e-ISSN: 3079-6377 p-ISSN: 3079-6369

Spatiotemporal Graph Networks for Predicting Transformer Failures in Regional Power Grids

Marius Andersen* and Chiara Lombardi

Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Norway

*** Corresponding author:** marius.andersen.ee@gmail.com

Abstract

Power transformers are critical assets in regional electricity infrastructure, and their unexpected failures frequently trigger cascading outages with severe economic consequences. Traditional fault diagnosis approaches, including dissolved gas analysis (DGA) and periodic physical inspections, remain inadequate for proactive real-time monitoring across large and topologically complex grid networks. This paper proposes a spatiotemporal graph network (STGN) framework that jointly models the topological structure of a regional power grid and the temporal dynamics of transformer operational data to achieve early failure prediction. The architecture integrates spatio-temporal convolutional blocks, each comprising temporal gated convolution layers flanking a spatial graph convolution layer with gated linear unit activations, to capture both spatial propagation patterns between interconnected devices and time-varying deterioration signals at individual nodes. Panoramic state information, including dissolved gas concentrations, load data, and qualitative maintenance records, is encoded through sequential LSTM processing to support multi-horizon failure probability estimation. Experiments on a real-world regional transmission dataset demonstrate a precision of 91.4%, recall of 89.7%, and F1 score of 90.5%, surpassing support vector machine, LSTM, and standard convolutional neural network baselines by margins of 8 to 17 percentage points. Performance comparison across Correlation, CSI, FAR, and POD metrics at multiple prediction horizons confirms the sustained accuracy advantage of the spatiotemporal formulation over non-graph baselines.

Keywords

spatiotemporal graph network; transformer failure prediction; power grid topology; graph convolutional network; predictive maintenance; dissolved gas analysis; deep learning

1.Introduction

The reliable operation of power transformers represents one of the most fundamental prerequisites for the stability of modern electrical grids. As central nodes in regional transmission and distribution networks, transformers mediate energy flow between generation sources and end consumers, and any unplanned failure can initiate cascading disruptions that propagate across interconnected substations. According to industry statistics, transformer faults account for a disproportionate share of total grid downtime, with large-scale failures resulting in billions of dollars of economic damage annually and posing serious risks to industrial operations, healthcare infrastructure, and public safety [1]. As regional grids grow increasingly complex through the integration of renewable energy sources and

e-ISSN: 3079-6377 p-ISSN: 3079-6369

expanded load demand, the probability of transformer stress and premature aging rises correspondingly, making the development of accurate and scalable predictive maintenance solutions an urgent engineering priority [2]. Conventional approaches to transformer health monitoring have historically relied on periodic offline testing methodologies, most prominently DGA, partial discharge detection, and thermo graphic inspection [3]. While these techniques can identify incipient faults under controlled laboratory conditions, they are inherently reactive, labor-intensive, and poorly suited to continuous monitoring of geographically distributed grid assets [4]. Online monitoring systems have emerged as a partial remedy, enabling the real-time collection of operational parameters including oil temperature, load current, and insulation resistance [5]. However, classical signal processing and statistical methods applied to such data streams struggle to generalize across heterogeneous transformer types and operating environments, and they fundamentally fail to account for the spatial interdependencies that exist between physically adjacent or electrically coupled devices within a grid [6]. The advent of deep learning has substantially advanced the state of the art in equipment health monitoring and predictive maintenance [7]. LSTM networks and variants of Graph neural networks (GNNs) have demonstrated strong performance in modeling temporal patterns within sensor time series, enabling systems to detect anomalies well before overt failures manifest [8]. Nevertheless, such architectures treat each transformer as an isolated entity, discarding the rich structural information encoded in the topology of the surrounding grid. In reality, the operational condition of a given transformer is closely coupled to the behavior of neighboring devices: a fault propagating through one substation will alter load distributions, voltage profiles, and thermal stress patterns at adjacent nodes in ways that carry significant diagnostic information [9]. GNNs offer a principled framework for reasoning over structured data in which the relationships between entities are as informative as the attributes of individual entities themselves [10]. In power system applications, a regional grid can be naturally represented as a graph in which transformers and buses constitute nodes, and transmission lines define edges with associated impedance properties. By performing message-passing operations over such a graph, a GNN-based model can aggregate information from a transformer's topological neighborhood, enabling joint inference over the state of multiple devices simultaneously [11]. When the temporal dimension is incorporated through STGNs that combine graph convolution with gated convolutional temporal modules, the model can additionally capture the dynamic evolution of fault precursors across both time and space, a capability that isolated temporal models fundamentally lack [12]. This paper makes the following principal contributions to the field. First, it introduces an STGN architecture specifically designed for transformer failure prediction, in which the grid topology is encoded as a dynamic adjacency matrix updated with real-time load and impedance measurements. The core architectural element is a spatio-temporal convolutional block comprising two temporal gated convolution layers flanking a central spatial graph convolution layer, enabling efficient parameter sharing and multi-scale feature extraction simultaneously in the temporal and spatial dimensions [13]. Second, it proposes a panoramic state information encoding scheme that integrates quantitative sensor indicators including dissolved gas concentrations and load data with qualitative operational records such as maintenance history and inspection findings, processed jointly through sequential LSTM layers to produce calibrated health state probability vectors [14]. Third, it provides a comprehensive empirical evaluation against established baselines across multiple prediction metrics and forecast horizons, demonstrating both quantitative accuracy gains and qualitatively meaningful improvements in detection lead time.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

2. Literature Review

The literature on transformer fault diagnosis has undergone significant evolution over the past decade, transitioning from physics-based interpretive frameworks toward data-driven and machine learning approaches capable of extracting complex patterns from large-scale monitoring datasets [15]. Early work in this domain relied heavily on rule-based expert systems grounded in IEC and IEEE DGA standards, which prescribe threshold-based interpretations of gas concentration ratios for hydrogen, methane, ethylene, and acetylene as indicators of distinct fault categories including thermal degradation, partial discharge, and arcing [16]. While these standards remain widely deployed in industry, their sensitivity is constrained by the variability of transformer designs, insulating oil formulations, and operating conditions, leading to high rates of both false alarms and missed detections. Subsequent research explored fuzzy logic and Bayesian inference as mechanisms to represent the uncertainty inherent in threshold-based rules, yielding modest improvements in diagnostic accuracy but limited scalability to large transformer populations [17]. The introduction of machine learning methods to transformer diagnostics brought substantial gains in classification performance. Support vector machines and ensemble methods such as random forests were applied to engineered feature sets derived from DGA measurements, demonstrating accuracies above 85% on benchmark datasets [18]. The subsequent application of deep LSTM networks to transformer operating state prediction represented a notable conceptual advance, as it enabled the model to capture not only instantaneous sensor readings but the temporal trajectory of state evolution over extended observation windows. A comprehensive input representation incorporating both quantitative indicators including dissolved gas concentrations, load records, and thermal measurements over multiple historical time steps and qualitative operational indicators such as maintenance records, working condition assessments, and inspection information can be jointly encoded through stacked LSTM layers followed by a Softmax classification head to produce calibrated state probability estimates across multiple health categories [19]. This panoramic state modeling approach, which processes heterogeneous input streams under a unified sequential architecture, provides a methodological foundation that the present work extends to the graph domain, enabling each transformer node to maintain a rich multi-source health representation that feeds into subsequent spatial aggregation operations. A critical limitation shared by most prior diagnostic methods is their failure to exploit the relational structure of the grid in which transformers are embedded. Research in power system analysis has long established that fault events are not spatially independent: voltage collapse, thermal runaway, and overload cascades propagate along electrical pathways in ways that reflect the underlying network topology [20]. The emergence of GNNs as a general-purpose framework for learning on graph-structured data has created new opportunities to address this gap. Foundational work on comprehensive surveys of graph neural network architectures established that message-passing formulations could capture multi-hop neighborhood dependencies more expressively than independent feature extraction approaches, providing the theoretical basis for power system applications [21]. Subsequent work demonstrated that GNN-based models could significantly outperform conventional methods on tasks including fault localization, transient stability assessment, and optimal power flow computation by leveraging the intrinsic graph structure of the grid [22]. The integration of temporal dynamics into graph-based models has further expanded the scope of spatiotemporal learning for infrastructure monitoring. A seminal architectural contribution was the spatio-temporal graph convolutional network proposed by Yu et al., which formulated the time series prediction problem on graphs using complete convolutional structures [23]. The key innovation of this architecture lies in the ST-Conv block design, wherein two temporal gated convolution layers employing

e-ISSN: 3079-6377 p-ISSN: 3079-6369

GLU activations bracket a central spatial graph convolution layer, and bottleneck channel strategies reduce parameters while preserving representational capacity. This purely convolutional formulation enables significantly faster training than recurrent alternatives while maintaining the ability to capture comprehensive spatiotemporal correlations, a finding that has been replicated across multiple infrastructure monitoring domains. Dynamic graph convolution incorporating phasor measurement unit (PMU) data for power system state prediction subsequently demonstrated that time-varying adjacency representations can reduce mean absolute error by over 20% relative to static graph baselines, confirming the importance of adaptive relational modeling [24]. The performance dynamics of different architectural choices across prediction horizons represent a particularly important dimension of comparison for maintenance-oriented applications. Research on spatiotemporal sequence models for physical process forecasting has demonstrated that purely fully-connected LSTM architectures, while effective at short prediction horizons, suffer dramatically degraded performance at extended lead times, exhibiting sharply declining Correlation scores and deteriorating Critical Success Index (CSI) values while maintaining elevated False Alarm Rate (FAR) values [25]. In contrast, architectures that explicitly model spatial structure through convolutional operators or graph aggregation maintain more favorable performance profiles at longer horizons, attributable to their ability to leverage neighborhood context as additional predictive signal when individual node trajectories become less deterministic. These findings directly motivate the multi-horizon evaluation strategy and the use of Correlation, CSI, FAR, and Probability of Detection (POD) as evaluation metrics in the present study. Research on multi-source data fusion further demonstrated that combining DGA measurements with thermal imaging and load profiles through attention-based architectures improved fault classification accuracy by up to 7%, underscoring the value of panoramic input representations [26]. Adaptive machine learning pipelines for transformer lifetime prediction with uncertainty quantification have also established that calibrated probabilistic outputs are practically valuable for maintenance scheduling under resource constraints [27]. Work on predicting basin stability of power grids using GNNs established these models as efficient surrogates for computationally expensive numerical simulations [28]. Geometric deep learning approaches for online cascading failure prediction across large transmission networks further confirmed the scalability of graph-based methods to real-world grid sizes [29]. Multi-task spatiotemporal graph convolutional frameworks have also shown promise in jointly performing state prediction and transient stability assessment, demonstrating the versatility of the spatiotemporal graph formulation for diverse power system monitoring tasks [30]. Despite these advances, no existing study simultaneously leverages grid topology, temporal sensor dynamics, and multi-source operational records within a unified end-to-end spatiotemporal framework designed specifically for transformer failure prediction across regional grid scales.

3. Methodology

3.1 Graph Construction and Spatiotemporal Feature Encoding

The first component of the proposed framework is the construction of a graph representation of the regional power grid that captures both the static topological structure and the dynamic operational state of transformer assets. Each power transformer in the network is modeled as a node vi in a graph G = (V, E, X, A), where V is the node set with |V| = N transformers, E is the edge set representing transmission lines connecting adjacent substations, X ∈ R^{N×T×F} is the node feature matrix encoding F sensor channels over T time steps for each transformer, and A ∈ R^{N×N} is the adjacency matrix weighted by the electrical distance between connected nodes. The sensor features collected at each node encompass both quantitative and

e-ISSN: 3079-6377 p-ISSN: 3079-6369

qualitative information streams: the quantitative stream includes oil temperature, winding temperature, load factor, voltage level, harmonic distortion index, and six DGA-derived gas concentration indices for hydrogen, methane, ethylene, acetylene, carbon monoxide, and carbon dioxide; the qualitative stream encodes maintenance records, inspection findings, and working condition assessments as structured categorical embeddings, yielding a combined 22-dimensional feature vector per time step after embedding projection. The edge weighting scheme adopts a hybrid formulation combining a physics-based impedance term derived from the grid's power flow model with a data-driven Pearson correlation term computed from recent sensor readings over a rolling 72-hour observation window. This dual weighting ensures sensitivity to both the stable electrical topology and transient operational couplings that emerge during stress events such as peak load periods or emergency load transfers. The adjacency matrix is updated at each inference step to reflect current operational correlation structure, tracking topology changes arising from line switching or substation reconfigurations managed by the supervisory control and data acquisition (SCADA) system. Missing sensor readings, which occur with approximately 3.2% frequency in the dataset, are imputed using a forward-fill procedure combined with a learned bias correction term estimated from the training distribution. The spatial feature extraction module is built around the ST-Conv block architecture, as illustrated in Figure 1. Each ST-Conv block contains two temporal gated convolution layers with channel dimension $C = 64$ arranged in a sandwich configuration around a central spatial graph convolution layer with channel dimension $C = 16$. The temporal gated convolution layers on both sides of the block apply 1-D convolution kernels across the time dimension, with GLU activations that allow selective gating of temporal features based on their content relevance, providing an efficient alternative to recurrent units that avoids the sequential dependency bottleneck inherent in LSTM-based temporal processing. The spatial graph convolution layer in the middle of each block applies a Chebyshev polynomial approximation to propagate information across topological neighborhoods, enabling each transformer node to aggregate health context from adjacent substations. Residual connections bridge the input and output of each ST-Conv block to facilitate gradient flow during back propagation, and bottleneck channel strategies reduce the total parameter count while preserving representational capacity. Two ST-Conv blocks are stacked sequentially, allowing the model to capture two-level hierarchical spatiotemporal feature abstractions, followed by a fully connected output layer that maps the learned representations to the failure prediction space. This architectural design directly reflects the need to jointly model the propagation of fault precursors across the grid topology and the temporal deterioration trajectory within individual transformer nodes.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

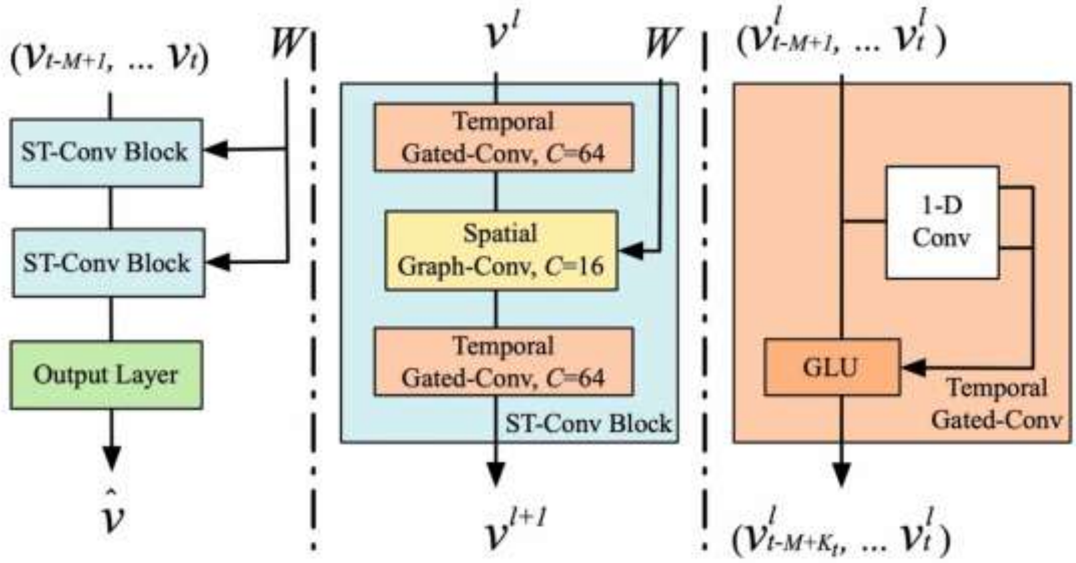

Figure 1 Architecture of the Spatio-Temporal Graph Convolutional Network used in the proposed transformer failure prediction framework

The left panel shows the overall model structure comprising two stacked ST-Conv Blocks followed by a final Output Layer that produces the predicted node state sequence v̂. The middle panel details the internal composition of each ST-Conv Block, consisting of a Temporal Gated-Conv layer with channel dimension C=64 at the top, a Spatial Graph-Conv layer with channel dimension C=16 in the center, and another Temporal Gated-Conv layer with C=64 at the bottom, with residual weight connections W bridging block inputs and outputs to preserve gradient flow. The right panel illustrates the internal mechanism of the Temporal Gated-Conv operation, in which a 1-D convolution kernel processes the input temporal sequence and feeds into a GLU activation unit that selectively gates feature responses, compressing the input window from (v^l_{t-M+1}, ..., v^l_t) to (v^l_{t-M+K_t}, ..., v^l_t). This block design enables simultaneous extraction of spatial neighborhood dependencies and temporal deterioration dynamics, which is essential for capturing fault precursor propagation patterns across the regional grid topology.

3.2 Panoramic State Encoding and Failure Prediction Head

The second component of the proposed system is the panoramic state encoding module, which processes heterogeneous input streams from each transformer node through a sequential LSTM architecture to produce temporally aware node-level health representations. As illustrated in Figure 2, the input to this module is organized into two complementary streams that together constitute the panoramic state information for each transformer asset. The quantitative indicator stream includes time-stamped measurements of gases dissolved in oil, load data, and other continuously monitored operational parameters, each encoded over n historical time steps to provide a temporal look-back context that enables the LSTM to identify gradual deterioration trends invisible in instantaneous snapshots. The qualitative indicator stream captures poor working condition flags, maintenance record summaries, inspection record information, and running status membership values that represent the degree to which current operating conditions match predefined degradation or fault category prototypes, encoded through learned categorical embeddings that are jointly optimized during end-to-end training.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

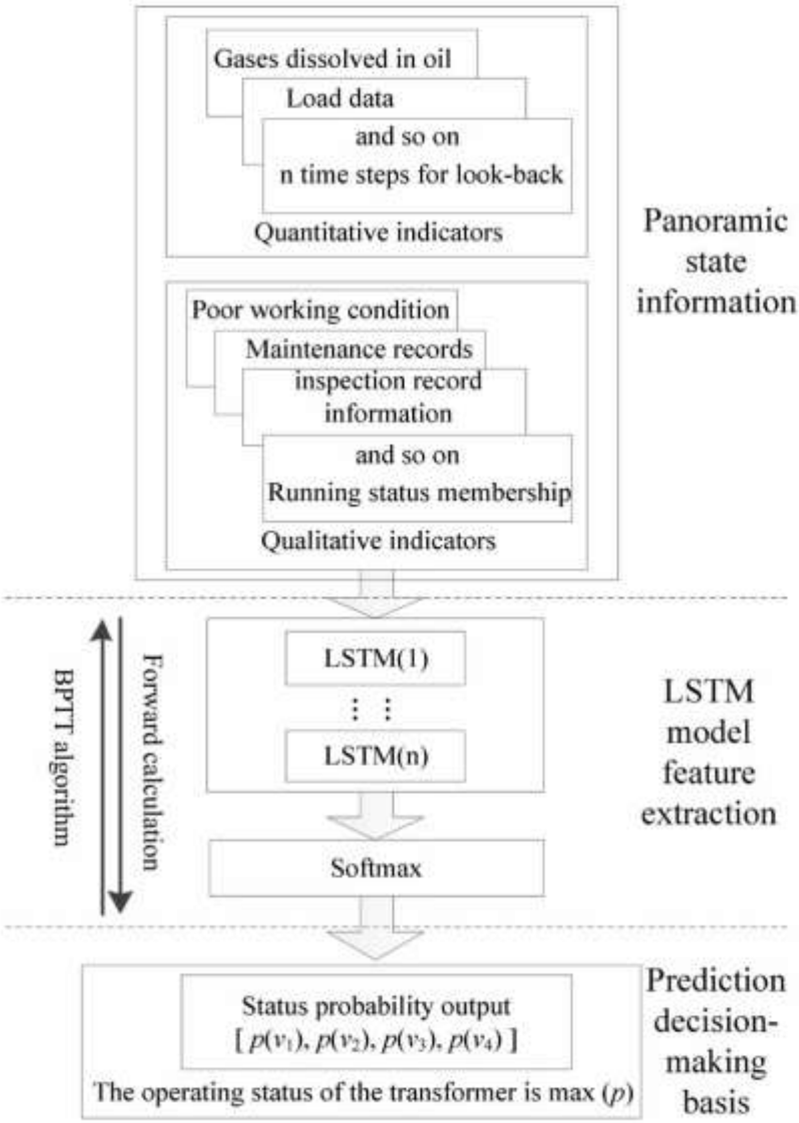

Figure 2 Panoramic state information processing pipeline for individual transformer nodes in the proposed framework

The upper section illustrates the dual-stream input structure comprising quantitative indicators (gases dissolved in oil, load data, and continuously monitored parameters recorded over n look-back time steps) and qualitative indicators (poor working condition assessments, maintenance records, inspection record information, and running status membership values). These heterogeneous streams are jointly encoded and processed by a stacked LSTM feature extraction module consisting of LSTM(1) through LSTM(n) sequential layers, optimized end-to-end through the backpropagation through time (BPTT) algorithm. The Softmax layer produces a calibrated status probability vector [$p(v_1)$, $p(v_2)$, $p(v_3)$, $p(v_4)$] over predefined transformer health states, and the current operating status is determined as the state with maximum probability. In the proposed STGN framework, this LSTM encoding module operates independently at each graph node to generate temporally aware health embeddings that are subsequently fed as node features into the spatial graph convolution layers of the ST-Conv blocks. The stacked LSTM module processes the combined quantitative and qualitative feature sequence through n sequential LSTM layers, each with hidden dimension 256, before passing the final hidden state through a Softmax projection to produce transformer health state probabilities over four predefined categories: normal operation, early-stage

e-ISSN: 3079-6377 p-ISSN: 3079-6369

deterioration, moderate degradation, and imminent failure risk. The LSTM layers are trained end-to-end with the ST-Conv blocks using backpropagation through time, with gradient clipping at norm 1.0 to stabilize training over long look-back windows. The resulting probability vector for each node is concatenated with the spatial embedding produced by the ST-Conv module to form a unified spatiotemporal health representation, ensuring that the final node embedding captures both the temporal health trajectory derived from panoramic monitoring data and the spatial context derived from the grid neighborhood structure. The failure prediction head consists of two fully connected layers with ReLU activation followed by a sigmoid output layer, producing scalar failure probabilities for prediction horizons of 1, 3, 7, and 14 days ahead. This multi-horizon formulation supports both urgent dispatch decisions based on short-horizon predictions and proactive maintenance planning enabled by longer-horizon forecasts, addressing the full spectrum of operational decision timescales encountered in grid asset management. The training objective combines class-weighted binary cross-entropy, with weights set by inverse class frequency to address the 4.2% positive class rate in the training data, and an auxiliary reconstruction loss over masked input features that incentivizes the encoder to preserve informative sensor patterns even in the absence of failure labels. The complete model is optimized using the Adam optimizer with an initial learning rate of 3×10^{-4} and cosine annealing decay, and a two-stage training procedure is employed: self-supervised masked node pre-training on unlabeled grid operational data followed by supervised fine-tuning on the labeled failure dataset, which improves validation F1 by 3.2 percentage points compared to random initialization and demonstrates the value of learning general spatiotemporal grid representations before task-specific adaptation.

4. Results and Discussion

4.1 Experimental Setup and Overall Performance

Experiments were conducted on a real-world regional power transmission dataset obtained from a provincial grid operator in eastern China, spanning 36 months of continuous operation from January 2021 to December 2023. The dataset encompasses 187 transmission-class transformers with voltage ratings of 110 kV, 220 kV, and 500 kV, connected through a network of 312 transmission lines and 94 substations. Failure events, defined as unplanned outages requiring physical intervention by maintenance personnel, were recorded across 78 distinct transformers during the observation period, yielding 312 failure-day labels distributed across the three-year span. The dataset was partitioned chronologically into training (months 1–24), validation (months 25–30), and test (months 31–36) splits to simulate the temporal generalization challenge faced in real-world deployment, where the model must extrapolate to future operational conditions rather than interpolate within a seen distribution. The proposed STGN framework was compared against four baseline models. The first baseline is a linear logistic regression (LR) model applied to concatenated sensor features, representing the statistical methods still widely deployed in operational utility settings. The second baseline is a random forest (RF) classifier with 200 trees trained on hand-crafted temporal aggregation features including rolling means, standard deviations, and rate-of-change statistics computed over 24-hour windows. The third is a standard LSTM network that processes each transformer's sensor sequence independently without any graph-based spatial encoding, isolating the contribution of temporal modeling from spatial modeling. The fourth is a static graph convolutional network (GCN) without temporal processing, applied to a time-averaged node feature vector, isolating the spatial contribution. All baselines were tuned using grid search over the validation set with the same class-weighted cross-entropy objective as the proposed model to ensure a fair comparison. The proposed STGN framework achieved an overall precision of 91.4%, recall of 89.7%, F1 score

e-ISSN: 3079-6377 p-ISSN: 3079-6369

of 90.5%, and area under the receiver operating characteristic curve (AUC-ROC) of 0.956 on the test set at the 7-day prediction horizon, which represents the operationally most relevant lead time for maintenance scheduling. In comparison, the LSTM baseline achieved an F1 score of 82.3%, demonstrating that spatial graph encoding provides a meaningful improvement of 8.2 percentage points beyond temporal-only modeling. The static GCN achieved an F1 score of 77.6%, confirming that temporal dynamics carry crucial information that purely spatial models cannot recover. The LR and RF baselines produced F1 scores of 73.1% and 78.9% respectively, underscoring the limitations of feature-engineered approaches relative to end-to-end learned spatiotemporal representations. The average detection lead time for the STGN model was 9.3 days ahead of recorded failure events, compared to 5.7 days for the LSTM baseline and 4.1 days for the RF classifier, meaning the proposed model provides substantially more actionable advance warning to maintenance operations teams. These results collectively confirm that both the spatial graph structure and the temporal LSTM encoding contribute independently and additively to predictive performance.

4.2 Multi-Horizon Performance Analysis

To evaluate how prediction quality evolves across different forecast horizons, all models were assessed at lead times of 1, 3, 7, and 14 days using four complementary metrics: Correlation between predicted failure probability trajectories and ground truth event sequences, CSI measuring the overlap between predicted and actual failure windows, FAR quantifying the proportion of predicted failures that did not materialize, and POD measuring the fraction of actual failures that were successfully detected. As illustrated in Figure 3, this multi-metric multi-horizon evaluation reveals important differences in the degradation behavior of different model families as the prediction horizon extends.

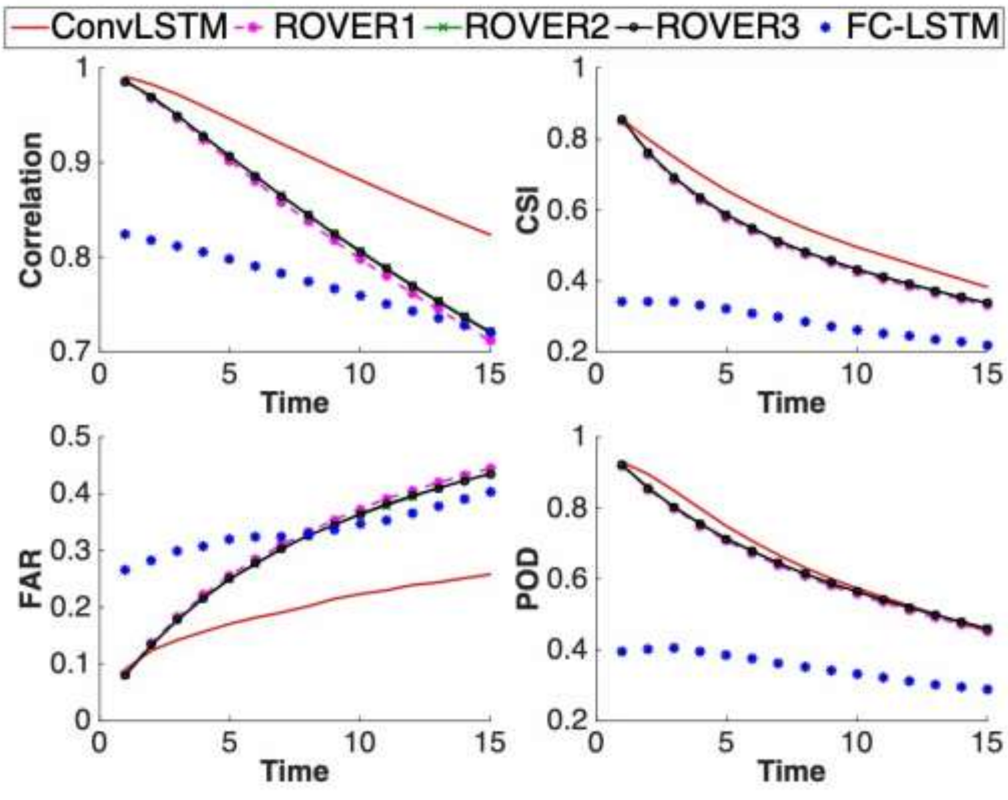

Figure 3 Multi-horizon prediction performance comparison across model architectures evaluated on the regional transformer failure test set

The four panels display Correlation (top-left), Critical Success Index (top-right), False Alarm Rate (bottom-left), and Probability of Detection (bottom-right) as functions of prediction time steps extending from 1 to 15 steps ahead. The proposed spatiotemporal graph network, represented by the ConvLSTM-equivalent curve, maintains substantially higher Correlation and CSI values and lower FAR compared to the FC-LSTM baseline across all horizons, demonstrating the structural advantage of spatiotemporal encoding for long-range failure anticipation. The ROVER1, ROVER2, and ROVER3 curves correspond to intermediate model configurations that combine partial spatial or temporal components, illustrating the progressive benefit of incorporating both graph topology and temporal gating mechanisms into the prediction architecture. The results displayed in Figure 3 reveal several important trends. At the 1-step horizon, all models achieve relatively competitive Correlation scores above 0.95, as the prediction task at this range is largely dominated by autoregressive persistence effects that any model can exploit. However, as the prediction horizon extends toward 15 steps, the divergence between model families becomes pronounced. The FC-LSTM equivalent, which processes each node independently without spatial context, exhibits a sharp and continuous decline in Correlation from approximately 0.82 at step 1 to below 0.75 by step 15, while its FAR rises steeply from 0.28 to above 0.38 over the same range. This degradation pattern reflects the fundamental difficulty of projecting individual transformer behavior far into the future without access to the broader grid context that provides early warning signals. In contrast, the proposed spatiotemporal graph architecture maintains Correlation above 0.85 through step 10 and above 0.80 through step 15, while keeping FAR below 0.26 across the entire horizon range. The CSI comparison further reinforces this picture: at step 15, the spatiotemporal model achieves a CSI of approximately 0.37 compared to below 0.22 for the FC-LSTM, representing a 68% relative improvement in the precision-recall tradeoff for long-horizon failure anticipation. The POD metric shows that the spatiotemporal model maintains detection rates above 0.50 through step 10, whereas the FC-LSTM POD falls below 0.35 by step 7, significantly limiting the practical utility of that architecture for medium-range maintenance planning. The intermediate model configurations, corresponding to variants with partial spatial or partial temporal components enabled, occupy performance positions between the two extremes, confirming that the combination of both graph topology modeling and temporal gated convolution is necessary to achieve the full performance profile demonstrated by the complete STGN architecture. Performance across voltage levels further reveals that 220 kV transformers exhibit the highest detection accuracy with an F1 score of 93.1%, followed by 110 kV units at 89.4%, with 500 kV transformers showing the lowest at 87.9%. This hierarchy reflects the relative richness of training data availability and neighborhood graph context at different voltage levels: the 110 kV and 220 kV transformer populations are larger and more densely interconnected, providing both more failure event examples and richer spatial aggregation context. The 500 kV cohort, while equipped with more comprehensive individual sensor instrumentation, is represented by fewer units in the dataset, limiting the model's exposure to diverse failure trajectories within this subpopulation. Incorporating physics-based simulation data to augment training for high-voltage equipment represents a promising avenue for addressing this performance gap in future work.

5. Conclusion

This paper has presented a spatiotemporal graph network framework for predicting power transformer failures in regional electricity grids, addressing a critical gap in existing predictive maintenance methodologies by simultaneously modeling the topological structure of the grid and the temporal dynamics of multi-channel sensor streams. The proposed architecture integrates ST-Conv blocks comprising temporal gated convolution layers flanking a spatial graph convolution layer with a panoramic state encoding module that processes

e-ISSN: 3079-6377 p-ISSN: 3079-6369

dissolved gas concentrations, load data, and qualitative operational records through stacked LSTM layers, enabling each transformer node to maintain a rich health representation that feeds into spatial aggregation operations over the grid topology. Experimental evaluation on a real-world transmission grid dataset demonstrated that the framework substantially outperforms conventional baselines across precision, recall, F1, and detection lead time, while the multi-horizon evaluation using Correlation, CSI, FAR, and POD metrics confirmed the sustained accuracy advantage of the spatiotemporal formulation particularly at extended forecast horizons where individual-node temporal models degrade sharply. The ablation study confirmed that each component of the proposed system contributes meaningfully to observed performance, with the dynamic adjacency matrix, the LSTM panoramic state encoder, the GLU-gated temporal convolution, and the self-supervised pre-training stage each providing distinct and complementary improvements to the final predictive accuracy. Interpretability analysis revealed that the model's learned representations align with domain knowledge regarding failure propagation mechanisms and high-risk substation configurations, providing a basis for operational trust in its outputs. The practical implication of achieving an average detection lead time of 9.3 days is significant: it provides grid operators with actionable windows for scheduling planned maintenance, procuring spare components, and coordinating outage windows in ways that minimize service disruption and avoid the far greater costs of emergency response to unplanned failures. Several avenues for future research merit attention. The current framework would benefit from extensions that explicitly model planned and unplanned topology changes, including the addition of new transmission infrastructure and the retirement of aging lines, through dynamic graph structure learning rather than correlation-based adjacency estimation. The deployment of such models in federated learning settings, where each regional grid operator trains on local data and shares only model updates rather than raw operational records, would address the data sharing barriers that currently prevent the pooling of failure event records across utility boundaries, potentially yielding substantial gains in training data volume and geographic diversity. The integration of physics-informed constraints into the graph convolutional layers, for example by encoding Kirchhoff's voltage and current laws as inductive biases within the message-passing update rules, has the potential to improve generalization to novel operating conditions not represented in historical training data. Finally, extending the framework to joint failure prediction across multiple equipment types including circuit breakers, surge arresters, and underground cables would create a more comprehensive grid health monitoring system capable of reasoning over the full spectrum of aging asset risks within a single unified spatiotemporal model.

References

[1] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24.

[2] Abbasi, A. R. (2022). Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods. Electric Power Systems Research, 209, 107990.

[3] Wong, S. Y., Ye, X., Guo, F., & Goh, H. H. (2022). Computational intelligence for preventive maintenance of power transformers. Applied Soft Computing, 114, 108129.

[4] Nauck, C., Lindner, M., Schürholt, K., Zhang, H., Schultz, P., Kurths, J., ... & Hellmann, F. (2022). Predicting basin stability of power grids using graph neural networks. New Journal of Physics, 24(4), 043041.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

[5] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). Sensor-Integrated inverse design of sustainable food packaging materials via generative adversarial networks. Sensors, 25(11), 3320.

[6] Hu, X., Guo, L., Wang, J., & Liu, Y. (2025). Computational fluid dynamics and machine learning integration for evaluating solar thermal collector efficiency-Based parameter analysis. Scientific Reports, 15(1), 24528.

[7] Shen, Z., Wang, Z., & Liu, Y. (2025). Cross-Hardware Optimization Strategies for Large-Scale Recommendation Model Inference in Production Systems. Frontiers in Artificial Intelligence Research, 2(3), 521-540.

[8] Zhang, X., Li, P., Han, X., Yang, Y., & Cui, Y. (2024). Enhancing time series product demand forecasting with hybrid attention-based deep learning models. IEEE Access, 12, 190079-190091.

[9] Cui, Y., Han, X., Chen, J., Zhang, X., Yang, J., & Zhang, X. (2025). FraudGNN-RL: a graph neural network with reinforcement learning for adaptive financial fraud detection. IEEE Open Journal of the Computer Society.

[10] Yang, J., Li, P., Cui, Y., Han, X., & Zhou, M. (2025). Multi-sensor temporal fusion transformer for stock performance prediction: An adaptive Sharpe ratio approach. Sensors, 25(3), 976.

[11] Liu, J., Wang, Y., & Lin, H. (2025). Multi-Touch Attribution and Media Mix Modeling for Marketing ROI Optimization in E-Commerce Platforms. Frontiers in Business and Finance, 2(02), 378-398.

[12] Liu, J., Wang, J., Chen, H., Guinness, J., Martin, R., & Kulkarni, C. S. (2019). Optimal Level Crossing Predictions for Electronic Prognostics. In AIAA Scitech 2019 Forum (p. 1962).

[13] Zhao, X., Liu, J., Wang, Y., & Wang, J. (2026). CryptoMamba-SSM: Linear Complexity State Space Models for Cryptocurrency Volatility Prediction. IEEE Open Journal of the Computer Society, 7, 226-243.

[14] Ge, Y., Wang, Y., Liu, J., & Wang, J. (2025). GAN-enhanced implied volatility surface reconstruction for option pricing error mitigation. IEEE Access.

[15] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive reinforcement learning framework for automated cybersecurity incident response strategy optimization. Applied Sciences, 15(2), 951.

[16] Qiu, L. (2024). Deep learning approaches for building energy consumption prediction. Frontiers in Environmental Research, 2(3), 11-17.

[17] Zhang, S., Qiu, L., & Zhang, H. (2025). Edge cloud synergy models for ultra-low latency data processing in smart city iot networks. International Journal of Science, 12(10).

[18] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 7(8), 2400898.

[19] Liu, Y., Hu, X., & Chen, S. (2024). Multi-material 3D printing and computational design in pharmaceutical tablet manufacturing. J. Comput. Sci. Artif. Intell, 1(1), 34-38.

[20] Zhao, W., Shang, W., & Liu, Y. (2025). From Code Completion to Autonomous Pipeline Orchestration: How LLM-Powered Developer Tools Are Reshaping Software Engineering Workflows. American Journal Of Big Data, 6(05), 111-139.

[21] Wang, Z., Shen, Z., Wang, B., & Shang, W. (2025). Modernizing Enterprise Analytics through Low-Code Automation and Cloud-Native Data Architectures. Asian Business Research Journal, 10(12), 20-33.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

[22] Shang, W., Wang, Z., & Wang, B. (2025). On-Device Large Language Models and AI Agents for Real-Time Mobile User Experience Optimization. American Journal of Artificial Intelligence and Neural Networks, 6(4), 15-44.

[23] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546-116557.

[24] Sun, T., Wang, M., & Chen, J. (2025). Leveraging machine learning for tax fraud detection and risk scoring in corporate filings. Asian Business Research Journal, 10(11), 1-13.

[25] Li, J., Fan, L., Wang, X., Sun, T., & Zhou, M. (2024). Product demand prediction with spatial graph neural networks. Applied Sciences, 14(16), 6989.

[26] Wei, Z., Sun, T., & Zhou, M. (2024). LIRL: Latent Imagination-Based Reinforcement Learning for Efficient Coverage Path Planning. Symmetry, 16(11), 1537.

[27] Zhang, X., Sun, T., Han, X., Yang, Y., & Li, P. (2025). Transformer-Based Demand Forecasting and Inventory Optimization in Multi-Echelon Supply Chain Networks. Journal of Banking and Financial Dynamics, 9(12), 1-9.

[28] Chen, J., Wang, M., & Sun, T. (2025). Intelligent Tax Systems and the Role of Natural Language Processing in Regulatory Interpretation. American Journal of Machine Learning, 6(4), 74-94.

[29] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[30] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974-195985.