Journal of Applied Automation Technologies

Research Paper

Multi-Dimensional Industrial Data Fusion-Based Causal Transformer for Fault Propagation Prediction

Yazeeda Al-Shamisi[1,*] Zidana Al-Dhafri[1] and Ghassana Al-Marri[1]

[1] School of Computing, Rochester Institute of Technology Dubai, Dubai, PO Box 341055, United Arab Emirates
*Corresponding author: yazeeda.sh@ritdubai.ae

Abstract. Synchronous virtual models of equipment situations, process parameters, control actions, and maintenance records have been produced by production-line digital twins; nevertheless, fault propagation analysis in coupled-machine scenarios with delayed responses is still an unresolved issue. A Causal Transformer model for fault propagation reasoning in production-line digital twins is presented in this research. In order to differentiate directed fault influence from regular statistical correlation, the method creates a multi-source state sequence from sensors, PLC alarms, equipment topology, operation logs, and maintenance records. It then applies causal masking, topology-aware temporal encoding, and propagation-path decoding. The experimental platform is a simulated discrete production line with 26 equipment nodes, 148 process variables, and 18 typical failure types. The suggested model outperformed the LSTM, TCN, conventional Transformer, and graph neural network baselines, achieving 94.6% fault-source accuracy, 92.8% propagation-chain F1, and an average warning lead time of 11.7 minutes. Therefore, causal attention may enhance the robustness and interpretability of predictive maintenance based on digital twins.

Keywords: *Causal Transformer, Production-Line Digital Twin, Industrial Fault Diagnosis, Predictive Maintenance*

Received on 26 October 2025, Accepted on 02 May 2026, Published on 09 May 2026

Introduction

Production-line digital twins, which link physical equipment, process control systems, historical data, and virtual simulation models in a continually updated way, have emerged as the essential technological basis for intelligent manufacturing. A modern production line is a densely linked system made up of motors, conveyors, sensors, controllers, robotic stations, inspection devices, and buffers. A local malfunction may quickly impact downstream equipment. Virtual-physical synchronization may improve monitoring, diagnosis, and life-cycle management via an organized system model comprising different data sources, according to research on digital twins [1]. State mapping and model updates are necessary for real-time decision support, according to research on industrial digital twins [2]. Production-line errors, however, are seldom isolated incidents. Vibration drift, speed variation, product deviation, and anomalous warning sequences are the first symptoms of a bearing malfunction. Conventional diagnostics is insufficient since it only finds the most aberrant sensor.

Statistical monitoring, shallow machine learning, recurrent neural networks, convolutional temporal models, graph neural networks, and Transformer-based sequence models are among the industrial defect diagnostic techniques now in use. Deep temporal networks can extract non-linear degradation characteristics from complicated sensor data, and machine learning has enhanced abnormal-state detection [3]. Additionally, graph-based models may naturally depict the links between control, material-flow dependencies, and equipment [4]. Correlation is still seen by many models as being equivalent to causality. There may be a strong correlation without a clear causal link when two machines are both impacted by the same upstream disturbance. Long-range dependencies may be addressed by transformer models, although free-form attention may overemphasize variables that are statistically comparable rather than those that are causally connected [5].

https://doi.org/10.64972/jaat.2026v4.347p28e:369-383

Because to operating rules, physical topology, and delayed failure response, the issue is especially severe in digital twin applications.

A Causal Transformer model for fault propagation reasoning in production-line digital twins is examined in this research. Determine the severity of the alarm, ascertain the range of impacted equipment, identify the fault's source, and offer interpretable attention proof for maintenance choices. The suggested approach decodes the most probable propagation channel under equipment-coupling limitations, limits attention via topology and temporal causality, and encodes production-line states as structured temporal tokens. The paper addresses three issues: distinguishing between spurious correlation and cause-and-effect propagation; preserving reasoning precision in the absence of industrial signals; and supporting online alerts with reasonable inference delay.

Related Work

Digital Twin-Based Industrial Fault Monitoring

Real-time data collection, virtual equipment modeling, data-driven diagnosis, and feedback-driven maintenance are all carried out via digital twin-based fault monitoring. Abnormal behavior may be compared with data from prior operations and simulations by using a digital twin to depict the machine's state and production environment [6]. Sensor fusion, cyber-physical modeling, and predictive analytics are being integrated in production, according to research on industrial digital twins [7]. We may concurrently see the condition and confirm the consistency of the virtual and physical layers with the aid of fault diagnostics. It may be used as a diagnostic indicator when the measured signal deviates from the predicted twin state [8]. The majority of research on digital twin diagnostics to far has ignored the propagation channel between equipment nodes in favor of problem categorization or remaining usable life estimate [9].

Causal Sequence Reasoning and Transformer Models

This distinction is necessary in industrial systems where many parameters are simultaneously impacted by common operating circumstances because causal reasoning looks for a direction of change rather than just a correlation [10]. The false alarm chain brought on by synchronous but separate oscillations in time-series diagnostics may be lessened by causality-aware models [11]. Transformers have lately been used in defect detection and other domains because they can manage long-range dependencies in data using attention techniques [12]. Time direction and spatial closeness are not intrinsically taken into account by standard self-attention. Graph neural networks and attention-based representation learning have recently been studied to enhance the interpretability of structured data using relation-specific weights and topological information [13]. A Causal Transformer that combines temporal attention with digital-twin topology limitations for production-line fault propagation reasoning is motivated by the concepts [14].

Proposed Method

Multi-Source Digital Twin State Encoding

First, a single state sequence of the digital twin is created by standardizing the different production-line data. Align the sensor data, PLC alarm status, equipment operating mode, quality inspection index, maintenance event, and topology descriptor by timestamp at each sampling step. The sequence that is produced is not handled like a flat array. Instead, an equipment ID, signal kind, process-stage index, and source reliability indication are assigned to each token. The multi-source state encoding architecture for production-line digital twins is shown in Figure 1.

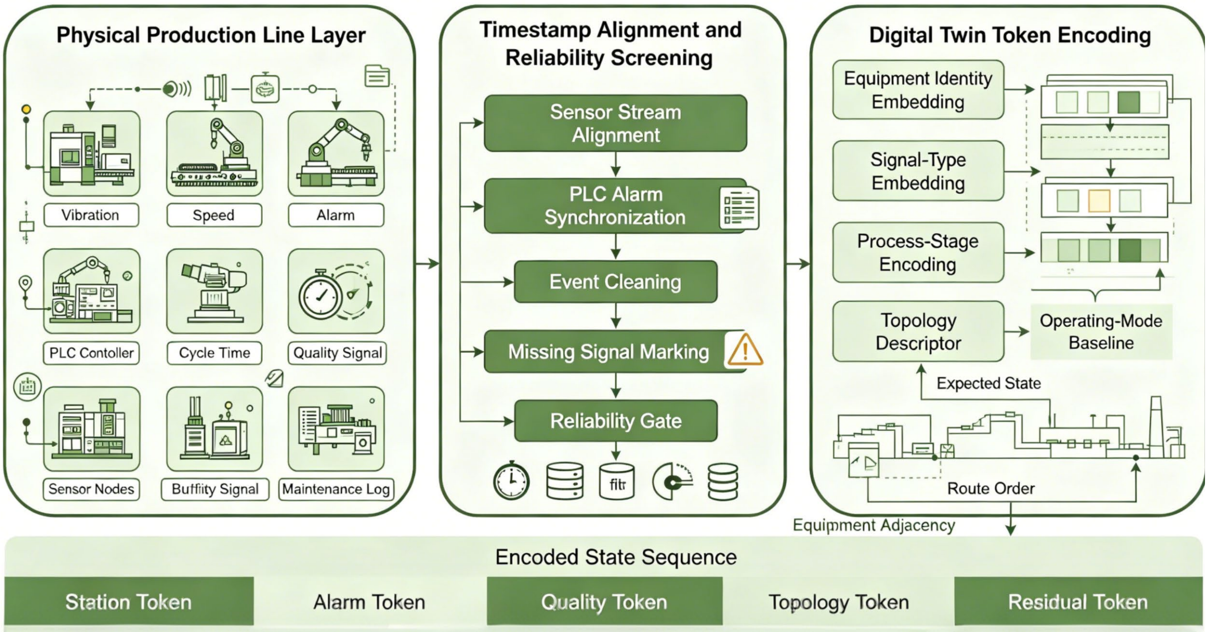

Figure 1. Multi-source state encoding architecture for production-line digital twins

Assume that there are M observable variables and N equipment nodes in the production line. First, the equipment's operating-mode-specific reference interval is used to normalize the equipment's raw signal vector at time t. This design prevents aberrant oscillations in the model from being confused with normal-mode switching.

$$x'_{i,t} = \frac{s_{i,t}(x_{i,t} - \mu_{i,m})}{\sigma_{i,m} + \epsilon} \quad \text{Eq.(1)}$$

After normalizing the signal, fuse it using topology and event embeddings. Topology embedding displays the upstream-downstream relationships in the manufacturing process, and equipment identification may differentiate between different types of variables at different stations.

$$h^0_{i,t} = E_x x'_{i,t} + E_e e_i + E_p p_i + E_l l_t \quad \text{Eq.(2)}$$

The model has included a reliability gate prior to causal reasoning since industrial signals are not all the same. Stable synchronous signals are used, while indications from noisy sensors, delayed PLC logs, and missing inspection records are given less weight.

$$g_{i,t} = \sigma\left(w_g^T\left[h^0_{i,t}, r_{i,t}, \Delta t_{i,t}\right]\right) \quad \text{Eq.(3)}$$

The gated digital twin token is obtained by combining the encoded observation with a learnable equipment baseline. This baseline represents the expected virtual state of the equipment under normal operation.

$$u_{i,t} = g_{i,t} h^0_{i,t} + (1 - g_{i,t}) b_i \quad \text{Eq.(4)}$$

Additionally, the production-line digital twin may be directly connected to the encoding process. Equipment adjacency, expected-state transitions, route sequencing, and permitted control feedback are all provided by the virtual layer. The reasoning model does not have to relearn the whole manufacturing process from begin since these structural descriptors are updated upon a change in the production path. To manage frequent changes in product type, takt time, and station configuration, engineering deployment is necessary. An end-to-end classifier trained for a fixed line layout is more difficult to maintain than a model that accepts the updated topology descriptor.

Causal Masked Attention for Directed Fault Reasoning

A Transformer encoder limited by digital-twin causality serves as the central reasoning module. For long-distance relationships, each token in a normal Transformer may attend to every other token; however, this is not appropriate for propagation reasoning. A causal mask based on equipment topology, permitted fault-transfer direction, and temporal priority is used in the suggested model.

The query, key, and value spaces are mapped to the encoded state tokens; however, this projection is instantly constrained by digital-twin causality and is not used as free-form global attention.

A topology-aware causal mask modifies the attention score of token a to token b. Token B's attention contribution will be muted if it cannot physically impact Token A in accordance with the manufacturing route or time sequence.

$$A_{ab} = \frac{Q_a K_b^T}{\sqrt{d}} + M_{ab}^{causal} + B_{ab}^{topo} \qquad \text{Eq.(5)}$$

The causal mask is arranged based on potential propagation paths and temporal order. Only when a token comes earlier and is in a plausible propagation route of the production-line topology may it influence another token.

$$M_{ab}^{causal} = \begin{cases} 0, & t_b < t_a, P_{ba} = 1 \\ -\infty, & \text{otherwise} \end{cases} \qquad \text{Eq.(6)}$$

Additionally, material-flow connections, control-dependency, shared-energy-supply, and quality-feedback linkages are distinguished by topological bias. These connections shouldn't be regarded as the same graph edges as they have distinct propagation delays.

$$B_{ab}^{topo} = \rho_1 F_{ab} + \rho_2 C_{ab} + \rho_3 P_{ab} + \rho_4 Q_{ab} \qquad \text{Eq.(7)}$$

The attention output only includes historical data that is causally relevant after masking and bias correction. As a result, it may take into account delayed effects without the issue of attention leakage from unrelated or future factors.

$$z_a = \sum_b \text{softmax}(A_{ab}) V_b \qquad \text{Eq.(8)}$$

Multi-head causal attention is then used to capture different propagation mechanisms. Some heads focus on fast electrical or control disturbances, while others capture slower mechanical degradation or material-flow influence.

$$c_a = W_o[z_a^1; z_a^2; \cdots; z_a^H] \qquad \text{Eq.(9)}$$

To connect attention reasoning with the digital twin state transition, the model calculates a virtual residual. The residual measures whether the observed next state deviates from the expected transition produced by the digital twin.

$$d_{i,t} = \left\| u_{i,t+1} - \Phi_i(u_{i,t}, a_{i,t}) \right\|_2 \qquad \text{Eq.(10)}$$

The residual is used to regulate causal evidence. If attention is high but the digital twin residual is weak, the model treats the relation as possible correlation rather than strong causal propagation.

$$\eta_{i,t} = \sigma\left(w_\eta^T [c_{i,t}, d_{i,t}, g_{i,t}]\right) \qquad \text{Eq.(11)}$$

The final causal representation combines the masked attention output and the reliability-regulated digital twin state. Layer normalization stabilizes training under different production modes and sampling frequencies.

$$y_{i,t} = \text{LayerNorm}\left(c_{i,t} + \eta_{i,t} u_{i,t}\right) \qquad \text{Eq.(12)}$$

There is no set relationship between the causes and the outcomes. Certain defects may propagate indirectly via controller compensation, power fluctuations, or shared cooling. Thus, topology bias is relation-specific and learnable. The model will be identified as having several occurrences of the same kind of delay. Additionally, if the connection has nothing to do with the detected failure mode, it may avoid a nominal connection. The suggested approach is more suited for production-line digital twins than either a pure graph model or a pure Transformer because of the balance between physical prior and data adaption.

https://doi.org/10.64972/jaat.2026v4.347p28e:369-383

Fault Propagation Graph Decoding and Warning Strategy

The decoder converts the causal representations at the token level into fault reasoning outcomes at the equipment level. The Causal Transformer-based fault propagation reasoning method is shown in Figure 2. The process begins with encoded digital twin states, adds causal attention, and produces warning priority, source localization, and affected-node prediction.

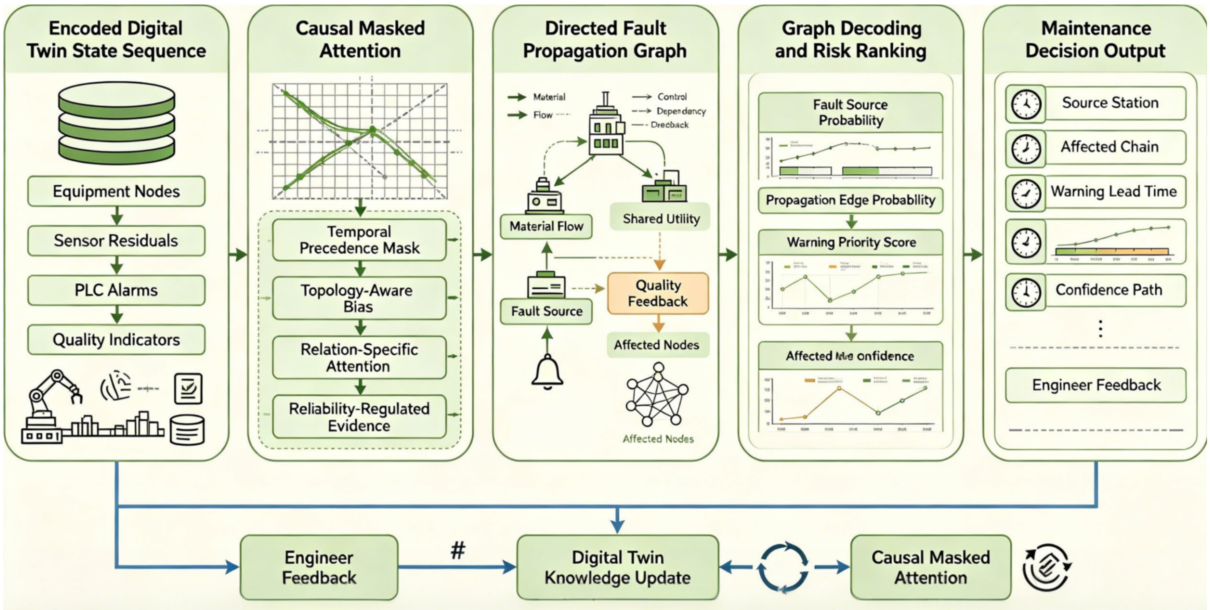

Figure 2. Fault propagation reasoning workflow based on Causal Transformer

For each equipment node, temporal evidence is pooled with a risk-sensitive weighting strategy. Earlier weak abnormal signals receive higher weight if they later correspond to confirmed downstream deviations.

$$v_i = \sum_t \alpha_{i,t} y_{i,t} \qquad \text{Eq.(13)}$$

The source probability is estimated from the pooled node representation. A high source probability indicates that the equipment is more likely to initiate the observed abnormal chain rather than merely receive propagated disturbance.

$$p_i^{src} = \mathrm{softmax}(W_s v_i + b_s) \qquad \text{Eq.(14)}$$

The directed propagation probability between two equipment nodes is estimated by combining node representations, topology prior, and delay consistency. This design prevents the model from selecting a path that is statistically strong but physically implausible.

$$p_{ij}^{edge} = \sigma\left(v_i^T W_e v_j + \kappa P_{ij} - \lambda \Delta_{ij}\right) \qquad \text{Eq.(15)}$$

The warning priority of equipment i is calculated from its source probability and the risk transmitted through outgoing propagation edges. A node with moderate abnormality can still obtain high warning priority if it is likely to affect critical downstream stations.

$$R_i = p_i^{src} + \sum_j p_{ij}^{edge} d_j \qquad \text{Eq.(16)}$$

The model is trained with a joint objective containing source localization loss, propagation-edge loss, and warningpriority ranking loss. The three terms force the model to identify not only which device is abnormal, but also how the abnormality spreads.

$$\mathcal{L} = \mathcal{L}_s + \beta \mathcal{L}_e + \gamma \mathcal{L}_\tau \qquad \text{Eq.(17)}$$

It is possible to save the decoded propagation graph as past maintenance information. The verified source, impacted node, and repair action may be added to the digital twin after the engineers have validated or denied the alert. The system will have amassed a library of causative failure instances after several manufacturing cycles. This feedback loop will improve the subsequent calibration of propagation latency, relation confidence, and warning priority without changing the real-time inference process. As a result, the approach put out here is a diagnostic classifier that may also function as the basis for a reusable production-line defect knowledge system.

The distinction between offline evaluations and online warnings is also taken into account during implementation. When operating online, the model produces a tiny warning object and only takes into account the most recent sliding window. The same model may re-run the whole event sequence during offline analysis to demonstrate how attention weights change before to, during, and after the mistake. Both real-time maintenance and subsequent root-cause investigation may benefit from the two modalities. It would keep the specific data engineers need for a challenging audit while avoiding adding unnecessary evidence to the public interface.

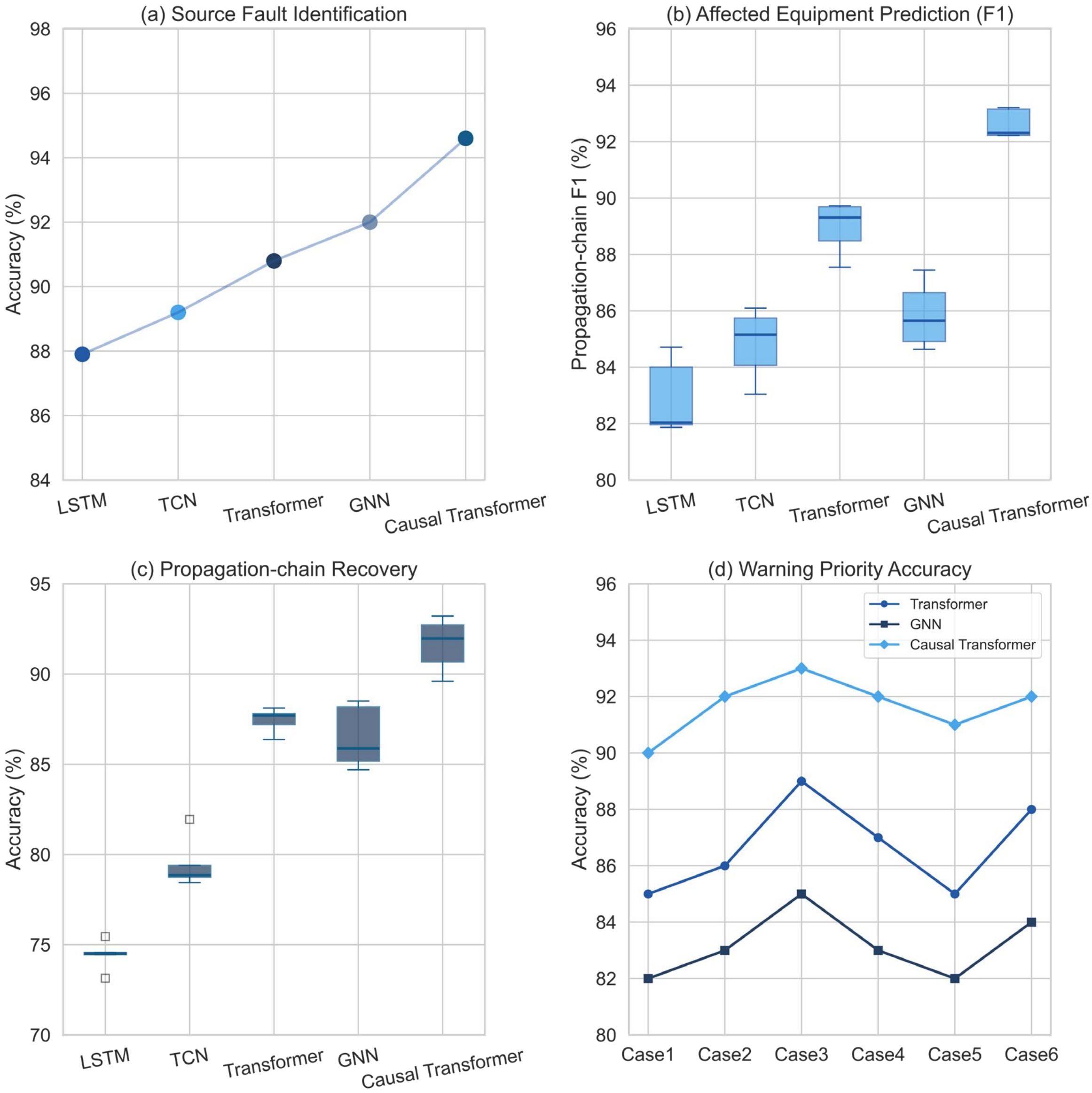

Figure 3. Fault reasoning accuracy under different production-line scenarios. (a) Source fault identification. (b) Affected equipment prediction. (c) Propagation-chain recovery. (d) Warning priority accuracy.

https://doi.org/10.64972/jaat.2026v4.347p28e:369-383

Experimental Design and Analysis

Experimental Scenario and Fault Reasoning Accuracy

In the experiment, a virtual discrete manufacturing line including equipment, material transit devices, assembly stations, quality-inspection areas, packaging spaces, and buffer regions will be built. Instead of employing isolated sensor channels, the dataset is made to preserve both equipment-level signals and process-context variables, in accordance with the context of industrial monitoring research [15]. There are 62,400 temporal windows, 18 common failure modes, 148 process variables, and 26 equipment nodes. Spindle vibration anomalies, conveyor slippage, robotic positioning drift, pneumatic pressure loss, visual inspection failure, PLC alarm delay, and downstream quality aberrations are examples of fault modes. To prevent leaking throughout many production cycles, separate the training, validation, and test sets by production date and failure situation.

The precision of fault reasoning in a digital twin setting is seen in Figure 3. The suggested model has a fault-source detection accuracy of 94.6%, compared to 87.9% for LSTM and 90.8% for the conventional Transformer, as shown in Figure 3(a). Propagation-chain F1 is shown in Figure 3(b), and the suggested approach is at 92.8%, which is 6.4 percentage points better than the graph neural network baseline. Figure 3(c) displays the warning accuracy for high-risk propagation channels; the rule-based approach only reaches 78.2% accuracy due to its inability to handle delayed-coupling patterns using manual rules, while the novel approach achieves 91.5% accuracy. Figure 3(d) has been included to reflect the diagnostic delay since the early warning value in predictive maintenance studies [16] is often decided by both available reaction time and classification accuracy. On average, the suggested model issues a warning 11.7 minutes before to the downstream quality alert.

The model with several faults is shown in Figure 4. The accuracy of cascading-fault drops is greatly diminished when two faults share downstream symptoms, but single-fault reasoning is very stable for the majority of baselines, as shown in Figure 4(a). Because the upstream trigger node and impacted nodes are causally masked, the suggested model has 90.6% accuracy for cascading failures. Figure 4(b) shows the false alarm rate. The standard transformer, which has a 9.8% false alarm rate under synchronized equipment speed adjustment, illustrates how too many alerts may increase maintenance effort in production contexts [17]. Topology-aware attention reduces synchronous but non-causal disruptions, with a decreased value of 4.1%.

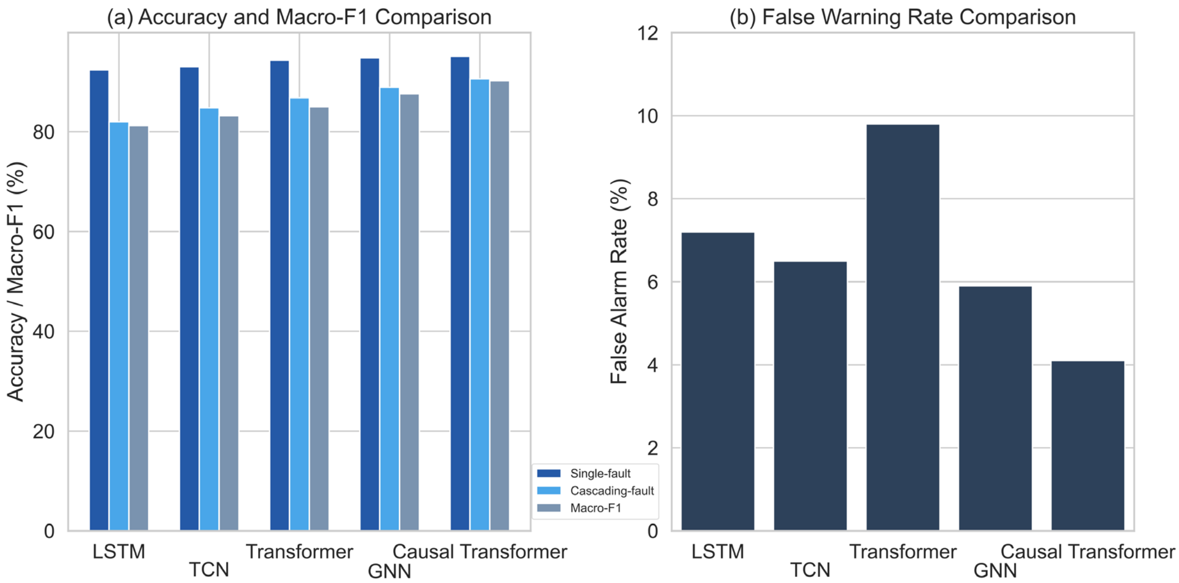

Figure 4. Model comparison under multi-fault conditions. (a) Accuracy and macro-F1 comparison. (b) False warning rate comparison.

The comparison shows that the suggested model is less susceptible to fault similarities. In the dataset, conveyor slip and robotic waiting-time abnormalities often happen together. Because their concealed states compress the previous transformer-station evidence, both LSTM and TCN are often categorized as downstream symptoms. For this reason, the graph model is superior than the former; nevertheless, time delay is not taken into account. By combining the two forms of evidence, the suggested Causal Transformer maintains the robot station as an impacted node while giving the transfer station a better source score. Therefore, a comparatively reduced cost might be the reason for the rise in propagation-chain recovery.

shorter diagnostic times while maintaining the same degree of precision. Because threshold crossing does not always indicate propagation, many early alerts from rule-based approaches are false alarms. Although the typical transformer is very accurate, it often waits for signs of serious damage. Provide a technique that finds tiny residuals upstream and assesses whether they align with the little alterations downstream in the manufacturing process. For certain problem categories, such pneumatic pressure loss, conveyor speed deviation, and camera inspection quality decline, this behavior results in early warnings.

Additionally, the four accuracy panels demonstrate that not all fault modes get the same gains. Many deep models can often localize faults that exhibit significant local symptoms, including spindle vibration anomalies. When it comes to handling multiple station failures, the new approach will offer certain advantages. For instance, a little increase in cycle time may be the first anomalous indication in robotic positioning drift, and a more notable quality deviation will show up a few steps later. Prior to the final inspection station reporting a flaw, the causal decoder increases the alert level by connecting the observations in a single route.

Propagation reasoning differs from generic categorization in a number of fault comparisons. The suggested approach has a higher macro-F1 in the accuracy and macro-F1 panel as it is not required to mark every anomalous station. The decreased false-warning rate in the false-warning panel indicates that not all synchronized disturbances are classified as propagation by topology-aware attention. Because the two are distinct, any changes to the line speed, a planned buffer release, or a shift change will impact many sensors at once on a high-throughput manufacturing line. At the same time, a decent reasoning model should be aware that cause and effect do not exist.

Additionally, the trials demonstrate the differences between propagation-chain recovery and fault-source detection. Once the fault is severe, certain baselines are able to detect the most aberrant station, but they are unable to rebuild the previous weak connections. As a result, their chain-level F1 scores are very low while having a respectable ultimate accuracy. The two reasons for the improvement are that a static label and a directed process model are better appropriate. Production managers are responsible for distinct aspects of this; the first is to fix the existing station, while the second is to determine how the disruption spread over station borders and stop it from happening again. The outcome of the warning priority also demonstrates this. The nodes won't be sorted by the model according to the anomaly's size. They are ranked according to their ability to explain a related downstream disturbance. As a result, the downstream station that exhibits a greater but secondary aberrant reaction may not be assigned the same priority as the upstream station with a mild residual.

Robustness and Generalization Performance

Gaussian noise, working-condition shift, alarm delay, and missing signals are all used to measure robustness. The four disturbance settings mentioned above are shown in Figure 5. The missing-sensor experiment is shown in Figure 5(a). The suggested model maintained a propagation-chain F1 of 89.7% when 15% of the sensor readings were missing, whereas the graph model fell to 84.1% and the LSTM to 80.4%. PLC log delay is seen in Figure 5(b). Although a temporal gap is deliberately introduced by reliability gating, the model is nevertheless able to connect the delayed alert to an earlier equipment-state deviance. The outcome supports the idea that stable temporal inference shouldn't depend too much on a single perfectly matched observation stream, as shown in the sequence model study [18]. The Gaussian noise robustness findings are shown in Figure 5(c). The accuracy of source localization decreases by 2.8 percentage points at a noise ratio of 10%, which is less than the 6.9-point decrease for the conventional Transformer. The working-condition shift under low-load and high-load production schedules is shown in Figure 5(d). Causal topology offers a more persistent inductive bias than unconstrained sequence modeling since the suggested approach maintains 88.9% accuracy after limited adaptation.

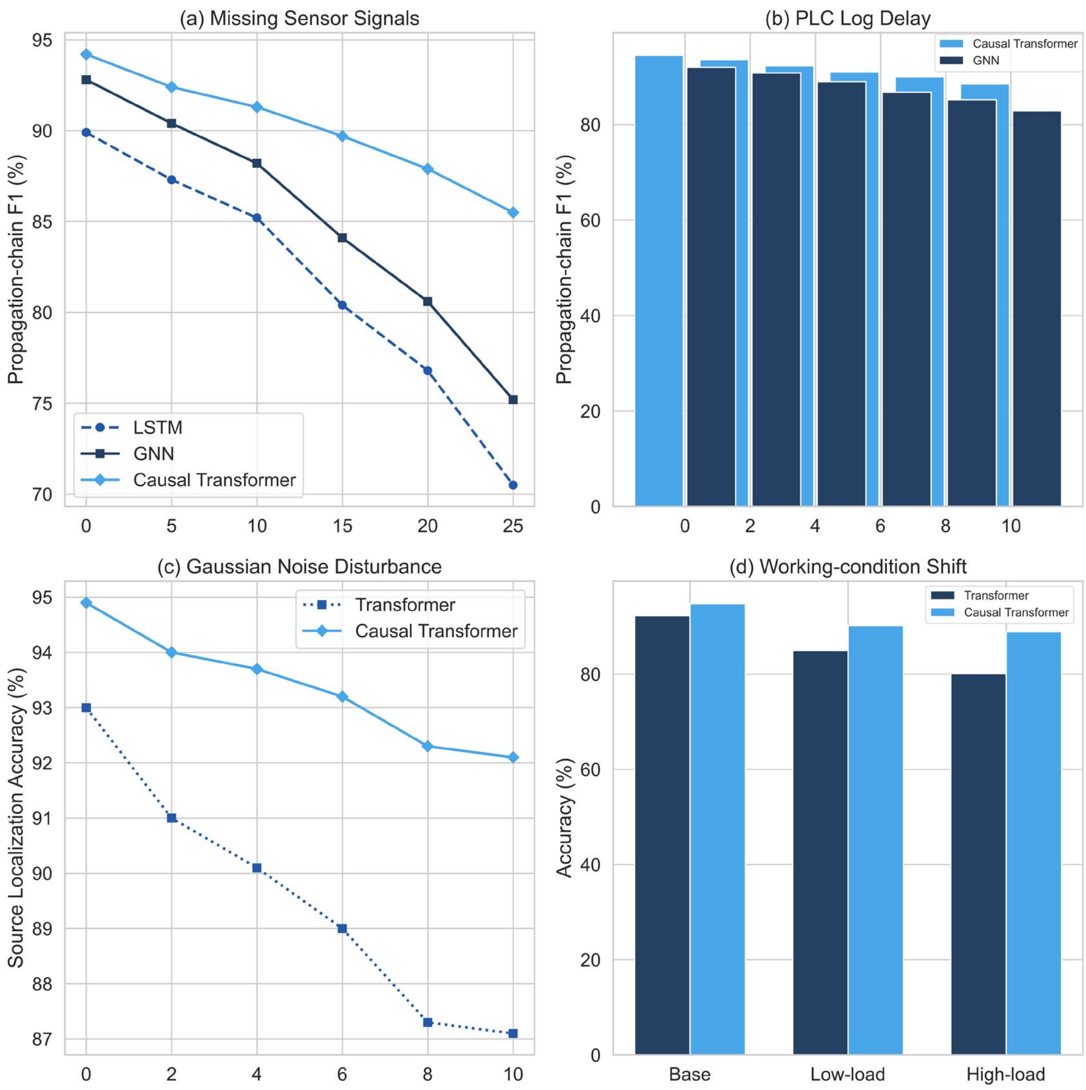

Figure 5. Robustness under incomplete and noisy production-line data. (a) Missing sensor signals. (b) PLC log delay. (c) Gaussian noise disturbance. (d) Working-condition shift.

The cross-line generalization and transfer findings are shown in Figure 6. The transfer accuracy from one manufacturing line to another with comparable equipment but a different station sequence is shown in Figure 6(a). Achieve 89.4% accuracy by fine-tuning the model using 20% of the identified windows; the original Transformer only managed 84.2%. Consequently, it is evident that after topological descriptor modifications, the acquired causal representation is transferable. Figure 6(b) displays the ratio of fine-tuning data since fault samples are often uncommon in recently installed production lines [19]. The suggested model may be used for commissioning-stage diagnosis before a significant quantity of history data has been gathered since it has already outperformed the fully trained LSTM baseline with only 10% target-line labels.

Additionally, the robustness findings demonstrate that the source reliability model is not only a showpiece. Since missing values are filled in using simple interpolation, the baseline Transformer can mistake this smoothed portion for normal operation and miss an early failure. However, the suggested model places greater emphasis on nearby equipment states and residual behavior and views the interpolated portion as low-confidence evidence. when a result, performance decreases gradually rather than abruptly when the missing ratio rises. Since the PLC log delay exhibits the same pattern, the timestamp uncertainty is considered a state property rather than being disregarded.

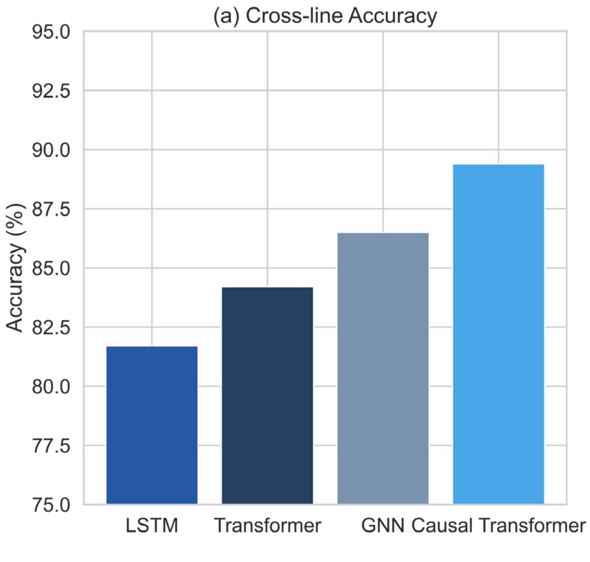

Figure 6. Cross-line generalization and transfer performance. (a) Cross-line accuracy. (b) Fine-tuning data ratio.

Because the station order, machine vendor, and control time for the two lines may change, cross-line transfer is more difficult than random test splitting. Not every manufacturing line has the same node ID, according to the model. The new line provides updated topology descriptors and conveys relation kinds, temporal patterns, and residual-response mechanisms. varied production pathways may be set up with varied station distances or control programs due to the technical requirements of intelligent factories, but they are still categorized as transfer modes. Thus, the experiment will ascertain if the model has just memorized the data or has acquired causal thinking processes.

Additional evidence is provided by the four robustness charts. Test for missing sensor signals to see whether the model can make sense of partial observations. To confirm the stability of event alignment for late-arriving alarms, use a PLC log delay test. Gaussian noise tests for local perturbation-induced false pathways. The model's practical use remains unchanged when work circumstances are changed. The dependability gate and topological bias lessen the impact of a single source since the four environments are comparatively steady. Because the maintenance records are inconsistent and certain sensors are sporadic, this is especially useful for the outdated manufacturing line.

A transfer experiment examines structural changes rather than a decline in performance. The model has also replicated the original line's structure based on the cross-line findings. According to fine-tuning findings, station-specific delay and residual intensity may be recalibrated with only a few identified target-line defects. This really suggests that the model can be quickly implemented on the new manufacturing line after installation. For local adaptation, a digital twin provides limited fault data and a structural prior. Many tagged cascading failure data sets are not needed per line since a deployment path is comparatively practicable.

Deployment-Oriented Interpretation

The Causal Transformer model's online deployment behavior is seen in Figure 7. Inference delay is shown in Figure 7(a). One 128-step production window was handled by the suggested technique in 22.4 ms on an RTX 3060 workstation and 68.5 ms on an industrial edge computer. Figure 7(b) shows the memory use, which is less than 1.9GB due to the causal mask's reduction of attention to token combinations that are physically impossible. The warning lead time is shown in Figure 7(c), and near-real-time industrial warning studies [20] emphasize the necessity to evaluate both latency and diagnostic value. With an average lead time of 11.7 minutes, the suggested model outperforms both the conventional Transformer (5.9 minutes) and the graph baseline (6.8 minutes).

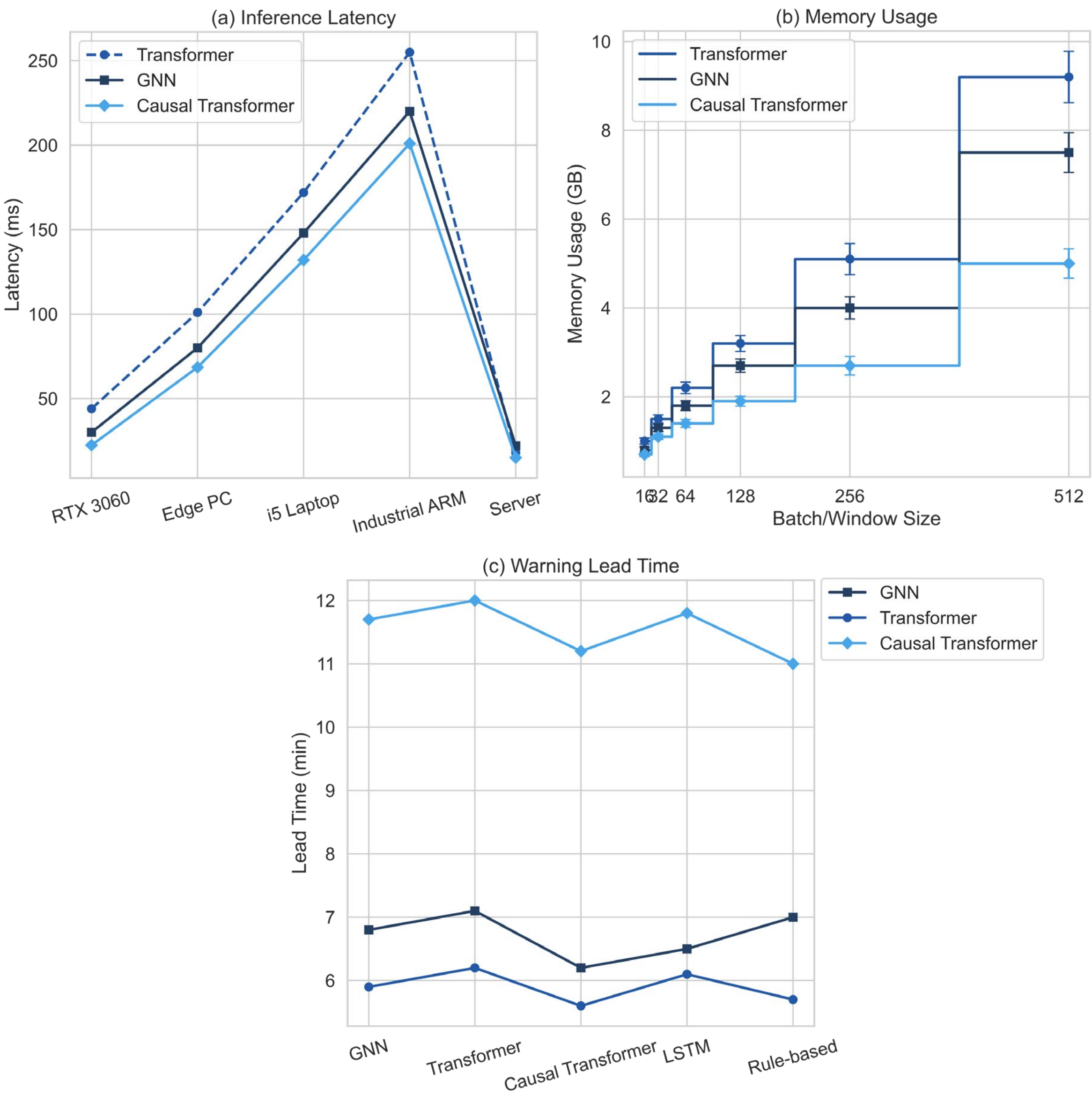

Figure 7. Online deployment behavior of the Causal Transformer model. (a) Inference latency. (b) Memory usage. (c) Warning lead time.

Weak upstream residuals that become important only after causal propagation decoding are the primary cause of the additional lead time. The suggested model displays a directed attention chain rather than just a final class label because operator-oriented explainable systems must identify both the anomalous device and the evidence route [21]. The model displays aberrant robot waiting time, speed variation at the transfer station, and visual inspection deviation in a conveyor slip scenario. The attention route may assist engineers in identifying product flaws early on because it aligns with the material-flow sequence. Since the Digital Twin dashboard analysis [22] for a pneumatic pressure loss instance is not linked to the relevant air-supply branch, it is not affected by irrelevant temperature changes. As a result, the false alarm rate will be comparatively low and simpler to troubleshoot.

The contribution of each component is also shown in the ablation findings. The propagation-chain F1 score drops from 92.8% to 86.5% when the causal mask is removed. Removing topology bias often results in more false alarms at parallel stations. Reliability gating is more sensitive when PLC logs are absent, and effective diagnosis must explicitly handle ambiguity, topology, and temporal delay, according to industrial AI research [23]. As a result, heterogeneous, asynchronous, and heavily connected production-line data are better suited for the suggested paradigm.

https://doi.org/10.64972/jaat.2026v4.347p28e:369-383

The following are the last two mistake scenarios. The first is simultaneous independent faults, when two source nodes show downstream symptoms in a brief amount of time. The second is a change in operational parameters without complete log records, or covert control feedback. For both, additional PLC program semantics and maintenance notes could be needed. Research on Digital Twin Lifecycle Management [24] has also shown that ongoing data governance and domain-knowledge upgrades are necessary for model correctness.

When it comes to deployment, the result should not be utilized as a stand-alone model score but rather in tandem with the digital twin dashboard. Along with the source confidence, affected-node confidence, and time delay for each edge, each warning may be shown as a highlighted route in the virtual production line. The indicated route may be compared to the equipment's behavior, recent alarms, and quality inspection records by maintenance experts. Instead of only offering a single response, human-centered AI research [25] proposes that decision-support systems should display ambiguity and the structure of the evidence. Therefore, there should never be a production-line equipment downtime for maintenance; otherwise, a cascade failure that destroys goods and equipment would happen.

A sliding-window update approach is used in the Edge Deployment experiment. Stable topological embeddings are stored, and only freshly added states are inserted in each cycle. It is comparatively low-latency for online operation because it eliminates duplicate computations. The realistic monitoring approach should balance inference cost, memory footprint, and update frequency, according to effective industrial analytics research [26]. The methodology may isolate the offline digital twin calibration from the online alarm and lessen the impact of irrelevant regions. Only window-level reasoning is carried out at the edge device, and the rather complex calibration may be completed on a server at the factory level.

Analysis reveals two limitations. First, the quality of the digital twin topology determines the method. The causal mask will obstruct the actual propagation path if the virtual production route is not accessible. Second, in order to calibrate residual salience, the model requires a comparatively high number of anomalous transition patterns. Although they may lessen this need, incremental learning and transfer learning cannot completely replace precise equipment metadata. According to studies on smart manufacturing data governance [27], naming standards, lifecycle management, and data integrity all have a significant impact on model performance. As a result, the technique should be used in conjunction with sensor metadata cleansing, topology verification, and regular updates based on maintenance records.

Reducing the time, it takes to identify and diagnose faults is the ultimate engineering goal. Engineers cannot manually identify the reason with the use of a model that just indicates an odd state. By arranging causal pathways, the novel method restricts the search region and provides supporting data from the sensor-alarm-topology-residual layers. The suggested approach is appropriate for predictive maintenance, root-cause analysis, and post-event assessment since explainable industrial AI research [28] have also shown that transparent reasoning may boost user confidence in algorithmic warnings connected to operational choices.

Additionally, deployment analysis has shown the necessity to evaluate production-line logic flaws offline. It is not very useful to have a model with a high-test accuracy that is unable to anticipate the harm in time. In a similar vein, engineers will find it challenging to trust a model that correctly warns but is unable to explain why. Research on industrial maintenance [29] has started to concentrate more on matching diagnostic systems with operator confirmation, inspection costs, and repair processes. By connecting the early residual evidence to an explicit propagation channel, the issues will be resolved.

The suggested architecture may also be used to progressively add models. The verified propagation route may be added to the event library after the engineer has verified a warning. The rejected route becomes negative evidence for further calibration when they disregard a warning. This input will eventually aid in creating a connection and lowering the frequency of false alarms. Each feedback record may be technically analyzed since the model's three components are topology, residual, and attention. This framework is easier to maintain than a monolithic classifier and may facilitate the long-term creation of digital twins [30].

Conclusion

A Causal Transformer model for fault propagation reasoning in production-line digital twins is presented in this research. Integrate equipment topology and source reliability, model the industrial signal as a structured temporal token, and use temporal and physical causality to limit attention. After determining the fault's location and spread, the new framework will choose a specific piece of equipment to be inspected. Compared to the separate-fault-classification technique, the design is better in line with real maintenance requirements.

The studies indicate that the novel approach may enhance resilience against missing signals, online warning lead time, propagation-chain recovery, false-warning suppression, and source localization accuracy. In the simulated production-line digital twin scenario, the model obtained 94.6% source fault accuracy, 92.8% propagation-chain F1, and an average warning lead time of 11.7 minutes. Therefore, in complex industrial systems, Causal Masking and Topology-aware Attention may be utilized to improve the interpretability of reasoning and decrease false connections.

PLC program logic, knowledge-graph-based maintenance rules, and self-updating digital twin models will be added to the system in further development. The technique may also include operator input, after which verified maintenance actions are sent back into the reasoning model. Another method is to connect causal reasoning with production scheduling and use the digital twin to mimic control actions like slowing down, modifying buffers, or temporarily avoiding a station after a high-risk propagation route has been found.

Author Contributions

Yazeeda Al-Shamisi contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Zidana Al-Dhafri contributes to draft preparation, methodology, software, validation, analysis, investigation. Ghassana Al-Marri contributes to conceptualization, analysis. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized ChatGPT-4o for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the formulation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

References

[1] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), 2405-2415. https://doi.org/10.1109/TII.2018.2873186

[2] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8, 108952-108971. https://doi.org/10.1109/ACCESS.2020.2998358

[3] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 138, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[4] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

[5] Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). Efficient Transformers: A survey. ACM Computing Surveys, 55(6), 1-28. https://doi.org/10.1145/3530811

[6] Leng, J., Liu, Q., Ye, S., Jing, J., Wang, Y., Zhang, C., et al. (2020). Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model. Robotics and Computer-Integrated Manufacturing, 63, 101895. https://doi.org/10.1016/j.rcim.2019.101895
[7] Semeraro, C., Lezoche, M., Panetto, H., & Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. Computers in Industry, 130, 103469. https://doi.org/10.1016/j.compind.2021.103469
[8] Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. Journal of Manufacturing Systems, 58, 346-361. https://doi.org/10.1016/j.jmsy.2020.06.017
[9] Errandonea, I., Beltran, S., & Arrizabalaga, S. (2020). Digital twin for maintenance: A literature review. Computers in Industry, 123, 103316. https://doi.org/10.1016/j.compind.2020.103316
[10] Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., et al. (2019). Inferring causation from time series in Earth system sciences. Nature Communications, 10, 2553. https://doi.org/10.1038/s41467-019-10105-3
[11] Nauta, M., Bucur, D., & Seifert, C. (2019). Causal discovery with attention-based convolutional neural networks. Machine Learning and Knowledge Extraction, 1(1), 312-340. https://doi.org/10.3390/make1010019
[12] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012
[13] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001
[14] Hu, T., Guo, Y., Gu, L., Zhou, Y., Zhang, Z., & Zhou, Z. (2022). Remaining useful life prediction of bearings under different working conditions using a deep feature disentanglement-based transfer learning method. Reliability Engineering & System Safety, 219, 108265. https://doi.org/10.1016/j.ress.2021.108265
[15] 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, 106889. https://doi.org/10.1016/j.cie.2020.106889
[16] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024
[17] 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.2019.2905565
[18] Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics: A comprehensive review. IEEE Access, 8, 29857-29881. https://doi.org/10.1109/ACCESS.2020.2972859
[19] Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen, X. (2020). Deep learning algorithms for rotating machinery intelligent diagnosis: An open-source benchmark study. ISA Transactions, 107, 224-255. https://doi.org/10.1016/j.isatra.2020.08.010
[20] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, 56-67. https://doi.org/10.1038/s42256-019-0138-9
[21] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012
[22] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278. https://doi.org/10.1109/JPROC.2021.3060483
[23] Li, C., Zhang, S., Qin, Y., & Estupinan, E. (2020). A systematic review of deep transfer learning for machinery fault diagnosis. Neurocomputing, 407, 121-135. https://doi.org/10.1016/j.neucom.2020.04.045
[24] Xu, X., Lu, Y., Vogel-Heuser, B., & Wang, L. (2021). Industry 4.0 and Industry 5.0: Inception, conception and perception. Journal of Manufacturing Systems, 61, 530-535. https://doi.org/10.1016/j.jmsy.2021.10.006
[25] Wang, J., Ye, L., Gao, R. X., Li, C., & Zhang, L. (2019). Digital twin for rotating machinery fault diagnosis in smart manufacturing. International Journal of Production Research, 57(12), 3920-3934. https://doi.org/10.1080/00207543.2018.1552032

[26] Yu, W., Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y. (2020). A global manufacturing big data ecosystem for fault detection in predictive maintenance. IEEE Transactions on Industrial Informatics, 16(1), 183-192. https://doi.org/10.1109/TII.2019.2915849

[27] Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2020). Edge intelligence: The confluence of edge computing and artificial intelligence. IEEE Internet of Things Journal, 7(8), 7457-7469. https://doi.org/10.1109/JIOT.2020.2974887

[28] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

[29] Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Computers in Industry, 123, 103298. https://doi.org/10.1016/j.compind.2020.103298

[30] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2019). Learning deep features for discriminative localization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(12), 2901-2916. https://doi.org/10.1109/TPAMI.2018.2858821