AI-Enabled Theory of Constraints: A Two-Case Study for Bottleneck Detection and Production Flow Improvement in Manufacturing

Ass. Prof. Dr. Thamer Okab Hawas

Country Department of Business Administration and Economics, University of Tikrit, Iraq.

DOI:10.37648/ijtbm.v16i01.004

[1]*Received: 27 November 2025; Accepted: 28 December 2025; Published: 29 January 2026*

Abstract

This research paper explores the role of artificial intelligence (AI) algorithms in bottleneck detection and constraint management according to the Tank Theory of Constraints (TOC) in manufacturing. One ongoing management problem is that our constraints are often observed after the fact, i.e., after throughput has already been lost, simply because signals of downtime can be scattered among products, shifts, operators and operating conditions. To close this gap, we switch to a two–case study where we connect operational data and TOC-relevant, decision-ready information. In Case 1, production and downtime data are combined to analyse constraint effect (e.g., total amount of downtime minutes, downtime share and efficiency loss) and pinpoint "vital few" causes of flow interruption via a Pareto-style loss analysis; predictive modelling is applied next for estimating downtime magnitude and ranking improvement actions. In Case 2, a predictive maintenance context is adopted to predict failure-induced perturbations endangering the constraint with a train-only balancing strategy and an untouched test set leveraging real-world class imbalance and decision realism; explainability delivers most interesting risk drivers allowing for targeted preventive actions. Results indicate that AI algorithms can improve TOC practice by increasing bottleneck visibility, earlier actions and prioritization of constraint-focused improvements systematic manner. The study contributes a business-oriented framework that would serve as an intermediate between AI outputs and executable TOC decisions that safeguard throughput and stabilise the production flow.

Keywords: *Theory of Constraints (TOC); Bottleneck Detection; Downtime Analysis; Predictive Maintenance; Machine Learning; Explainable AI (XAI); Decision Support.*

1. Introduction

Manufacturing companies face a continuous operations paradox: even with continued investments in automation and lean initiatives, production flow is still subject to limiting few capacity constraint paradoxes (bottlenecks), that propagate delays, increase work-in-process (WIP) and decrease delivery reliability. (Hopp & Spearman., 2011; Thürer et al., 2017). The Theory of Constraints (TOC) makes a case for this phenomenon as it suggests that the performance of an entire system is heavily dependent on both the current constraint from which the performance is derived and the

---

[1] *How to cite the article:* Hawas T.O. (January 2026); AI-Enabled Theory of Constraints: A Two-Case Study for Bottleneck Detection and Production Flow Improvement in Manufacturing; *International Journal of Transformations in Business Management,* Vol 16, Issue 1, 40-68, DOI: http://doi.org/10.37648/ijtbm.v16i01.004

organization's ability to identify, exploit and elevate the latter. (Gupta & Boyd, 2008; Goldratt, 2016) In practice though, many production environments are not static, and constraints are dynamic: the ideal bottleneck can move from machine to machine, product to product, shift to shift or upstream disruption to upstream disruption, making managerial diagnosis non-trivial and frequently reactive. (Kumbhar et al., 2022; Lai et al., 2021; Li et al., 2009).

In business and operational perspective, late serial bottleneck detection has an apparent implication on throughput, unit cost and service level. (Buer, et al., 2018; Ivanov, et al., 2025) When bottlenecks are discovered after queues form and schedules are missed, managers tend to react expensively: with expedited orders, overtime work or rework that may help secure output for the time being, but which can not only impact profitability, but also workforce sustainability. (Helo & Hao, 2017; Tarafdar et al., 2019) Therefore, not only technical problems but also the operational governance problems associated with the performance management and decision quality under uncertainty are the increase of bottleneck visibility and strengthening of early-warning capability (Mikalef et al 2019; Wamba et al., 2017).

At the same time, the adoption of Industry 4.0 has provided greater access to shop-floor data (machine signals, downtime logs, contextual information in the sense of products and operators) to facilitate a transition to an evidence-based constraint management approach (as opposed to experience-based) (Frank et al., 2019; Xu et al., 2018) Recent operations and manufacturing operations display though that data driven approaches can be used to increase the transparency of operations by transforming varied production logs into action indicators of loss, threat and flow subversion (Bokrantz et al., 2020; Ghobakhloo, 2020) Nonetheless, numerous contributions are still technology-oriented (e.g., focusing on algorithms and accuracy), without addressing the managerial challenge that pushes in TOC practice that asks how AI-based analytics can be converted into constraint-based decisions and routines of sustainable improvement (Dubey et al., 2020; Sony & Naik, 2020). Figure 1 visually outlines the conceptual framework of the research by bridging the two concepts that are TOC-based constraint management and AI-based diagnostics, early warning systems, and actionable decision support in manufacturing activities.

Figure 1: AI-driven constraint management in manufacturing: integrating TOC logic with data-driven diagnosis, early warning, and actionable insights to support flow stability and throughput protection.

Artificial intelligence (AI) and machine learning (ML) are being adopted more in manufacturing, i.e., forecast or schedule support, anomaly detection and predictive maintenance. (Lee et al., 2018; Zhang et al., 2019) For predictive maintenance in particular, systematics of reviews demonstrate that ML can identify precursors to a failure down the road as weak signals and hence diminish unplanned downtime (directly related to TOC, where defections often "create" or "move" constraints through sudden capacity reductions). (Carvalho et al., 2019; Zonta et al., 2020) However, deployment of them in operations is often limited by the "last-mile problem", i.e., even good models not being able to impact decisions if outputs are uninterpretable, do not align with operational KPIs or cannot be easily plugged into a decision making adjoined system (Rudin, 2019; Rai, 2020).

Explainable AI (XAI) has gained traction as a manager enable that could allow for transparent model outputs in support of accountability, learning and action. (Arrieta et al., 2020; Vilone & Longo, 2020) Interpretability is particularly important in industry as managers need to justify interventions (maintenance decision, line balancing and change-over policies, training or standard work update) and shop floor decisions are assessed with respect to operational performance such as throughput, stability or delivery. (Moosavi et al., 2024; Holzinger et al., 2020) As a result, the operational sense explicit spanning the visualization of AI is bigger than "predicting better", but also concerns enabling diagnosis (which drives the constraint), prioritization (which causes dominate problems), and governance (how decisions are standardized and monitored) (Barredo et al., 2019; Samek et al., 2019).

This research paper extends a managerial framing of AI within TOC and locates the AI analytics as supportive decision-making tool for constraint identification and constraint protection, rather than mere technical classification or regression exercise (Kumbhar et al., 2022; Subramaniyan et al., 2021). In particular, we define bottleneck management as a two-fold managerial capacity: (1) diagnostic ability to objectively assess and quantify the operation implications of bottlenecks and their most detrimental loss contributors; (2) setting up early-warning signals for events that endanger constraint stability (e.g. spiritual machine failure) (Ivanov & Dolgui, 2021; Wenzel et al., 2020) This resource-based perspective is consistent with operations strategy research that prioritizes the conversion of digital analytics into more effective routines for improving flow, reliability and resource utilization (Buer et al., 2018; Tang et al., 2024).

From a methodological point of view, the research is set up as a two-case study to illustrate how AI makes compatible TOC contributions at different managerial questions and data realities. (Voss et al., 2010) Case 1 makes use of operating production and downtime logs to measure bottleneck losses and ascertain the "vital few" downtimes causes according to a TOC-logical loss focus (e.g., Pareto-type patterns). (Mourtzis et al., 2020) Case 2 builds an early-warning risk signal for machine failure and interprets the drivers with strongest associated on failure, which enable to perform proactive constraint protection using large predictive maintenance datasets.

Our primary contribution is managerial in nature: We demonstrate how AI can reinforce ToC practice by; (a) accelerating and making more reliable the bottleneck detection processions, (b) quantifying the operational impact of constraint-related losses and (c) allowing proactive interventions which will stabilize the constraint variable and preserve throughput. The contribution attends to requests in operations and Industry 4.0 research for the unification of analytics with improvement logics (not showcasing analytics as an end to itself), and draw out implications for operational decision-making, governance. In this way, the research seeks to assist academics and practitioners who are interested in deploying AI in a manner that directly enhances flow, which is conducive for removing bottlenecks and stabilizing production in manufacturing systems.

1.1 Research Questions

RQ1. How can AI-enabled analytics be structured into a TOC-aligned decision-support pathway that makes bottlenecks operationally visible and quantifies flow loss in manufacturing?

RQ2. To what extent can AI-based early warning protect constrained (or near-constrained) resources under rare-event conditions, while preserving deployment realism through an untouched test set?

RQ3. How can explainability outputs be translated into governance-ready operational levers (dashboards, SOP triggers, escalation rules) that support constraint protection and intervention prioritization?

1.2 Contribution

- TOC decision contribution proposes a business-oriented pathway that links heterogeneous shop-floor data to executable TOC steps (identify–exploit–subordinate–elevate) through quantified flow-loss indicators and prioritization logic.
- Deployment-realistic analytics contribution demonstrates a rare-event evaluation design (train-only balancing with an untouched test set) that aligns early-warning assessment with operational prevalence and managerial generalizability.
- Actionability contribution translates predictive outputs into intervention design via explainability, providing a governance-ready basis for monitoring policies and preventive actions that protect throughput-critical resources.

2. Literature Review

2.1 Bottlenecks and the Theory of Constraints in contemporary manufacturing

In the field of operations performance, bottlenecks (often presented as system constraints) are important because they control throughput and lead times and contribute to exacerbating variability in the production process (Telles et al., 2020; de Jesus Pacheco et al.,2021). In the TOC, the constraint is not only a technological limitation of a machine but also a managerial focus which leads to understand where efforts for improvement can produce more system level returns (e.g., enhancement of throughput rather than increasing local efficiency) (Telles et al., 2020; de Jesus Pacheco et al., 2021). Recent research in operations highlights that under Industry 4.0 constraints are still important, but now they are more volatile as a result of product mix volatility, shorter runs and changeovers, multi-skilling labor demand (Telles et al., 2020; de Jesus Pacheco et al., 2021; Luiz et al., 2025).

Post-2018 literature has consistently pointed the difficulty on "finding the constraint" in flow where losses are decomposed into micro-stoppages, set-up loss, quality rework and material interruption deteriorating flows even when no single station is permanently saturated (Telles et al., 2020; de Jesus Pacheco et al., 2021). As a result, TOC practice and operation analytics have more frequently met halfway, via data traces (time stamps, downtime logs, sensor streams) that flow through the enterprise to empirically locate where flow is being blocked and why (Telles et al., 2020; de Jesus Pacheco et al., 2021). This development is consistent with restated contemporary operations management views that emphasize end-to-end flow and delivery reliability, over disconnected resource-availability measures, especially in settings where service guarantee, cost and responsiveness are traded-offs (Telles et al., 2020; Luiz et al. 2025).

2.2 From "static constraints" to data-driven bottleneck detection

Classical bottleneck identification methodologies (e.g. based on utilization heuristics, queue length inspection or deterministic line balancing) are still useful but can be misleading in stochastic and high-mix environments where the bottleneck might "travel" over periods and product families (Subramaniyan et al., 2020; Mahmoodi et al., 2022). In return, the 2018-2025 literature has evolved to data-driven bottleneck detection that combines event logs, down time factorization and time-based measures including active duration, waiting time and loss accumulation (Mahmoodi et al., 2022). Such methods have especially been in harmony with TOC since they make the concept of the constraint an observable mechanism of loss (most causes of downtime) and not in the form of a capacity-like abstract concept (Subramaniyan et al., 2020; Mahmoodi et al., 2022; Roser et al., 2015).

A different stream highlights the applicability of a Pareto-style breakdown of losses, that is, to find a small set of major causes with high proportion of all lost time since it would facilitate managerial priorities and it is compatible with the focusing logic of TOC (Subramaniyan et al., 2020; Telles et al., 2020). In a practical operation sense, bottleneck detection turns into a decision-support task:(i) quantify loss, (ii)assign loss to sources and context (product, shift, operator effect, calendar effects), (iii) select improvement interventions with the highest expected throughput impact (Subramaniyan et al., 2020; Mahmoodi et al., 2022).

2.3 AI-enabled operations analytics and predictive maintenance as "constraint protection"

Against this background, research is beginning to emerge that posits AI as a technology that develops operational resilience predicting disruptions risks and enforcing preventive operations ahead of axiological performance collapses (Carvalho et al., 2019; Zhang et al., 2019; Zonta et al., 2020). Predictive maintenance (PdM) is especially important for TOC since it can be seen as "protecting the constraint" against downtime, which helps to stabilize flow and ensure throughput (Carvalho et al., 2019; Zhang et al., 2019). Two recent systematic reviews stress that ML-based PdM has evolved from PoC models to decision-centric pipelines consisting of pre-processing, balancing and the evaluation strategies which better capture operational rarity (rarity for failures but also costly) (Carvalho et al., 2019; Zhang et al., 2019).

In terms of operation, the product is not its model but the management capacity it brings early signs, planning and prioritizing maintenance plan (shocks) emergency stops and improved delivery performance (Carvalho et al., 2019; Zonta et al., 2020). (Papazachos et al., 2018) This literature is becoming more and more conclusive about the necessity for evaluation to focus on metrics that measure decision quality under class imbalance (such as precision–recall behaviour), since false negatives can result in the costly interruption of lines and missed delivery commitments (Carvalho et al., 2019; Zhang et al., 2019). Consequently, there is a strong tendency to visualize PdM not as a classical classification problem but rather as a risk management tool from the production flow perspective (Zhang et al., 2019; Carvalho et al., 2019; Cheng et al.,2022).

2.4 Explainable AI and adoption in manufacturing decision contexts

A pervasive blockage to putting AI into use operationally is the "trust–action gap": managers may trust that models predict well, but still unable to act on model outputs without a sensible explanation that aligns with operational logic and accountability needs (Puthanveettil Madathil et al. 2025; Arrieta et al. 2020; Rudin, 2019; Ahangar et al., 2025). Research in XAI claims that explanations are not "icing on the cake;" instead, they form governance and decision support mechanisms to enable stakeholders verify if a model's behaviour is aligned with the domain reality, safety boundaries and improvement interests (Ahangar et al., 2025; Arrieta et al., 2020; Rudin, 2019; Tzionis et al., 2025; Pashami et al., 2023).

In smart manufacturing, recent literature goes on to suggest that explainability is especially important as a misclassification can result in unnecessary stoppages (over-maintenance) or undetected failures (under-maintenance), both of which may negatively impact the throughput and customer service (Puthanveettil Madathil et al., 2025; Arrieta et al., 2020). Translated to TOC, explainability makes the translation from analytics to focus decisions stronger: it helps rationale why this constraint-protective treatment is recommended and which operational levers are most strongly associated with risk (Moosavi et al., 2024; Rudin, 2019). Finally, the literature warns against judging the quality of explanation on trustworthiness alone and to evaluate explanation quality in terms of reliability due to misleading explanations causing ill-advised managerial action (Arrieta et al., 2020; Rudin, 2019; Ahangar et al., 2025).

2.5 Synthesis and research gap: toward TOC-aligned AI that is business-oriented

Throughout these streams, a consensus is building that the most impactful contributions are achieved when AI is delivered as part of an operational logic able to (i) identify and quantify flow constraints, (ii) prioritize improvement causes by impact(iii) and secure early warning and interpretable diagnosis that protect throughput (Telles et al., 2020; Carvalho et al., 2019). Nevertheless, even in many of those papers bottleneck analytics and predictive maintenance are considered as independent technical fields to be studied independently (de Jesus Pacheco et al., 2021; Arrieta et al., 2020). This provides a rationale for a TOC based framing where AI is evaluated by its contribution to managerial action constraint identification, constraint exploitation, and constraint protection, rather than by accuracy alone (Telles et al., 2020; Rudin, 2019).

3. Methodology

The present research paper takes the two-case research design and reports how AI may be employed to modernize TOC in manufacturing by better identifying the bottlenecks and enabling a faster managerial response to them. The methodological logic is decision support rather than purely technical: AI is considered as an operational analytics capability helping managers (i) to identify a constraint, (ii) measure its impact on the flow, and (iii) anticipate disruptions which will spread congestion and throughputs loss.

The two cases are intentionally complementary: Case 1 focuses on diagnosing constraint based on downtime and productivity evidence (identifying and quantifying the bottleneck and its drivers), whereas Case 2 concentrates on early warning and prevention (predicting failures that threaten the constraint and destabilize flow), which destabilize flow. As a set, the cases make TOC's managerial cycle of identify → exploit → subordinate → elevate into an operational tool by giving concrete and quantified inputs to strive for at each step.

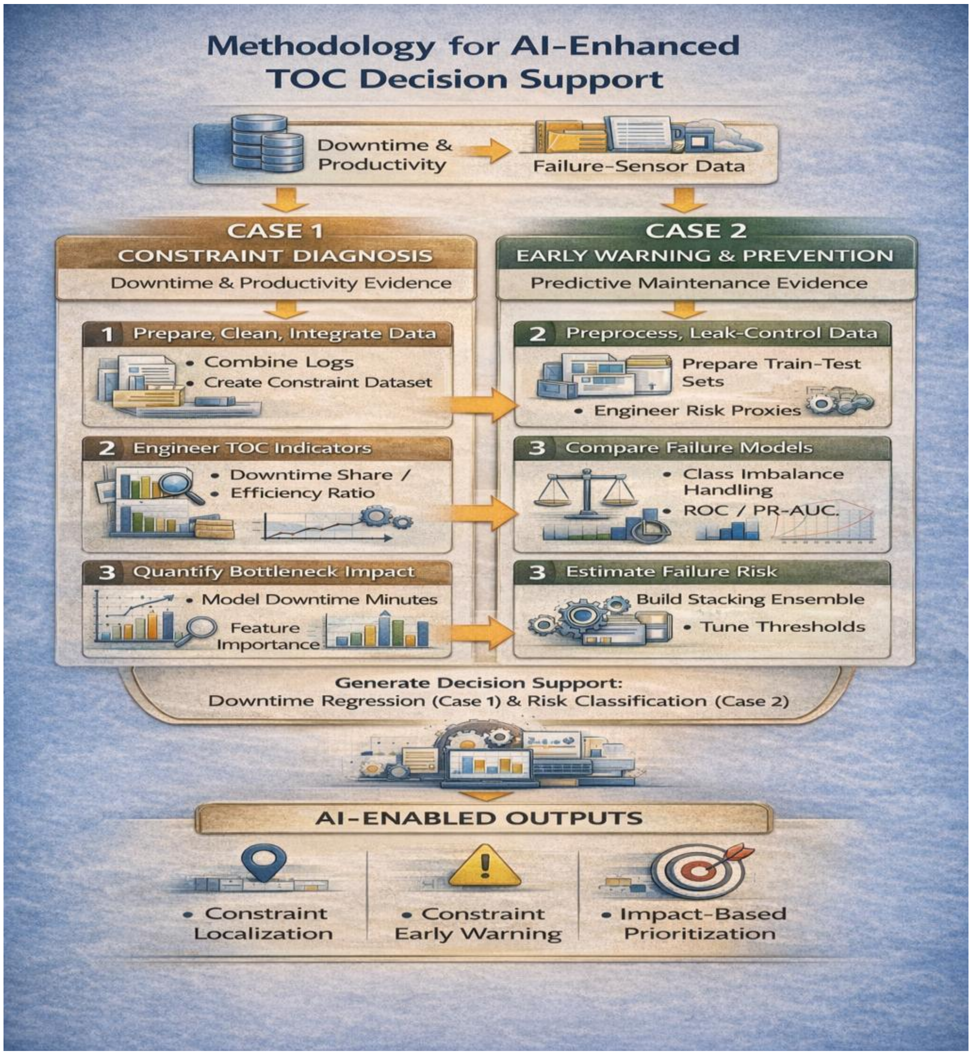

Figure 2: Two-case methodology for AI-enhanced TOC decision support.

Figure 2 gives a full process overview of the research workflow and shows the two cases complementarily facilitating TOC decision making. In Case 1, the constraint loss is measured by downtime and productivity data from (constraint diagnosis) while in Case 2 an early warning signal for failure risk is constructed to protect constraint stability (prevention). The aggregated results allow the localization of constrains early warning and impact-based prioritization.

3.1 Context and Data Sources (Two Operational Cases):

Case 1: TOC-Oriented Downtime and Productivity Evidence (Small-Sample Operational Case)

Case 1 is based on shop-floor operational information recording line productivity and downtime at the batch level. This dataset is created by combining multi operational files that are acquired into (i) line productivity logs (batch timing and operator, shift news context), (ii) downtime logs (downtime minutes, by factor, cause per batch), (iii) a dictionary of the downtime factors (cause descriptions and whether operator error is involved), and (iv) product master file (product attributes and reference-standard batch time). The resultant integrated dataset is an operational "constraint analytics" view at the batch level where every batch is tied to its observed downtime losses and related contexts (Pambudi, 2025). Case 1 deliberately chooses a small sample plant- level diagnostic framework due to the need of performing not population wise prognostic but constraint-oriented loss localisation and prioritisation information under data constraints typical of applied operations settings. Both downtime burden and Pareto-style concentration of dominant causes, being decision-oriented TOC outputs by nature, can be directly calculated from the batch-level table. Although the number of samples restricts generalizability, it enhances the external (contextual) validity as well because results are rooted in shop-floor loss dimensions and replication logic rather than sampling logic.

While the size of the sample is small (consistent with limited or local access to plant data that is typical in applied Operations Research (OR)), this case fits within the scope of managerially diagnosable in that it allows for: (i) ranking highest damage downtime causes (loss concentration); (ii) comparison of current performance versus standard, and (iii) quantification of bottleneck losses in time.

Case 2: Predictive Maintenance Evidence for Constraint Protection (Large-Sample Operational Case)

Case 2 is based on a larger manufacturing dataset (AI4I 2020 predictive maintenance) with operational sensor/process measurements and machine failure (Matzka, 2022); it represented through a binary measure. The setting provided by the dataset corresponds naturally to constraint protection: emergences are relatively rare (class imbalance), as in actual manufacturing plants where disruptive breakdowns do not often occur, but they result in disproportionate flow loss. The case is intended to enable manager's prevention by predicting the risk of failure as an early warning that can be used in scheduling maintenance, setting operating regions, weigh resource allocation and which resources are more likely to fail - especially among bottleneck resources.

Case 2 uses a widely studied predictive-maintenance dataset as a controlled evidence base to test the governance and evaluation logic required for constraint protection under rare-event conditions. The value of this case is methodological and managerial: it demonstrates how an early-warning system can be assessed under deployment-realistic class imbalance (untouched test prevalence), how thresholds operationalize risk appetite, and how explainability can anchor intervention design. Accordingly, the case is presented as transferable decision logic rather than a direct representation of any single plant's sensor ecosystem or cost structure.

3.2 Data Preparation and Integration Procedures (Managerial Rationale)

Case 1: Data Cleansing, Integration and TOC Variables

Case 1 performs data preparation in accordance with sound operational analytics practice: (re)producing comparability across logs, resolving key mismatches and creating interpretable TOC indicators.

i. Data cleaning and validation: column names were normalised, duplicates were removed, next the downtime factor columns were converted to numeric minutes.
ii. Downtime consolidation: total downtime minutes per batch was calculated as the sum of all different downtime causes; "top downtime cause" by batch for managerial rank-ordering (Pareto).
iii. Time alignment and duration measurement: start and end timestamps were created from date and time due fields; hours in warm-up, batch duration was calculated in minutes, incorporating an overnight adjustment when the end time extends beyond midnight. Including shift of end-time overnight.

iv. Operational integration: productivity records were merged with product attributes and downtime causes using batch and product IDs (standardized to consistent formats).

v. TOC-relevant feature engineering: managerial indicators were created:

- Total downtime of system minutes: (proxy for constraint impact / lost productive time).
- Downtime fraction (downtime per processing time giving how far from the flow capacity value).
- Efficiency factor ((standard/minimum batch time relative to actual duration, indicating productivity loss).
- Temporal context variables (hour, day-of-week, month) for shift/seasonality insights.

These procedures generate a coherent dataset that addresses two management decisions: (i) constraint diagnosis involving Pareto of downtimes and operational segmentation; and (ii) impact quantification by predicting downtime magnitude from operational conditions.

Case 2: Data Preparation, Risk-Oriented Features, and Leakage Control

Preparation Case 2 is motivated by a business requirement: the predictions should be deployable to serve as decision support; therefore, preprocessing is performed only on training data to excludes overfitting and artificially inflated performance.

i. Removal of non-actionable identifiers: ID-like features were removed to avoid learning spurious patterns and facilitate understanding for managers.

ii. Train–test split: the dataset was split into a training and held-out test using stratification to maintain the inherent low prevalence of failure events for both sets. More importantly, the test set is left completely untouched, as in real operational deployment.

iii. Operational feature engineering: to increase manager interpretability and strengthen the link between process physics and failure risk, a small number of explainable, operations-relevant derived variables were included (e.g., temperature gap, power proxy, wear-speed interaction).

iv. Outlier treatment (train-only): numerical features were minorized/clipped using quantiles from the training data to reduce distortion caused by extreme values; transformation was designed in a deployment-acceptable change.

v. Handling of mixed data types: Numeric features have been imputed and scaled; Category features have been imputed and encoded (one-hot) so that its representation remains consistent across the folds.

3.3 Modelling as Decision Support:

Case 1: Quantifying Constraint Impact via Downtime Regression

Case 1 represents downtime minutes as an indicator for impact of a constraint. A set of candidate prediction models were cross-validation and performance presented in terms of MAE, RMSE and R2. The approach assists managers by making scenario-like estimations of downtime impact for different operational conditions (product/operator-time context), and so in prioritising improvement actions where losses are likely to be the largest.

To strengthen reliability of the predictive models, an ensemble approach (stacking regression) was implemented using the top-performing learners as base estimators and a simple final estimator. In business terms, the ensemble model represents a robust forecasting mechanism that reduces dependence on one model's assumptions and improves stability for managerial use.

Case 2: Predicting Failure Risk to Protect Flow (Classification)

In case 2, the failure risk is modelled as an early warning signal to protect throughput, i.e. reduce unplanned stoppages in particular if the at-risk machine is a "bottle neck" machine.

An important methodological decision is how to deal with class imbalance. Instead of balancing total dataset (which would distort deployment reality), we balance only within training folds. This keeps the test distribution as the performance standard by which learning algorithms are judged while providing sufficient data to allow an algorithm to learn rare failure patterns.

Cross-validation was used to compare candidate models with metrics selected for managerial implications under imbalance:

• PR-AUC (average precision) to measure the quality of decision making when failure is a rare event.

• ROC-AUC for ranking quality.

• F1-Recall-Precision to characterize trade-offs between missing failures and false alarms.

A stacking of those best candidates formed a model that generalizes well. Furthermore, an out-of-fold threshold selection technique was applied to determine a classification threshold that maximizes Training F1, indicating managerial preference for trading missed failures with false alarms.

3.4 Visual Analytics and Interpretability for Managerial Use (XAI)

In order to transform output into operational action (instead of abstractions such as accurate statistics), the study also provides interpretable visual evidence:

• Case 1: histogram views of downtime and duration, and a Pareto-like ranking of top downtime causes to support TOC prioritization ("vital few causes").

• Case 2: a training vs. test class distribution plot to visualize that balancing is performed only during the training stage; performance curves (ROC and PR) to remark on risk discrimination; and an assessment of permutation importance in the held-out testing set to identify the most influential operational drivers of failure risk. Feature importance is reported with actual feature names to assist in actionable interpretation (e.g., wear-based, torque-based, or failure mode indicators), to help managers turn findings into maintenance and operational practices.

Methodological Steps:

To present the methodology in a structured, business-ready way, the study follows these steps:

a. Formulate TOC decision problem: enhance flow by identifying and controlling constraints and their disturbances.
b. Select complementary cases: diagnostic (downtime/bottleneck) and preventive (the risk of failure).
c. Prepare datasets: clean, validate, integrated (Case 1) and pre-processed with leakage controlled (Case 2).
d. Engineer TOC-relevant indicators: the share of downtime, the ratio of efficiency, and interpretable risk proxies.
e. Establish evaluation logic: cross-validation for reliability; held-out test for deployment realism.
f. Address operational imbalance: apply balancing only inside the train folds to preserve real test conditions.
g. The study adheres to the following phases to offer the methodology in a systematic, business-ready format:
h. Compare candidate decision models: compare many decision algorithms using imbalance-cognizant metrics.
i. Build robust decision support: create stacking ensembles for both stability and generalization.
j. Select decision thresholds: tune threshold based on out-of-fold prediction to match managerial trade-off.
k. Explain drivers for action: translating outputs to operational priorities with Visual Analytics and Permutation Importance.

In this paper, the methodology is business oriented. Instead of considering AI as an independent end, this methodology treats AI as a type of measurement tool and decision-support mechanism integrated into TOC: it improves (i) The visibility of where flowage is impeded; (ii) The quantification of impact of impediments in loss-of-operational-time terms, and (iii) Advance warning about disturbances that threaten throughput. Such a framing can allow managerial interpretation in terms of priority-setting, resource allocation, preventive planning and flow improvement, serving to "ground" the value contribution in operational and business performance results rather than merely computational novelty.

4. Result and Discussion

This section presents results from two complementary case studies that use Theory of Constraints (TOC) instrumented as a decision support system to optimize production flow. The results are organized in such a manner that they facilitate the implementation of TOC: (i) revealing the constraint, (ii) (ii) measuring the loss of capacity caused by flow interruptions, (iii) guarding the constraint using early warning and (iv) converting the insight into intervention priorities. Therefore, the emphasis here is placed on bottleneck visibility and throughput maintenance and support of management as opposed to algorithmic details.

4.1 Case Study 1: TOC Diagnostics Bottleneck Visibility and Downtime Impact Quantification

4.1.1 Flow-loss visibility: where capacity is being consumed

The combined operational dataset (productivity logs, downtime cause records and product master data) created a final analytical table of 30 batches × 22 variables. Two critical TOC-related indicators were selected to quantify flow loss:

• Total minutes of downtime, which reflects the actual direct non-productive time that costs available capacity.

• Downtime: which is the proportion of downtime minutes to total minutes in an actual batch cycle provides a convenient management proxy for "how much of a batch's time was consumed by non-value activity".

In TOC terms, downtime translates to actual loss of effective capacity at or near the constraint, a failure loss throughput potential and an increase in flow variety. Such losses due to the constraint affect not only the constraint machine itself but also have a ripple effect along the shop in different downstream stages as schedule instability and increase of WIP and deterioration on delivery reliability.

Downtime intensity and variability: The distribution of downtime minutes reveals whether a system is governed by numerous short disruptions (systemic waste) or by a few severe ones (episodic shocks):

• Figure 3 (refer to Figure 3) shows the pattern and dispersion of downtime minutes per batch. A fat right tail corresponds to "few high-impact breakdowns", which is consistent with the TOC emphasis on eliminating the most throughput-reducing losses first.

• The distribution of actual batch duration can be seen on Figure 4 (see Figure 4). Large variations in realised durations imply the flow is unstable, and therefore (in operational language) planning is more difficult and requires provision of buffer space.

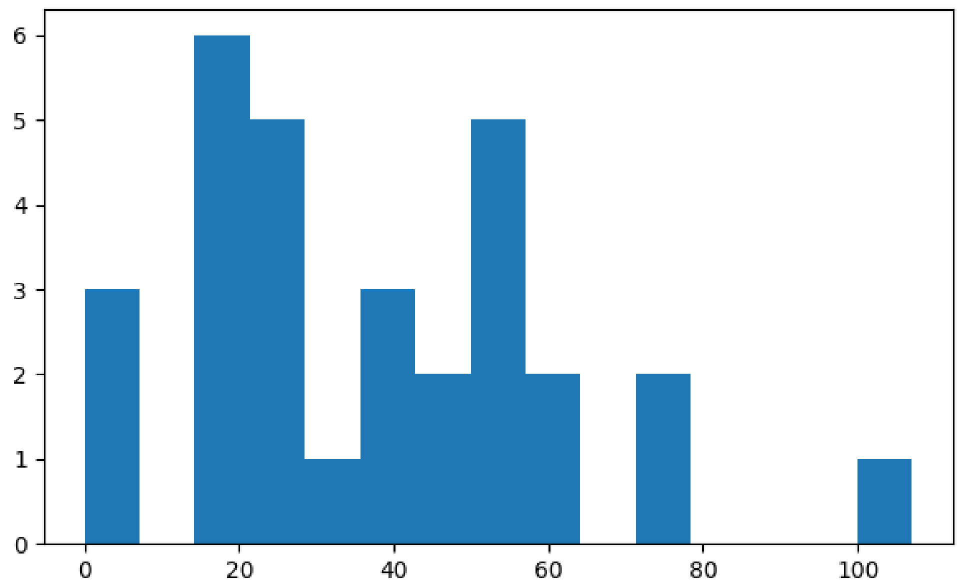

Figure 3: Downtime distribution (minutes).

The distribution of the total downtime in minutes per batch is depicted for Case 1 in Figure 3. The distribution is right skewed, meaning that most batches have relatively low-down time but some of the batches have very high down time (have a fat right tail). From a TOC point of view, such extreme downtime incidents are managerially Decisive, they dramatically reduce available effective capacity in or near the constraint and they induce flow variability to be abundant, which may translate into congestion and WIP. Thus, the figure also begins to validate the logic of prioritization for effect-based improvement: investigate and remove the handful of largest downtime batches (and their key causes) to save throughput and stabilize delivery performance.

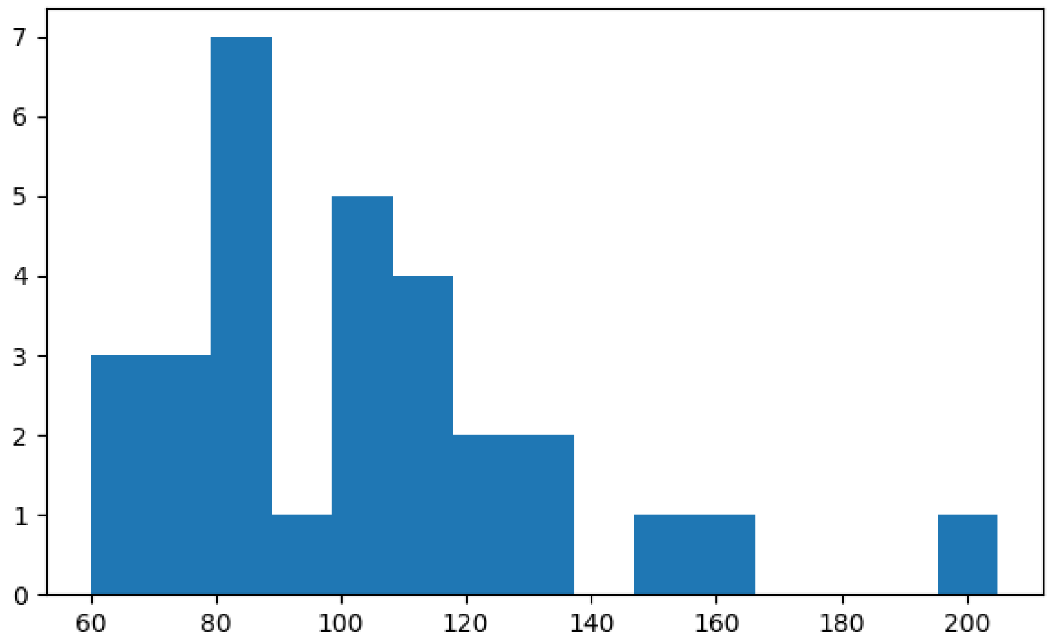

Figure 4: Actual batch duration distribution (minutes).

In case 1, actual batch durations (minutes) are illustrated in Figure 4. The variance between realized processing times shows significant randomness across runs, as some batches run for far longer than the average time taken in a cycle. From a TOC standpoint, this variation is operationally significant since batches that are longer than normal eat up precious capacity at the constraint (or quasi-constraint), increase flow instability, and increase the degree of protection that we need to ensure on-time delivery (time/WIP). Then, the figure also encourages a management target on identifying either product mix or shift effects and disruptive conditions as the root causes behind those longest duration batches in order to have a handle towards stabilizing throughput by cutting down propagation of congestion.

4.1.2 Concentration of losses: "few vital causes" for improvement focus

A Pareto-style sum of downtime minutes by the "top factor" in each batch of downtime minutes shows whether the downtime is spread across many causes or concentrated into a few top drivers of disruption.

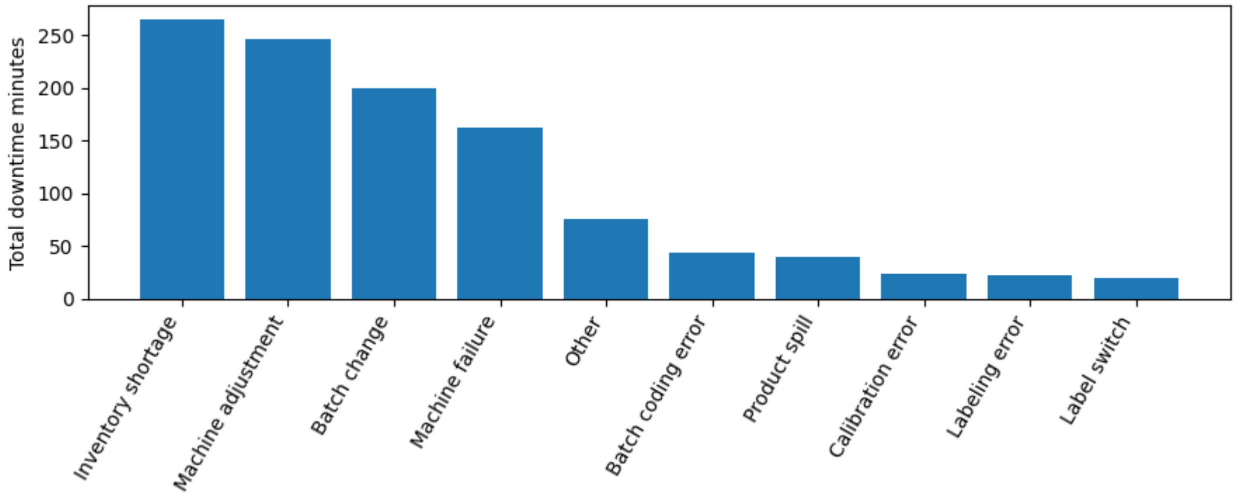

Figure 5: Proxy Pareto chart of the top 10 downtime causes in Case 1 (total minutes).

Figure 5 summarizes how downtime losses are distributed across the most frequent stoppage categories, showing a clear concentration of lost minutes in a small number of causes. Inventory shortage and machine adjustment account for the largest share of downtime, followed by batch change and machine failure, while the remaining causes contribute comparatively smaller losses. This concentration pattern aligns with TOC's "vital few" logics: reducing downtime in the dominant categories offers the fastest route to reclaim constrained capacity and stabilize flow. Managerially, the chart provides a practical prioritization map, directing improvement resources first toward the highest-loss causes (e.g., material availability discipline, setup and adjustment standardization, and reliability-focused actions) before addressing lower-impact issues such as calibration, labelling, or minor handling errors.

4.1.3 Estimating downtime burden to prioritize interventions

The study focused on validation of performance across recross, not technical superiority, to aid managerial prioritization (e.g., informing which products, shifts or operating contexts will likely experience high downtime), while estimating the minutes lost to downtime by approaches predictive in sign only.

Table 1 condenses performance for downtime estimation into interpretable error metrics (minutes of downtime error) for operations management.

Table 1: Case 1 Predictive performance of alternative models for estimating batch-level downtime minutes

Model	MAE	RMSE	R^2
Ridge	3.1043	4.3163	0.9469
ExtraTrees	3.3408	6.0080	0.8911
RandomForest	4.8201	7.3456	0.8126
HistGB	20.2972	23.4773	-0.3538

In Case 1, Table 1 compares the accuracy of each of four regression approaches in estimating total downtime minutes per batch, using MAE and RMSE as directly interpretable "minutes-of-error" measures relevant to operational decision-making, along with R^2 as an explanatory power measure. it compares the accuracy of four regression approaches to estimating total downtime minutes per batch in Case 1. The best overall performance (MAE ≈ 3.10 mins; RMSE ≈ 4.32 mins) and goodness-of-fit ($R^2 \approx 0.95$, indicating that most of systematic variation in downtime is captured by a relatively simple, regularized linear specification) was achieved by the Ridge model. ExtraTrees and RandomForest do well, too, but at higher error and lower R^2, imply some benefit from non-linear modelling, but at less stable gains with Ridge in this setting. Conversely, The Histogram Gradient Boosting model exhibits suboptimal performance (MAE ≈ 20.30 minutes; RMSE ≈ 23.48 minutes) and achieves a negative R^2, meaning that it is overfitting versus naïve, so it cannot be used in a reliable manner for managerial purposes here. Collectively, these results provide for the decision-support rationale of the study: the burden of downtime can be estimated with high fidelity (to ~95% explained variance), allowing for the anticipation of expected capacity loss across different products, shifts and operating conditions, and the ability to rank interventions and buffering based on where the potential burden of downtime is forecasted to be greatest.

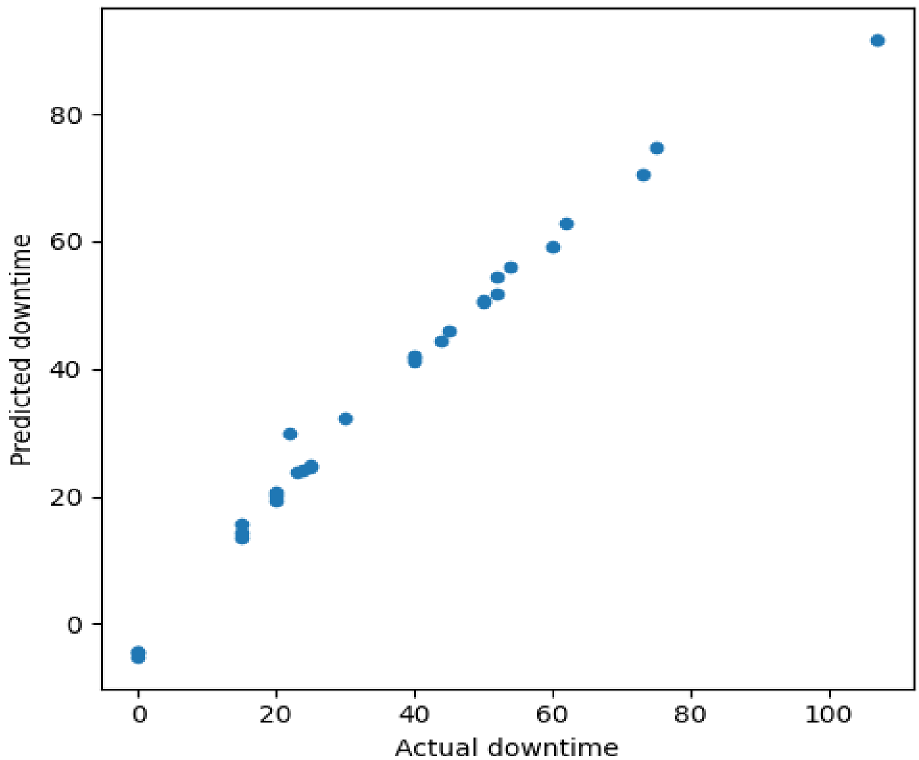

Figure 6: Actual vs. predicted downtime minutes for Case 1.

Plotting actual versus model-predicted downtime minutes for Case 1 (left plot in Figure 6) gives a visual check that regression model adequately captures the capacity loss at the batch level in terms of the magnitude. This is because the points are densely clustered around the implied line of equality (i.e. where predicted ≈ actual), meaning that the model replicates observed downtimes very closely, with little deviation across the spectrum of the batches, even for those that resulted in much greater loss. This alignment is important from a TOC decision-support perspective because it indicates that downtime burden can be predicted ex ante based on operational context (e.g., product/shift conditions), empowering the organization to know when the constraint is likely to take larger time hits so that they can prepare buffering, staffing changes, or targeted corrective maintenance before flow instability moves downstream.

4.1.4 Managerial implications for TOC practice (Case 1): bottleneck improvement sequencing

These findings strengthen TOC execution in two ways. First, the downtime distributions and Pareto concentration provide a defensible basis for deciding where to focus first a core TOC requirement when improvement resources are limited. Second, estimating the expected downtime expected by a loss, impact-based prioritization (which losses to remove first) and supports the "exploit/subordinate" steps by improving planning around expected constraint losses. In practical terms, managers can convert downtime information into a prioritized improvement backlog and align staffing, changeover planning, and operational controls to reclaim capacity where it yields the highest throughput benefit.

4.2 Case Study 2: Constraint Protection Early Warning for Disruptive Failures as Flow Assurance

4.2.1 Baseline risk structure: rare events with disproportionate operational impact

The predictive maintenance dataset illustrates a genuine operational characteristic: faults are infrequent yet significantly impactful to flow.

- Overall distribution: **0 = 9661**, **1 = 339**
- Training split: **0 = 7729**, **1 = 271**
- Untouched test split: **0 = 1932**, **1 = 68**

Figure 7 illustrates the highly imbalanced situation of predictive maintenance, where non-failure observations are majority and failure events are minority. This low-frequency phenomenon is analogous to realistic plant operations where disruptions are infrequent but may lead to incommensurate throughput losses if the disrupted asset had been or becomes a bottleneck (constraint). Hence, the figure motivates a focus on evaluation beyond overall accuracy focusing on decision-relevant performance under scarcity (e.g., ability to detect failures with controlled false-alarm rates) for TOC-aligned constraint protection and flow stability.

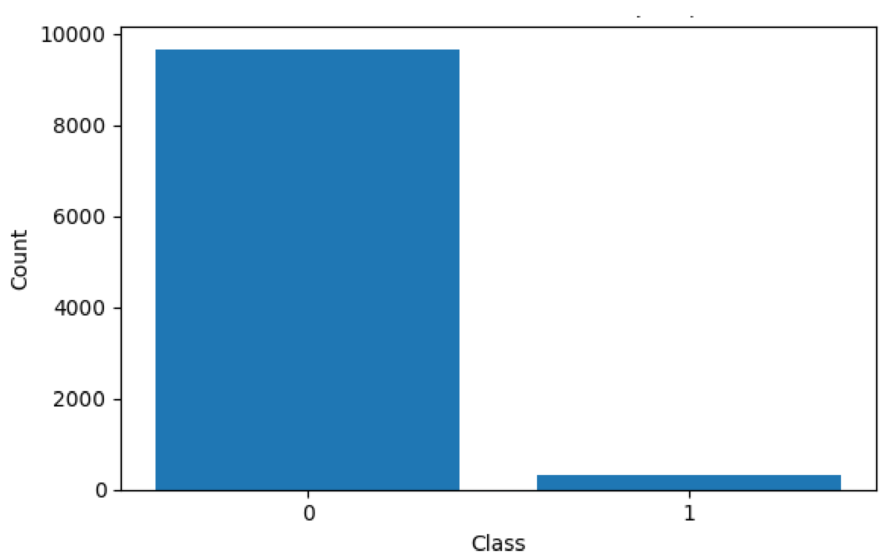

Figure 7: Case 2 Raw class distribution of machine failure events (AI4I 2020).

4.2.2 Training-only balancing: improving learning while preserving real deployment realism

To strengthen learning without inflating performance, balancing was applied within training only, while the test set remained untouched.

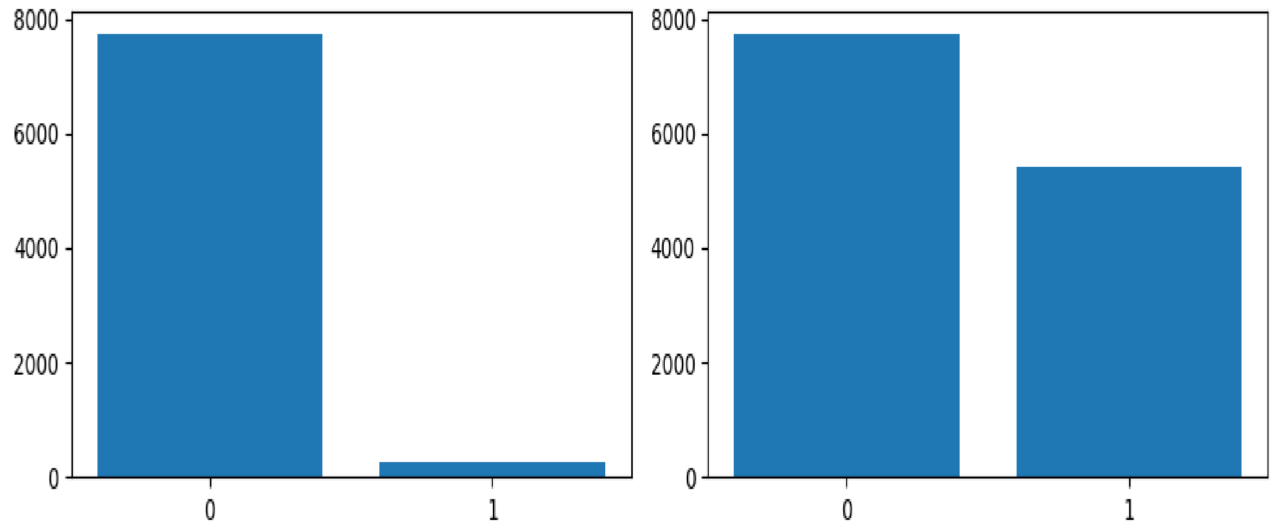

Figure 8: Training distribution before vs. after balancing.

Class rebalancing (in training set only) (Figure 8) Prior to balancing, the training set is heavily imbalanced with respect to the non-failure class, and learning algorithms may overfit the majority pattern thus under-detecting rare non-existence, costly failure events. A large increase in the minority failure class relative to the majority class occurs post balance (strategy = 0.7) to provide a richer exposure to failure signatures and more informative decision boundaries for the learner. In TOC-minded view, this step enhances the early-warning method we use to safeguard the constraint, it minimizes the chance a rare disruption is systematically ignored and keeps the underlying methodology intact because there is no adjustment of the training only phase and into the evaluation environment.

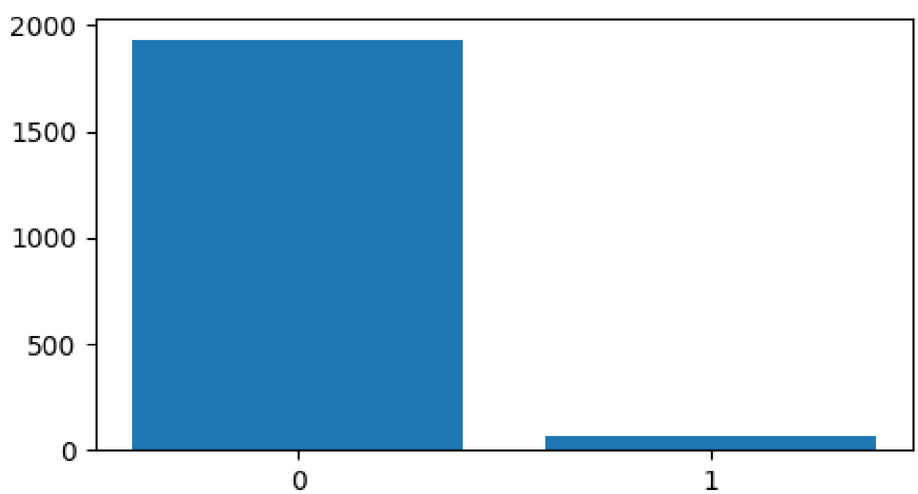

Figure 9: Test distribution (untouched).

The test-set class distribution is still highly imbalanced as the number of non-failure observations extremely outnumber that of failure cases. Keeping the test set unchanged is also a key methodological design choice that has been made intentionally: to maintain the true operational prevalence of breakdown events, since performance should be reported in deployment conditions rather than an artificially "easier" evaluation environment. In a TOC sense, this is important, because if rare failures occur at (or propagate to) the constraint, they can still lead to outsized throughput losses. As a result, Figure 9 lends support to the robustness and managerial generalizability of the findings: model metrics and threshold selections are evaluated under the same limited abundance of failures confronting managerial practice, thus creating a credible foundation for early-warning policies that preserve constrained capacity by avoiding inflated, test-data manipulation-based advantages.

4.2.3 Comparative model performance: decision-support quality

Instead of stressing on algorithms, our study compared alternative predictive approaches as decision-support mechanisms and assessed them based on metrics that are relevant to uncommon-event operations (e.g. PR-AUC, recall, and precision).

Table 2: Case 2 Comparative decision-support quality (sorted by PR-AUC).

Model	ROC-AUC (mean)	PR-AUC (mean)	F1 (mean)	Recall (mean)	Precision (mean)
LogReg	0.9819	0.9182	0.7952	0.8672	0.7437
HistGB	0.9788	0.9139	0.7689	0.8819	0.6832
ExtraTrees	0.9790	0.9094	0.7674	0.8708	0.6914
RF	0.9803	0.9026	0.7734	0.8819	0.6915

The cross-validated performance of four alternative classifiers on the prediction of failure risk in Case 2, evaluated by rare-event decision-context appropriate measures directly related to TOC-style constraint protection, is shown in Table 2. Discrimination is quite good with all models (ROC-AUC ≈ 0.979–0.982), suggesting that they can rank-riskier observations higher than less risky ones. More crucially for operational use under class imbalance, PR-AUC is still high (≈ 0.903–0.918) implying that substantial precision–recall quality is retained by each model because failures are rare. Logistic Regression gives the best all-round score profile with the highest PRAUC (0.9182) and F1 (0.7952) representing a balance in finding failures (Recall = 0.8672) and limiting false alarms (Precision = 0.7437). The tree-based models display similar recall (≈ 0.871–0.882), but slightly lower precision (≈ 0.683–0.692), suggesting a greater volume of operational "noise" per unit of capture, a critical management consideration, as false alarms initiate expensive inspections, stoppages, or maintenance interventions. Altogether the results suggest failure risk can be modelled as providing more consistently strong decision quality, which supports the TOC-oriented aim of the study: to provide credible early-warning signals that enable managers to intervene before disruption events convert into throughput loss and thereby protect constrained capacity and stabilize flow.

As shown in Figure 10, Precision–Recall (PR) curve for the Case 2 failure-risk classifier calculated on the unseen test split reveals a very good PR-AUC of 0.9066. In the predictive maintenance context where failure events are rare, this is manifested in PR analysis presenting a more actionable perspective than accuracy-based summaries, by directly delineating the precision–recall tradeoff between capturing disruptions (recall) and limiting false alarms (precision). The curve shows that high precision is maintained over a wide range of recall levels before falling off at high recall, meaning an accessible early-warning footprint can be sustained without burdening operations with superfluous interventions. In terms of TOC, such a performance reinforces protection of the constraint: it provides managers with the ability to target proactive measures towards only the most credible high-risk signals, reducing the risk that rare disruptions will result in outsized throughput loss and flow disturbance.

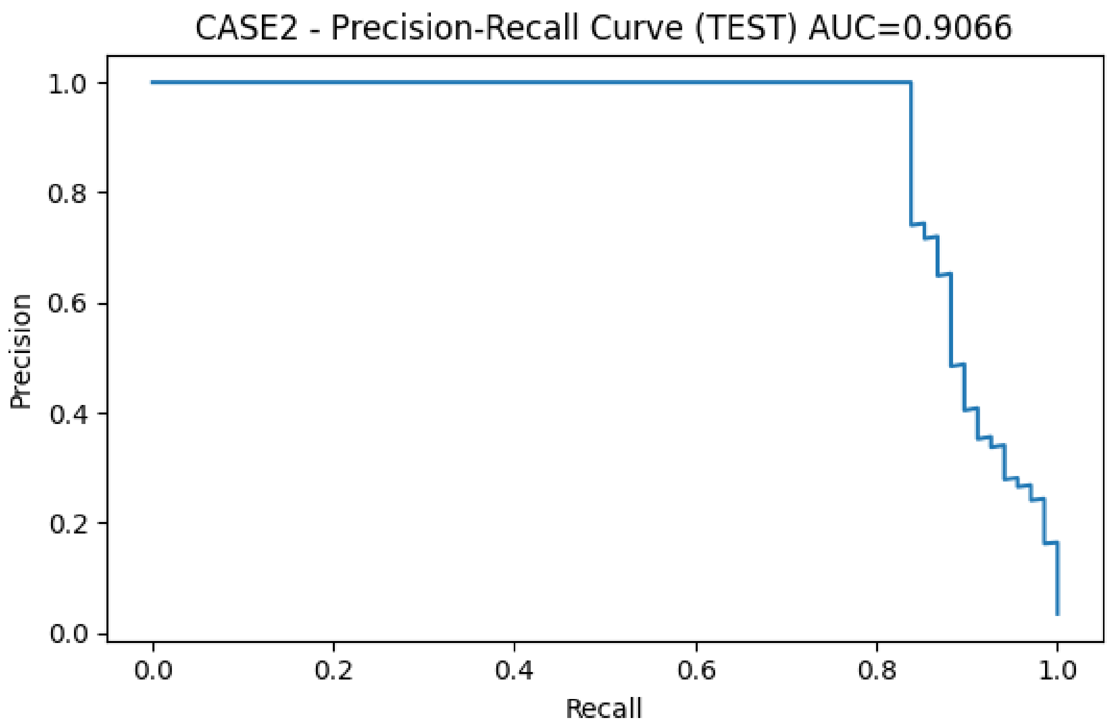

Figure 10: Precision–Recall performance on the untouched test set (Case 2).

4.2.4 Ensemble risk scoring: governance-ready early warning and threshold choice

An ensemble from the best models created a more stable risk-scoring mechanism. Since both accuracy and F1 score are balanced measures, the decision threshold was tuned on the out of fold training predictions in a way that reflects managerial preference for balanced detection performance.

- OOF ROC-AUC: **0.9807**
- OOF PR-AUC: **0.9168**
- Selected threshold (max F1 on training-only OOF): **0.9864**

The decision threshold versus F1 corresponding out-of-fold (training-only) of the classification in Case 2 is shown in Figure 11. This came out from the curve showing the extreme sensitivity of predictive performance to threshold in rare-event settings: thresholds too low inflate type I errors with false alarms, thresholds too high inflate type II errors with missed failures, both jeopardizing constraint protection. The horizontal dashed marker marks the chosen operating point, which maximizes out-of-fold F1 and therefore makes the target level of aggressiveness of failure risk flagging more transparent and reproducible. The notion of "risk appetite" here translates into managerial governance from a TOC decision-support perspective; this visualization makes clear the "risk appetite" choice, permitting practitioners, when deploying early-warning policies, to operationalize the relative cost of a preventive intervention with the throughput loss and schedule instability risk stemming from an unexpected breakdown at (or near) the constraint.

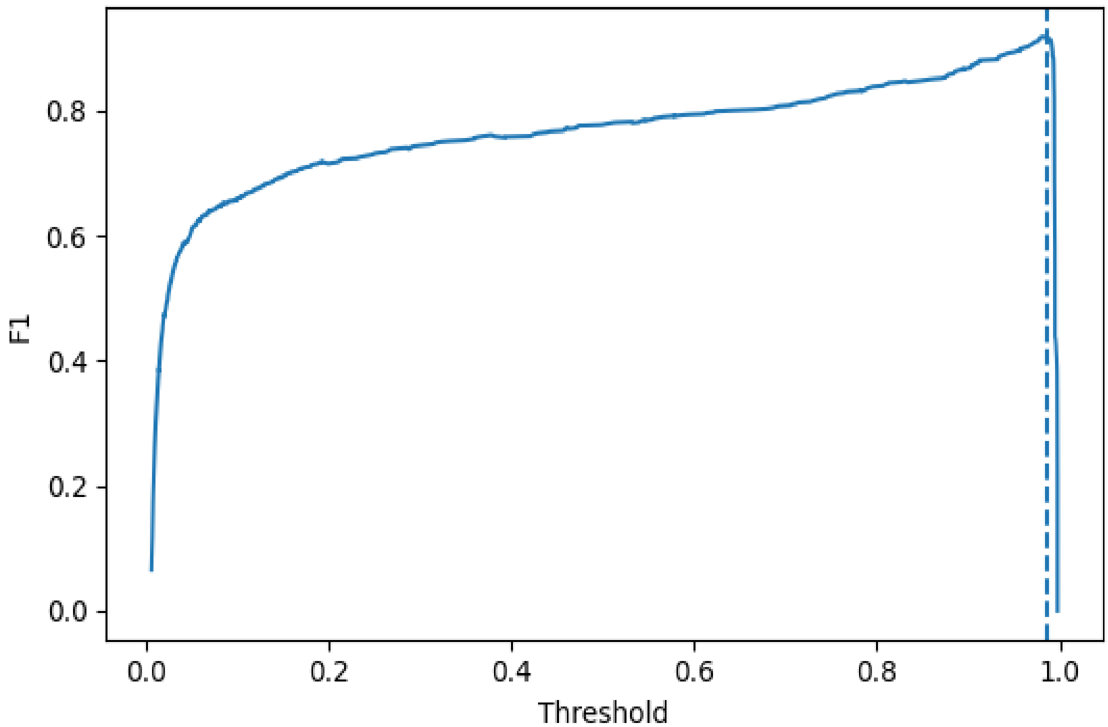

Figure 11: F1 vs. threshold curve (OOF).

4.2.5 Untouched test performance: operationally realistic validation

We performed the final evaluation on the untouched test set as this more accurately represents realistic deployment conditions.

Table 3 shows the performance of the ensemble on the unseen test set, which is a more realistic evaluation of the early-warning model in Case 2 from an operational perspective. The discrimination is still very high (ROC-AUC = 0.9896), meaning that the relative model ranks of high and low risk states are still reliable across thresholds, and the PR-AUC of 0.9066 confirms that model alerts maintain high quality under the extreme class imbalance representative of real maintenance environments. Overall accuracy is high (0.9945), however, the more decision-relevant failure-class metrics are the critical managerial signals: precision for Class 1 is 1.0000, suggesting that when failure alerts are issued they may be highly trusted as it is unlikely they would lead to pointless stoppages and inspections, while recall of 0.8382 shows that the model is able to identify large proportion of failure events before they develop. This results in an F1-score (0.9120) that reflects a very good compromise between detection coverage and operational noise. These results directly speak to constraint protection in TOC - With a high precision and good recall, such a model presents an opportunity to generate low false-alarm, high credibility warning signals; and the proactive maintenance scheduling and resource allocation would therefore reduce unanticipated downtime at (or propagating to) the constraint, helping to stabilize flow, contain WIP, and protect throughput and delivery reliability.

Table 3: Case 2 Test performance summary

Metric	**Value**
ROC-AUC	0.9896
PR-AUC	0.9066
Accuracy	0.9945
Failure-class precision (Class 1)	1.0000
Failure-class recall (Class 1)	0.8382
Failure-class F1 (Class 1)	0.9120

The confusion matrix on the completely untouched test set is reported in Figure 12, to summarize the classification results obtained by the ensemble of the labelled data on deployment-realistic conditions. The model attains 0 false positives (no-failure cases wrongly flagged as failures), meaning that the alerts generated are highly actionable and unlikely to create operational noise through unnecessary machine stops or inspections. Simultaneously, it successfully detects 57 out of 68 failure events even when 11 failures are not detected. On TOC lens, this pattern supports to protect the constraint because when system can execute targeted preventive action from breakdowns with minimal disruption to routine execution, while still intercepting majority of high impact breakdowns, flow will be stable as well flow is not going to be destabilized due to depleting constrained capacity. From a managerial perspective, these residual false negatives reinforce the necessity of mission alignment by aligning thresholding on the risk appetite (e.g. slightly more alerts acceptable if failure cost at the constraint is higher) and reinforce governance with recalibration intervals as the operating condition drift.

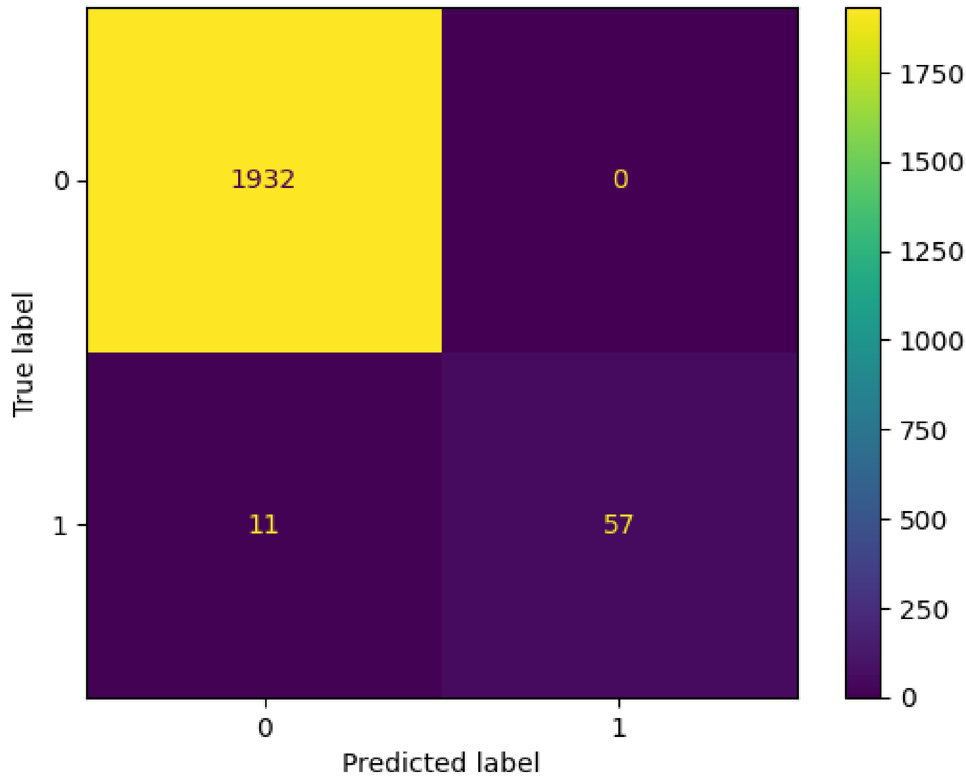

Figure 12:Confusion matrix on the untouched test set (Case 2).

Managerial implications for TOC practice (Case 2): constraint protection and maintenance governance

In Case 2, we discuss how TOC can benefit from an AI-enabled early warning that protects the constraint from disruptive stoppages increasing flow stability and throughput protection. The achieved recall (~84%) suggests that most failure events are detected, but the very low false-positive tendency makes it possible to trust the alerts and really operationalize them. From TOC perspective, this fortifies exploitation and subordination by minimizing schedule variability and avoiding unintentional constraint starvation or blocking. From a managerial perspective, the output is amenable for embedding within maintenance governance as a risk-based prioritisation layer: informing inspection intensity, maintenance scheduling and escalation rules according to disruption likelihood, rather than a reactive breakdown response return agenda.

4.3 Explainability as Action Guidance: Translating Signals into Interventions

In order to turn predictive results into actionable operational lines (instead of "black-box" outputs), permutation importance was calculated upon the test set within the final model input space (i.e., after preprocessing and variable selection). This also fits with business/operations sense where: explainability should drive intervention, not just provide description of statistical relationships.

Table 4: Case 2 Permutation importance (named features, test set).

Feature	Permutation importance (mean)
HDF	0.0270
OSF	0.0153
PWF	0.0115
power_proxy	-0.0037
Torque [Nm]	-0.0046
Rotational speed [rpm]	-0.0084
Tool wear [min]	-0.0134
wear_speed	-0.0143

The permutation importance results that translate the failure-risk model of Case 2 into TOC-relevant, actionable levers are shown in Table 4. The most positive signals, HDF (0.0270), OSF (0.0153) and PWF (0.0115), are also the most decision-relevant ones, permutations of this result in highest decrease in model performance with respect to baseline. Operationally, these indicators can anchor an early-warning governance layer (dashboards, escalation rules, and preventive actions) to protect constrained or near-constrained resources from disruption-driven capacity loss, thereby supporting constraint protection and throughput stability. The other features display very little negative importance, and this should be seen in terms of correlation or redundancy (or interaction) rather than an indication for a bad effect. Management focus should, therefore, be on control and SOP triggers related to HDF/OSF/PWF with lower ranked variables informing management as context signals which do not add much extra value in this dataset.

The most decision-relevant drivers of the Case 2 failure-risk model are outlined in Fig.13, determined by permutation importance on the untouched test set to facilitate deployment-realistic interpretation. The strongest positive contributors due to the best contributors (HDF, OSF, PWF) exhibit the largest decrease in predictive quality when shuffled implying that they carry more actionable signal for TOC-aligned constraint protection, i.e., information indicating where it is currently most important to take actions that remove variation potentially leading to capacities lost from disruptions at or close to the bottleneck. On the other hand, power proxy, torque, rotational speed, tool wear and wear–speed interaction have small but negative importances, which are likely due to a feature redundancy/correlation rather than operational irrelevance; in this sense managerial focus should be on monitoring and escalation rules addressed to these high-positive indicators while dealing with the rest of the variables as contextual support for diagnosis and governance.

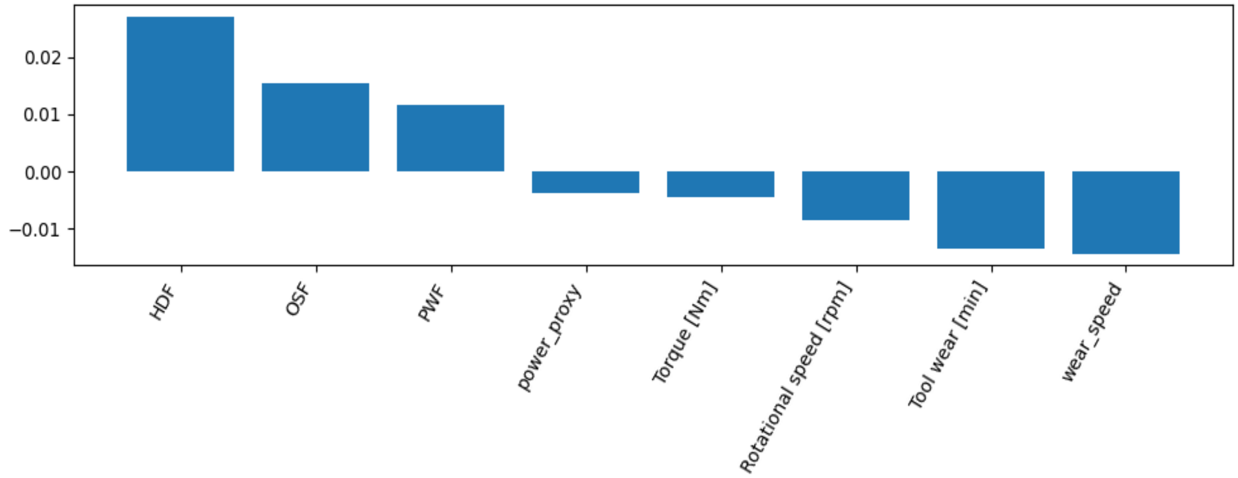

Figure 13: Permutation importance on the untouched test set (named features).

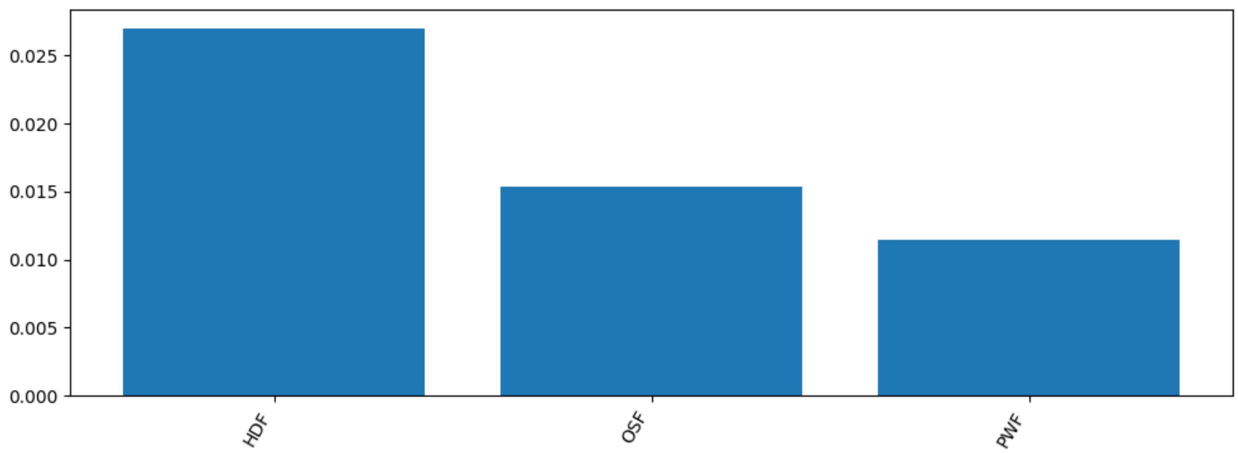

Figure 14:Positive-only permutation importance on the untouched test set (named features).

Figure 14 shows a simplified permutation-importance perspective, keeping only positive contributors on the unbalanced test set for which factors consistently contribute towards improving failure-risk prediction under deployment-realistic prevalence. The superiority of HDF over OSF and PWF suggests that these signals provide the most reliable (i.e., less conflicting) and decision-critical information for the early-warning mechanism, hence forming a more robust basis for TOC-aligned constraint protection, that is, aligning monitoring, control limits, and PM triggers to reduce disruption-driven capacity loss at or near the system bottleneck. By removing negative importances the figure provides a theory rather than just a chart managers could use for designing governance: it tells them what metrics to base their dashboards and escalation rules on, in order to ensure throughput is protected and production flow stabilised.

Managerial implications for TOC practice (Explainability): dashboards, SOP triggers, and escalation rules

These explainability results enable operationalization. First, they facilitate the dashboard design by focusing on the most decision-relevant signals and thereby helping managers manage cognitive load. Second, they drive the design of SOP triggers (like monitoring thresholds and escalation policies) by cantering governance around the most stable indicators. Finally, they provide support for TOC constraint protection by offering early, auditable and explainable interventions crucial for successful organizational adoption in production context where trust and accountability are of utmost importance.

4.4 Cross-Case Synthesis: How AI Strengthens TOC Execution End-to-End

Across both cases, the results support a unified business contribution: AI strengthens TOC execution by converting operational data into constraint-focused decisions.

i. **Constraint visibility (Case 1):** downtime patterns and cause concentration clarify where and why flow is being restricted.

ii. **Impact quantification (Case 1):** estimating downtime burden supports impact-based prioritization and improvement sequencing.

iii. **Constraint protection (Case 2):** early warning provides a practical mechanism to reduce disruptive stoppages under realistic prevalence.

iv. **Actionable guidance (Explainability):** named importance results translate prediction into governance levers (dashboards, SOP triggers, maintenance prioritization).

In combination, these findings frame AI not as some technical terminus but as a managerial capability that can enhance the speed, accuracy, and consistency by which TOC decisions are taken to support stability in throughput, reliability over schedule expectations, and more effective deployment of effort on the operation's improvement.

5. Conclusion

This study advances a business-oriented perspective on AI-enabled Theory of Constraints (TOC) by demonstrating how machine-learning analytics can be structured into executable, constraint-focused decisions for manufacturing flow improvement. Using two complementary cases, the paper shows that AI can (i) make constraints operationally visible through quantification of downtime burden and flow-loss indicators, (ii) support managerial prioritization by revealing the "vital few" drivers of lost capacity via Pareto-style loss concentration, and (iii) extend TOC from reactive diagnosis toward proactive constraint protection through rare-event early-warning for failure risk under deployment-realistic prevalence. Importantly, the contribution is not positioned as algorithmic novelty, but as a decision-support pathway that connects heterogeneous shop-floor data to TOC execution, including improvement sequencing (exploit–subordinate–elevate), throughput protection, and governance-ready intervention design supported by explainability. The study is limited by the small and context-specific dataset in Case 1 and the use of a benchmark predictive-maintenance dataset in Case 2, as well as by emphasizing operational performance metrics rather than fully monetized throughput-accounting outcomes. Future research should validate the framework across multiple sites and sectors, incorporate richer production-system evidence (e.g., queues, WIP dynamics, changeovers, and maintenance work orders), and explicitly model economic trade-offs linking early-warning and downtime reduction to throughput, service penalties, and intervention cost. Further work should also examine implementation governance, such as threshold policies, recalibration under drift, and human–AI routines for acting on explanations, to clarify how AI-enabled TOC can be sustained as an organizational capability within continuous improvement practice.

References

Ahangar, M. N., Farhat, Z. A., & Sivanathan, A. (2025). AI trustworthiness in manufacturing: Challenges, toolkits, and the path to Industry 5.0. *Sensors, 25*(14), Article 4357. **https://doi.org/10.3390/s25144357**

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82-115. **https://doi.org/10.1016/j.inffus.2019.12.012**

Barredo Arrieta, A., Tabik, S., García López, S., Molina Cabrera, D., Herrera Triguero, F., & Díaz Rodríguez, N. A. (2019). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI.

Bokrantz, J., Skoogh, A., Berlin, C., Wuest, T., & Stahre, J. (2020). Smart maintenance: A research agenda for industrial maintenance management. *International Journal of Production Economics, 224*, Article 107547. **https://doi.org/10.1016/j.ijpe.2019.107547**

Buer, S. V., Strandhagen, J. O., & Chan, F. T. (2018). The link between Industry 4.0 and lean manufacturing: Mapping current research and establishing a research agenda. *International Journal of Production Research, 56*(8), 2924-2940. **https://doi.org/10.1080/00207543.2018.1442945**

Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering, 137*, Article 106024. **https://doi.org/10.1016/j.cie.2019.106024**

Cheng, X., Chaw, J. K., Goh, K. M., Ting, T. T., Sahrani, S., Ahmad, M. N., ... Ang, M. C. (2022). Systematic literature review on visual analytics of predictive maintenance in the manufacturing industry. *Sensors, 22*(17), Article 6321. **https://doi.org/10.3390/s22176321**

de Jesus Pacheco, D. A., Junior, J. A. V. A., & de Matos, C. A. (2021). The constraints of theory: What is the impact of the Theory of Constraints on operations strategy? *International Journal of Production Economics, 235*, Article 107955. **https://doi.org/10.1016/j.ijpe.2020.107955**

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., & Roubaud, D. (2020). Upstream supply chain visibility and complexity effect on focal company's sustainable performance: Indian manufacturers' perspective. *Annals of Operations Research, 290*(1), 343-367. **https://doi.org/10.1007/s10479-018-3014-7**

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics, 210*, 15-26. **https://doi.org/10.1016/j.ijpe.2019.01.004**

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production, 252*, Article 119869. **https://doi.org/10.1016/j.jclepro.2019.119869**

Goldratt, E. M., & Cox, J. (2016). *The goal: A process of ongoing improvement*. Routledge.

Gupta, M. C., & Boyd, L. H. (2008). Theory of constraints: A theory for operations management. *International Journal of Operations & Production Management, 28*(10), 991-1012. **https://doi.org/10.1108/01443570810903122**

Helo, P., & Hao, Y. (2017). Cloud manufacturing system for sheet metal processing. *Production Planning & Control, 28*(6-8), 524-537. **https://doi.org/10.1080/09537287.2017.1309711**

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2020). Explainable AI methods—A brief overview. In A. Holzinger et al. (Eds.), *Explainable AI (XAI): Beyond opening the black box* (pp. 13-38). Springer. **https://doi.org/10.1007/978-3-030-28954-6_1**

Hopp, W. J., & Spearman, M. L. (2011). *Factory physics* (3rd ed.). Waveland Press.

Ivanov, D., & Dolgui, A. (2021). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control, 32*(9), 775-788. **https://doi.org/10.1080/09537287.2020.1768450**

Ivanov, D., Dolgui, A., Das, A., & Sokolov, B. (2025). Digital supply chain twins: Managing the ripple effect, resilience, and disruption risks by data-driven optimization, simulation, and visibility. In *Handbook of ripple effects in the supply chain* (pp. 407-432). Springer Nature Switzerland. **https://doi.org/10.1007/978-3-031-67630-9_18**

Kumbhar, M., Ng, A. H., & Bandaru, S. (2022). Bottleneck detection through data integration, process mining and factory physics-based analytics. In *10th Swedish Production Symposium (SPS2022)* (pp. 737-748). IOS Press. **https://doi.org/10.3233/ATDE220115**

Lai, X., Shui, H., Ding, D., & Ni, J. (2021). Data-driven dynamic bottleneck detection in complex manufacturing systems. *Journal of Manufacturing Systems, 60*, 662-675. **https://doi.org/10.1016/j.jmsy.2021.08.004**

Li, L., Chang, Q., & Ni, J. (2009). Data driven bottleneck detection of manufacturing systems. *International Journal of Production Research, 47*(18), 5019-5036. **https://doi.org/10.1080/00207540802267393**

Luiz, J. V. R., Souza, F. B. D., & Luiz, O. R. (2025). Theory of Constraints and Industry 4.0: Mutual contributions and research perspectives. *Production, 35*, e20250032. **https://doi.org/10.1590/0103-6513.20250032**

Mahmoodi, E., Fathi, M., & Ghobakhloo, M. (2022). The impact of Industry 4.0 on bottleneck analysis in production and manufacturing: Current trends and future perspectives. *Computers & Industrial Engineering, 174*, Article 108801. **https://doi.org/10.1016/j.cie.2022.108801**

Matzka, S. (2022, November 6). *Predictive maintenance dataset (AI4I 2020)*. Kaggle. **https://www.kaggle.com/datasets/stephanmatzka/predictive-maintenance-dataset-ai4i-2020**

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management, 30*(2), 272-298. **https://doi.org/10.1111/1467-8551.12343**

Moosavi, S., Farajzadeh-Zanjani, M., Razavi-Far, R., Palade, V., & Saif, M. (2024). Explainable AI in manufacturing and industrial cyber-physical systems: A survey. *Electronics, 13*(17), Article 3497. **https://doi.org/10.3390/electronics13173497**

Mourtzis, D. (2020). Simulation in the design and operation of manufacturing systems: State of the art and new trends. *International Journal of Production Research, 58*(7), 1927-1949. **https://doi.org/10.1080/00207543.2019.1636328**

Pambudi, A. (2025a, January 27). *Manufacturing efficiency in downtime operations*. Kaggle. **https://www.kaggle.com/datasets/agungpambudi/predict-manufacturing-downtime-performance-dataset**

Pashami, S., Nowaczyk, S., Fan, Y., Jakubowski, J., Paiva, N., Davari, N., ... Gama, J. (2023). Explainable predictive maintenance. *arXiv*. **https://doi.org/10.48550/arXiv.2306.05120**

Puthanveettil Madathil, A., Luo, X., Liu, Q., Walker, C., Madarkar, R., & Qin, Y. (2025). A review of explainable artificial intelligence in smart manufacturing. *International Journal of Production Research*. Advance online publication. **https://doi.org/10.1080/00207543.2025.2446820**

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137-141. **https://doi.org/10.1007/s11747-019-00710-5**

Roser, C., & Nakano, M. (2015). A quantitative comparison of bottleneck detection methods in manufacturing systems with particular consideration for shifting bottlenecks. In *IFIP international conference on advances in production management systems* (pp. 273-281). Springer. **https://doi.org/10.1007/978-3-319-22756-6_34**

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215. **https://doi.org/10.1038/s42256-019-0048-x**

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). *Explainable AI: Interpreting, explaining and visualizing deep learning* (Vol. 11700). Springer Nature. **https://doi.org/10.1007/978-3-030-28954-6**

Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. *Technology in Society, 61*, Article 101248. **https://doi.org/10.1016/j.techsoc.2020.101248**

Subramaniyan, M., Skoogh, A., Bokrantz, J., Sheikh, M. A., Thürer, M., & Chang, Q. (2021). Artificial intelligence for throughput bottleneck analysis—State-of-the-art and future directions. *Journal of Manufacturing Systems, 60*, 734-751. **https://doi.org/10.1016/j.jmsy.2021.08.003**

Tang, J., Dai, Z., Jiang, W., Wu, X., Zhuravkov, M. A., Xue, Z., & Wang, J. (2024). A comprehensive review of theories, methods, and techniques for bottleneck identification and management in manufacturing systems. *Applied Sciences, 14*(17), Article 7712. **https://doi.org/10.3390/app14177712**

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. *MIT Sloan Management Review, 60*(4), 1-5.

Telles, E. S., Lacerda, D. P., Morandi, M. I. W. M., & Piran, F. A. S. (2020). Drum-buffer-rope in an engineering-to-order system: An analysis of an aerospace manufacturer using data envelopment analysis (DEA). *International Journal of Production Economics, 222*, Article 107500. **https://doi.org/10.1016/j.ijpe.2019.107500**

Thürer, M., Stevenson, M., Silva, C., & Qu, T. (2017). Drum-buffer-rope and workload control in high-variety flow and job shops with bottlenecks: An assessment by simulation. *International Journal of Production Economics, 188*, 116-127. **https://doi.org/10.1016/j.ijpe.2017.03.010**

Tzionis, G., Mouratidis, P., Kougka, G., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I., & Vlachopoulou, M. (2025). A review of explainable AI methods and their application in manufacturing systems. *Discover Applied Sciences*. Advance online publication. **https://doi.org/10.1007/s42452-025-06565-9**

Vilone, G., & Longo, L. (2020). Explainable artificial intelligence: A systematic review. *arXiv*. **https://doi.org/10.48550/arXiv.2006.00093**

Voss, C. (2010). Case research in operations management. In *Researching operations management* (pp. 176-209). Routledge.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356-365. **https://doi.org/10.1016/j.jbusres.2016.08.009**

Wenzel, M., Stanske, S., & Lieberman, M. B. (2020). Strategic responses to crisis. *Strategic Management Journal, 41*(S1), 3161. **https://doi.org/10.1002/smj.3161**

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. *International Journal of Production Research, 56*(8), 2941-2962. **https://doi.org/10.1080/00207543.2018.1444806**

Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal, 13*(3), 2213-2227. **https://doi.org/10.1109/JSYST.2019.2897802**

Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., Da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering, 150*, Article 106889. **https://doi.org/10.1016/j.cie.2020.106889**