselves establish a closed-loop methodology that systematically transforms operational event data into a current-state representation, diagnoses the operational and structural mechanisms associated with poor performance, evaluates candidate interventions, and translates the selected configuration into a quantitatively supported future-state

Value Stream Design (VSD). Closing this loop is particularly relevant because identifying where a process performs poorly is analytically different from determining which intervention should be implemented and estimating its system-level consequences before deployment.

Accordingly, this study addresses the following research question:

> *How can operational event data be systematically transformed into evidence-based current-state diagnosis and quantitatively evaluated future-state Value Stream Designs?*

To address this question, this paper proposes a data-driven VSM/VSD decision-support framework that connects Process Mining, Complex Network Analysis, Explainable Machine Learning, and Digital Twin simulation within a closed-loop improvement methodology. Operational event logs are first used to reconstruct actual process behavior and automatically derive quantitative current-state Value Stream Maps. Process and network-derived indicators are then integrated into predictive models to identify operational criticalities and explain the factors associated with undesirable process outcomes. These diagnostic results guide the definition of candidate improvement interventions, which are evaluated through Digital Twin simulation. The resulting performance evidence is subsequently used to select and represent an improved future-state Value Stream Design.

The main contributions of this work are threefold:

1. **A closed-loop data-driven VSM/VSD methodology** that systematically connects operational event data, current-state reconstruction, analytical diagnosis, intervention evaluation, and future-state Value Stream Design, extending digital VSM beyond process visualization toward evidence-based process redesign.
2. **A graph-enhanced and explainable diagnostic mechanism** that combines operational indicators derived from Process Mining and VSM with network-based structural features and Explainable Machine Learning. This integration enables the identification and interpretation of bottlenecks, delays, defects, rework patterns, and structurally critical process elements that may not be captured by conventional value-stream indicators alone.
3. **A simulation-based transition from diagnosis to future-state design** in which analytically identified process criticalities are translated into candidate interventions and evaluated through a calibrated Digital Twin. This provides a quantitative mechanism for comparing alternative configurations before generating the corresponding future-state VSM/VSD.

The proposed framework is evaluated using one controlled synthetic manufacturing event log and three real-world event logs with heterogeneous process characteristics. The experimental design combines ablation analysis, predictive evaluation, structural network analysis, explainability mechanisms, and repeated Digital Twin simulation to assess the incremental and cumulative performance associated with progressively incorporating the analytical capabilities of the framework, as well as their combined ability to support data-driven process redesign.

The remainder of this paper is organized as follows. Section 2 reviews the evolution from conventional and digital VSM toward Process Mining-supported value-stream analysis, predictive and structural process diagnosis, and simulation-assisted future-state design. Section 3 presents the proposed methodology and its analytical components. Section 4 describes the experimental design and validation protocol. Section 5 presents the experimental results. Section 6 discusses the findings, implications, and limitations of the proposed approach. Finally, Section 7 summarizes the main conclusions and outlines directions for future research.

2 Related Work

The digitalization of production systems has progressively transformed Value Stream Mapping (VSM) from a predominantly manual Lean technique into a data-supported approach for analyzing increasingly complex manufacturing processes. Rather than considering Process Mining (PM), Machine Learning (ML), Digital Twins (DTs), and graph-based analysis as independent technological streams, this section reviews how these capabilities contribute to the broader transition from conventional VSM toward data-driven current-state analysis and evidence-based

future-state Value Stream Design (VSD). Particular attention is given to recent approaches that directly combine VSM with Process Mining, predictive analytics, and simulation-based process improvement.

2.1 From Conventional to Data-Driven Value Stream Mapping

Value Stream Mapping is an established Lean Manufacturing methodology for representing material and information flows, identifying waste, and supporting the transition from current-state processes toward improved future-state configurations [17]. Its visual and end-to-end perspective has made VSM particularly useful for communicating process inefficiencies and supporting continuous improvement initiatives.

Nevertheless, conventional VSM relies substantially on shop-floor observations, interviews, time studies, and expert interpretation. The resulting map generally represents a static snapshot of a production system and may therefore overlook execution variability, alternative process paths, rework, dynamic queues, and time-dependent resource interactions. These limitations become increasingly important as manufacturing environments become more complex and operational data become continuously available through ERP, MES, IIoT, and cyber-physical production infrastructures.

Digitalized VSM has emerged in response to these limitations. Horsthofer-Rauch et al. [10] reviewed approaches for digitalizing value-stream analysis and highlighted Process Mining as a particularly promising mechanism for exploiting production data while reducing the manual effort required by conventional VSM. This transition is conceptually important because digital VSM does not merely replace paper-based visualization: it creates the possibility of continuously reconstructing value-stream behavior from execution data and connecting Lean analysis with computational decision-support techniques.

More recently, Teriete et al. [18] systematically reviewed the combination of VSM and Process Mining in production. Their review identified 21 relevant studies and confirmed that this is an emerging but increasingly established research stream. At the same time, the review showed that many existing contributions remain conceptual or case-specific and that systematic tool support and deeper methodological integration are still limited. Important open challenges include the aggregation of event-level information toward the abstraction level required by VSM, the representation of multiple process perspectives, end-to-end value-stream reconstruction, and the formal connection between value-stream semantics and operational event data [18].

2.2 Process Mining for Data-Driven Value-Stream Analysis

Process Mining bridges process management and data science by extracting process knowledge from event logs [1]. Process discovery techniques reconstruct actual execution paths, while performance analysis can identify process variants, waiting-time patterns, bottlenecks, rework loops, and resource-related inefficiencies. These capabilities complement the top-down and aggregated perspective of conventional VSM with a bottom-up representation of observed process executions.

The increasing availability of production event data has consequently motivated more direct integration between Process Mining and value-stream analysis. Horsthofer-Rauch et al. [11] proposed a sustainability-integrated VSM approach supported by Process Mining, illustrating how event data can reduce manual analysis effort and extend value-stream assessment beyond conventional operational indicators. Their work demonstrates the potential of PM-supported VSM while also emphasizing the importance of appropriately connecting event-level process information with the abstractions required for value-stream analysis.

Khakpour et al. [12] introduced the concept of Lean Process Mining (LPM), explicitly connecting process mining with Lean improvement. Their methodology comprises an as-realized process state, improvement strategies, and a reengineered process state, so moving beyond process discovery toward improvement-oriented analysis. This contribution is particularly relevant because it demonstrates that PM can support not only the reconstruction of existing processes but also the identification and prevention of waste.

At a broader production-network level, Kroeger et al. [13] proposed a multidimensional

data model for process-mining-based value-stream planning. Their approach addresses the need to represent multiple dimensions such as material flow, information flow, and inventory, and combines data-driven VSM with scenario-oriented production network planning. This work further demonstrates the progression from isolated process discovery toward richer value-stream representations.

These studies establish an important foundation for data-driven VSM. Nevertheless, PM-supported approaches primarily emphasize process reconstruction, waste identification, sustainability assessment, or value-stream planning. A remaining challenge is to systematically connect these current-state analytical capabilities with predictive diagnosis and quantitatively evaluated future-state redesign.

2.3 Predictive and Explainable Analytics for Process Diagnosis

While Process Mining describes how processes have been executed, predictive analytics can estimate the likelihood of undesirable future outcomes. Machine Learning methods have been extensively applied in industrial environments for quality prediction, anomaly detection, predictive maintenance, delay prediction, and process optimization. Tree-based models such as Random Forest and XGBoost are particularly suitable for heterogeneous operational data and can capture nonlinear interactions among process variables [4, 6].

For continuous improvement applications, however, predictive performance alone is insufficient. Process engineers must also understand which operational conditions contribute to predicted bottlenecks, delays, defects, or rework. Explainable Artificial Intelligence (XAI) methods such as SHAP provide a mechanism for decomposing model predictions into feature-level contributions, thereby supporting the interpretation of complex predictive models.

Within a value-stream context, this distinction is important. Conventional VSM identifies visible waste and performance losses, whereas predictive and explainable models can provide evidence about the variables associated with those losses. Consequently, the role of ML/XAI in data-driven VSM is not simply to increase prediction accuracy, but to provide an analytical bridge between observed process behavior and potential improvement interventions.

Nevertheless, the integration of explainable predictive analytics into PM-supported VSM/VSD remains comparatively underdeveloped. Existing data-driven VSM approaches predominantly focus on reconstructing process behavior and computing performance indicators, leaving an opportunity to incorporate predictive and interpretable diagnostic mechanisms directly into the value-stream improvement cycle.

2.4 Structural Process Analysis through Complex Networks

Event logs also contain structural information regarding how activities, resources, and process stages interact. Graph-based representations provide a natural mechanism for modeling these relationships. Complex Network Analysis (CNA) offers quantitative indicators such as degree, betweenness, closeness, eigenvector centrality, and network density, which can characterize connectivity, intermediary roles, structural influence, and criticality [8, 15].

Graph-based representations have increasingly been incorporated into manufacturing and Digital Twin research. Mortlock et al. [14] demonstrated the potential of graph-learning approaches for capturing dependencies among manufacturing entities, while Braun et al. [3] explored graph-based representations for Digital Twin development and explicit modeling of manufacturing relationships.

From the perspective of VSM/VSD, network analysis provides information that is complementary to conventional Lean indicators. An activity with moderate cycle time, for example, may nevertheless occupy a structurally critical position if it connects multiple process paths or resources. Combining operational and network-derived indicators therefore offers the possibility of distinguishing local performance problems from systemically influential process elements.

Despite this potential, structural network indicators have received comparatively limited attention within data-driven VSM. In particular, their use as predictive features and as evidence for prioritizing value-stream interventions remains insufficiently explored.

2.5 Digital Twins and Simulation-Assisted Future-State Design

The transition from process diagnosis to future-state design requires a mechanism for estimating the consequences of candidate interventions before physical implementation. Simulation has long complemented Lean Manufacturing by allowing alternative production configurations to be evaluated under controlled conditions. Digital Twins extend this capability by connecting virtual system representations with operational data and enabling more dynamic monitoring, experimentation, and decision support [9].

Recent research has begun to connect Digital Twins directly with VSM. Reslan et al. [16] demonstrated the integration of a Digital Twin with VSM in a manufacturing cell, using machine data to support real-time visualization and the identification of improvements aimed at reducing non-value-added time. Their study illustrates the potential of Digital Twins to transform VSM from a static representation into a more dynamic decision-support mechanism.

Other research has integrated Process Mining, Artificial Intelligence, and Digital Twins for industrial monitoring and optimization [2, 5, 7]. Collectively, these developments demonstrate that simulation and Digital Twin technologies can provide the experimental layer required to test improvement alternatives before implementation.

However, simulation capability alone does not establish a systematic VSD methodology. A key unresolved issue is how candidate interventions should be derived from current-state evidence and how simulation outcomes should be translated back into an explicit future-state value-stream representation. This diagnosis–intervention–evaluation–redesign connection constitutes a central requirement for closed-loop data-driven VSD.

2.6 Synthesis of the Literature and Research Gap

Table 1 summarizes representative recent approaches that are directly related to the methodological capabilities considered in this study. Unlike comparisons based primarily on foundational works or general-purpose analytical tools, the table focuses on contributions that explicitly advance digital or data-driven value-stream analysis, PM-supported Lean improvement, or simulation-assisted VSM/VSD.

The literature reveals a clear progression from conventional VSM toward digitalized value-stream analysis. Process Mining has enabled increasingly data-driven reconstruction of current-state processes [11, 18]; Lean Process Mining has extended this perspective toward improvement-oriented analysis [12]; multidimensional approaches have begun to connect process data with value-stream planning [13]; and Digital Twins have introduced simulation and dynamic evaluation capabilities into VSM [16].

Therefore, the research gap is not the absence of integration between VSM and digital technologies. Rather, the remaining gap concerns the *closed-loop integration of current-state reconstruction, interpretable diagnosis, structural analysis, intervention evaluation, and explicit future-state Value Stream Design*. Existing approaches address important subsets of this cycle, but limited attention has been given to a reproducible methodology that systematically connects all five stages using operational event data.

This study addresses this gap through a framework in which Process Mining reconstructs the observed process and supports automatic current-state VSM generation; Complex Network Analysis characterizes structural dependencies; Explainable Machine Learning identifies and interprets operational criticalities; Digital Twin simulation evaluates candidate interventions; and the selected configuration is translated into a quantitative future-state VSM/VSD. The resulting architecture therefore focuses not on technological integration as an end in itself, but on establishing a traceable analytical path from operational evidence to process redesign.

3 Materials and Methods

This section presents the proposed data-driven framework for Value Stream Mapping (VSM) and Value Stream Design (VSD) integrating Process Mining (PM), Explainable Artificial Intelligence (XAI), Digital Twins (DT), and Complex Network Analysis (CNA). The methodology was designed to transform event-level operational data into

Table 1 Comparison of representative approaches directly related to data-driven Value Stream Mapping and Value Stream Design.

Study	VSM	PM	Predictive	XAI	Structural	Simulation/DT	Future-state design
Horsthofer-Rauch et al. [10]	✓	✓	×	×	×	×	△
Horsthofer-Rauch et al. [11]	✓	✓	×	×	×	×	△
Khakpour et al. [12]	✓	✓	✓	×	×	×	✓
Kroeger et al. [13]	✓	✓	×	×	×	✓	✓
Reslan et al. [16]	✓	×	△	×	×	✓	✓
Teriete et al. [18]	✓	✓	×	×	×	×	△
Proposed framework	✓	✓	✓	✓	✓	✓	✓

Note: ✓ = explicitly addressed; △ = partially or conceptually addressed; × = not a primary methodological capability.

actionable process improvement recommendations through a closed-loop analytical workflow.

3.1 Datasets and Experimental Materials

The proposed data-driven Value Stream Mapping and Value Stream Design (VSM/VSD) framework was evaluated using both synthetic and real-world event logs to assess its applicability under controlled and heterogeneous operational settings. The experimental study considered one synthetic manufacturing event log together with three publicly available real-world event logs: the BPI Challenge 2017, the BPI Challenge 2019, and the Data Quality Issues (DQI) event log.

The synthetic manufacturing event log was designed to validate the complete methodological workflow under controlled conditions, enabling systematic evaluation of Process Mining, Complex Network Analysis, Explainable Machine Learning, Digital Twin calibration, future-state scenario generation, and cumulative ablation experiments. The real-world event logs were subsequently employed to assess the generalizability, scalability, and robustness of the proposed framework across heterogeneous business and manufacturing processes.

Table 2 summarizes the datasets considered throughout the experimental evaluation.

All datasets were transformed into the standardized event-log schema presented in Table 4, ensuring a consistent representation of cases, activities, timestamps, resources, and operational attributes across domains. Subsequently, every dataset followed the same experimental workflow, including event-log standardization, Process Mining, Value Stream Mapping extraction, activity-network construction, predictive modeling, explainability analysis, and, where applicable, Digital Twin-based scenario evaluation. This unified pipeline enables a consistent assessment of the proposed framework across heterogeneous operational environments.

3.1.1 Synthetic Manufacturing Event Log Generation

The synthetic manufacturing event log was developed to provide a controlled experimental environment for validating the complete VSM/VSD framework under reproducible conditions. Unlike real-world event logs, where operational conditions cannot be manipulated, the synthetic dataset enables systematic evaluation of bottlenecks, waiting times, resource utilization, quality issues, and alternative improvement scenarios while preserving realistic manufacturing behavior.

The simulated production system models a discrete manufacturing process composed of sequential production stages, including material preparation, machining, initial inspection, assembly, functional testing, packaging, and product release. Process instances were generated according to the predefined production routing while allowing stochastic variations in both processing and waiting times.

For every event, the simulator generated all attributes required by the proposed framework, including the case identifier, executed activity, responsible resource, timestamps, processing time, waiting time, work-in-process level, defect occurrence, and rework indicator, following the standardized event-log schema presented in Table 4. Processing and waiting times were sampled from predefined statistical distributions calibrated to emulate realistic manufacturing variability.

Table 2 Datasets employed to evaluate the proposed framework across different operational domains and validation objectives.

Dataset	Domain	Type	Purpose within the framework
Synthetic Manufacturing Event Log	Manufacturing	Synthetic	Controlled experimentation, framework validation, Digital Twin calibration, future-state Value Stream Design evaluation, and cumulative ablation analysis.
BPI Challenge 2017	Business Processes	Real	Generalization assessment across heterogeneous business processes through process discovery, predictive analytics, and activity-network analysis.
BPI Challenge 2019	Procurement Processes	Real	Scalability assessment using a large procurement event log with complex operational workflows.
Data Quality Issues Event Log	Manufacturing	Real	Robustness evaluation under noisy, incomplete, and quality-impaired event-log conditions.

Additional stochastic mechanisms were incorporated to reproduce queue formation, temporary resource congestion, quality defects, and rework operations. To emulate realistic production dynamics, bottlenecks were intentionally introduced in selected workstations by increasing processing times and resource utilization. Likewise, defect and rework events were generated according to predefined probabilities, producing representative quality-related patterns while preserving sufficient class variability for predictive modeling.

The resulting synthetic event log follows the same standardized schema adopted by the real-world datasets, ensuring complete compatibility with the Process Mining, Complex Network Analysis, Machine Learning, Explainable AI, and Digital Twin modules. All stochastic processes were initialized using fixed random seeds to facilitate experimental reproducibility.

The complete experimental pipeline was implemented in Python. Process discovery and conformance analysis were performed using PM4Py, activity-network construction with NetworkX, predictive modeling with Scikit-Learn and XGBoost, model interpretability with SHAP, and discrete-event simulation with SimPy. Fixed random seeds were employed throughout all experimental stages to facilitate reproducibility.

3.2 Problem Formulation and Research Architecture

The literature reviewed in Section 2 shows that the digitalization of Value Stream Mapping (VSM) has progressively enabled the use of operational data for process reconstruction, performance analysis, and simulation-assisted improvement. However, a methodological challenge remains in systematically connecting these capabilities across the complete improvement cycle. In particular, reconstructing the current state of a process does not by itself determine which operational mechanisms are responsible for poor performance, which interventions should be considered, or how alternative interventions will affect the future state of the value stream.

Accordingly, the problem addressed in this study is formulated as the transformation of operational event data into an evidence-based decision-support process that connects current-state reconstruction, diagnosis, intervention evaluation, and future-state Value Stream Design (VSD). The research question is therefore stated as follows:

> *How can operational event data be systematically transformed into evidence-based current-state diagnosis and quantitatively evaluated future-state Value Stream Designs?*

To address this question, the proposed framework establishes a closed-loop analytical architecture in which Process Mining (PM), Complex

Table 3 Main characteristics of the synthetic manufacturing event log used for controlled experimentation and framework validation.

Characteristic	Description
Production system	Discrete manufacturing process with sequential production stages.
Activities	Material Preparation, Machining, Inspection, Assembly, Testing, Packaging, and Release.
Process routing	Sequential workflow incorporating stochastic operational variability.
Generated attributes	Case ID, Activity, Resource ID, Start Timestamp, End Timestamp, Cycle Time (CT), Waiting Time (WT), Work-in-Process (WIP), Defect Flag, and Rework Flag.
Processing times	Randomly sampled from statistical distributions calibrated for each activity.
Waiting times	Queue-dependent stochastic delays representing operational congestion.
Quality events	Probabilistic generation of defects and rework activities.
Bottlenecks	Artificially introduced through resource-capacity constraints and congestion effects.
Output format	CSV event log compatible with PM4Py and the proposed experimental pipeline.
Reproducibility	Fixed random seeds and deterministic generation procedure.

Network Analysis (CNA), Explainable Machine Learning (XAI/ML), and Digital Twin (DT) simulation perform complementary functions rather than acting as independent analytical modules. Process Mining reconstructs observed process behavior and provides the operational basis for the current-state VSM; Complex Network Analysis characterizes structural dependencies among activities, resources, and value flows; Explainable Machine Learning identifies and interprets operational criticalities; and Digital Twin simulation evaluates candidate interventions before they are translated into a future-state VSD.

Let

$$L_t = \{e_1, e_2, \ldots, e_n\}$$

denote the operational event log available at improvement cycle t. The proposed analytical transformation is defined as

$$\begin{aligned}\mathcal{F}(L_t): \quad & L_t \to VSM_{c,t} \to \mathcal{D}_t \\ & \to \Omega_t \to \mathcal{E}_t \to VSM_{f,t},\end{aligned}$$

where:

- $VSM_{c,t}$ denotes the data-driven current-state Value Stream Map reconstructed from the event log;
- $\mathcal{D}_t$ denotes the diagnostic evidence obtained from operational indicators, structural network analysis, predictive models, and their explanations;
- $\Omega_t = \{S_1, S_2, \ldots, S_q\}$ denotes the set of candidate improvement scenarios derived from the diagnosed process criticalities;
- $\mathcal{E}_t$ denotes the quantitative evaluation of candidate scenarios obtained through Digital Twin simulation; and
- $VSM_{f,t}$ denotes the future-state Value Stream Design associated with the selected improvement scenario.

This formulation distinguishes between the analytical technologies employed by the framework and the decision-support stages they enable. Specifically, the internal diagnostic evidence can be expressed as

$$\mathcal{D}_t = \{\mathbf{K}_{c,t}, \mathcal{G}_t, \mathcal{M}_t, \mathcal{X}_t\},$$

where $\mathbf{K}_{c,t}$ contains current-state operational indicators, $\mathcal{G}_t$ represents the set of process and resource networks, $\mathcal{M}_t$ denotes the predictive models, and $\mathcal{X}_t$ contains the corresponding explainability information.

The operational state of the value stream is characterized through the KPI vector

$$\mathbf{K} = [LT, TH, WIP, DR, RR],$$

where LT denotes Lead Time, TH Throughput, WIP Work-In-Process, DR Defect Rate, and RR Rework Rate.

Rather than aggregating these indicators into a single optimization objective, the framework evaluates future-state configurations through their individual operational performance dimensions. For each candidate scenario $S_i \in \Omega_t$, the Digital Twin estimates a future-state performance vector

$$\mathbf{K}_f^{(i)} = \left[LT_f^{(i)}, TH_f^{(i)}, WIP_f^{(i)}, DR_f^{(i)}, RR_f^{(i)}\right].$$

The objective is to evaluate the operational consequences of candidate configurations relative to the current-state performance vector $\mathbf{K}_c$. Because the considered indicators have different optimization directions, candidate configurations are assessed directly through their individual KPI changes. Favorable configurations seek to increase throughput while reducing lead time, work-in-process, defect rate, and rework rate. This formulation avoids aggregating heterogeneous performance indicators into a single objective function and preserves the individual operational trade-offs available to decision makers.

Figure 1 presents the resulting research architecture. The framework begins with operational event logs rather than a manually constructed process representation. Process Mining is first used to reconstruct actual process behavior, extract operational indicators, and generate the current-state VSM. The discovered process is subsequently represented through complementary activity, resource, and value-flow networks, from which structural indicators are obtained.

Operational and network-derived features are then integrated into Explainable Machine Learning models to identify undesirable process outcomes and characterize the factors associated with them. Importantly, these diagnostic outputs are not treated as the final result of the analysis. Instead, they provide evidence for defining candidate interventions Ω_t. The Digital Twin subsequently evaluates the expected system-level consequences of these alternatives under controlled simulation conditions.

Finally, the simulation outcomes associated with favorable candidate configurations are translated into a future-state Value Stream Design $VSM_{f,t}$. This establishes the central analytical loop of the proposed methodology:

$$\begin{gathered}\text{Operational data} \rightarrow \text{Current state} \\ \text{Diagnosis} \rightarrow \text{Intervention} \\ \text{Evaluation} \rightarrow \text{Future-state design.}\end{gathered}$$

The resulting architecture therefore uses Process Mining, Complex Network Analysis, Explainable Machine Learning, and Digital Twin simulation as complementary mechanisms for establishing a traceable transition from observed operational behavior to evidence-based process redesign.

3.3 Event Log Modeling and Process Mining

The first stage of the proposed framework transforms raw operational records into structured event logs suitable for Process Mining analysis. Event logs constitute the primary data source from which process behavior, execution patterns, and performance indicators are extracted.

Let

$$L = \{e_1, e_2, \ldots, e_n\}$$

denote an event log composed of events generated during process execution. Each event is represented as

$$e_i = (c_i, a_i, r_i, t_i, x_i),$$

where c_i is the case identifier, a_i the executed activity, r_i the associated resource, t_i the timestamp, and x_i a vector of optional operational attributes.

To ensure compatibility across heterogeneous data sources, all datasets were transformed into a standardized event-log structure containing the attributes shown in Table 4.

After preprocessing, Process Mining techniques were applied to discover process structures and extract operational indicators. Directly-Follows Graphs (DFGs) were employed to identify activity transitions and execution frequencies.

The process structure is represented as

$$P = (A, E),$$

where A denotes the set of activities and

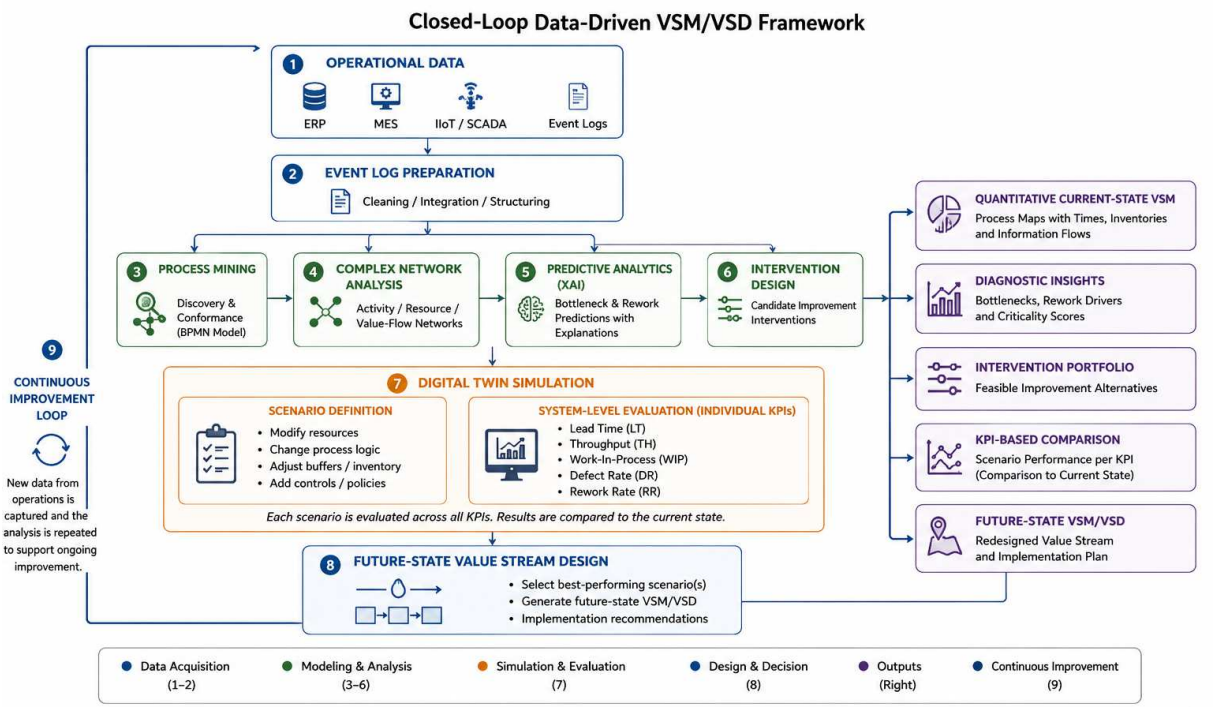


Fig. 1 Research architecture of the proposed closed-loop data-driven VSM/VSD framework. Operational event logs are transformed into a quantitative current-state VSM through Process Mining. Operational and network-derived indicators provide complementary evidence for explainable predictive diagnosis. The identified process criticalities guide the definition of candidate improvement interventions, which are evaluated through Digital Twin simulation. The selected scenario is subsequently translated into a quantitatively supported future-state Value Stream Design, thereby establishing a traceable path from operational evidence to process redesign.

Table 4 Standardized event-log schema adopted across all datasets to ensure consistent process mining, machine learning, digital twin calibration, and network-analysis workflows.

Attribute	Description
`case_id`	Unique identifier of the process instance (case).
`activity`	Name of the executed process activity.
`resource`	Resource, operator, or machine responsible for the activity.
`timestamp`	Execution timestamp associated with the event.
`cycle_time`	Processing time required to complete the activity.
`waiting_time`	Queueing or idle time elapsed before activity execution.
`wip`	Work-in-process level observed at the event occurrence.
`defect_flag`	Binary indicator identifying defect occurrence.
`rework_flag`	Binary indicator identifying rework occurrence.

$$E \subseteq A \times A$$

represents the set of directly-following relationships identified from the event log.

For each transition (a_i, a_j), the execution frequency is computed as

$$w_{ij} = \sum_{c \in C} I(a_i \rightarrow a_j),$$

where $I(\cdot)$ is an indicator function equal to one when the transition occurs within a process instance and zero otherwise.

The discovered process model enables the estimation of key operational indicators required for Value Stream Mapping, including average cycle time, waiting time, throughput, defect rate, and rework frequency.

For an activity a, the average cycle time is defined as

$$CT(a) = \frac{1}{N_a} \sum_{k=1}^{N_a} ct_k,$$

where N_a denotes the number of executions of activity a.

Similarly, the average waiting time is computed as

$$WT(a) = \frac{1}{N_a} \sum_{k=1}^{N_a} wt_k.$$

These metrics constitute the quantitative foundation for the automatic generation of current-state Value Stream Maps described in the next section.

For the predictive tasks considered in this study, bottleneck labels were derived from the operational indicators extracted from the event log. An activity was classified as a bottleneck when its average cycle time or waiting time exceeded a predefined percentile threshold within the process. Formally, for each activity $a \in A$, the bottleneck indicator is defined as

$$B(a) = \begin{cases} 1, & \text{if } CT(a) \geq Q_{0.90}(CT(A)) \text{ or} \\ & \quad WT(a) \geq Q_{0.90}(WT(A)), \\ 0, & \text{otherwise.} \end{cases}$$

where $Q_{0.90}(\cdot)$ denotes the 90th percentile of the corresponding activity-level indicator distribution. The 90th percentile was selected to identify the most critical process activities while avoiding excessive sensitivity to isolated extreme observations.

3.4 Automated Current-State Value Stream Mapping

The process model obtained through Process Mining is subsequently transformed into a Current-State Value Stream Map (VSM) through the automatic aggregation of operational metrics extracted from the event log. Unlike traditional VSM approaches that rely on manual observation and expert estimation, the proposed method generates quantitative value stream representations directly from process execution data.

Let

$$P = (A, E)$$

be the process model discovered from the event log, where A represents the set of activities and E the set of directly-following relationships. Each activity $a \in A$ is associated with a set of operational indicators estimated during the Process Mining stage.

The average work-in-process level is estimated as

$$WIP(a) = \frac{1}{N_a} \sum_{k=1}^{N_a} wip_k.$$

Quality-related indicators are computed as

$$DR(a) = \frac{D_a}{N_a},$$

and

$$RR(a) = \frac{R_a}{N_a},$$

where D_a and R_a denote the number of defective and reworked executions associated with activity a, respectively.

To preserve compatibility with traditional Value Stream Mapping practices, activities discovered through Process Mining are grouped into value-stream stages using a predefined aggregation scheme. The mapping function

$$\phi : A \to M$$

assigns each low-level process activity $a \in A$ to exactly one macro-process $m \in M$, such that

$$A_m = \{a \in A : \phi(a) = m\}.$$

The aggregation scheme may be defined using domain knowledge, organizational process models, or existing production-routing structures. This aggregation preserves the hierarchical structure commonly employed in traditional Value Stream Mapping while enabling quantitative indicator estimation directly from event logs.

For each macro-process $m \in M$, the corresponding indicators are calculated as

$$CT_m = \frac{1}{|A_m|} \sum_{a \in A_m} CT(a),$$

$$WT_m = \frac{1}{|A_m|} \sum_{a \in A_m} WT(a),$$

$$WIP_m = \frac{1}{|A_m|} \sum_{a \in A_m} WIP(a),$$

$$DR_m = \frac{1}{|A_m|} \sum_{a \in A_m} DR(a),$$

and

$$RR_m = \frac{1}{|A_m|} \sum_{a \in A_m} RR(a).$$

The resulting metrics are incorporated into an automatically generated Current-State Value Stream Map represented as

$$VSM_c = (M, F, K),$$

where M denotes the set of value stream stages, F the material and information flows connecting them, and K the associated operational indicators.

At the global process level, Lead Time is estimated as

$$LT = \sum_{m \in M} (CT_m + WT_m),$$

while the Value-Added Ratio is computed as

$$VAR = \frac{\sum_{m \in M} CT_m}{\sum_{m \in M} (CT_m + WT_m)}.$$

The automatically generated VSM provides a quantitative representation of process performance and serves as the foundation for subsequent network analysis, predictive modeling, and Digital Twin simulation.

3.5 Complex Network Modeling and Feature Engineering

Although Process Mining and Value Stream Mapping provide valuable process-level insights, they are often insufficient for capturing the structural dependencies and interaction patterns that emerge among activities, resources, and value flows. To address this limitation, the proposed framework incorporates Complex Network Analysis (CNA) to model the operational system from multiple complementary perspectives.

Three complementary graph representations are constructed from the event log and the automatically generated Value Stream Map:

$$\mathcal{G} = \{G_A, G_R, G_V\},$$

where G_A represents the Activity Network, G_R the Resource Network, and G_V the Value-Flow Network. The Activity Network captures directly-following relationships among process activities, the Resource Network models interactions among operators, machines, and organizational resources, and the Value-Flow Network represents material, information, and product transfers across value-stream stages.

To characterize the structural behavior of these networks, standard graph-theoretic indicators were computed, including degree, betweenness, closeness, and eigenvector centrality. These metrics quantify node connectivity, intermediary influence, accessibility, and structural importance, respectively. In addition, network density was calculated to characterize the overall connectivity of the process structure and provide a global indicator of interaction complexity.

Detailed mathematical definitions of these metrics can be found in the complex-network literature [8, 15].

The resulting network-derived attributes were integrated with operational indicators extracted from Process Mining and Value Stream Mapping to create a unified feature space:

$$\mathbf{X} = \{\mathbf{X}_{PM}, \mathbf{X}_{VSM}, \mathbf{X}_{CN}\},$$

where $\mathbf{X}_{PM}$, $\mathbf{X}_{VSM}$, and $\mathbf{X}_{CN}$ denote Process Mining, Value Stream Mapping, and Complex Network features, respectively.

The resulting feature matrix combines operational and structural information, enabling the identification of critical activities, process bottlenecks, and interaction patterns that may not be observable through conventional process indicators alone. These features are subsequently used by the Explainable Artificial Intelligence layer described in the following section.

3.6 Explainable Machine Learning Layer

The operational indicators extracted from Process Mining and Value Stream Mapping, together with

the network-derived features described in the previous section, are integrated into an Explainable Artificial Intelligence (XAI) layer designed to predict process outcomes and identify the factors that most strongly influence system performance.

Depending on the application scenario, predictive tasks may include defect occurrence, rework probability, process delays, bottleneck identification, throughput degradation, or other operational performance indicators. Multiple supervised learning algorithms are evaluated, and the best-performing model is selected according to standard classification or regression metrics, including Accuracy, Precision, Recall, F1-score, ROC-AUC, MAE, RMSE, and R^2, depending on the prediction task.

To ensure interpretability, the framework incorporates SHAP (SHapley Additive exPlanations), which decomposes model predictions into additive feature contributions. For an observation x, the prediction can be expressed as

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j,$$

where ϕ_0 denotes the baseline prediction and ϕ_j represents the contribution of feature j.

The resulting explanations enable the identification of operational factors associated with undesirable outcomes, such as excessive waiting times, elevated defect rates, resource overloads, or structural bottlenecks. Consequently, the Explainable Machine Learning layer not only provides predictive capabilities but also generates actionable insights for process improvement.

The generated predictions and explanations are subsequently incorporated into the Digital Twin environment to evaluate future-state scenarios under different operational conditions.

The previous stages transform raw event logs into current-state Value Stream Maps, network-based features, and predictive insights. Algorithm 1 summarizes the analytical pipeline preceding the Digital Twin stage.

Input: Event log L.
Output: Current-State VSM VSM_c, feature matrix X, predictive model M^*, and explanations E.

1. Preprocess and standardize L; discover the process model; compute operational indicators; and generate VSM_c.
2. Construct the activity, resource, and value-flow networks G_A, G_R, and G_V.
3. Compute network- and node-level metrics, including degree, betweenness, closeness, eigenvector centrality, and density.
4. Build X, train the candidate Machine Learning models, and select the best-performing model M^*.
5. Compute SHAP values and generate the explanations E.
6. Return VSM_c, X, M^*, E.

Algorithm 1: Data-Driven Current-State VSM Generation

3.7 Digital Twin-Based Future-State Design

The diagnostic evidence generated through Process Mining, Complex Network Analysis, and Explainable Machine Learning is subsequently incorporated into a Digital Twin (DT) environment to evaluate alternative process configurations and support data-driven Future-State Value Stream Design (VSD).

Within the proposed framework, the Digital Twin acts as a simulation-based decision-support mechanism capable of reproducing process behavior under alternative operating conditions. The preceding analytical layers identify operational and structural criticalities, whereas the Digital Twin estimates the system-level consequences of improvement actions before physical implementation.

The calibrated Digital Twin is represented as

$$DT = (A, R, \Theta, \Omega),$$

where A denotes the set of process activities, R represents the available resources, Θ contains the operational parameters estimated from the event log, and Ω denotes the set of candidate process configurations evaluated through simulation.

The parameter set Θ includes processing times, waiting times, arrival patterns, defect probabilities, rework behavior, and process-flow characteristics derived from the current-state analysis. In this way, the simulation environment is parameterized using the same operational evidence used to construct the current-state VSM.

The diagnostic outputs obtained from Process Mining, network analysis, and predictive modeling are used to identify process elements where interventions may provide operational benefits. These analytical outputs support the definition of improvement actions but do not automatically generate them. Candidate interventions may involve:

- capacity increases at critical resources;
- workload redistribution or workforce reassignment;
- reductions in waiting time or flow-related delays;
- quality-oriented actions intended to reduce defects and rework; and
- combined interventions affecting multiple process dimensions.

For experimental validation, the contribution of the analytical layers is assessed through the six configurations defined in the ablation study:

$$\Omega_{\text{exp}} = \{G_0, G_1, G_2, G_3, G_4, G_5\},$$

where G_0 represents the baseline process configuration, G_1 manual VSM, G_2 Process Mining-assisted VSM, G_3 PM-VSM with Machine Learning, G_4 PM-VSM with Machine Learning and Digital Twin support, and G_5 the complete framework including Complex Network Analysis.

Each configuration is evaluated through repeated independent simulation replications. For every configuration G_j, the Digital Twin estimates the performance vector

$$\mathbf{K}_j = [LT_j, TH_j, WIP_j, DR_j, RR_j],$$

where LT, TH, WIP, DR, and RR denote Lead Time, Throughput, Work-In-Process, Defect Rate, and Rework Rate, respectively.

The performance of each configuration is evaluated relative to the baseline configuration G_0. For an indicator to be minimized, $k \in \{LT, WIP, DR, RR\}$, the relative improvement is computed as

$$I_k^{(j)} = \frac{k_{G_0} - k_{G_j}}{k_{G_0}} \times 100.$$

For throughput, for which larger values are preferred, the improvement is computed as

$$I_{TH}^{(j)} = \frac{TH_{G_j} - TH_{G_0}}{TH_{G_0}} \times 100.$$

This formulation allows the incremental performance associated with progressively enriched configurations to be quantified without aggregating heterogeneous KPIs into a single optimization objective. A configuration is considered operationally preferable when it improves throughput while simultaneously reducing lead time, work-in-process, defect rate, and rework rate.

The simulation outputs associated with an improved configuration are subsequently mapped back to the value-stream representation. Let

$$\mathbf{K}_{f,m} = \left[CT_m^f, WT_m^f, WIP_m^f, DR_m^f, RR_m^f\right]$$

denote the future-state indicators estimated for value-stream stage $m \in M$.

The corresponding Future-State Value Stream Map is defined as

$$VSM_f = (M, F, \mathbf{K}_f),$$

where M represents the set of value-stream stages, F denotes the material and information flows connecting them, and

$$\mathbf{K}_f = \{\mathbf{K}_{f,m} : m \in M\}$$

contains the stage-level performance indicators obtained from simulation.

The expected future-state Lead Time is computed as

$$LT_f = \sum_{m \in M} \left(CT_m^f + WT_m^f\right),$$

and the future-state Value-Added Ratio is

$$VAR_f = \frac{\sum_{m \in M} CT_m^f}{\sum_{m \in M} \left(CT_m^f + WT_m^f\right)}.$$

Accordingly, the role of the Digital Twin within the proposed framework is to provide a quantitative experimental mechanism for assessing how analytical capabilities and process interventions affect future-state performance. The Digital Twin therefore establishes the transition

Diagnosis → Intervention
Simulation → Performance assessment
Future-state VSD.

This formulation preserves a traceable connection between current-state evidence, simulated improvement actions, and the resulting future-state Value Stream Design.

3.8 Closed-Loop Data-Driven VSM/VSD Framework

Building upon the problem formulation introduced in Section 3.2, the complete framework establishes a traceable analytical path from operational event data to evidence-based Future-State Value Stream Design. Rather than representing the framework as a sequence of independent technologies, its analytical stages correspond to the main decision-support functions required to transform observed process behavior into an evaluated redesign alternative.

For an improvement cycle t, the framework is represented as

$$L_t \rightarrow VSM_{c,t} \rightarrow \mathcal{D}_t \rightarrow \Omega_t \rightarrow \mathcal{E}_t \rightarrow VSM_{f,t},$$

where L_t denotes the operational event log, $VSM_{c,t}$ is the reconstructed current-state Value Stream Map, $\mathcal{D}_t$ represents the diagnostic evidence obtained from operational, structural, predictive, and explainability analyses, Ω_t denotes the set of candidate improvement configurations, $\mathcal{E}_t$ represents their simulation-based evaluation, and $VSM_{f,t}$ is the resulting future-state Value Stream Design.

The framework therefore performs four main decision-support functions. First, operational event data are transformed into a quantitative current-state representation through Process Mining and automated VSM generation. Second, operational indicators, Complex Network Analysis, and Explainable Machine Learning provide complementary evidence for identifying process criticalities and their associated factors. Third, this diagnostic evidence supports the definition of candidate interventions, whose expected system-level consequences are evaluated through Digital Twin simulation. Finally, the simulation outputs are mapped back to the value-stream representation to construct and assess the corresponding future-state design.

The effect of a candidate future-state configuration is assessed by comparing its performance with the observed current state. Let

$$\mathbf{K}_c = [LT_c, TH_c, WIP_c, DR_c, RR_c]$$

and

$$\mathbf{K}_f = [LT_f, TH_f, WIP_f, DR_f, RR_f]$$

denote the current-state and simulated future-state KPI vectors, respectively. Their difference is

$$\Delta\mathbf{K} = \mathbf{K}_f - \mathbf{K}_c.$$

Because the performance indicators have different optimization directions, the expected redesign effect is interpreted individually for each KPI. An intervention provides simultaneous improvement over the current state when

$$\begin{aligned} &LT_f < LT_c, \quad WIP_f < WIP_c, \quad DR_f < DR_c, \\ &RR_f < RR_c, \quad TH_f > TH_c. \end{aligned}$$

However, these conditions are not imposed as mandatory feasibility constraints on every candidate configuration. Instead, they provide a reference for identifying configurations that simultaneously improve all five operational dimensions. When trade-offs occur, the individual KPI changes remain available to decision makers rather than being obscured by aggregation into a single performance score.

Algorithm 2 summarizes the analytical workflow evaluated in this study.

The workflow described above corresponds to the analytical cycle evaluated in the experimental study. It does not require online connectivity between the operational system and the Digital Twin; instead, the Digital Twin is calibrated from historical event-log evidence and used for offline experimentation with alternative configurations.

In an industrial deployment, the framework can be extended into a recurrent improvement

Input: Operational event log L_t.
Output: Evaluated Future-State Value Stream Design $VSM_{f,t}$.

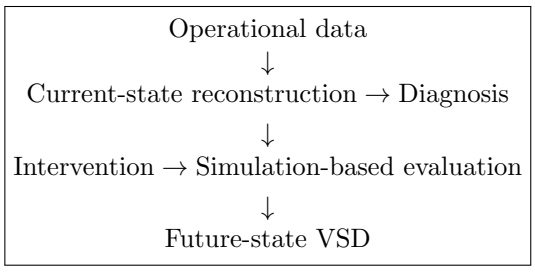
1. Reconstruct observed process behavior from L_t, generate the Current-State VSM $VSM_{c,t}$, and extract operational indicators.
2. Construct the activity, resource, and value-flow networks and compute their structural indicators.
3. Train predictive models with operational and network features, generate explanations, and construct diagnostic evidence $\mathcal{D}_t$.
4. Define candidate configurations Ω_t and calibrate the Digital Twin using current-state evidence.
5. For each $S_i \in \Omega_t$, perform repeated simulations, estimate $\mathbf{K}_f^{(i)}$, and compare it with $\mathbf{K}_c$.
6. Identify favorable configurations, map the selected configuration to $VSM_{f,t}$, and return $VSM_{f,t}$.

Algorithm 2: Data-driven VSM/VSD analytical workflow

cycle. After an evaluated future-state design is implemented, new operational observations may generate an updated event log L_{t+1}, from which the analytical workflow can be executed again:

$$VSM_{f,t} \xrightarrow{\text{implementation}} L_{t+1} \rightarrow VSM_{c,t+1}.$$

This recurrent extension represents a deployment mechanism rather than an experimentally validated component of the present study. The empirical evaluation reported in this paper focuses on the traceable transformation



thus establishing the closed analytical loop between observed process behavior and quantitatively supported process redesign.

4 Experimental Design

This section describes the experimental protocol used to validate the proposed data-driven VSM/VSD framework. The evaluation focuses on three dimensions: operational performance improvement, predictive capability, and structural process analysis. To quantify the incremental and cumulative performance associated with the progressive incorporation of the methodological capabilities, a systematic ablation study is conducted using both synthetic and real-world event logs.

4.1 Experimental Objectives and Datasets

The experimental evaluation aims to assess the effectiveness, robustness, and generalizability of the proposed data-driven Value Stream Mapping and Value Stream Design (VSM/VSD) framework. The evaluation is designed to determine whether the integration of Process Mining, Complex Network Analysis, Explainable Machine Learning, and Digital Twin simulation provides measurable improvements in operational analysis, predictive modeling, and future-state process redesign when compared with conventional Value Stream Mapping approaches.

The experiments were conducted using the datasets described in Section 3.1, comprising one synthetic manufacturing event log and three real-world event logs. The synthetic manufacturing dataset was employed for controlled experimentation, Digital Twin calibration, simulation-based evaluation of cumulative framework configurations, and statistical ablation analysis under reproducible operating conditions. The real-world event logs were subsequently used to evaluate the generalizability, scalability, and robustness of the proposed framework across heterogeneous operational domains, including business processes, procurement workflows, and manufacturing environments affected by data-quality issues.

Three complementary dimensions were considered throughout the experimental evaluation.

- **Operational evaluation** analyzes the capability of the proposed framework to improve Lean-oriented performance indicators, including Lead Time (LT), Throughput (TH), Work-In-Process (WIP), Cycle Time (CT), Defect Rate (DR), and Rework Rate (RR), through

simulation-supported future-state Value Stream Design.

- **Predictive evaluation** assesses the ability of Explainable Machine Learning models to predict operational criticalities, specifically delay, bottleneck, and rework events, using both conventional process features and network-derived structural indicators.
- **Structural evaluation** examines the contribution of Complex Network Analysis to identifying structurally relevant activities, resource dependencies, connectivity patterns, and process characteristics that complement conventional Value Stream Mapping indicators.

Although the datasets originate from different operational domains, all experiments follow the same analytical workflow. Each event log is transformed into the standardized schema described in Section 3.1, followed by Process Mining, automatic Value Stream Mapping generation, activity-network construction, predictive modeling, explainability analysis, and, where applicable, Digital Twin-based scenario evaluation. This unified experimental protocol enables a consistent comparison of the proposed framework across heterogeneous process environments while preserving methodological reproducibility.

4.2 Ablation Study Design

To quantify the incremental and cumulative performance associated with the analytical capabilities incorporated into the proposed framework, a cumulative ablation study was designed in which methodological components are progressively introduced and evaluated. Unlike conventional component-removal ablation studies, the objective of the present design is not to isolate the individual causal contribution of each analytical component, but rather to evaluate the progressive improvements obtained as additional decision-support capabilities are incorporated into an increasingly comprehensive analytical workflow.

Accordingly, each experimental configuration extends the previous one by adding a new analytical capability while preserving all previously incorporated components. Consequently, performance differences between consecutive configurations are interpreted as incremental framework-level improvements associated with the newly incorporated analytical capability, whereas comparisons with the baseline quantify the cumulative operational benefit achieved by the complete framework.

The cumulative experimental configurations considered in this study are summarized in Table 5.

The role of each cumulative configuration is summarized as follows.

- **G0 (Baseline).** Represents the original operational process without advanced analytical support and serves as the reference configuration for all cumulative comparisons.
- **G1 (Manual VSM).** Represents the conventional Lean Manufacturing approach based on manually constructed Value Stream Maps.
- **G2 (PM-VSM).** Introduces Process Mining to automatically reconstruct the observed process and generate data-driven current-state Value Stream Maps.
- **G3 (PM-VSM + XAI).** Incorporates Explainable Machine Learning to support predictive diagnosis of operational criticalities using process-derived features.
- **G4 (PM-VSM + XAI + DT).** Extends the previous configuration through Digital Twin simulation, enabling quantitative evaluation of alternative future-state Value Stream Design configurations before implementation.
- **G5 (Complete Framework).** Integrates Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis into a unified decision-support framework for data-driven Value Stream Mapping and Value Stream Design.

Operational performance is evaluated across all six cumulative configurations using Lead Time, Throughput, Work-In-Process, Defect Rate, and Rework Rate. Predictive performance is evaluated separately on the real-world event logs through temporal train-validation-test partitions, whereas the contribution of Complex Network Analysis is assessed by comparing predictive models trained with conventional process features alone against models additionally incorporating network-derived structural features.

Because the cumulative configurations are not statistically independent, the observed differences should be interpreted as incremental framework-level improvements rather than isolated causal

Table 5 Cumulative experimental configurations employed in the ablation study. Each successive configuration incorporates an additional analytical capability while preserving all components included in the previous configurations.

Group	Description
G0	Baseline operational process without analytical decision-support mechanisms.
G1	Traditional manually constructed Value Stream Mapping (VSM).
G2	Automatically generated Process Mining-based Value Stream Mapping (PM-VSM).
G3	PM-VSM enriched with Explainable Machine Learning for predictive process diagnosis.
G4	PM-VSM combined with Explainable Machine Learning and Digital Twin-based future-state scenario evaluation.
G5	Complete framework integrating Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis.

effects attributable to individual analytical components. This cumulative experimental design reflects the intended practical deployment of the proposed framework, in which analytical capabilities are progressively integrated to support increasingly informed operational decision making.

4.3 Evaluation Metrics

The proposed framework is evaluated from three complementary perspectives: operational performance, predictive capability, and structural process behavior. Operational indicators quantify Lean-oriented process performance, predictive metrics evaluate the effectiveness of the Explainable Machine Learning layer, and structural indicators characterize process topology through Complex Network Analysis.

Table 6 summarizes the metrics employed throughout the experimental evaluation.

Operational performance is evaluated on the cumulative experimental configurations (G0–G5) using the Digital Twin simulation environment. Predictive performance is evaluated independently on the real-world event logs using temporal train-validation-test partitions, whereas structural analysis quantifies the additional information provided by network-derived features beyond conventional process indicators.

Because the proposed framework integrates complementary analytical components, no single metric is sufficient to characterize its overall performance. Instead, the combined use of operational, predictive, and structural metrics provides a multidimensional evaluation of the framework's capability to support current-state analysis, predictive diagnosis, simulation-based intervention assessment, and future-state Value Stream Design.

4.4 Model Training and Simulation Protocol

The experimental evaluation combines predictive modeling with simulation-based experimentation to assess complementary aspects of the proposed framework. Explainable Machine Learning is employed to identify operational criticalities from historical event logs, whereas the Digital Twin is used to evaluate the operational consequences of alternative future-state Value Stream Design (VSD) configurations before implementation.

For the predictive experiments, real-world event logs are partitioned using a temporal train–validation–test strategy to preserve the chronological order of process executions and reduce the possibility of information leakage. The adopted partition allocates 70% of the observations for training, 15% for validation, and 15% for independent testing.

Random Forest is employed as an interpretable baseline model, while XGBoost is adopted as the primary predictive algorithm because of its ability to model nonlinear relationships and complex interactions among operational and network-derived features.

To further reduce target leakage, proxy labels requiring threshold estimation are generated only after the temporal partition. In particular, thresholds used to define delay-related targets are estimated exclusively from the training subset and subsequently applied to the validation and testing partitions. Likewise, predictor variables directly

Table 6 Evaluation metrics employed to assess operational performance, predictive capability, and structural process characteristics.

Category	Metrics	Purpose
Operational	Lead Time (LT), Throughput (TH), Work-In-Process (WIP), Cycle Time (CT), Defect Rate (DR), Rework Rate (RR)	Evaluation of process efficiency, operational performance, and Lean-oriented improvement achieved through successive cumulative framework configurations.
Predictive	Accuracy, Precision, Recall, F1-score, ROC-AUC	Assessment of predictive models for delay, bottleneck, and rework prediction using conventional process features and network-derived structural features.
Structural	Degree Centrality, Betweenness Centrality, Closeness Centrality, Eigenvector Centrality, Network Density	Characterization of process topology and identification of structurally relevant activities, resource dependencies, and connectivity patterns through Complex Network Analysis.

involved in constructing a target variable are excluded from the corresponding predictive models. For example, *cycle time* is excluded when predicting delay-related targets, and *betweenness centrality* is excluded when predicting bottleneck-related targets because it is used to define the corresponding proxy labels.

The validation subset is reserved for methodological consistency and model selection, whereas all predictive results reported in this study correspond exclusively to the independent testing partition. Fixed random seeds are employed throughout the complete predictive pipeline to facilitate experimental reproducibility.

The Digital Twin is calibrated using operational parameters extracted from the event logs together with the current-state analytical outputs produced by Process Mining and Explainable Machine Learning. Simulation-based evaluation is subsequently performed over the cumulative experimental configurations defined in Section 4.2, ranging from the baseline process (G0) to the complete framework (G5).

For each cumulative configuration, 30 independent simulation replications are executed to account for stochastic variability in processing times, waiting times, quality events, routing decisions, and resource utilization. The resulting configurations are compared using Lead Time, Throughput, Work-In-Process, Defect Rate, and Rework Rate. Mean KPI values, standard deviations, 95% confidence intervals, and relative improvements with respect to the baseline configuration are reported.

The Digital Twin supports the evaluation of multiple classes of process interventions, including capacity adjustments, workload balancing, waiting-time reduction, quality-oriented actions, and combined improvement strategies. In the present study, however, statistical comparisons are performed across the cumulative framework configurations (G0–G5) rather than through optimization-based ranking of independently generated intervention alternatives.

Unlike the simulation experiments, the predictive evaluation on the real-world event logs does not involve repeated stochastic replications. Consequently, predictive performance is reported only for the independent testing partition, whereas measures of statistical variability, including standard deviations and 95% confidence intervals, are reported exclusively for the Digital Twin simulation experiments, where repeated replications permit statistically meaningful estimation of variability.

To quantify the contribution of Complex Network Analysis, every predictive task is executed twice. The first experiment uses only conventional process and Value Stream Mapping features, whereas the second incorporates additional network-derived structural features. Improvements observed under the second configuration quantify the additional predictive information provided by structural process analysis beyond conventional operational indicators.

4.5 Statistical Validation and Research Hypotheses

Statistical analyses were conducted to determine whether the operational differences observed across the cumulative experimental configurations exceeded the variability introduced by stochastic simulation. Because the Digital Twin experiments involve repeated stochastic replications, inferential statistical analyses were performed exclusively for the simulation-based operational evaluation.

When the assumptions of normality and homoscedasticity were considered appropriate, differences among the cumulative configurations (G0–G5) were evaluated using one-way Analysis of Variance (ANOVA). Subsequently, Tukey's Honestly Significant Difference (HSD) test was employed to identify statistically significant pairwise differences. For performance indicators that did not satisfy the assumptions required for parametric analysis, the Kruskal–Wallis nonparametric test was adopted. Statistical significance was evaluated using a confidence level of 95% ($\alpha = 0.05$).

Unlike the simulation experiments, predictive evaluation on the real-world event logs does not involve repeated stochastic replications. Consequently, predictive performance is assessed through out-of-sample evaluation on the independent testing partition using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Improvements associated with Complex Network Analysis are quantified by comparing predictive models trained with conventional process features against models additionally incorporating network-derived structural features.

The experimental evaluation addresses the following research hypotheses.

- **H1.** The complete data-driven VSM/VSD framework achieves statistically significant improvements in operational performance with respect to the baseline and conventional Value Stream Mapping configurations, as measured by Lead Time, Throughput, Work-In-Process, Defect Rate, and Rework Rate.
- **H2.** Incorporating network-derived structural features improves the predictive performance of Machine Learning models for delay, bottleneck, and rework prediction compared with models based exclusively on conventional process and Value Stream Mapping features.
- **H3.** The cumulative integration of Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis enables the identification and quantitative evaluation of future-state Value Stream Design configurations that provide simultaneous improvements across multiple operational performance indicators.

For the simulation experiments, the null hypothesis assumes that no statistically significant differences exist among the evaluated cumulative configurations, whereas the alternative hypothesis assumes that at least one configuration produces significantly different operational performance. For the predictive experiments, H2 is evaluated through direct comparison of out-of-sample predictive performance obtained with and without network-derived structural features.

It is important to emphasize that the cumulative experimental design evaluates progressively enriched analytical workflows rather than isolated analytical components. Consequently, statistically significant differences among cumulative configurations should be interpreted as evidence of incremental framework-level improvements and not as independent causal estimates of the individual contribution of each analytical capability.

4.6 Threats to Validity

As with any empirical study integrating multiple analytical methodologies, several factors may influence the validity and generalizability of the obtained results. The principal threats to validity considered in this study are summarized below.

Internal validity. All cumulative experimental configurations were evaluated using identical datasets, preprocessing procedures, feature engineering pipeline, temporal data partitioning strategy, evaluation metrics, and simulation protocol. Fixed random seeds were employed throughout the predictive and simulation experiments to reduce variability introduced by stochastic model training and Digital Twin execution. Furthermore, proxy labels requiring threshold estimation were generated exclusively from the training partition, and predictor variables directly involved in defining a target variable were excluded from the

corresponding predictive models in order to reduce target leakage.

Construct validity. Operational, predictive, and structural evaluation metrics were selected according to widely adopted practices in Lean Manufacturing, Process Mining, Machine Learning, and Complex Network Analysis. Nevertheless, some proxy variables, including delay, bottleneck, and rework indicators, are operational approximations derived from event-log information rather than directly observed organizational labels. Consequently, their interpretation depends on the quality and granularity of the available process data.

External validity. Although the experimental evaluation combines one controlled synthetic manufacturing event log with three heterogeneous real-world event logs from manufacturing, procurement, and business-process domains, additional validation using industrial production environments, logistics systems, healthcare processes, and service operations would further strengthen the generalizability of the proposed framework. Likewise, future studies should evaluate the framework under additional process complexities, organizational settings, and event-log characteristics.

Model and data validity. The predictive performance of the proposed framework depends on the quality, completeness, and representativeness of the underlying event logs. Likewise, Digital Twin simulations necessarily rely on calibration assumptions regarding processing times, waiting times, routing behavior, resource utilization, and operational variability. These risks were mitigated through standardized preprocessing, reproducible data preparation, multiple predictive models, repeated simulation replications, and statistical validation procedures.

Methodological validity. The cumulative ablation design adopted in this study evaluates progressively enriched analytical workflows rather than isolated analytical components. Consequently, statistically significant differences among cumulative configurations should be interpreted as incremental framework-level improvements rather than independent causal estimates of each analytical capability. In addition, although the proposed framework defines a closed analytical loop connecting Process Mining, Explainable Machine Learning, Complex Network Analysis, Digital Twin simulation, and Future-State Value Stream Design, the present study evaluates this workflow in an offline analytical setting. Continuous online deployment, recurrent Digital Twin recalibration, and streaming event-log processing remain important directions for future research.

Despite these limitations, the combination of controlled synthetic experimentation, heterogeneous real-world datasets, temporal validation protocols, statistical analysis, and cumulative framework evaluation provides a rigorous and reproducible basis for assessing the effectiveness of the proposed data-driven VSM/VSD framework.

5 Results

This section presents the experimental evaluation of the proposed data-driven Value Stream Mapping and Value Stream Design (VSM/VSD) framework. The results are organized into four complementary parts. First, the framework is evaluated under controlled conditions using the synthetic manufacturing event log, allowing quantitative analysis of the cumulative framework configurations through Digital Twin simulation. Second, the framework is assessed on three heterogeneous real-world event logs to evaluate its generalizability, predictive capability, and structural analysis across different operational domains. Third, the cumulative ablation study quantifies the incremental operational improvements obtained as additional analytical capabilities are progressively incorporated into the framework. Finally, statistical analyses determine whether the observed operational improvements exceed the variability introduced by stochastic simulation.

5.1 Experimental Results on the Synthetic Manufacturing Event Log

The synthetic manufacturing event log provides a controlled environment for evaluating the complete analytical workflow proposed in this study. Unlike the real-world event logs, which are primarily employed to assess predictive performance and framework generalizability, the synthetic dataset enables controlled experimentation through Digital Twin simulation, cumulative ablation analysis, and quantitative comparison of alternative framework configurations under reproducible operating conditions.

Six cumulative experimental configurations (G0–G5) were evaluated, ranging from the baseline operational process to the complete framework integrating Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis. As described in Section 4.2, each configuration extends the previous one by incorporating an additional analytical capability while preserving all previously introduced components. Consequently, the reported improvements should be interpreted as cumulative framework-level gains rather than isolated causal effects attributable to individual analytical components.

Figure 3 summarizes the principal experimental results obtained on the synthetic manufacturing event log. The dashboard integrates complementary evidence describing the cumulative behavior of the proposed framework, including operational performance, predictive modeling, simulation-based evaluation, and structural process analysis.

5.2 Synthetic Manufacturing Dataset

The proposed data-driven VSM/VSD framework was first evaluated using a synthetic manufacturing event log designed to reproduce realistic production flows, waiting times, rework activities, quality events, and resource interactions under controlled experimental conditions. Six analytical configurations were considered, ranging from a baseline representation of the process to the complete framework integrating Process Mining (PM), automated Value Stream Mapping (VSM), Machine Learning (ML), Digital Twin (DT) simulation, and Complex Network Analysis (CNA). The main results are summarized in Figure 2.

Figure 2A shows a progressive increase in the overall framework score as additional analytical capabilities are incorporated into the decision-support process. The baseline configuration achieved the lowest score (11.6), whereas the complete framework reached the highest performance level (50.1). Intermediate configurations exhibited incremental gains, indicating that Process Mining, Machine Learning, Digital Twins, and Complex Network Analysis provide complementary capabilities when integrated within a unified VSM/VSD environment.

The comparison between the Full Framework and the baseline configuration (Figure 2B) reveals substantial improvements across all operational indicators. As summarized in Table 7, lead time was reduced by 32.4%, work-in-process inventory by 45.1%, defect rate by 27.6%, and rework rate by 45.5%, while throughput increased by 71.7%. These results indicate that the combined analytical framework not only improves process visibility but also supports the identification and evaluation of operational improvement opportunities.

The evolution of lead time and throughput across the six experimental configurations is illustrated in Figure 2C. A monotonic reduction in lead time is observed as more analytical layers are incorporated, while throughput consistently increases. The largest incremental performance gains are observed in the PM-VSM-XGB-DT and Full Framework configurations. These results indicate that configurations incorporating simulation-based evaluation and structural network information are associated with additional performance improvements beyond those obtained from descriptive and predictive process analysis.

Furthermore, the ablation analysis demonstrates that each cumulative configuration is associated with incremental performance gains. The progression from Manual VSM to PM-VSM, PM-VSM-XGB, PM-VSM-XGB-DT, and finally the Full Framework indicates that the proposed architecture benefits from the complementary capabilities of descriptive, predictive, simulation-based, and network-oriented analytics.

The automatically generated current-state VSM (Figure 2D) identified *Functional Testing* and *Assembly* as the activities contributing the largest proportion of total process time. In particular, waiting times represent a substantial component of overall lead time, indicating the existence of improvement opportunities associated with queue management and flow synchronization.

The Machine Learning layer demonstrated strong predictive capabilities for bottleneck and delay identification tasks (Figure 2E). Both Random Forest and XGBoost achieved high F1-scores for bottleneck prediction, while strong performance was obtained for delay prediction. The defect prediction task exhibited considerably lower performance, indicating either limited class representation or reduced predictive separability within the synthetic dataset.

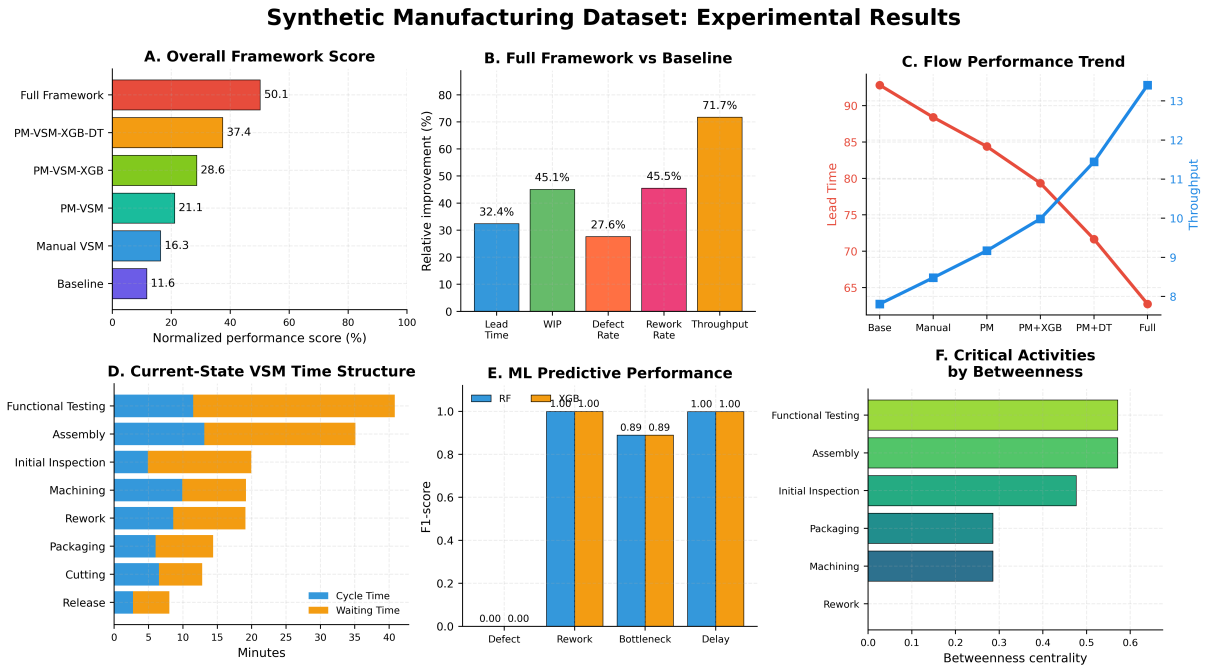


Fig. 2 Experimental results obtained using the synthetic manufacturing dataset. (A) Composite performance score achieved by each experimental configuration, showing the progressive benefits associated with incorporating Process Mining, Machine Learning, Digital Twins, and Complex Networks into the proposed framework. (B) Relative improvement of the Full Framework (G5) with respect to the baseline configuration (G0). (C) Evolution of lead time and throughput across the evaluated configurations. (D) Current-state Value Stream Map (VSM) time decomposition, highlighting value-added and non-value-added activities. (E) Predictive performance of the Random Forest and XGBoost models. (F) Critical process activities identified using betweenness centrality, revealing structurally influential flow-control points within the production system.

Table 7 Performance improvement achieved by the complete framework (G5) relative to the baseline configuration (G0) using the synthetic manufacturing dataset.

Metric	Improvement (%)
Lead Time Reduction	32.4
Work-in-Process Reduction	45.1
Defect Rate Reduction	27.6
Rework Rate Reduction	45.5
Throughput Increase	71.7

The Complex Network Analysis results (Figure 2F) identified *Functional Testing*, *Assembly*, and *Initial Inspection* as the most influential activities according to betweenness centrality. These activities also appear among the largest contributors to process time in the generated VSM, providing complementary evidence from both flow-oriented and network-oriented perspectives. This convergence suggests that these activities represent structurally relevant elements whose improvement may influence system-wide process behavior.

Overall, the synthetic manufacturing experiments show that the progressive incorporation of Process Mining, Machine Learning, Digital Twin simulation, and Complex Network Analysis is associated with incremental improvements across the evaluated operational indicators. The

complete configuration achieves the strongest performance among the evaluated alternatives, supporting the complementary role of these analytical capabilities within the proposed VSM/VSD workflow.

5.3 Results on Real-World Event Logs

To evaluate the generalizability of the proposed framework beyond controlled experimentation, the methodology was applied to three heterogeneous real-world event logs exhibiting different operational characteristics, process complexity, and organizational contexts: BPI2017, BPI2019, and the Data Quality Issues (DQI) dataset. Collectively, these datasets provide a challenging evaluation environment that includes large-scale business processes, procurement workflows, and manufacturing operations affected by quality-related issues.

Table 8 summarizes the principal characteristics of the evaluated event logs. BPI2019 represents the largest dataset in terms of cases, activities, and organizational resources, whereas DQI corresponds to a substantially smaller manufacturing process characterized by pronounced quality-related issues. The considerable differences in scale, process structure, and operational behavior make these datasets appropriate for evaluating the robustness and generalizability of the proposed framework.

Figure 3A reveals substantial differences in process behavior across the evaluated datasets. BPI2017 and BPI2019 exhibit relatively low Value-Added Ratios of 28.0% and 21.2%, respectively, indicating that a considerable proportion of the observed lead time is associated with waiting, coordination, synchronization delays, or other non-value-added activities. Their corresponding mean lead times of 387.6 and 373.6 hours further illustrate the long execution horizons that characterize these administrative processes.

In contrast, the DQI dataset presents a Value-Added Ratio close to 100%. This value should not be interpreted as evidence of near-perfect Lean performance. Rather, it reflects the timestamp semantics and waiting-time information available in the original event log, which record relatively little measured waiting time. This interpretation is particularly important because the same dataset simultaneously exhibits a rework rate of 74.8%, revealing substantial process instability that cannot be inferred from conventional time-based Value Stream Mapping indicators alone. Accordingly, this dataset illustrates the importance of complementing traditional VSM analysis with quality-related, predictive, and structural information.

Complex Network Analysis was subsequently applied to characterize structural relationships among process activities. The resulting activity-network properties are summarized in Table 9.

As illustrated in Figure 3B, BPI2017 exhibits the highest activity-network density (0.5862), indicating comparatively strong connectivity among process activities. BPI2019 presents an intermediate density (0.3556), whereas DQI shows the sparsest network structure (0.2429). The structurally relevant activities identified through betweenness centrality analysis occupy strategically important positions within the observed activity-flow networks and provide complementary structural information beyond conventional flow-oriented Value Stream Mapping indicators.

To evaluate the predictive contribution of structural information, Machine Learning models were trained using two matched experimental configurations: one employing only operational process features and another augmenting the same feature set with Complex Network Analysis descriptors. Following the leakage-controlled protocol described in Section 4.4, all temporal partitions, proxy-label thresholds, and predictor exclusions were derived exclusively from training data before evaluation on the independent testing partition.

Table 10 summarizes the resulting predictive performance.

Table 10 together with Figure 3C summarize the predictive performance obtained on the three real-world event logs. For every prediction task, two matched experimental configurations were evaluated. The first configuration employed only conventional process-derived operational features, whereas the second incorporated additional network-derived structural descriptors obtained through Complex Network Analysis. Consequently, the reported differences quantify the incremental predictive information provided by structural process characteristics beyond conventional Value Stream Mapping indicators.



Fig. 3 Summary of the experimental results obtained on the BPI2017, BPI2019, and DQI datasets. The dashboard integrates dataset characteristics, activity-network density, leakage-controlled predictive performance with and without network-derived features, cumulative framework evaluation, ablation improvements, and structurally relevant activities identified through Complex Network Analysis.

Table 8 Descriptive characteristics of the real-world event logs employed for external validation of the proposed framework.

Dataset	Events	Cases	Activities	Resources	Observation Hours
BPI 2017	249,533	31,469	26	147	9,530.57
BPI 2019	250,000	155,771	39	563	622,190.53
Data Quality Issues	8,601	271	21	15	5,977.72

For the delay prediction task, the testing partitions of BPI2017 and BPI2019 contained a single target class; consequently, F1-scores were not reported for those datasets. The DQI dataset was evaluable and its F1-score increased from 0.121 to 0.271 for Random Forest and from 0.000 to 0.232 for XGBoost after network-derived features were incorporated. These results suggest that structural relationships among activities can provide useful complementary information in manufacturing environments exhibiting substantial process variability and quality-related interactions.

For bottleneck prediction, BPI2017 and BPI2019 also contained single-class testing targets and were therefore not evaluable. On DQI, both Random Forest and XGBoost reached an F1-score of 0.995 after network-derived structural descriptors were incorporated, compared with 0.643 and 0.000, respectively, using operational features alone. This behavior should be interpreted carefully. The bottleneck proxy employed in this study is defined from activity-level structural properties, and the corresponding predictive models intentionally exclude betweenness centrality itself from the predictor set to avoid direct target leakage. Consequently, the observed performance reflects the predictive value of the remaining network-derived descriptors together with conventional operational attributes rather than trivial reconstruction of the proxy label.

For the rework prediction task, the influence of Complex Network Analysis depended on the

Table 9 Structural properties of the activity networks derived from the evaluated real-world event logs. Reported activities correspond to the nodes with the highest structural relevance according to betweenness centrality.

Dataset	Nodes	Edges	Density	WCC	Structurally Relevant Activity
BPI 2017	26	381	0.5862	1	`W_Call incomplete files`
BPI 2019	39	527	0.3556	1	`Record Invoice Receipt`
Data Quality Issues	21	102	0.2429	1	`/vgr/pick_up_and_transport`

Table 10 Leakage-controlled predictive performance on the real-world event logs. F1-scores compare models using operational features only (No CNA) against the same models augmented with Complex Network Analysis (CNA) features.

Dataset	Target	Model	F1 (No CNA)	F1 (CNA)	ΔF1	Status
BPI2017	Delay	Random Forest	N/A	N/A	N/A	Single-class target
BPI2017	Delay	XGBoost	N/A	N/A	N/A	Single-class target
BPI2017	Bottleneck	Random Forest	N/A	N/A	N/A	Single-class target
BPI2017	Bottleneck	XGBoost	N/A	N/A	N/A	Single-class target
BPI2017	Rework	Random Forest	0.2243	1.0000	0.7757	Evaluable
BPI2017	Rework	XGBoost	0.0000	1.0000	1.0000	Evaluable
BPI2019	Delay	Random Forest	N/A	N/A	N/A	Single-class target
BPI2019	Delay	XGBoost	N/A	N/A	N/A	Single-class target
BPI2019	Bottleneck	Random Forest	N/A	N/A	N/A	Single-class target
BPI2019	Bottleneck	XGBoost	N/A	N/A	N/A	Single-class target
BPI2019	Rework	Random Forest	0.4658	0.7451	0.2793	Evaluable
BPI2019	Rework	XGBoost	0.0062	0.7453	0.7391	Evaluable
Data Quality Issues	Delay	Random Forest	0.1209	0.2714	0.1505	Evaluable
Data Quality Issues	Delay	XGBoost	0.0000	0.2320	0.2320	Evaluable
Data Quality Issues	Bottleneck	Random Forest	0.6432	0.9951	0.3519	Evaluable
Data Quality Issues	Bottleneck	XGBoost	0.0000	0.9951	0.9951	Evaluable
Data Quality Issues	Rework	Random Forest	0.9362	0.9469	0.0107	Evaluable
Data Quality Issues	Rework	XGBoost	0.9399	0.9455	0.0055	Evaluable

characteristics of each dataset. BPI2019 exhibited substantial improvements after incorporating structural features, with the F1-score increasing from 0.466 to 0.745 for Random Forest and from 0.006 to 0.745 for XGBoost. In contrast, the DQI dataset already achieved very high predictive performance using conventional operational features alone (F1 $\approx$ 0.94), leaving only limited room for additional improvement. This behavior is consistent with the event-log characteristics of DQI, where rework is strongly associated with specific operational activities and is therefore largely identifiable from process-derived attributes.

Overall, the predictive experiments demonstrate that network-derived structural information provides complementary information beyond conventional operational indicators. However, the magnitude of the observed improvements depends on the operational characteristics of each process and the prediction task under consideration. Rather than uniformly improving all models, Complex Network Analysis contributes additional predictive evidence when structural dependencies among activities contain information that is not fully captured by traditional Value Stream Mapping features alone.

These observations support the role of Complex Network Analysis as a complementary analytical layer within the proposed framework. Instead of replacing conventional Lean indicators or Process Mining outputs, network-derived descriptors enrich the representation of process behavior by capturing structural interactions among activities, thus improving predictive diagnosis under appropriate operational conditions.

5.4 Cumulative Ablation Study

The cumulative ablation study quantifies the progressive operational improvements obtained as increasingly rich analytical capabilities are incorporated into the proposed framework. Unlike

conventional component-removal ablation studies, the objective of the present evaluation is not to estimate the isolated causal contribution of individual analytical modules, but rather to assess the cumulative operational benefits achieved by progressively enriching the decision-support workflow.

Figure 4 summarizes the operational performance of the six cumulative experimental configurations (G0–G5), whereas Table 11 reports the corresponding quantitative improvements relative to the baseline configuration.

Across all evaluated operational indicators, the proposed framework exhibits a consistent pattern of progressive improvement as additional analytical capabilities are incorporated. The baseline configuration (G0) represents the original operational process without advanced analytical support, whereas G5 corresponds to the complete framework integrating Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis.

The transition from manual Value Stream Mapping (G1) to Process Mining-based Value Stream Mapping (G2) provides the first substantial operational improvement by replacing manually constructed process representations with data-driven current-state models extracted directly from operational event logs. This automated process reconstruction establishes the analytical foundation for the predictive and simulation-based components incorporated in the subsequent cumulative configurations.

Introducing Explainable Machine Learning (G3) further improves operational performance by enabling predictive identification of delay, bottleneck, and rework conditions before evaluating alternative process interventions. Rather than directly modifying process behavior, this analytical capability supports more informed selection of candidate improvement actions that are subsequently evaluated through Digital Twin simulation.

The incorporation of Digital Twin simulation (G4) enables quantitative evaluation of future-state Value Stream Design alternatives before implementation. Unlike conventional Value Stream Mapping, which typically relies on qualitative assessment of proposed improvements, the Digital Twin provides a simulation-based mechanism for estimating the operational consequences of alternative interventions while explicitly accounting for process variability, resource utilization, waiting times, and stochastic operational behavior.

The complete framework (G5) consistently achieves the best operational performance among the evaluated cumulative configurations. Relative to the baseline process, Lead Time is reduced by 32.40%, Work-In-Process decreases by 45.09%, Defect Rate decreases by 27.63%, and Rework Rate decreases by 45.52%, while Throughput increases by 71.73%. Collectively, these improvements demonstrate that the proposed framework supports simultaneous enhancement of multiple Lean-oriented operational indicators rather than optimizing a single performance measure.

Figure 4 further illustrates that these improvements are cumulative rather than abrupt. Each successive configuration contributes additional analytical capabilities that extend the previous workflow, resulting in progressively improved operational performance throughout the experimental sequence. Consequently, the observed gains should be interpreted as framework-level cumulative improvements rather than isolated causal effects attributable to individual analytical components.

5.5 Statistical Validation of the Simulation Results

To determine whether the operational improvements observed throughout the cumulative ablation study exceed the variability introduced by stochastic simulation, inferential statistical analyses were performed using the 30 independent Digital Twin replications executed for each cumulative configuration (G0–G5). The analyses focused on the principal Lean-oriented performance indicators, namely Lead Time, Throughput, Work-In-Process, Defect Rate, and Rework Rate.

For each performance indicator, the assumptions required for parametric analysis were evaluated prior to statistical testing. When appropriate, one-way Analysis of Variance (ANOVA) was employed to determine whether significant differences existed among the cumulative configurations, followed by Tukey's Honestly Significant Difference (HSD) test for pairwise comparisons. For indicators that did not satisfy the assumptions

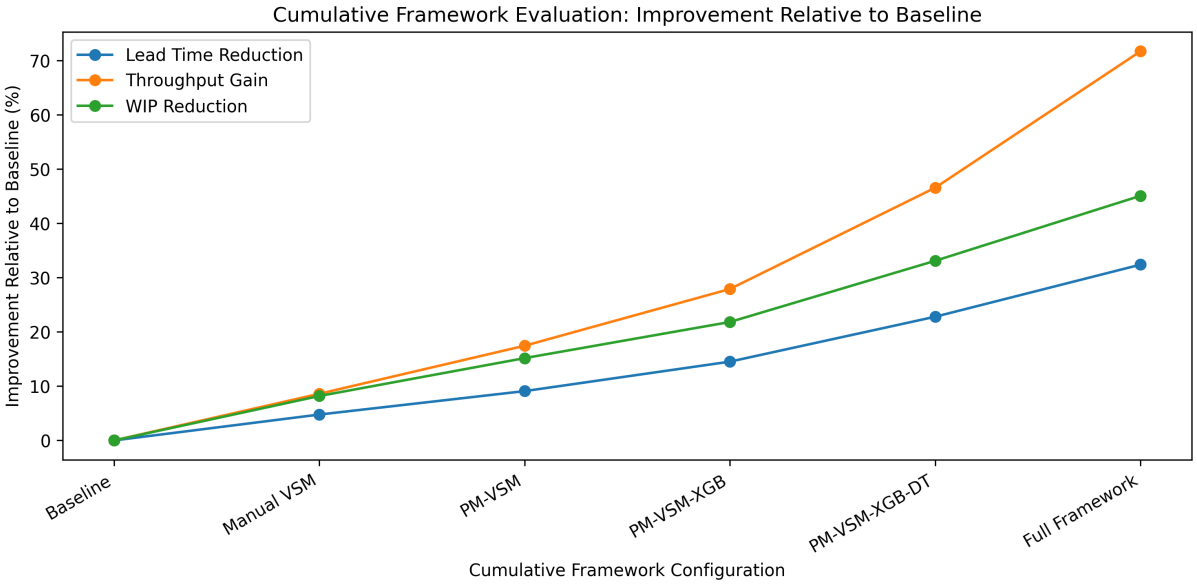


Fig. 4 Progressive operational improvements obtained across the cumulative framework configurations relative to the baseline process. The cumulative experimental design progressively incorporates Process Mining (PM), Explainable Machine Learning and Explainable Artificial Intelligence (XAI), Digital Twin (DT) simulation, and Complex Network Analysis (CNA), illustrating their combined contribution to Lean performance.

Table 11 Operational improvements obtained across the cumulative framework configurations relative to the baseline process.

Configuration	**LT reduction (%)**	**TH gain (%)**	**WIP reduction (%)**
G0 – Baseline	0.0	0.0	0.0
G1 – Manual VSM	4.8	8.6	8.2
G2 – PM-VSM	9.1	17.4	15.2
G3 – PM-VSM + XAI	14.5	27.9	21.8
G4 – PM-VSM + XAI + DT	22.8	46.4	33.1
G5 – Complete framework	32.4	71.7	45.1

required for parametric inference, the Kruskal–Wallis non-parametric test was used instead. Statistical significance was assessed using a confidence level of 95% ($\alpha = 0.05$).

The statistical analyses consistently indicate that the cumulative framework configurations produce significantly different operational performance. In particular, the complete framework (G5) exhibits statistically significant improvements over the baseline configuration (G0) for all evaluated operational indicators, supporting the hypothesis that progressively enriching the analytical workflow leads to measurable operational benefits.

The observed statistical differences are consistent with the cumulative performance trends illustrated in Figure 4. Rather than producing isolated improvements restricted to individual indicators, the proposed framework simultaneously reduces Lead Time, Work-In-Process, Defect Rate, and Rework Rate while increasing Throughput. Such behavior is particularly relevant in Lean-oriented process improvement because operational interventions frequently improve one indicator at the expense of another. The proposed framework instead demonstrates balanced improvements across multiple performance dimensions.

It is important to emphasize that these statistical results should be interpreted within the context of the cumulative experimental design. Because each configuration progressively incorporates additional analytical capabilities while preserving those introduced previously, the observed differences represent incremental framework-level improvements rather than independent causal estimates of individual components.

Collectively, the statistical analyses support Hypothesis H1 by showing that the complete framework achieves statistically significant improvements in operational performance relative to conventional Value Stream Mapping approaches. Likewise, the predictive experiments reported in Section 5.3 support Hypothesis H2 by demonstrating that network-derived structural features improve predictive performance under the leakage-controlled evaluation protocol. Finally, the cumulative ablation study together with the Digital Twin simulation results support Hypothesis H3 by demonstrating that progressively integrating Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis enables quantitatively supported Future-State Value Stream Design.

6 Discussion

The experimental results demonstrate that Value Stream Mapping can be extended beyond static process visualization by establishing a traceable analytical workflow that integrates operational event data with current-state reconstruction, predictive process diagnosis, Digital Twin simulation, and Future-State Value Stream Design. Rather than treating Process Mining, Explainable Machine Learning, Digital Twin simulation, and Complex Network Analysis as independent analytical technologies, the proposed framework organizes them into a coherent decision-support pipeline that progressively transforms raw event logs into quantitative evidence supporting process redesign.

Across both the synthetic manufacturing environment and the three real-world event logs, the proposed methodology consistently preserved the logical sequence of Lean improvement activities. Process Mining automatically reconstructed the current-state process, Explainable Machine Learning identified operational risk factors while maintaining model interpretability, Complex Network Analysis revealed structurally relevant activities within the process, and Digital Twin simulation evaluated alternative improvement scenarios before implementation. This sequential integration substantially reduces the subjectivity traditionally associated with manual Value Stream Mapping while preserving the decision logic employed during Lean improvement initiatives.

The cumulative ablation study further demonstrates that progressively enriching the analytical workflow produces consistent operational improvements across all evaluated Lean performance indicators. Relative to the baseline configuration, the complete framework reduced Lead Time by 32.40%, Work-In-Process by 45.09%, Defect Rate by 27.63%, and Rework Rate by 45.52%, while increasing Throughput by 71.73%. Moreover, the complete configuration consistently achieved the best overall operational performance among the evaluated cumulative configurations, and the corresponding statistical analyses indicate that these improvements substantially exceed the variability introduced by stochastic simulation.

Importantly, these improvements should not be interpreted as the isolated contribution of individual analytical modules. The cumulative experimental design intentionally evaluates progressively richer decision-support capabilities, meaning that each configuration preserves all previously incorporated analytical components while extending the overall workflow. Consequently, the reported improvements represent the operational benefit of increasing analytical integration rather than the independent effect of Process Mining, Explainable Machine Learning, Digital Twin simulation, or Complex Network Analysis alone. This distinction is essential for correctly interpreting the experimental findings and avoiding causal attributions that are not supported by the experimental design.

From a methodological perspective, this cumulative interpretation differentiates the proposed framework from many existing studies that evaluate analytical technologies independently. Previous research has largely focused on isolated contributions such as automated process discovery, predictive maintenance, anomaly detection, or simulation-based optimization. In contrast, the

present framework integrates these complementary analytical capabilities into a unified decision-support workflow spanning current-state analysis, predictive diagnosis, structural process analysis, simulation-based intervention assessment, and Future-State Value Stream Design. The contribution therefore lies not only in combining multiple analytical techniques, but also in organizing them into a coherent methodology that supports continuous process improvement from data acquisition to operational redesign.

An additional observation emerging from the experimental evaluation is that operational improvements are not exclusively associated with higher predictive accuracy. Although Explainable Machine Learning contributes valuable predictive information, its effectiveness increases considerably when combined with Process Mining, Complex Network Analysis, and Digital Twin simulation. Machine Learning estimates operational risks, Process Mining provides process transparency, Complex Network Analysis identifies structurally influential activities, and Digital Twin simulation evaluates the operational consequences of alternative interventions. Their integration therefore enables evidence-based decisions that cannot be obtained by optimizing each analytical component independently.

The evaluation performed on heterogeneous real-world event logs further suggests that the proposed methodology is not restricted to manufacturing environments. Despite representing different application domains, including business processes, procurement workflows, and manufacturing quality management, all evaluated datasets could be processed using the same standardized event-log representation and analytical pipeline. This cross-domain applicability indicates that the proposed framework constitutes a general methodology for data-driven process improvement rather than a domain-specific solution, increasing its potential applicability across a broad range of industrial and organizational settings.

From an industrial perspective, the proposed framework provides a practical mechanism for supporting Lean transformation initiatives under Industry 4.0 and Industry 5.0 paradigms. Traditional Value Stream Mapping frequently depends on manual observations, expert knowledge, and static process representations that rapidly become outdated as production systems evolve. By automatically reconstructing operational processes from event logs and continuously integrating predictive analytics, structural network information, and Digital Twin simulation, the proposed methodology enables organizations to maintain an up-to-date digital representation of operational performance while quantitatively evaluating improvement alternatives before their implementation. Such capabilities can substantially reduce the uncertainty and cost associated with continuous improvement initiatives.

Despite these promising results, several limitations should be acknowledged. First, although the proposed framework was evaluated using multiple real-world datasets from different domains, the Digital Twin optimization experiments were conducted using a synthetic manufacturing environment to enable controlled experimentation and reproducible comparisons across cumulative framework configurations. Additional validation using large-scale industrial Digital Twins would further strengthen the generalizability of the proposed methodology. Second, although the cumulative experimental design effectively demonstrates the operational value of progressively integrating complementary analytical capabilities, it does not estimate the isolated causal contribution of each analytical component. Future studies may therefore complement the present evaluation with factorial experimental designs or sensitivity analyses that explicitly quantify individual module contributions. Finally, the current implementation focuses on offline event-log analysis; extending the framework toward real-time streaming process monitoring and online Digital Twin synchronization represents an important direction for future research.

Overall, the experimental evidence suggests that integrating Process Mining, Explainable Machine Learning, Complex Network Analysis, and Digital Twin simulation into a unified Value Stream Mapping and Design workflow constitutes an effective strategy for supporting evidence-based process improvement. Rather than replacing traditional Lean methodologies, the proposed framework augments them with quantitative analytical capabilities that improve process understanding, reduce decision uncertainty, and support the systematic design and evaluation of future operational scenarios.

7 Conclusions and Future Work

This paper presented a data-driven framework for Value Stream Mapping (VSM) and Value Stream Design (VSD) that establishes a traceable analytical workflow connecting operational event data with current-state reconstruction, predictive process diagnosis, simulation-based evaluation, and Future-State Value Stream Design. Rather than combining multiple analytical techniques in isolation, the proposed methodology integrates Process Mining, Complex Network Analysis, Explainable Machine Learning, and Digital Twin simulation into a unified decision-support framework supporting successive stages of Lean process improvement.

The experimental evaluation combined a controlled synthetic manufacturing event log with three heterogeneous real-world event logs to assess both operational improvement and methodological generalizability. The predictive experiments demonstrated that network-derived structural descriptors provide complementary information beyond conventional operational variables under a leakage-controlled temporal evaluation protocol. However, the magnitude of the observed improvements depended on the prediction task and process characteristics, indicating that the contribution of structural information is context dependent rather than universally beneficial.

The cumulative simulation-based evaluation showed that progressively enriching the analytical workflow consistently improves operational performance. Relative to the baseline configuration, the complete framework reduced Lead Time by 32.40%, Work-In-Process by 45.09%, Defect Rate by 27.63%, and Rework Rate by 45.52%, while increasing Throughput by 71.73%. Statistical analyses confirmed that these improvements are significant across the evaluated cumulative configurations. Because the experimental design is cumulative, these results should be interpreted as evidence of incremental framework-level improvements rather than isolated causal contributions of individual analytical components.

The principal contribution of this work is therefore methodological rather than algorithmic. The proposed framework establishes a closed analytical workflow linking Process Mining, structural process analysis, interpretable predictive analytics, and Digital Twin simulation into a coherent methodology for supporting evidence-based Future-State Value Stream Design. In doing so, it extends conventional Value Stream Mapping from a predominantly descriptive Lean tool toward a reproducible, data-driven decision-support methodology capable of quantitatively evaluating process improvement alternatives before implementation.

Several limitations define opportunities for future research. First, although the predictive framework was validated using multiple real-world event logs, the complete diagnosis–simulation–redesign cycle was evaluated within a controlled synthetic manufacturing environment. Industrial case studies involving implemented process improvements remain necessary to validate the framework under operational conditions. Second, the bottleneck prediction task relies on a structurally defined proxy, motivating future validation using independently observed operational bottleneck labels derived from production systems. Third, the cumulative ablation strategy demonstrates the value of progressively integrating complementary analytical capabilities but does not isolate their individual causal contributions. Future studies may therefore employ factorial experimental designs, component-removal analyses, or sensitivity studies to quantify the independent contribution of each analytical module.

Additional research directions include extending the framework toward real-time operation through MES, IIoT, and streaming event-data infrastructures, enabling continuous Digital Twin recalibration and recurrent VSM/VSD analysis. Furthermore, graph neural networks, temporal learning architectures, optimization algorithms, and reinforcement learning may provide promising mechanisms for automatically generating and evaluating candidate future-state process redesign strategies while explicitly considering operational constraints, implementation costs, and managerial objectives.

Overall, the proposed framework provides a reproducible foundation for integrating operational data, structural process analysis, predictive analytics, and Digital Twin simulation into a unified Lean decision-support methodology. The

results demonstrate that combining complementary analytical perspectives enables quantitatively supported process redesign while preserving the logic of conventional Value Stream Mapping, then contributing toward the development of digitally enabled production systems under Industry 4.0 and Industry 5.0 paradigms.

Statements and Declarations

Competing Interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Author Contributions

Edwin Montes-Orozco: Conceptualization, Methodology, Software, Formal Analysis, Writing – Original Draft, Visualization. Mariam Dopslaf: Conceptualization, Supervision, Validation, Writing – Review & Editing. Roman Anselmo Mora-Gutiérrez: Validation, Formal Analysis, Writing – Review & Editing. Sergio Gerardo de-los-Cobos-Silva: Methodology, Validation, Writing – Review & Editing. Eric Alfredo Rincón-García: Formal Analysis, Validation, Writing – Review & Editing. Pedro Lara-Velázquez: Validation, Supervision, Writing – Review & Editing. Miguel Ángel Gutiérrez-Andrade: Validation, Supervision, Writing – Review & Editing.

Data Availability

The datasets used in this study are publicly available through their respective repositories. The code and implementation materials supporting this study are available from the corresponding author upon reasonable request and will be deposited in a public repository upon acceptance.

Ethics Approval and Consent to Participate

Not applicable. This study did not involve human participants, human data, or animals.

Consent for Publication

Not applicable.

References

[1] van der Aalst WMP (2016) Process Mining: Data Science in Action, 2nd edn. Springer

[2] Anwer N, Stark R, Tao F, et al (2025) Developing and leveraging digital twins in engineering design. CIRP Annals 74(2):843–868

[3] Braun D, Müller T, Sahlab N, et al (2022) A graph-based knowledge representation and pattern mining supporting the digital twin creation of existing manufacturing systems. In: 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, pp 1–4

[4] Breiman L (2001) Random forests. Machine Learning 45:5–32

[5] Chen L, Yao X, Liu K, et al (2023) Multisensor fusion-based digital twin in additive manufacturing for in-situ quality monitoring and defect correction. Proceedings of the Design Society 3:2755–2764

[6] Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794, https://doi.org/10.1145/2939672.2939785

[7] Chen YP, Karkaria V, Tsai YK, et al (2025) Real-time decision-making for digital twin in additive manufacturing with model predictive control using time-series deep neural networks. Journal of Manufacturing Systems 80:412–424

[8] Freeman LC (1978) Centrality in social networks: Conceptual clarification. Social Networks 1(3):215–239

[9] Grieves M, Vickers J (2017) Digital twin: Mitigating unpredictable, undesirable emergent

behavior in complex systems. In: Transdisciplinary Perspectives on Complex Systems. Springer

[10] Horsthofer-Rauch J, Schumann M, Milde M, et al (2022) Digitalized value stream mapping: Review and outlook. Procedia CIRP 112:244–249. https://doi.org/10.1016/j.procir.2022.09.079

[11] Horsthofer-Rauch J, Guesken SR, Weich J, et al (2024) Sustainability-integrated value stream mapping with process mining. Production and Manufacturing Research 12(1):2334294. https://doi.org/10.1080/21693277.2024.2334294

[12] Khakpour R, Ebrahimi A, Seyed-Hosseini SM (2025) Lean process mining: Adopting process mining in lean manufacturing for dynamic process mapping and avoiding waste occurrence in real time. International Journal of Lean Six Sigma 16(1):231–255. https://doi.org/10.1108/IJLSS-03-2024-0059

[13] Kroeger S, Rafles A, Jordan P, et al (2025) Data model to enable multidimensional process mining for data farming based value stream planning in production networks. Production Engineering 19(2):307–327. https://doi.org/10.1007/s11740-024-01308-9

[14] Mortlock T, Muthirayan D, Yu SY, et al (2021) Graph learning for cognitive digital twins in manufacturing systems. IEEE Transactions on Emerging Topics in Computing 10(1):34–45

[15] Newman M (2018) Networks. Oxford University Press

[16] Reslan M, Triebe MJ, Venketesh R, et al (2025) Automation of value stream mapping: A case study on enhancing lean manufacturing tools through digital twins. Procedia CIRP 134:455–460. https://doi.org/10.1016/j.procir.2025.02.159

[17] Rother M, Shook J (2003) Learning to See: Value Stream Mapping to Add Value and Eliminate Muda. Lean Enterprise Institute

[18] Teriete T, Erlach K, Bauernhansl T, et al (2026) Combining value stream mapping and process mining in production: A systematic literature review. Discover Applied Sciences 8:285. https://doi.org/10.1007/s42452-026-08446-y