Temporal Dynamics and Causal Feature Integration for Predictive Maintenance in Manufacturing Systems: A Causality-Informed Framework

Seyed Iman Hosseini
iman.hosseini@ijs.si
Jožef Stefan Institute
Ljubljana, Slovenia
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Klemen Kenda
klemen.kenda@ijs.si
Jožef Stefan Institute
Ljubljana, Slovenia
Qlector
Ljubljana, Slovenia

Dunja Mladenič
dunja.mladenic@ijs.si
Jožef Stefan Institute
Ljubljana, Slovenia
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

ABSTRACT

Predictive maintenance is increasingly central to manufacturing, where the goals are to reduce unplanned downtime and extend asset lifetimes. Conventional models often rely on correlations that insufficiently capture temporal dynamics and causal dependencies underlying failures. This study proposes a causality-informed feature-engineering pipeline that combines cross-correlation-derived lags with VARLiNGAM to construct lag-aware features from multivariate sensor streams, and evaluates it against standard time-series models using a time-aware split. Three machine-learning models—Random Forest, XGBoost, and Gradient Boosting—were trained and assessed by F1-score (rather than accuracy) on a single-machine subset of the Microsoft Azure Predictive Maintenance dataset (8,708 samples; 26 failures, ≈0.3% prevalence). XGBoost trained on raw temporal features achieved F1 ≈ 0.94 for longer prediction horizons (≥10 h) under time-series–aware cross-validation, with performance declining at shorter horizons as temporal context diminishes. In this setting, causality-informed features did not improve results over the raw-feature baseline. These findings indicate that, with data from a single machine, causal discovery is susceptible to overfitting and may suppress informative temporal patterns; broader, multi-machine datasets are likely required for causality-enhanced representations to yield consistent gains.

KEYWORDS

Predictive Maintenance, Causality, Time-Series Analysis, Machine Learning, VARLiNGAM, Manufacturing Systems

1 INTRODUCTION

The rising complexity and interconnectivity of industrial systems have accelerated the need for intelligent maintenance strategies that move beyond reactive and preventive paradigms. Predictive maintenance, driven by sensor data and machine learning, has emerged as a transformative approach to minimize unplanned downtime and optimize asset life cycles [1]. Traditional predictive maintenance models, however, often rely on statistical correlations that fail to capture the directionality and temporal dynamics inherent in real-world system failures [6].

To address these limitations, this study proposes a causality-informed framework for predictive maintenance that leverages temporal causal discovery techniques, such as Vector Autoregressive LiNGAM (VARLiNGAM), to engineer predictive features from multivariate sensor data. Our approach integrates cross-correlation analysis and lag-optimized causal graphs to detect failure precursors and identify their optimal predictive windows.

We hypothesize that the observed lack of competitive advantage for causality-informed models, especially when applied to data from a single machine, arises from the limited operational diversity and failure variability. This limitation may cause models to overfit to machine-specific correlations and exclude informative temporal features, thereby hindering their generalizability. Testing this hypothesis through multi-machine datasets will be a key focus of future work.

2 RELATED WORK

Causality in time series analysis has become increasingly critical in predictive maintenance, particularly within industrial and manufacturing domains, where early failure detection plays a pivotal role in minimizing operational disruptions and financial losses [5]. Classical statistical models have been widely used to infer causal relationships between sensor measurements and machine states, yet they often fail to capture complex temporal dynamics and the nonlinear relationships inherent in real-world system failures.

Recent studies have explored advanced causal inference techniques to enhance fault prediction. Wang S. *et al.* proposed a framework for fault diagnosis that integrates spatiotemporal dependencies, demonstrating improved predictive accuracy in chemical manufacturing systems [9]. While their work advances reliability in industrial diagnostics, it lacks the flexibility to generalize across diverse application domains. On the other hand, Cui et al. introduced a deep learning framework that enhances predictive maintenance by integrating causal reasoning and long-sequence multivariate time-series data, significantly improving predictive performance and interpretability [3]. Despite this, the challenge of automating temporal feature engineering and seamlessly deploying models across different domains remains.

Yang X. *et al.* contributed to the growing literature on data-driven causal analysis by incorporating dynamic latent variables and probabilistic graphical models into causal modeling frameworks [10]. However, these models have yet to fully address the temporal feature extraction required for scalable deployment in real-world predictive maintenance applications. Furthermore, more recent work by Wang Q. *et al.* introduced a Causal Graph

Information Society, 2025, Ljubljana, Slovenia

https://doi.org/https://doi.org/10.70314/is.2025.sikdd.12

Convolution Module that adapts causal discovery within time-series prediction [8], but their approach is still dependent on complex model adjustments across domains.

In this study, we propose a novel framework that integrates lagged correlation with causal analysis techniques to detect failure precursors and quantify their lead times. This framework automates temporal feature engineering and is designed for diverse real-world applications across manufacturing settings, without requiring extensive architectural modifications. The automation of temporal feature engineering and its seamless deployment across comparable manufacturing environments remains a significant challenge, and extending generalization beyond this domain is left for future work.

3 EXPERIMENT

Our experimental methodology followed a sequential four-stage process to construct and validate a robust failure prediction model, as shown in Figure 1. The first stage involved performing a cross-correlation analysis between each sensor's time-series data and the target failure events to determine the optimal predictive time lag, which guided the subsequent steps. In the second stage, the identified optimal lag was used to parameterize a Vector Autoregressive LiNGAM (VARLiNGAM) model, which generated a directed acyclic graph (DAG) representing the causal relationships and effect strengths between sensor variables and the failure event. The third stage focused on creating a causality-informed feature vector by integrating standard statistical metrics from rolling time windows along with advanced features informed by the causal analysis, using the correlation strengths and causal effect strengths derived from the VARLiNGAM model to select and weight features based on their respective optimal and causal lags. Finally, in the fourth stage, the enriched feature set was fed into a machine learning pipeline, employing a time-based data split to prevent look-ahead bias, and training several classification models, including Random Forest, XGBoost, and Gradient Boosting, to assess the effectiveness of the causality-informed approach for predictive maintenance. This integrated approach enhances the predictive capabilities of machine learning models, offering a robust solution for failure prediction in industrial settings.

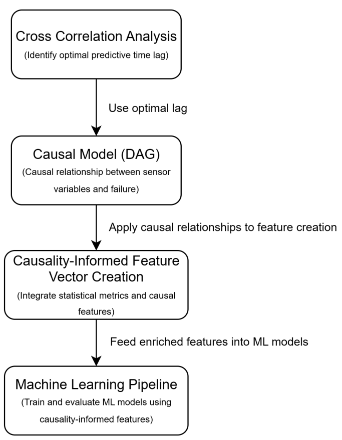

Figure 1: proposed framework

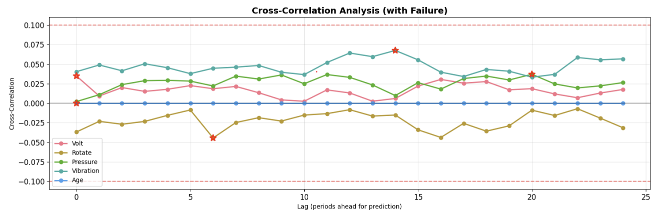

Figure 2: Cross correlation analysis

3.1 Dataset and Preprocessing

We used the Microsoft Azure Predictive Maintenance Dataset [2], which provides hourly telemetry (voltage, rotation, pressure, vibration) plus maintenance records, failure events, incident reports, and machine metadata for 100 machines over 12 months in 2015 (over 800k hourly summaries and thousands of non-failure error entries). For this study, we restricted the analysis to machine ID 98; after cleaning and merging the sources, we constructed a causality-informed feature vector and standardized features across modalities. Cross-correlation suggested predictive lags of 1–24 hours, so we derived lagged/statistical features from six primary variables (voltage, rotation, pressure, vibration, age, error type). The final dataset comprised 8,708 samples with 26 failures (≈0.3%), indicating strong class imbalance [7, 2]. The feature set comprised 150 causality-informed features and 36 features without causal information.

3.2 Cross-correlation Analysis

Cross-correlation analysis examines the correlation between two time series as a function of the time lag applied to one of them [11][12]. Unlike simple correlation, which measures linear relationships at a single point in time, cross-correlation reveals how variables relate across different time delays, making it particularly valuable for identifying lead-lag relationships and temporal dependencies. The initial phase of our experimental framework involved a cross-correlation analysis to empirically determine the predictive temporal relationships between sensor signals and equipment failures. For each sensor, we computed the Pearson correlation coefficient between its time series and the binary failure time series across a range of discrete time lags. This procedure was executed by systematically shifting the failure signal backward in time, which allowed for the correlation of sensor readings at a given time t with failure events at a future time t + lag. The optimal predictive lag for each sensor was then identified as the time lag that yielded the maximum absolute correlation value. This analysis is critical as it quantifies the time window in which each sensor's data is most informative for forecasting an impending failure, thereby providing an empirical foundation for the subsequent causal discovery and feature engineering stages.

In the cross-correlation plot shown in Figure 2, the red star annotated on each sensor's curve denotes the optimal predictive lag—20 hours for Pressure, 14 hours for Vibration, and so forth. This marker identifies the specific time lag, measured in hours, at which the sensor's signal exhibits the highest absolute Pearson correlation with the future failure event. Consequently, the red star highlights the most influential temporal offset for each variable, effectively quantifying the sensor's most informative predictive window within the 24-hour forecasting horizon.

3.3 Causal Graph Construction

To elucidate the causal interdependencies between sensor signals and equipment failures, a causal graph was constructed using VARLiNGAM. This methodology first employs a Vector Autoregression (VAR) model to capture the linear, time-lagged relationships among the multivariate sensor time series. The optimal lag for the VAR model was adaptively informed by the preceding cross-correlation analysis to focus on the most predictive temporal window. Following the VAR estimation, the LiNGAM algorithm is applied to the resulting model residuals, or innovations. By exploiting the non-Gaussian nature of these innovations, LiNGAM uniquely identifies the contemporaneous causal structure—the instantaneous effects between variables—and determines the direction of influence, thereby producing a directed acyclic graph (DAG). The final output is a set of adjacency matrices representing the causal graph, where each non-zero entry quantifies the strength and direction of a causal link from one variable to another at a specific time lag. Our approach constructs a directed causal graph from time-series sensor data using the following steps:

(1) **Data Sorting and Integrity:** Chronologically sort sensor data, verifying integrity and noting irregular intervals.
(2) **Variable Definition:** Define variables which are vibration, rotation, pressure, voltage, and a binary failure indicator as the target node.
(3) **Causal Model Setup:** Configure a VARLiNGAM [4] model with a specified lag order and BIC-based pruning.
(4) **Model Fitting:** Fit the model to the prepared data matrix, applying regularization—by adding small Gaussian noise (e.g., 10^{-6})—when numerical instability arises during VARLiNGAM causal graph construction due to ill-conditioned matrices.
(5) **Adjacency Extraction:** Extract adjacency matrices to identify directed edges, effect strengths, and corresponding lags.
(6) **Graph Assembly:** Assemble the causal graph, categorizing edges by their relation to the target and between sensor variables.

This workflow ensures that temporal ordering is respected and that detected causal links most likely represent meaningful relationships for predictive maintenance and further analytical investigations. Figure 3 presents the causal graph generated by the VARLiNGAM algorithm, illustrating the network of causal relationships between sensor telemetry (volt, pressure, vibration, rotate), machine properties (age), and the target failure event. In this graph, nodes represent the variables, and the directed edges (arrows) signify the direction of causality, with edge thickness corresponding to the strength of the effect. The labels on each edge quantify the causal strength and the time delay (lag) in hours. The analysis reveals a complex web of interactions, prominently highlighting that machine age is the most significant causal driver of failure, with an exceptionally strong effect strength at a lag of 6 hours. Other notable, though weaker, causal pathways are also identified, such as the influence of rotate on failure. This causal structure provides critical insights into the system's dynamics, identifying the key variables and time-delayed interactions that precede a failure event.

3.4 Causality-Informed Feature Engineering

We prepared the data by building a *causality-informed feature vector* grounded in the paper's causal graph and a temporal causality analysis that selects per-sensor optimal prediction windows. Using a sliding feature window (typically 72 h), samples are formed from historical data only to avoid leakage. Feature construction proceeds in four stages: (1) **basic statistics** (mean, standard deviation, min/max, latest/earliest within the window); (2) **causality-aligned temporal features** computed at the optimal lags identified by causal analysis; (3) **dynamics** via trend slopes (linear regression), rolling volatility (standard deviation), and rates of change; and (4) **cross-feature terms** implied by the causal graph (e.g., voltage/rotation ratios and pressure–vibration correlations). Targets are defined for multiple horizons (1, 6, 12, and 24 h ahead) to enable early warnings at different lead times. The resulting dataset contains 150 features that integrate causal dependencies with temporal patterns.

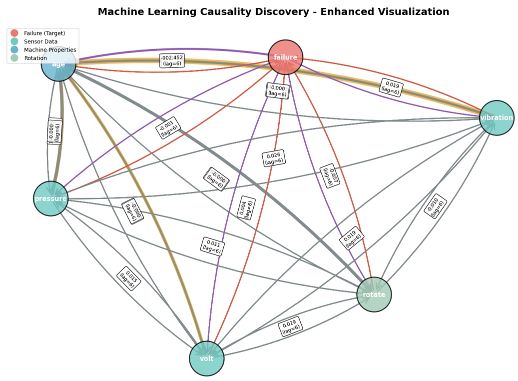

Figure 3: Causal Graph

3.5 Machine Learning Models

Three classification algorithms, each configured with default hyperparameters, were evaluated using time-based data partitioning to mitigate the risk of data leakage.

- **Random Forest (RF)**: Ensemble method with 200 estimators, maximum depth of 15, and balanced class weights
- **XGBoost (XGB)**: Gradient boosting with 200 estimators, learning rate of 0.1, and automatic scale balancing
- **Gradient Boosting (GB)**: Scikit-learn implementation with 200 estimators and 0.8 subsample ratio

Model performance was assessed using F1 Score metric appropriate for imbalanced classification:

- F1-Score: Harmonic mean of precision and recall

A time-series–aware data partitioning strategy was implemented using scikit-learn's *TimeSeriesSplit*, which generates folds in chronological order by progressively expanding the training set with earlier observations and reserving subsequent periods for testing. This procedure ensures that all training data temporally precedes the corresponding test data. To approximate stratification and preserve class balance between rare failure and more frequent non-failure events, the folds were constructed to proportionally distribute failure cases across splits without introducing randomization. This design maintains the temporal integrity of the sensor data while supporting reliable model evaluation.

4 RESULTS AND DISCUSSION

Figure 4 presents the comprehensive F1-score evaluation of all three models, while Figure 5 provides a comparative analysis

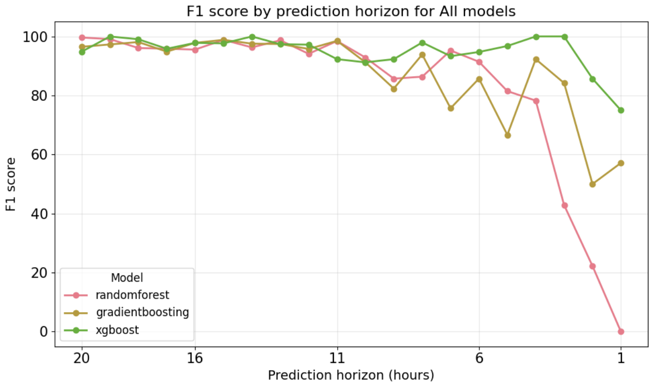

Figure 4: F1-score evaluated over a 20-hour prediction horizon

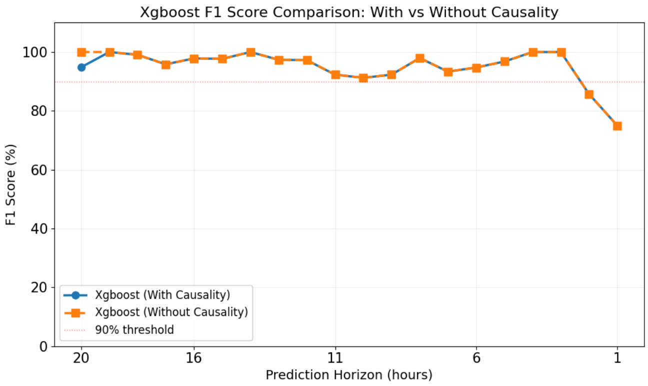

Figure 5: The XGBoost F1-score across a 20-hour prediction horizon, evaluated with and without a causality-informed feature vector

of the XGBoost model with and without the causality-informed feature vector. Standard time-series models, particularly those trained on raw temporal data, consistently outperform causality-informed approaches in predictive maintenance tasks, especially at extended prediction horizons. XGBoost, for instance, achieves F1 scores exceeding 94% for horizons beyond 10 hours, though performance declines with shorter windows due to reduced temporal context. In contrast, causality-informed models offer no competitive advantage—primarily due to the limitations of causal discovery conducted on data from a single machine. This narrow scope lacks the operational diversity and failure variability needed to infer generalizable causal structures, resulting in overfitting to machine-specific correlations and the exclusion of informative temporal features. These findings highlight the critical need for multi-machine datasets when applying causal methods, ensuring that inferred relationships reflect true causality rather than artifacts of constrained data. In addition, Longer prediction horizons (e.g., 20 hours) afford models access to extended historical windows (e.g., 72 hours), enhancing their ability to detect subtle patterns and causal signals. In contrast, short horizons (e.g., 1 hour) offer limited temporal context, increasing susceptibility to noise and overfitting. Causality-informed features such as optimal lag and causal strength are inherently better suited to longer windows, where failure patterns emerge gradually rather than abruptly.

5 FUTURE WORKS

While this study establishes a robust, domain-agnostic framework for failure prediction, future work will focus on enhancing its transparency and causal reasoning capabilities. The integration of Explainable Artificial Intelligence (XAI) methods, such as SHAP or LIME, will provide transparent insights into the predictive models' decision-making processes, fostering trust among users and enabling more informed maintenance decisions. Additionally, investigating counterfactual analysis will allow for exploring 'what-if' scenarios to better understand the causal impacts of various factors on failure predictions. Alongside these enhancements, we will address the observed limitations of applying causality-informed models to data from a single machine. Specifically, we hypothesize that the lack of competitive advantage stems from the limited operational diversity and failure variability of a single-machine dataset, leading to overfitting. Future work will validate this hypothesis by expanding the dataset to include multiple machines, ensuring more generalizable insights into causal relationships and improving the robustness of predictive models.

ACKNOWLEDGEMENTS

We gratefully acknowledge the European Commission for its support of the Marie Skłodowska-Curie program through the Horizon Europe DN APRIORI project (GA 101073551).

REFERENCES

[1] Abdeldjalil Benhanifia, Zied Ben Cheikh, Paulo Moura Oliveira, Antonio Valente, and José Lima. 2025. Systematic review of predictive maintenance practices in the manufacturing sector. *Intelligent Systems with Applications*, 26, 200501. DOI: https://doi.org/10.1016/j.iswa.2025.200501.

[2] Arnab Biswas. 2025. Microsoft azure predictive maintenance. Accessed: 2025-05-20. (2025). https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance/data.

[3] Qing'an Cui, Jiao Lu, and Xianhui Yin. 2025. Causality enhanced deep learning framework for quality characteristic prediction via long sequence multivariate time-series data. *Measurement Science and Technology*, 36, (Mar. 2025), 3, (Mar. 2025). DOI: 10.1088/1361-6501/adb05a.

[4] LiNGAM Developers. 2025. VARLiNGAM — LiNGAM 1.10.0 documentation. https://lingam.readthedocs.io/en/latest/tutorial/var.html. Accessed: 2025-06-25. (2025).

[5] Karim Nadim, Ahmed Ragab, and Mohamed Salah Ouali. 2023. Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining. *Journal of Intelligent Manufacturing*, 34, (Jan. 2023), 57–83, 1, (Jan. 2023). DOI: 10.1007/s10845-021-01903-y.

[6] P. Nunes, J. Santos, and E. Rocha. 2023. Challenges in predictive maintenance – a review. *CIRP Journal of Manufacturing Science and Technology*, 40, 53–67. DOI: https://doi.org/10.1016/j.cirpj.2022.11.004.

[7] Margarida Da Rocha and Faísca Moreira. 2024. FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO Data-Driven Predictive Maintenance for Component Life-Cycle Extension. Tech. rep.

[8] Qipeng Wang, Shoubo Feng, and Min Han. 2025. Causal graph convolution neural differential equation for spatio-temporal time series prediction. *Applied Intelligence*, 55, (May 2025), 7, (May 2025). DOI: 10.1007/s10489-025-06287-7.

[9] Sheng Wang, Qiang Zhao, Yinghua Han, and Jinkuan Wang. 2023. Root cause diagnosis for complex industrial process faults via spatiotemporal coalescent based time series prediction and optimized granger causality. *Chemometrics and Intelligent Laboratory Systems*, 233, (Feb. 2023). DOI: 10.1016/j.chemolab.2022.104728.

[10] Xing Yang, Tian Lan, Hao Qiu, and Chen Zhang. 2025. Nonlinear causal discovery via dynamic latent variables. *IEEE Transactions on Automation Science and Engineering*. DOI: 10.1109/TASE.2024.3522917.

[11] Tanja Zerenner, Marc Goodfellow, and Peter Ashwin. 2021. Harmonic cross-correlation decomposition for multivariate time series. *Physical Review E*, 103, (June 2021), 6, (June 2021). DOI: 10.1103/PhysRevE.103.062213.

[12] XIAOJUN ZHAO, PENGJIAN SHANG, and JINGJING HUANG. 2017. Several fundamental properties of dcca cross-correlation coefficient. *Fractals*, 25, 02, 1750017. eprint: https://doi.org/10.1142/S0218348X17500177. DOI: 10.1142/S0218348X17500177.