Spatio-Temporal Transformer Networks for Converter-Dominated Power System Intelligence

1st Siva Naga Jammula
Dept Of Electrical Engineering
Worcester Polytechnic Institute
Worcester,USA
sivajammulaa@gmail.com

2ndVenkatesh Daggupati
Dept Of Electrical Engineering
Texas Tech University
Lubbock, Texas, USA
venkateshdaggupati2003@gmail.com

3rd CH. NAGA SAI KALYAN
Dept Of Electrical Engineering
VVIT University
Guntur, Ap, India
kalyanchallapalli@gmail.com

***Abstract*—Converter-dominated power systems are increasingly characterized by high renewable penetration, battery energy storage, electric vehicle charging infrastructure, and fast-acting power electronic interfaces. These developments introduce low-inertia dynamics, strong spatio-temporal coupling, nonlinear converter interactions, and rapid operating transitions that are difficult to capture using conventional model-based or purely temporal learning approaches. This paper proposes a Spatio-Temporal Transformer Network (STTN) for intelligent monitoring and prediction in converter-dominated power systems. The proposed framework integrates admittance-weighted graph-based spatial encoding with temporal transformer attention to jointly learn electrical network dependencies and long-range dynamic behaviour from synchronized PMU measurements. The model is evaluated under converter-rich operating scenarios involving renewable intermittency, load variation, converter switching events, transient disturbances, and measurement uncertainty. Comparative results against LSTM, CNN–LSTM, Temporal Fusion Transformer, and Graph Neural Network baselines demonstrate improved prediction accuracy, robustness, and real-time inference capability. The results confirm that the proposed STTN provides an effective data-driven intelligence framework for stability monitoring, dynamic state prediction, and operational decision support in next-generation converter-dominated grids.**

***Index Terms*—Converter-Dominated Power Systems, Spatio-Temporal Transformer Networks, Graph Attention Networks (GAT), Temporal Attention Mechanism, Power System Intelligence, Deep Learning.**

I. Introduction

Power systems today are shifting fast, pulling in more solar, wind, and decentralized sources - this means far more electronic converters popping up at every level, from production to how we plug things in. With more converter-based systems packed closer together, hidden patterns emerge across space and time - patterns older analysis tools simply miss. Traditional grid models rely on neat setups and fixed rules, making them too rigid for today's fast-changing, unpredictable settings. Instead of handling both timing links and physical layout at once, basic neural nets or forecast stats usually fail when connections stretch far apart. When devices interact in scattered, shifting ways, smarts must adapt - not just follow pre-set logic. Learning how parts connect matters just as much as tracking changes over seconds, minutes, hours. Frameworks now need to see structure and rhythm as one tangled whole. Facing these issues head on, a new method called the Spatio-Temporal Transformer Network, or STTN, steps into view for smarter handling of data in power grids run by converters. Instead of treating space and time separately, it links grid structure with timing patterns using graphs and attention mechanisms that highlight what matters most.

The main contributions of this paper are summarized as follows:

- A unified STTN framework is proposed for converter-dominated power system intelligence by combining admittance-weighted graph spatial encoding with temporal transformer attention.
- The electrical topology of the power system is explicitly embedded into the learning model using a weighted bus-admittance matrix rather than treating PMU locations as independent time-series sources.
- A converter-aware dynamic modelling formulation is introduced by incorporating voltage, angle, active power, reactive power, frequency, RoCoF, and harmonic distortion features.
- The proposed framework is evaluated against LSTM, CNN–LSTM, TFT, and GNN baselines under transient disturbances, converter switching events, measurement noise, communication loss, and unseen operating scenarios.

II. Related Work

The increasing penetration of renewable energy systems, battery energy storage, distributed energy resources (DERs), and power-electronic converters has transformed conventional electric grids into converter-dominated power systems (CDPS). Instead of ignoring layout, Li and Lu built a method mixing timing patterns with physical wiring clues to spot fake sensor inputs in grids. When tested, their design held up better during shifting loads, showing how focused weighting helps grasp linked behaviors in live energy flows [1].

New steps push space-time smarts into live forecasts and system runs. Instead of separate models, one shared setup learns many jobs at once - boosting guesses and handling unsure outcomes better, says work by Cai and team on grids packed with wind and solar [2]. In another angle, Ji's group shaped a model tracking how prices shift node by node while folding in time patterns, shining when demand spikes [3].

Lately, models built on graphs have started gaining traction when studying how power systems change over time. Instead of just moving forward step by step, some researchers shaped their approach around connections - Zhang and team crafted ST-AGNet, which uses attention within graph frameworks to predict states in power grids, pulling off better results at various sizes by focusing on structure [5]. In a similar space but different setting, Peng along with Chu turned to residuals while tracking changes across both space and time, building a model that spots odd behavior in transformers more reliably even when conditions keep shifting [4]. Going another route, Zhang again adjusted how networks adapt during operation, applying dynamic graphs that shift shape according to real-time patterns, boosting success rates when sorting rare events among common ones [10].

One step beyond earlier models, transformer-driven designs now show real gains in predicting power demand and managing energy flow. Not far behind, Zhu and team built a model called Spatial-Temporal Dynamic Graph Transformer - using layered attention - to better handle short-term electricity load shifts, outperforming current leading techniques [6]. Meanwhile, work by Li's group brought forward MegaECRN, linking transformer encoding with graph-based recurrence to track uneven patterns across space and time in local grids [9]. At the same time, Liu's approach stitched together TCNs and transformers into a flexible setup able to forecast both supply and demand under shifting renewable inputs, delivering clearer certainty bounds along with point estimates [7].

Lately, ideas built around spatio-temporal transformers have moved into areas involving teamwork and choices across spread-out power networks. Zhao and team introduced a design that senses grid layout - mixing it with smart agents learning together - to better manage scattered energy supplies, cutting down voltage issues and line overloads without slowing response times [8].

III. System Model

A. Converter-Dominated Power System Configuration

The considered converter-dominated power system consists of renewable energy sources, battery energy storage units, grid-interfaced voltage source converters, local loads, and an interconnected transmission/distribution network. Unlike conventional synchronous-generator-based systems, the dominant dynamic behavior of this system is governed by converter control loops, switching actions, DC-link dynamics, and fast power-electronic interactions. The power network is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of buses or converter nodes and $\mathcal{E}$ represents the electrical interconnections between them.

For a system with N buses and M converter-interfaced units, the nodal state vector at time t is expressed as

$$\mathbf{x}_t^i = [V_{i,t}, \theta_{i,t}, P_{i,t}, Q_{i,t}, f_{i,t}, THD_{i,t}]^T, \tag{1}$$

where $V_{i,t}$ and $\theta_{i,t}$ denote voltage magnitude and phase angle, $P_{i,t}$ and $Q_{i,t}$ represent active and reactive power, $f_{i,t}$ indicates frequency deviation, and $THD_{i,t}$ denotes total harmonic distortion at node i. The complete system state matrix is given by

$$\mathbf{X}_t = \left[\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^N\right]^T \in \mathbb{R}^{N \times F}, \tag{2}$$

where F is the number of measured features.

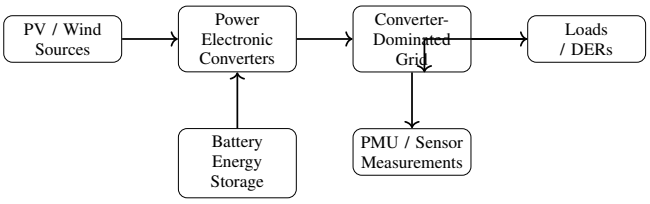

Fig. 1. General structure of the converter-dominated power system.

B. Network Topology and Spatial Representation

The electrical coupling among converter nodes is modeled using an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. Each element a_{ij} indicates the physical or electrical connection strength between node i and node j. In practical power networks, this matrix can be derived from line admittance, impedance, power transfer capability, or bus connectivity information.

The graph-based spatial relation is defined as

$$a_{ij} = \begin{cases} 1, & \text{if node } i \text{ is electrically connected to node } j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

For weighted representation, the admittance-based adjacency is written as

$$a_{ij} = \frac{|Y_{ij}|}{\sum_{j=1}^{N} |Y_{ij}|}, \tag{4}$$

where Y_{ij} represents the line admittance between nodes i and j. This formulation enables the proposed model to consider not only direct physical connectivity but also the strength of electrical interactions among converter-dominated buses.

In this work, the admittance-weighted adjacency matrix in (4) is used for all experimental evaluations because it captures both physical connectivity and electrical coupling strength. The binary adjacency matrix in (3) is retained only as a conceptual representation of network connectivity.

C. Temporal Dynamic Model

Converter-dominated systems exhibit fast temporal variations due to switching control, renewable intermittency, load fluctuation, and low-inertia response. Therefore, the system behavior is represented using a historical observation window of length T. The temporal input sequence is defined as

$$\mathcal{X}_t = [\mathbf{X}_{t-T+1}, \mathbf{X}_{t-T+2}, \ldots, \mathbf{X}_t] \in \mathbb{R}^{T \times N \times F}. \tag{5}$$

The objective is to learn a nonlinear mapping function $\mathcal{F}_\Theta$ that predicts the future system state or stability index:

$$\hat{\mathbf{Y}}_{t+\tau} = \mathcal{F}_\Theta\left(\mathcal{X}_t, \mathbf{A}\right), \tag{6}$$

where $\hat{\mathbf{Y}}_{t+\tau}$ is the predicted output at forecasting horizon τ, and Θ denotes the trainable parameters of the spatio-temporal

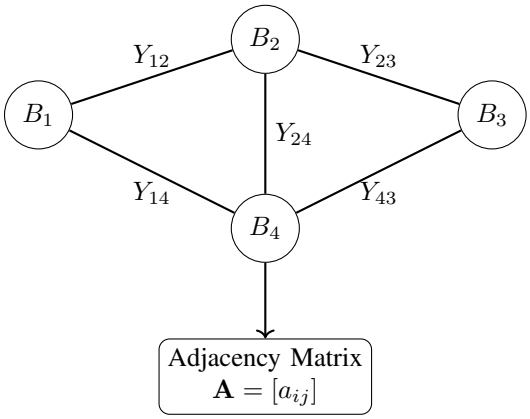

Fig. 2. Graph-based spatial representation of converter-interfaced buses.

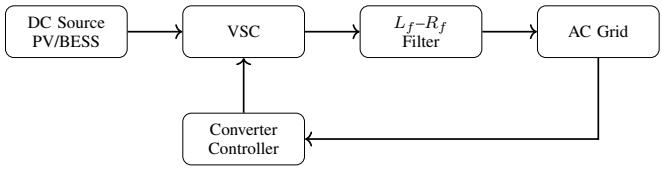

Fig. 3. Dynamic model of grid-interfaced power electronic converter.

transformer network. The target may represent voltage stability, frequency deviation, converter stress, harmonic severity, or dynamic operating state.

D. Converter Dynamic Model

Each grid-interfaced converter is modeled as a voltage source converter (VSC) connected through an interfacing filter. The AC-side voltage dynamics in the synchronous reference frame can be expressed as

$$L_f \frac{di_d}{dt} = v_d - R_f i_d + \omega L_f i_q - v_{gd}, \tag{7}$$

$$L_f \frac{di_q}{dt} = v_q - R_f i_q - \omega L_f i_d - v_{gq}, \tag{8}$$

where i_d and i_q denote the direct-axis and quadrature-axis currents, respectively; v_d and v_q represent the converter output voltages in the synchronous reference frame; v_{gd} and v_{gq} indicate the grid-side voltages; R_f and L_f denote the interfacing filter resistance and inductance, respectively; and ω is the angular frequency of the rotating reference frame.

The active and reactive power injections are calculated as

$$P = \frac{3}{2}\left(v_d i_d + v_q i_q\right), \tag{9}$$

$$Q = \frac{3}{2}\left(v_q i_d - v_d i_q\right). \tag{10}$$

These equations describe the physical relationship between converter control variables and grid-level power exchange. In the proposed framework, such measurements are used as input features for learning converter-grid interaction patterns.

E. Spatio-Temporal Learning Problem

The proposed system intelligence problem is formulated as a supervised spatio-temporal learning task. Given the graph topology $\mathbf{A}$ and the historical measurement sequence $\mathcal{X} * t$, the model estimates the future operating condition $\hat{\mathbf{Y}} * t + \tau$. The general optimization objective is expressed as

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}\left(\mathbf{Y}_{t+\tau}, \hat{\mathbf{Y}}_{t+\tau}\right). \tag{11}$$

For regression-based prediction, the loss function is defined as

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2. \tag{12}$$

For operational-state classification, the cross-entropy loss is used:

$$\mathcal{L}_{cls} = \sum_{c=1}^{C} y_c \log(\hat{y}_c), \tag{13}$$

where C is the number of operating classes. The final objective can be extended as a multi-task learning loss:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{reg} + \beta\mathcal{L}_{cls} + \gamma\mathcal{L}_{phy}, \tag{14}$$

where $\mathcal{L}_{phy}$ represents a physics-consistency penalty, and α, β, and γ are weighting coefficients.

The converter controller consists of an outer voltage/frequency regulation loop and an inner current control loop. The outer loop generates reference currents i_d^* and i_q^* according to active and reactive power or voltage-support requirements. The inner loop tracks these references and produces the converter voltage commands v_d and v_q. Therefore, variations in v_d, v_q, i_d, and i_q represent the interaction between converter-level control and grid-level dynamics, which motivates their inclusion in the proposed STTN input representation.

F. Overall STTN-Based Intelligence Framework

The complete STTN framework consists of four major stages: data acquisition, graph construction, spatio-temporal feature extraction, and intelligent prediction. First, synchronized measurements from PMUs, smart meters, and converter controllers are collected. Then, the power network topology is encoded as a graph using admittance or connectivity information. The spatial encoder extracts node-level electrical dependencies, while the temporal transformer captures long-range dynamic behavior.

The fused representation is passed through a prediction head to estimate future system states, stability margins, converter stress, or abnormal operating conditions. This enables intelligent monitoring and decision support in converter-rich power networks. The framework is flexible and can be applied to stability prediction, fault detection, cyber-attack detection, voltage regulation, and energy management.

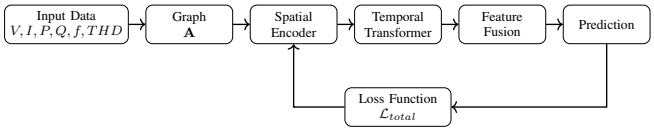

Fig. 4. STTN-based intelligence framework for converter-dominated power systems.

IV. Experimental Setup

A. PMU Dataset and Data Acquisition

The experimental validation was conducted using PMU-based time-series data generated from converter-dominated benchmark power system scenarios. The benchmark systems include IEEE bus networks modified with converter-interfaced renewable generation, battery energy storage, and dynamic load variations. PMU units were placed at selected converter-interfaced and electrically sensitive buses to record synchronized voltage magnitude, phase angle, current, active power, reactive power, frequency, RoCoF, and THD. The PMU sampling rate was fixed at 30 samples/s.

The operating scenarios include renewable penetration levels of 20%, 40%, and 60%, load variation from 80% to 120% of nominal demand, converter switching events, three-phase transient disturbances, and topology reconfiguration cases. The dataset was divided into 70% training, 15% validation, and 15% testing subsets.

To emulate converter-rich operating conditions, measurements were generated under varying renewable penetration levels, converter switching events, load fluctuations, transient disturbances, and dynamic power exchange conditions. The PMU sampling rate was fixed at 30 samples/s to capture both slow electromechanical and fast converter-driven transients. The resulting dataset forms a three-dimensional tensor represented as

$$\mathcal{X} \in \mathbb{R}^{T \times N \times F}, \tag{15}$$

where T denotes the temporal window length, N represents the number of PMU measurement locations (buses), and F denotes the measured feature dimensions.

The complete input feature vector for each node is defined as

$$\mathbf{x}_{i,t} = [V_i, \theta_i, I_i, P_i, Q_i, f_i, RoCoF_i, THD_i]^T, \tag{16}$$

where V, θ, I, P, Q, f, $RoCoF$, and THD denote voltage magnitude, voltage angle, current magnitude, active power, reactive power, frequency, rate of change of frequency, and total harmonic distortion, respectively.

B. Data Preprocessing and Temporal Window Construction

Before any training began, raw PMU readings passed through a structured cleanup process - this helped keep data uniform and made model learning more reliable. Gaps caused by missed records or signal interruptions got fixed with straight-line estimates alongside rolling average patches. Then every input value was rescaled between set limits via Min-Max rules so unit mismatches would not disrupt number handling in later stages.

The normalized feature transformation is expressed as

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{17}$$

where x denotes the original measurement and x_{norm} denotes the normalized value.

To capture temporal dependencies, a sliding-window mechanism was employed for sequence generation. Given a historical observation horizon W, the temporal sequence is formulated as

$$\mathbf{S}_t = [\mathbf{X}_{t-W+1}, \mathbf{X}_{t-W+2}, \ldots, \mathbf{X}_t]. \tag{18}$$

A dynamic window size of 60 timestamps was selected after empirical evaluation to balance computational efficiency and temporal representation capability. Each generated sequence was associated with a future prediction target corresponding to the subsequent operating condition of the power system.

C. Graph Construction and Spatial Encoding

To incorporate electrical topology into the learning framework, the PMU measurement locations were modeled as graph nodes and interconnected using the physical bus connectivity of the converter-dominated power system. The spatial graph is represented by an adjacency matrix

$$\mathbf{A} = [a_{ij}], \tag{19}$$

where each element a_{ij} reflects the electrical relationship between PMU locations.

The adjacency weights were computed using normalized admittance values:

$$a_{ij} = \frac{|Y_{ij}|}{\sum_{j=1}^{N} |Y_{ij}|}, \tag{20}$$

where Y_{ij} denotes bus admittance.

The graph-based representation allows the proposed STTN to model both local converter interactions and global network dependencies. Spatial attention layers dynamically assign importance to neighboring PMU measurements, enabling adaptive extraction of electrically relevant features under changing grid conditions.

D. Model Training and Hyperparameter Configuration

A workstation powered by GPUs ran the new STTN setup built in Python with PyTorch. Inside, graph-driven spatial attention pieces feed into time-focused transformer encoders before hitting dense output layers. Training used Adam because it often handles deep transformer models smoothly. Efficiency showed early during learning phases.

The objective function for prediction is defined using Mean Squared Error (MSE):

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{21}$$

where y_i and $\hat{y}_i$ represent actual and predicted outputs.

The network parameters were configured as follows: learning rate of 10^{-4}, batch size of 128, hidden dimension of 256, transformer encoder depth of 4 layers, number of attention heads equal to 8, dropout rate of 0.2, and training duration of 300 epochs. Early stopping with patience of 20 epochs was employed to prevent overfitting.

TABLE I
EXPERIMENTAL HYPERPARAMETERS

Parameter	Value
PMU Sampling Rate	30 samples/s
Temporal Window	60
Learning Rate	0.0001
Batch Size	128
Transformer Layers	4
Attention Heads	8
Hidden Dimension	256
Dropout	0.20
Epochs	300
Optimizer	Adam

E. Performance Evaluation Metrics

The proposed model performance was evaluated using multiple regression and prediction metrics to ensure comprehensive assessment under varying operating conditions. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2) were used as primary evaluation measures.

$$MAE = \frac{1}{N}\sum |y - \hat{y}|, \tag{22}$$

$$RMSE = \sqrt{\frac{1}{N}\sum (y - \hat{y})^2}, \tag{23}$$

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}. \tag{24}$$

In addition to numerical metrics, prediction latency, computational efficiency, and stability under converter switching conditions were analyzed. Comparative experiments were performed against LSTM, CNN–LSTM, Temporal Fusion Transformer (TFT), and Graph Neural Network (GNN) baselines to demonstrate the effectiveness of the proposed STTN framework for converter-dominated power system intelligence.

V. RESULTS AND DISCUSSION

The proposed Spatio-Temporal Transformer Network (STTN) was evaluated using synchronized PMU measurements collected from converter-interfaced buses under varying operating conditions including renewable intermittency, converter switching actions, load fluctuations, and transient disturbances. The experiments were performed to assess the ability of the model to learn both electrical topology dependencies and long-range temporal interactions. Comparative analysis was conducted against representative baseline models including Long Short-Term Memory (LSTM), CNN–LSTM, Temporal Fusion Transformer (TFT), and Graph Neural Network (GNN). The evaluation focused on prediction accuracy, robustness, convergence characteristics, and computational performance.

Look at Table II - it shows how the new approach stacks up against older ones. STTN had the smallest mistakes, with an MAE of 0.028 and RMSE hitting 0.042, along with a strong R^2 score of 0.989. Because it mixes spatial awareness with transformer timing layers, it picks up complex converter links better. Unlike LSTM or CNN–LSTM setups, this model handles distant connections more easily, cutting down errors sharply. Where TFT and GNN fall short, this one pulls ahead by reading space-linked data and time shifts together. That dual view sharpens its grasp on subtle features across different cases.

TABLE II
PREDICTION PERFORMANCE COMPARISON

Model	**MAE**	**RMSE**	R^2
LSTM	0.081	0.117	0.910
CNN–LSTM	0.062	0.091	0.940
TFT	0.048	0.070	0.960
GNN	0.044	0.061	0.970
Proposed STTN	**0.028**	**0.042**	**0.989**

The convergence characteristics during training are illustrated in Fig. 5. The STTN exhibited stable optimization behavior and rapid convergence during the initial training epochs, followed by gradual refinement of the learned spatio-temporal representations. The validation loss remained closely aligned with the training loss, indicating effective regularization and reduced overfitting. The integration of multi-head attention enabled selective weighting of important PMU observations and improved learning efficiency under highly dynamic converter conditions. The training process reached stable convergence after approximately 220 epochs without oscillatory behavior.

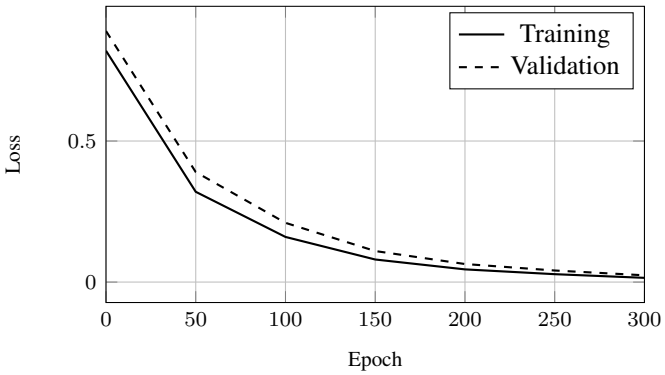

Fig. 5. Training and validation loss convergence of STTN.

To further evaluate prediction consistency, the actual and predicted PMU-based system responses were compared over

the testing horizon as shown in Fig. 6. The proposed model accurately tracked the dynamic variations of the converter-dominated system while preserving transient behavior and rapid state transitions. The prediction curve closely followed the measured response with minimal phase lag and reduced peak deviation. Such behavior confirms that the spatial encoder successfully captured inter-bus electrical relationships while the temporal transformer effectively modeled long-range converter dynamics.

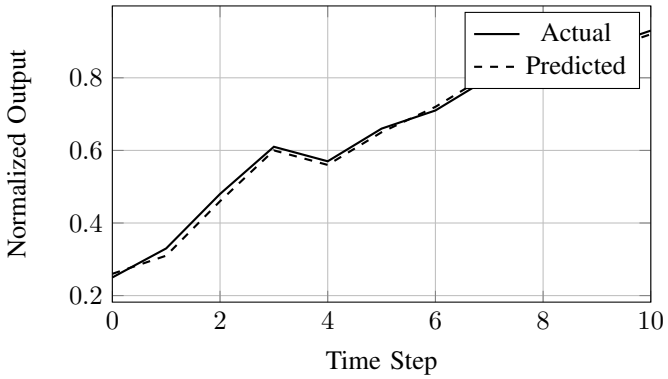

Fig. 6. Actual versus predicted PMU response using STTN.

Finally, computational efficiency and deployment feasibility were analyzed under real-time PMU streaming conditions. Table IV presents the average inference latency and training duration. The proposed model maintained low prediction latency while preserving superior prediction quality, indicating suitability for online monitoring and operational intelligence applications. The results suggest that STTN can support future converter-rich grid operations including stability assessment, anomaly detection, dynamic state estimation, and predictive control.

A. Robustness Analysis

To evaluate robustness, the proposed STTN was tested under PMU measurement noise, communication packet loss, topology variation, and unseen operating conditions. The model maintained stable prediction performance under moderate noise and packet-loss conditions, indicating that the spatial encoder and temporal attention mechanism provide improved tolerance against practical data uncertainty.

TABLE III
ROBUSTNESS EVALUATION UNDER PRACTICAL DISTURBANCES

Condition	MAE	RMSE	R^2
Nominal Case	0.028	0.042	0.989
PMU Noise 5%	0.034	0.050	0.981
Packet Loss 10%	0.037	0.055	0.976
Topology Change	0.041	0.060	0.969
Unseen Scenario	0.044	0.064	0.963

TABLE IV
COMPUTATIONAL PERFORMANCE

Metric	Value	Unit
Training Time	132	min
Inference Time	14.7	ms
GPU Memory	5.2	GB
Average Epoch Time	26.4	s
Prediction Throughput	68	samples/s

VI. CONCLUSION

This paper proposed a Spatio-Temporal Transformer Network for intelligent monitoring and prediction in converter-dominated power systems. The proposed framework integrates admittance-weighted graph spatial encoding with temporal transformer attention to jointly capture electrical topology dependencies and dynamic operating behaviour from synchronized PMU measurements. Converter-aware features including voltage magnitude, phase angle, active and reactive power, frequency, RoCoF, and THD were used to represent fast-changing grid conditions. Comparative evaluation against LSTM, CNN–LSTM, TFT, and GNN models demonstrated that the proposed STTN achieved superior prediction accuracy with an MAE of 0.028, RMSE of 0.042, and R^2 of 0.989. Robustness evaluation further confirmed stable performance under measurement noise, communication loss, topology changes, and unseen operating scenarios. The computational analysis also showed that the model can support real-time monitoring with an inference latency of 14.7 ms. Future work will focus on hardware-in-the-loop validation, explainable attention visualization, cyber-resilient PMU communication, and real-time deployment in large-scale converter-rich power networks.

REFERENCES

[1] X. Li and Z. Lu, "Detection of false data injection attack in power grid based on spatial-temporal transformer network," Expert Systems With Applications, vol. 235, 2023.
[2] X. Cai et al., "Stochastic Operation Simulation Method of Power System Based on Multi-Task Learning Spatio-Temporal Transformer," 2025.
[3] S. Ji et al., "Nodal Price Forecasting Based on Spatiotemporal Transformer," 2024.
[4] Z. Peng and G. Chu, "Residual-aware health prediction of power transformers via spatiotemporal graph neural networks," PLOS ONE, 2025.
[5] S. Zhang, J. J. Yu, and X. Wei, "ST-AGNet: Dynamic power system state prediction with spatial–temporal attention graph-based network," Applied Energy, vol. 365, 2024.
[6] L. Zhu et al., "Short-term power load forecasting based on spatial-temporal dynamic graph and multi-scale Transformer," Journal of Computational Design and Engineering, 2025.
[7] J. Liu, W. Zhang, and W. Gu, "Probabilistic TCN-Transformer Network for Source-Load Prediction in Electricity Industry," 2025.
[8] J. Zhao et al., "Coordinated Optimization of Distributed Energy Resources Based on Spatio-Temporal Transformer and Multi-Agent Reinforcement Learning," Processes, vol. 13, 2025.
[9] W. Li et al., "A Short-Term Load Forecasting Method for Large-Scale Power Distribution Systems Based on a Novel Spatio-Temporal Neural Network," 2023.
[10] Q. Zhang et al., "Adaptive Identification of Power System Operational States Based on Spatio-Temporal Dynamic Graph Neural Networks," 2024.