Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatiotemporal CNN LSTM Based Multi Class Stability Assessment of Renewable Integrated Power Systems Using PMU Measurements

Shashank Kumar
Symbiosis International University
Mangal Singh
Symbiosis International University
Mahesh Singh
mahesh.singh@sitpune.edu.in

Symbiosis International University

Article

Keywords: Power System Stability, Deep Learning, CNN-LSTM, PMU Data Stream, Voltage Stability, Frequency Stability, Transient Stability, WAMS, FACTS Devices, Smart Grid, Spatiotemporal Learning

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11572281/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatiotemporal CNN LSTM Based Multi Class Stability Assessment of Renewable Integrated Power Systems Using PMU Measurements

Shashank Kumar,[1] Mangal Singh,[2] Mahesh Singh[3*]

1 Research Scholar, Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India

2 Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India

3 Department of Robotics and Automation, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India

**Corresponding Author: mahesh.singh@sitpune.edu.in*

Abstract

The increasing penetration of renewable energy resources and the resulting variability in power-system operating conditions have created significant challenges for reliable real-time stability monitoring and corrective control. Conventional stability assessment methods often require extensive computation and may not satisfy the low-latency requirements of Wide Area Monitoring Systems (WAMS). To address these challenges, this paper proposes a spatiotemporal deep learning framework that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for multi-class stability assessment using synchronized Phasor Measurement Unit (PMU) data streams. The CNN layers learn spatial correlations among measurements from different buses, while the LSTM layers capture the temporal evolution of disturbance dynamics. The proposed framework simultaneously classifies four stability conditions: Stable, Transient Instability, Voltage Instability, and Frequency Instability. The methodology is evaluated on the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus benchmark systems under diverse disturbances, including three-phase faults, generator disconnections, line contingencies, and sudden load changes. A total of 5,150 operating and disturbance cases are considered. The proposed CNN-LSTM model achieves an overall classification accuracy of 97.4%, with an average model inference latency of 17.3 ms for each 500 ms PMU observation window, outperforming standalone SVM, ANN, CNN, and LSTM models. Because the classifier requires a complete observation window, a stability decision becomes available 519.2 ms after disturbance onset, and the corrective action is applied at 0.649 s for STATCOM, 0.769 s for SVC, and 1.019 s for battery dispatch, excluding communication delay. All closed-loop results reported in this paper use these actuation instants. For transient instability on the IEEE 39-bus system, STATCOM support limits the critical generator to a peak rotor angle of 75.0 degrees, against loss of synchronism at 1.33 s without control, and synchronism is retained for communication delays up to approximately 334 ms. For voltage instability on the IEEE 118-bus system, SVC support limits the weak-bus voltage to 0.894 p.u. and the excursion below the 0.90 p.u. ANSI limit to 0.18 s, against collapse to 0.621 p.u. without control. For frequency instability following a 650 MW generator trip, the decision delay places the battery dispatch after the rate-of-change-of-frequency peak, so the maximum ROCOF is unchanged at 1.48 Hz/s and the nadir improves only from 48.79 Hz to 48.89 Hz; the measurable benefit is a shorter recovery, 4.3 s against 7.5 s to return above 49.8 Hz. The results therefore establish a mechanism-

dependent latency limit: measurement-window-based deep-learning control is effective for rotor-angle and voltage mechanisms, which evolve over the first one to two seconds, and cannot act on the inertial phase of a frequency event. The proposed framework provides a scalable approach for data-driven stability monitoring and coordinated corrective control in renewable-integrated power systems, with the latency constraint quantified for each mechanism.

Keywords: Power System Stability, Deep Learning, CNN-LSTM, PMU Data Stream, Voltage Stability, Frequency Stability, Transient Stability, WAMS, FACTS Devices, Smart Grid, Spatiotemporal Learning.

1. Introduction

The reliability and continuity of electric power systems is the foundation of modern society, supporting key areas like telecommunication networks, industrial manufacturing, transportation infrastructure, healthcare and digital economies. The issue of system stability is, broadly speaking, the ability of an interconnected electric power system to remain synchronously operated and return to an acceptable equilibrium after a physical disturbance, and continues to be one of the most basic objectives of system planning and operational dispatch decisions, as well as real-time system control [1]. In the past, stability could be assured relatively easily in synchronous-generator-dominated grids with predictable load conditions.

Today's power systems, on the other hand, are going through a radical restructuring with the addition of massive Renewable Energy Sources (RES) such as solar photovoltaic (PV) and wind power plants [2]. These resources are essential to reach the goal of global carbon neutrality and sustainable energy transitions, but they also have a stochastic output and intermittent nature, which poses tremendous operational challenges. The conventional synchronous generators provide a significant rotational inertia which automatically reduces the frequency deviation after a generation-load imbalance. The contribution of the renewable energy converter, via power electronics, has a very small inertia constant, which decreases the overall system inertia constant and makes the grid extremely sensitive to large-signal transients, voltage collapse and quick frequency variations [3]. Further, the increasing presence of distributed generation, electric vehicles, demand-side flexible loads and battery energy storage systems has created a more complex bidirectional power flow and load behaviour, further adding to the operational uncertainty.

The traditional tools for power system stability assessment include time-domain numerical integration, energy function approaches based on Lyapunov theory, eigenvalue analysis based on linearization, continuation power flow, and security-constrained optimal power flow, and have proven to offer analytical tools for decades [4]. Although these classical methods are theoretically sound, they are typically very computationally intensive, especially in large-scale networks with many or thousands of potential contingencies and stochastic renewable generation sources with rapidly changing operating points. These computation requirements make them essentially unfit for real-time applications, where stability assessment and appropriate decision making must be performed in time scales of a few tens of milliseconds, to meet the operational requirements of modern Wide Area Monitoring Systems (WAMS) and protection coordination schemes [5].

At the same time, the progressive loss of stability margins due to aging electricity grid, growing electricity demand, lower spinning reserves in deregulated markets and higher cross-regional interconnections have significantly increased the likelihood of cascading failures and wide-area blackouts [6]. The importance of intelligent monitoring platforms that are able to predict the conditions that could lead to instability before

they culminate in a system-wide emergency, has been reinforced by operational experience gained during large scale disturbance events across the continents.

The Phasor Measurement Units (PMUs) have revolutionized power system observability by being deployed at a growing scale on transmission networks [7]. PMUs are able to measure voltage magnitudes, voltage phase angles, active and reactive power flow, system frequency, and rate-of-change-of-frequency (ROCOF) from geographically distributed substations at sampling speeds ranging from 30 to 120 frames per second and with GPS-synchronized timing [8]. The resulting high-resolution time-synchronized measurement streams provide rich repositories of spatiotemporal information, that are well suited for data-driven stability assessment, if proper analytical frameworks that take advantage of the spatial correlation among monitoring nodes and temporal evolution of dynamic system behavior are used.

Artificial Intelligence (AI) and Deep Learning (DL) are now emerging as a powerful analytical framework to tackle these problems in a radical way [9]. Deep learning architectures learn hierarchical representations and dynamics of the system from raw high dimensional measurement data without requiring any handcrafted features [10]. CNNs are especially good at learning spatial correlation patterns between network buses and generator groups, while Long Short-Term Memory (LSTM) networks are a type of recurrent network that is better at learning long-range temporal dependencies in sequential measurement streams [11]. A shared spatiotemporal context that combines both capabilities offers a promising prospect to significantly improve the state of the art of real-time stability intelligence.

To overcome these critical issues, this paper introduces a spatiotemporal deep learning approach combining CNN-based spatial feature extraction and LSTM-based temporal sequence modelling to achieve a comprehensive, multi-class stability classification from synchronized PMU measurement streams. In addition to stability prediction, the proposed framework incorporates a closed-loop corrective-control strategy that coordinates FACTS-based reactive-power compensation with active-power frequency-response mechanisms following instability detection [12]. Using various disturbance scenarios, the proposed framework is validated on IEEE 9 bus, IEEE 39 bus and IEEE 118 bus benchmark systems, extending recent deep-learning-based stability assessment, such as the deep transfer learning approach of Kim et al. [13] for transient stability, from a single stability mechanism to a unified multi-class formulation with closed-loop corrective control.

Initially, Artificial Neural Networks (ANNs) were used as pattern classifiers in the exploration of AI techniques for power systems stability evaluations. Sobajic and Pao [14] showed the first pioneering evidence that ANN-based approaches could significantly decrease the computational effort of transient security evaluation without compromising prediction accuracy. Following the SVM introduction, it became evident that it could provide better generalization capabilities in high-dimensional feature spaces and Moulin et al. [15] have shown it to be superior to ANN counterparts for transient stability classification. However, these traditional machine learning approaches had a major drawback – they required the creation of a feature extraction pipeline, which required significant domain knowledge and was inflexible when it came to adapting to new grid configurations.

Deep learning architectures solved the feature engineering problem by allowing to automatically learn hierarchical representations. Ren et al. [16] used an ensemble of randomized learning algorithms to perform temporal adaptive voltage stability assessment based on PMU derived measurements, resulting in better classification accuracy and avoiding the manual feature selection. CNN architectures were then shown to

be effective in capturing spatial correlation patterns, and Shi et al. [17] achieved good classification accuracy using CNN architectures directly applied to PMU voltage phasor trajectories following the disturbance for the transient stability assessment and instability mode prediction in IEEE 39-bus and WECC 179-bus systems. The long-term short-term memory (LSTM) networks, which have the ability to model the temporal dynamics sequentially, were seen to be suitable for near real-time voltage stability prediction using post-fault PMU trajectories by Zhang et al. [18]. Zhu et al. [19] also added the spatial-attention-rectified RNN learning to improve the temporal discrimination, which enabled them to focus on the critical time intervals in the disturbance observation windows adaptively.

Even with these developments, however, most current research into the spatial or temporal learning task has considered it as separate modelling goals. A more principled approach is hybrid architectures that make use of both feature dimensions. Singhal et al. [20] proposed a hybrid CNN-BiLSTM network for stability forecasting in the smart grid, which showed better forecasting accuracy than a single-modality network, and Wu et al. [26] combined convolutional, bidirectional recurrent and attention modules for the localization of forced oscillations, further indicating the value of joint spatial-temporal feature extraction in measurement-driven power-system applications. Topology-aware alternatives, such as Graph Neural Network (GNN) based approaches, have emerged, with Zhang et al. [21] showing that an unsupervised transfer learning method based on a graph-attention layer can be used for transient stability analysis to enhance the cross-scenario generalization capability. Sun et al. [22] then further developed spatiotemporal correlation modelling using probabilistic techniques based on copula specially developed to handle uncertainty in renewable energy.

Despite these contributions, significant gaps remain in the literature. The possibility of having a common framework for classification of all major stability categories, such as transient instability, voltage instability and frequency instability, is still not available. In addition, most of the published articles consider stability prediction as a goal to achieve instead of incorporating automatic corrective control for operational implementation of the predictions. Real-time deep-learning inference coupled with FACTS device actuation for closed-loop stability management has been under-researched. Finally, it is often difficult to see a thorough validation that covers a variety of standard benchmark systems, and an extensive variety of forms of disturbances.

Recent studies have begun to address parts of this gap. Arévalo-Cordero et al. [12] coupled an LSTM predictor with FACTS-based control for voltage and frequency stability in EV-penetrated microgrids, and Shahriyari et al. [25] proposed an LSTM architecture for multi-domain stability-boundary prediction. However, [12] targets a microgrid setting and does not assess rotor-angle (transient) stability, whereas [25] addresses stability-boundary prediction rather than mechanism-specific corrective actuation. The novelty of the present work lies in the combination of three elements that have not been reported together: (i) a single PMU-based spatiotemporal model that discriminates four mutually exclusive classes (Stable, Transient Instability, Voltage Instability, and Frequency Instability); (ii) the use of the predicted instability mechanism to select the corrective resource (STATCOM, SVC, or governor/AGC/BESS support); and (iii) validation on three transmission benchmark systems of increasing size, including renewable-integrated operation, robustness stress tests, and scenario-level generalization tests.

The principal contributions are:

1. A spatiotemporal deep-learning framework is developed for PMU-based power-system stability monitoring, in which convolutional layers extract spatial correlations among synchronized bus measurements while stacked LSTM layers capture their temporal evolution over a short online observation window.

2. A unified four-class stability assessment is formulated, distinguishing Stable, Transient Instability, Voltage Instability, and Frequency Instability from synchronized PMU measurements of voltage magnitude, voltage phase angle, active power, reactive power, and frequency. Unlike approaches that focus on a single stability mechanism or a single scalar stability index, the proposed framework provides mechanism-specific classification from the same PMU data stream.

3. The framework is coupled with closed-loop mitigation actions, where the identified instability mechanism determines the corresponding corrective response through STATCOM, SVC, and active-power support based on turbine-governor/AGC and fast-support models. This establishes an integrated monitoring-to-mitigation workflow rather than treating stability classification as an isolated prediction task.

4. Comprehensive validation is performed across IEEE 9-bus, 39-bus, and 118-bus systems, including renewable-integrated operating conditions and time-domain disturbance scenarios. The evaluation further considers reduced PMU coverage, measurement noise, missing data, communication delay, and synchronization uncertainty to examine robustness under non-ideal monitoring conditions.

5. The proposed framework is evaluated against conventional machine-learning baselines and alternative deep-learning architectures using classification and deployment-oriented metrics, including accuracy, precision, recall, F1-score, AUC, inference latency, and computational cost. Scenario-level generalization tests are additionally used to assess performance under unseen contingencies and operating conditions.

This paper is organized as follows: Power system dynamics, feature representation of PMUs and the proposed CNN-LSTM architecture are presented in the mathematical modelling framework of Section 2. In Section 3, the proposed deep learning method and adaptive corrective control strategy are explained in detail. Section 4 explains the simulation environment, benchmark test systems, and how the datasets were created. The results and discussion are discussed in detail in Section 5. The key findings and future research directions are summarized in Section 6.

2. Mathematical Modelling of System Dynamics

Proposed spatiotemporal deep learning approach is mathematically justified. This section introduces the governing equations of power system dynamic behaviour, the representation of power system measurements by PMUs, the convolutional feature extraction, LSTM based temporal processing, multi-class power system stability classification and adaptive FACTS control. These formulations contribute to the theoretical underpinning for the computational framework presented in this study.

2.1 Power System Swing Equation

The swing equation, which describes the dynamics of a synchronous generator after a disturbance, is the relation between the rotor acceleration and the difference between electrical output and mechanical input power [23]:

$$\frac{2H_i}{\omega_s}.\frac{d^2\delta_i}{dt^2} = P_{mi} - P_{ei} - D_i\frac{d\delta_i}{dt} \tag{1}$$

where H_i is the per-unit inertia constant (seconds), ω_s is the synchronous angular velocity (rad/s), δ is the rotor angle (rad), P_{mi} and P_{ei} are mechanical input and electrical output power in per-unit, respectively, and D_i is the damping coefficient. Transient synchronism after fault events is reflected by the rate of change of rotor angle, called speed deviation $\Delta\omega_i = \frac{d\delta_i}{dt}$.

2.2 PMU Measurement Feature Vector

At each monitored bus *i,* the PMU measurement vector at time step t is expressed as [24]:

$$X = [V, \theta, P, Q, f] \tag{2}$$

The voltage magnitude, voltage phase angle, injected active power, injected reactive power and local system frequency, V, θ, P, Q, and *f,* respectively. The input matrix $X \in \mathbb{R}T\times N\times 5$ is a three-dimensional multivariate PMU tensor for N monitored buses over T time steps that is fed into the proposed deep learning-based architecture.

Rotor-angle information is not included in the online PMU feature vector used by the proposed CNN-LSTM model [25]. The online model uses only synchronized PMU measurements of voltage magnitude, voltage phase angle, active power, reactive power, and frequency. Generator rotor-angle trajectories are obtained from the time-domain simulation model and are used exclusively for offline ground-truth transient-stability labeling and subsequent validation. Therefore, the proposed online stability-monitoring framework does not require rotor-angle measurements as an input feature.

2.3 Convolutional Spatial Feature Extraction

The CNN module extracts the spatial features of the PMU measurement tensor by a series of convolutions over the bus dimension [32]. The output feature map at the lth convolutional layer is obtained by computing the following:

$$H_1 = \mathrm{Re}\,LU(W_1 * X + b_i) \tag{3}$$

Here, W_1 is the learnable convolutional kernel, b_i is the bias vector, $*$ denotes convolution, and ReLU(·) denotes the rectified linear activation function. The successive layers of convolutional networks are able to learn about increasingly sophisticated, inter-bus spatial dependencies such as generator coherency patterns, voltage profile gradients, and power flow patterns across the entire network.

2.4 LSTM Temporal Sequence Modelling

The hidden state update of the LSTM network at each time step t is presented as:

$$h_t = LSTM(h_{t-1}, x_t) \tag{4}$$

Where x_t is the CNN extracted spatial feature vector at time *t*, h_t is the previous hidden state, and LSTM denotes the learnable parameter set of input gate, forget gate, output gate, and cell gate weights. The LSTM memory cells selectively hold information on the disturbance propagation history, allowing for a good characterization of stability evolution over the observation window following the disturbance.

2.5 Multi-Class Stability Classification

An output layer with a Softmax activation function projects the final temporal representation *hT* on the LSTM onto the posterior probability distribution over *K* stability classes [27]:

$$P(y=k) = \frac{e^{zk}}{\sum_{j=1}^{k} e^{zj}} \tag{5}$$

where $K = 4$ denotes the number of stability classes, namely Stable, Transient Instability, Voltage Instability, and Frequency Instability. The predicted stability class is obtained as:

$$y = \arg\max P(y=k) \tag{6}$$

2.6 Adaptive FACTS Corrective Control

The control module automatically sends reactive power compensation signals to the FACTS devices on instability classification. The injection of reactive power Q_{FACTS} at bus *i* is controlled by:

$$Q_{inj} = K_c(V_{ref} - V) \tag{7}$$

where V_{ref} is the voltage set point, V is the measured voltage magnitude of the bus and K_c is the proportional gain of the controller. The objective of the global control is to minimize a deviation function weighted by a weighting function:

$$J = \alpha(\Delta\delta)^2 + \beta(\Delta f)^2 + \gamma(\Delta V)^2 \tag{8}$$

The symbols $\Delta\delta$, Δf, and ΔV denote the deviations of rotor angle, frequency and voltage from their respective reference values, while the α, β and γ are the weighting coefficients set to give the higher priority to the control objects according to the detected instability class.

3. Proposed Spatiotemporal Deep Learning Methodology

Based on the mathematical concepts of the previous section, this work introduces a hybrid CNN-LSTM spatiotemporal deep learning framework with an adaptive FACTS-based corrective control module [28]. The proposed methodology aims to be able to account for the spatio-temporal correlation features from the

synchronized PMU measurement streams, classify the stability classes accurately and intervene in real time with corrective measures.

3.1 System Architecture Overview

The proposed end-to-end framework includes four main functional stages: synchronized PMU data acquisition and preprocessing, CNN-based spatial feature extraction, LSTM-based temporal sequence learning, and adaptive stability classification and corrective control actuation. Prior to deep learning inference, raw PMU measurements are collected from N monitored buses over a sliding observation window of 500 ms, then normalized and organized into three-dimensional tensors [29].

Accordingly, rotor-angle trajectories used for transient-stability assessment are part of the offline simulation-based labeling process and are not included among the five PMU features supplied to the online CNN-LSTM classifier.

It is important to distinguish the online observation window from the offline ground-truth assessment horizon. The 500 ms window is used exclusively as the model input to enable early stability classification and does not constitute the complete stability assessment period. The ground-truth labels are generated offline from the complete post-disturbance time-domain trajectory. Because the classifier requires a complete 500 ms window of post-disturbance measurements, the earliest classification decision is available approximately 519 ms after disturbance onset (500 ms observation window, 1.9 ms preprocessing, and 17.3 ms inference; Section 5.10), excluding communication delay.

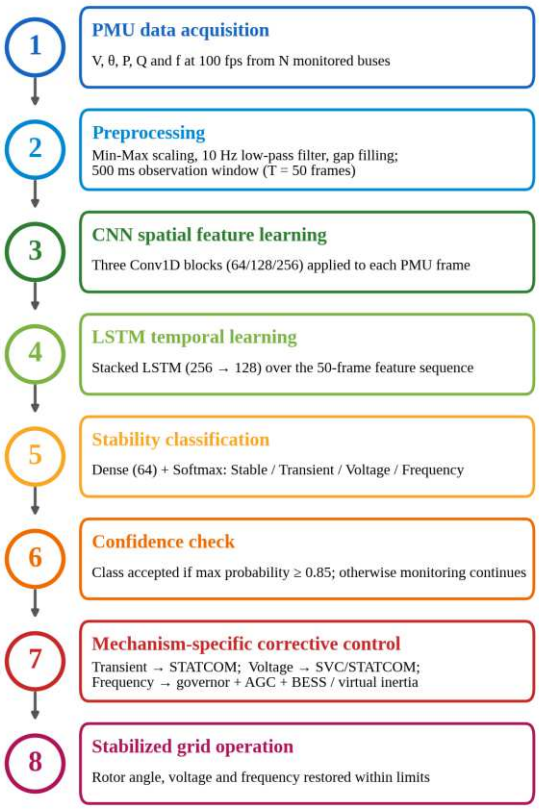


Figure 1. End-to-end workflow of the proposed CNN-LSTM spatiotemporal stability assessment and adaptive corrective control framework.

The overall information flow concept of the proposed system is shown in Figure 1. The CNN module first processes pre-processed PMU tensors to obtain spatial operating state representations throughout the network. The 3D spatial feature sequences are then fed into the LSTM network to learn the temporal dynamics and the disturbance propagation properties. The LSTM output is connected to a Softmax classification layer that classifies the current stability class; when instability is detected with a confidence level higher than a predetermined decision threshold, a command is sent to the FACTS control module.

3.2 PMU Data Preprocessing

For models to converge and predict under different operating conditions, the preprocessing must be robust. Each PMU measurement channel is independently scaled to the range [0, 1] using Min-Max scaling, which eliminates scale differences between voltage, power, and frequency measurements. High-frequency measurement noise is filtered out by a fourth-order Butterworth low-pass filter with a cut-off frequency of 10 Hz before constructing features. Measurement gaps caused by communication loss of up to three consecutive frames are filled by linear interpolation [30] during offline dataset preparation; in online operation, missing frames are forward-filled because future samples are not available (Section 5.13.3). The 500 ms observation window was selected as a compromise between early detection and sufficient temporal information. A shorter observation window reduces decision latency but may provide insufficient temporal information to characterize the initial evolution of voltage, frequency, and electromechanical disturbance responses. Conversely, a substantially longer window provides additional temporal information but increases the decision latency and computational burden. Therefore, the 500 ms window, corresponding to 50 PMU frames at a reporting rate of 100 fps, a rate that IEC/IEEE 60255-118-1 [31] permits above its specified 10, 25 and 50 fps rates for 50 Hz systems, was selected to support early stability classification while retaining sufficient temporal information for the CNN-LSTM model.

The simulation cases of each benchmark system are split separately into training, validation, and testing sets, while preserving the proportion of each stability class within each of the three sets; the overall metrics are computed on the pooled test sets of the three systems. Training set is sufficiently large, and held-out sets are representative enough for unbiased model evaluation to be done with a partitioning ratio of 70:15:15.

3.3 CNN Spatial Feature Extraction Module

A hierarchical CNN is used to extract spatial features. It contains three convolutional blocks, each consisting of a 1D convolutional layer followed by batch normalization and ReLU activation, as shown in Figure 2 [32]. The first convolutional layer has 64 filters, second has 128, and third has 256 filters, with all filters having a kernel size of 3 and a stride of 1. To decrease the dimensionality of feature maps while maintaining the discriminative spatial patterns, the first and second convolutional blocks are followed by max-pooling layers with a pool size of 2. The hierarchical design gradually refines the features extracted in the lower levels, which are correlation features between different buses, while the deeper levels represent higher-level network operating state features.

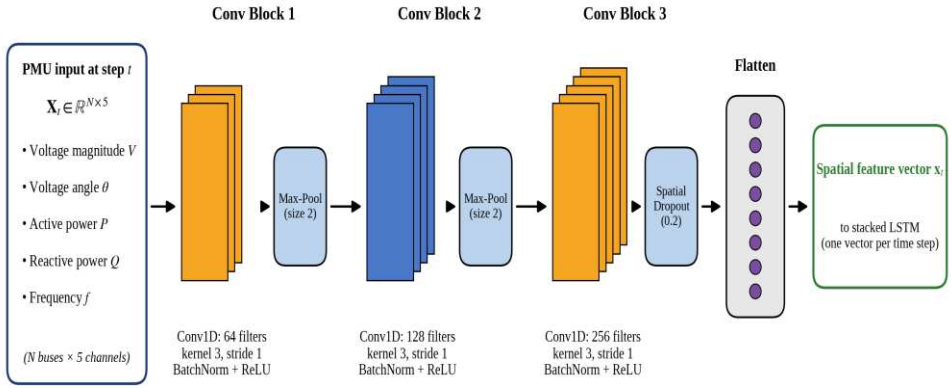


Figure 2. Architecture of the CNN spatial feature extraction module comprising three hierarchical convolutional blocks with batch normalization and max-pooling layers.

Convolutional feature maps learn physically meaningful representations such as generator coherency groupings, voltage profile changes along transmission corridors and patterns of power flow redistribution for various types of disturbances. The third convolutional block followed by a spatial dropout with dropout rate of 0.2 is used for better robustness in training. The CNN output is flattened into a series of dense feature vectors at each time step of the observation window, which is the input to the next LSTM module.

Spatiotemporal data flow. At each time step t of the observation window, the PMU snapshot X_t ∈ ℝ^(N×5) is treated as a one-dimensional sequence over the bus index, with the five measured quantities as input channels. The same CNN (shared weights) is applied to all T = 50 snapshots, so that each snapshot is mapped to a spatial feature vector x_t encoding the instantaneous inter-bus pattern of voltage, angle, power, and frequency. The sequence x_1, …, x_T is then processed by the stacked LSTM, whose final hidden state h_T summarizes how these spatial patterns evolve over the 500 ms window, and h_T is passed to the dense classification head. Spatial and temporal features are therefore integrated hierarchically: the CNN captures inter-bus correlations within each frame, and the LSTM captures their evolution across frames. Because one-dimensional convolution operates over the bus index, the buses are ordered in ascending order of their standard IEEE bus numbers (1 to N) as defined in the benchmark test-case data, and this ordering is kept identical for training, validation, and testing.

Handling of different network sizes. Because the flattened CNN output dimension depends on the number of monitored buses N, a separate CNN-LSTM model with the same architecture and hyperparameters is trained for each benchmark system. The input and output sizes of each layer and the corresponding parameter counts are reported for each benchmark system in Table A2.

3.4 LSTM Temporal Sequence Learning Module

The model is a two-layer stacked LSTM network to capture the temporal changes during the online model-input window shown in Figure 3 [33]. The first LSTM layer takes the CNN generated spatial feature sequences and outputs 256-dimensional hidden state representations capturing the characteristics of disturbance onset, the propagation of the fault and the dynamics of the initial generator response. The second LSTM block which receives the output of the first LSTM block has 128 hidden units and models

the longer-range temporal dependency related with the development of inter area oscillations, frequency recovery trajectories and voltage recovery behaviour.

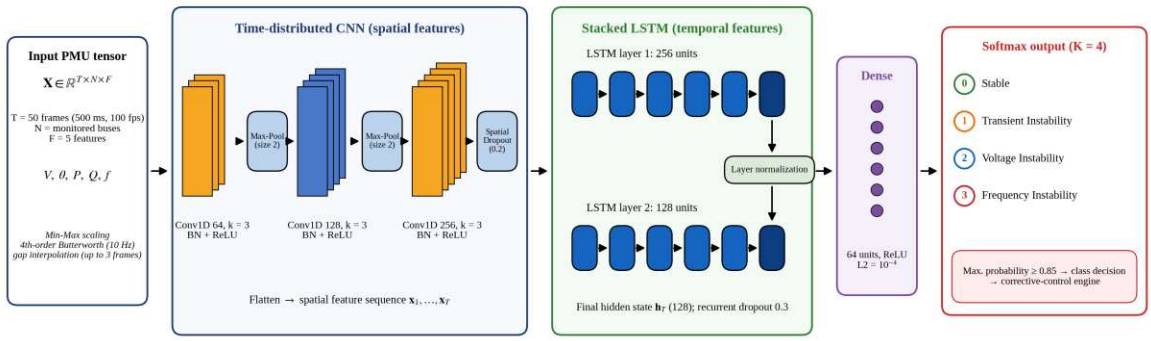


Figure 3. Stacked CNN-LSTM hybrid architecture showing the integration of spatial convolutional feature extraction with temporal LSTM sequence learning for multi-class stability classification.

To mitigate the variations of magnitude of the gradients and accelerate the converging process, layer normalization is applied between the LSTM layers. To avoid co-adaptation of temporal memory features and better generalize over unseen disturbance scenarios, recurrent dropout with a rate of 0.3 is embedded into each LSTM cell. Overall, the CNN–LSTM architecture first extracts spatial correlations from synchronized PMU measurements through the hierarchical CNN blocks and then captures their temporal evolution using two stacked LSTM layers with 256 and 128 units. The resulting spatiotemporal representation is processed by a 64-neuron fully connected layer and a four-unit Softmax output layer corresponding to Stable, Transient Instability, Voltage Instability, and Frequency Instability.

The complete architectural and training configuration of the proposed CNN-LSTM model, including layer dimensions, dropout settings, optimizer, learning rate, batch size, loss function, and early-stopping criterion, is summarized in Sections 3.3, 3.4, and 3.7 and in Table 10 for reproducibility.

3.5 Adaptive FACTS-Based Corrective Control

The proposed adaptive corrective-control module coordinates FACTS-based reactive-power support with active-power frequency-response mechanisms following the identification of voltage, transient, or frequency instability. For voltage and transient instability, SVC and STATCOM provide bounded reactive-power support, whereas frequency-instability events additionally activate the modeled turbine-governor and AGC pathways for active-power balancing.

3.5.1 FACTS Device Placement, Sizing, and Dynamic Control Parameters

To ensure the reproducibility of the proposed closed-loop corrective-control framework, the installation locations, ratings, operating ranges, controller gains, and response characteristics of the SVC and STATCOM are explicitly defined for all three benchmark systems. The FACTS-device locations were selected by engineering judgement close to the principal disturbance locations of the case studies (Section 3.5.2), and the resulting locations were fixed for the corresponding benchmark-system simulations. The final configurations used in the simulations are summarized in Table 1.

Table 1. FACTS Device Configuration for the Benchmark Systems

System	Device	Installation Bus	Rating	Operating Range	Control Gain	Response Time
IEEE 9-bus	STATCOM	Bus 6	50 MVAr	−50 to +50 MVAr	$K_p = 20, K_i = 5$	<30 ms
IEEE 9-bus	SVC	Bus 9	30 MVAr	−30 to +30 MVAr	$K_{svc} = 50$	~150 ms
IEEE 39-bus	STATCOM	Buses 16 and 19	150 MVAr / 100 MVAr	±150 MVAr / ±100 MVAr	$K_p = 20, K_i = 5$	<30 ms
IEEE 39-bus	SVC	Bus 8	300 MVAr	−300 to +300 MVAr	$K_{svc} = 50$	~150 ms
IEEE 118-bus	STATCOM	Bus 95	100 MVAr	−100 to +100 MVAr	$K_p = 20, K_i = 5$	<30 ms
IEEE 118-bus	SVC	Bus 44	100 MVAr	−100 to +100 MVAr	$K_{svc} = 50$	~150 ms

As shown in Table 1, the STATCOM and SVC are installed at the locations defined in Section 3.5.2 and provide voltage regulation and coordinated corrective-control actions. The device ratings are scaled according to the size and reactive-power requirements of each benchmark system, while the internal controller tuning is kept consistent across the systems. The finite response characteristics and operating-range limits are incorporated into the dynamic simulations to avoid assuming instantaneous or unlimited reactive-power compensation.

The internal dynamic-control parameters used by the FACTS devices are kept fixed across all benchmark systems, except for the device-specific rating limits. The parameters governing the STATCOM PI controller, voltage-measurement dynamics, SVC voltage regulator, susceptance saturation, and voltage droop are summarized in Table 2**.**

Table 2. Unified Dynamic Control Parameters of SVC and STATCOM

Parameter	Device	Value
Proportional gain K_p	STATCOM	20
Integral gain K_i	STATCOM	5
Firing-control delay T_0	STATCOM	0.01 s
Voltage measurement lag T_{v1}	STATCOM	0.05 s
Current limit I_{max}	STATCOM	1.2 p.u.
Regulator gain K_{svc}	SVC	50
Regulator time constant T_{svc}	SVC	0.02 s
Susceptance limits $B_{max} / B_{\min}$	SVC	+1.0 / −1.0 p.u.
Voltage droop	SVC	4%

The STATCOM is modeled using the WECC Generic SVSM03T2 representation, whereas the SVC is represented using a variable-shunt-susceptance voltage-regulation model. The controller outputs are constrained by the specified current, susceptance, voltage, and reactive-power limits. Thus, the corrective-control action remains bounded by the physical operating constraints of the modeled FACTS devices. The corresponding control formulations are given in the following subsection.

3.5.2 Voltage-Weakness Identification and Reactive-Power Allocation

The reactive-power compensation command is determined using a voltage-sensitivity-based criterion rather than by applying a fixed amount of compensation. For each operating condition, the voltage-reactive-power sensitivity of the load buses is evaluated from the power-flow Jacobian. The local sensitivity is expressed as

$$S_{VQ,i} = \frac{\partial V_i}{\partial Q_i} \tag{9}$$

where V_i and Q_i denote the voltage magnitude and reactive-power injection at bus i, respectively. A higher positive $S_{VQ,i}$ indicates that a given amount of reactive-power support produces a larger voltage improvement at the corresponding bus. During online operation, the monitored load buses are therefore ranked according to their voltage sensitivity together with their measured post-disturbance voltage deviation, and buses exhibiting low voltage magnitude and high voltage-reactive-power sensitivity are classified as critical voltage-support regions.

The required reactive-power compensation is determined from the voltage deviation and the corresponding voltage sensitivity as

$$Q_{req,i} = \frac{V_{ref} - V_i}{S_{VQ,i}} \tag{10}$$

where V_{ref} is the desired voltage reference.

For simultaneous voltage depressions at multiple buses, the control engine evaluates the voltage deviation at all monitored load buses and identifies the most critical voltage-support regions. Each available FACTS device is assigned to the electrically closest critical region for which it provides effective voltage support. The compensation command is independently calculated for each assigned device and coordinated through the common control objective while enforcing the individual MVAr and voltage limits. If the required compensation of one region exceeds the available rating of its assigned device, the remaining corrective requirement is transferred to the next available device serving the same electrically coupled region.

The calculated compensation command is constrained by the rated reactive-power capability of the corresponding FACTS device:

$$Q_{cmd,i} = sat(Q_{req,i}, Q_{min,i}, Q_{max,i}) \tag{11}$$

Thus, the controller does not apply a fixed or arbitrary amount of reactive power. The compensation level is determined according to the local voltage deviation and the voltage-reactive-power sensitivity and is subsequently limited by the physical operating range of the SVC or STATCOM. Equation (7) represents this voltage-regulation action in simplified proportional form; in the dynamic simulations, the commands are realized through the STATCOM PI controller and the SVC voltage-regulator dynamics listed in Table 2.

For multiple weak-voltage areas, the critical buses are first ranked according to their voltage severity and V−Q sensitivity. FACTS devices are then assigned to the most critical electrically distinct voltage-support areas. When more than one FACTS device is available, their reactive-power commands are coordinated using the measured voltage deviations of the corresponding critical areas, while respecting the individual device ratings and saturation limits. This coordinated allocation prevents excessive reactive-power injection at a single location and distributes the available compensation according to the severity and sensitivity of the voltage disturbances.

FACTS placement procedure. The FACTS locations were selected by engineering judgement close to the principal disturbance locations of the case studies: in the IEEE 39-bus system, the STATCOM at Bus 16 coincides with the three-phase fault location of Case T (Section 5.2), the second STATCOM is installed at Bus 19, a non-generator bus directly connected to the faulted Bus 16 that experiences significant post-disturbance voltage depression during the Case T three-phase fault; installing a second STATCOM at Bus 19 distributes the reactive-power reserve across two locations, reduces the reactive-power burden on the Bus 16 STATCOM, and improves voltage recovery and the transient-stability margin, and the SVC at Bus 8 coincides with the load-increase location of the frequency study (Section 5.3). In the IEEE 9-bus system (standard MATPOWER bus numbering, with load buses 5, 7, and 9), the SVC is installed at Bus 9, which is also the most sensitive load bus in the base case (Table A1), and the STATCOM at Bus 6, a critical transmission bus adjacent to load Buses 5 and 7 and to the terminal bus of generator 3; this location provides fast reactive-power injection on the opposite side of the network to the SVC and covers the fault locations considered in the 9-bus case studies. In the IEEE 118-bus system, the SVC is installed at Bus 44, the critical bus of the voltage-stability case study, which experiences the deepest voltage depression following the double-circuit outage of the Bus 38–Bus 65 line (Figure 17) and is the second most sensitive load bus in the base case ($dV/dQ = 0.127$ p.u./p.u.; Table A1). Bus 44 is located four branches from both terminal buses of the outaged line. The STATCOM is installed at Bus 95, a load bus with moderate base-case V–Q sensitivity (rank 27 of 64), located seven branches from Bus 44 in an electrically distinct part of the network, so that the two devices provide reactive-power support to different voltage-support areas. For reference, Table A1 reports the base-case V–Q sensitivity of these buses computed from the reduced power-flow Jacobian (Eq. (9)); the selected buses are not necessarily the most sensitive buses in the base case (e.g., Buses 16 and 19 rank 26th and 28th of 29 load buses in the IEEE 39-bus system). A systematic sensitivity-based placement is left for future work.

For clarity, the FACTS installation locations are fixed offline for each benchmark system, whereas the online control engine determines the reactive-power command from the voltage sensitivity of Eq. (9) and the measured voltage deviation at the corresponding monitored critical region. Thus, the physical installation location of each FACTS device is not changed during online operation; only its reactive-power output is adaptively adjusted within its operating limits.

3.5.3 Active-Power Support for Frequency Recovery

Frequency instability is primarily associated with an imbalance between active-power generation and demand. Therefore, the proposed frequency-remedial control does not attribute frequency recovery solely to the reactive-power support provided by SVC and STATCOM. For frequency-instability events, the CNN-LSTM decision engine activates coordinated active-power support through the available turbine-governor/AGC mechanism, while SVC and STATCOM provide supplementary voltage and reactive-power support.

The active-power corrective command is generated from the measured frequency deviation and is constrained by the available generation headroom and governor operating limits. The commanded generation increment is represented as.

$$\Delta P_{G,i}^{cmd} = \text{sat}\,(K_f(f_{ref} - f), 0, \Delta P_{G,i}^{max}) \tag{12}$$

where K_f is the frequency-response gain, f_{ref} is the nominal frequency, and $\Delta P_{G,i}^{max}$ represents the available generation headroom of generator *i*. The saturation operator constrains the commanded active-power increment between zero and the available generation headroom. In the present implementation, a governor droop of $R = 0.05$ p.u. is adopted, corresponding to a primary frequency-response gain of $K_f = \frac{1}{R} = 20$ in per-unit frequency coordinates. The commanded active-power increment is subsequently passed through the turbine-governor dynamics before contributing to the generator mechanical-power response.

The active-power command generated by the primary frequency-response controller is subsequently passed through the governor and turbine dynamics to determine the change in generator mechanical input. The corresponding simplified transfer relationship is expressed as

$$\Delta P_{m,i}(s) = \frac{1}{(1 + sTg)(1 + sTt)} \Delta P_{Gi}^{p}(s) \tag{13}$$

where $T_g = 0.2$ s is the governor time constant, $T_t = 0.5$ s is the turbine time constant, and $\Delta P_{m,i}$ represents the resulting change in generator mechanical-power input. The reheat characteristic of the turbine is represented using a reheat time constant of $T_r = 7$ s. For compactness, Eq. (13) shows the non-reheat form; the reheat stage is included in the time-domain simulations. The governor output is constrained between $G_{min} = 0$ p.u. and $G_{max} = 1.0$.

In addition to primary governor response, secondary frequency regulation is provided through the automatic generation control (AGC) loop. The AGC mechanism uses the area control error (ACE), which combines the tie-line power deviation and frequency deviation, to generate the secondary active-power correction. The ACE is defined as

$$ACE(t) = \Delta P_{tie}(t) + B_f \Delta f(t) \tag{14}$$

Because each benchmark system is represented as a single control area, the tie-line term in Eq. (14) is zero and the ACE reduces to the frequency-bias term. The AGC active-power command is then obtained using a proportional-integral control law:

$$\Delta P_{AGC}(s) = -\left(K_P^{AGC} + \frac{K_I^{AGC}}{s} \right) ACE(s) \tag{15}$$

where K_P^{AGC} =2.0 p.u./Hz and K_I^{AGC} =1.0 p.u./s are the proportional and integral gains of the AGC controller, respectively. The AGC time constant is $T_{AGC} = 5.0$ s, the frequency-bias coefficient is $B_f = 0.01$ p.u./Hz, and the maximum AGC ramp rate is limited to 0.1 p.u./min. The AGC command modifies the active-power operating point of participating generators while respecting their individual generation limits and available headroom.

When more than one generator is available to provide frequency support, the required active-power correction is distributed according to the available generation headroom of the participating generators. The allocation for generator ii is expressed as

$$\Delta P_{Gi}^{cmd} = \frac{\Delta P_{G_i}}{\sum_{j=1}^{N_G} \Delta P_{G_j}^{max}} \Delta P_{required} \tag{16}$$

where N_G is the number of participating generators and ΔP_{G_i} is the available generation headroom of generator i. The individual command is saturated at the corresponding generator headroom to prevent operation beyond the available active-power reserve. Thus, the total corrective active-power response is given by

$$\Delta P_{total} = \sum_{i=1} \Delta P_{Gi}^{cmd} \tag{17}$$

For the standard IEEE 39-bus generator data used in the simulations (MATPOWER case data), the total available upward generation headroom is 1,101 MW (Table 4).

The active-power frequency-control mechanism requires explicit specification of the governor, turbine, and AGC parameters to ensure reproducibility of the frequency-response simulations. The primary frequency response is governed by the droop characteristic and the governor–turbine dynamics, while the secondary response is provided through the AGC loop. The parameters adopted for these control mechanisms are summarized in Table 3.

Table 3. Active-Power Frequency-Control Parameters Used in the Simulation

Parameter	Symbol	Value	Unit
Governor droop	R	0.05	p.u.
Frequency-response gain	K_f	20	p.u./p.u.

Governor time constant	T_g	0.2	s
Turbine time constant	T_t	0.5	s
Reheat time constant	T_r	7.0	s
Maximum gate opening	G_{max}	1.0	p.u.
Minimum gate opening	G_{min}	0.0	p.u.
AGC proportional gain	K_P^{AGC}	2.0	p.u./Hz
AGC integral gain	K_I^{AGC}	1.0	p.u./s
AGC time constant	T_{AGC}	5.0	s
Frequency bias	B_f	0.01	p.u./Hz
Maximum AGC ramp rate	$\dot{P}_{AGC,\max}$	0.1	p.u./min

As shown in Table 3, a governor droop of $R = 0.05$ is used, corresponding to a primary frequency-response gain of $K_f = 20$ in per-unit frequency coordinates. The governor and turbine time constants are set to 0.2 s and 0.5 s, respectively, with a reheat time constant of 7.0 s. The AGC controller uses proportional and integral gains of 2.0 p.u./Hz and 1.0 p.u./s, respectively. The specified gate limits and AGC ramp-rate constraint ensure that the corrective active-power command remains within the operating capability of the generators.

In addition to the governor and AGC parameters, the available active-power headroom of each generator is explicitly considered when allocating the frequency-support command. The available headroom represents the maximum additional active power that can be supplied by a generator above its current operating point. The generator operating limits and corresponding headroom used in the simulations are given in Table 4**.**

Table 4. Generator Active-Power Limits and Available Generation Headroom

Generator	$P_{G_i}^{\max}$ (MW)	$P_{G_i}^{current}$ (MW)	$P_{G_j}^{\max}$ (MW)	Headroom Fraction
G1	1040	250	790	0.72
G2	646	678	0†	0.00
G3	725	650	75	0.07
G4	652	632	20	0.02
G5	508	508	0	0.00
G6	687	650	37	0.03
G7	580	560	20	0.02
G8	564	540	24	0.02
G9	865	830	35	0.03
G10	1100	1000	100	0.09
Total	7367	6298	1101	1.00

As indicated in Table 4**,** the total available upward generation headroom is 1,101 MW. The slack unit G2 (†) operates at 678 MW, i.e., 32 MW above its 646 MW rating, and is therefore

assigned zero headroom; the net capacity margin of the system is 7,367 − 6,298 = 1,069 MW. During a frequency-instability event, the required active-power correction is distributed among participating generators in proportion to their available headroom.

3.6 Stability Classification and Corrective Control Decision Engine

The output from the LSTM is fed into a fully connected dense layer of size 64 and ReLU activation, which is subsequently passed to a classification layer (Softmax). This layer produces posterior probabilities between the four stability classes. The classification decision threshold is put at 0.85: If the highest class probability is greater than the threshold, the decision engine will classify the sample to the corresponding stability classification; else the sample is tagged for additional monitoring. If the detection engine detects any instability class, it will choose and activate the corresponding corrective control action in the FACTS system. When voltage instability events occur, voltage magnitude is restored to within acceptable limits by injecting reactive power from SVC. When the transient instability conditions are met, STATCOM ensures reactive power support to improve synchronizing torque and suppress electromechanical oscillations. When frequency instability is detected, the active-power imbalance is addressed through the turbine-governor and AGC response available in the benchmark system, supplemented by fast BESS and virtual-inertia active-power support where available (Section 4.3.2), while the SVC and STATCOM provide supplementary reactive-power and voltage support. The governor response provides the primary active-power response to frequency deviation, whereas the AGC adjusts the generator operating point to support frequency recovery. The FACTS devices do not provide the primary active-power response; instead, their reactive-power compensation supports the network voltage profile during the active-power recovery process.

3.7 Model Training Configuration

Training is implemented in TensorFlow 2.12 and Keras using the Adam optimizer with an initial learning rate of 0.001, an exponential decay rate of 0.96 every 10 epochs, and a batch size of 64. As per the multi-class classification objective, the categorical cross-entropy is used as a loss function. Early stopping callback is used to check validation loss with patience of 20 epochs, which will stop the training when there is no improvement in validation loss. Model checkpointing is used to save the best configuration based on validation accuracy for further evaluation. To further prevent overfitting, L2 regularization with a coefficient of 10^{-4} is applied to the weights of all fully connected layers. Hyperparameters were selected as follows. The observation-window length and the PMU reporting rate were chosen from the sensitivity analyses in Section 5.12 (Tables 21 and 22). The numbers of convolutional filters and LSTM units, the dropout rates, and the learning rate were selected by comparing candidate configurations on the validation set. The decision threshold of 0.85 was chosen to balance missed and false detections.

The key evaluation measures are classification accuracy and macro-averaged precision, recall, and F1-score. Macro averaging weights the four stability classes equally, so that the metrics are not dominated by the majority Stable class. Per-class performance assessment is complemented by performance assessment using confusion matrix analysis and Receiver Operating Characteristic (ROC) curves with Area Under the Curve (AUC) values. Per-sample inference latency is used to capture computational efficiency in the context of standardized hardware platforms.

Model Complexity: Because a separate model is trained for each benchmark system, the number of parameters depends on N through the first LSTM layer. The models contain 1,119,684, 2,954,692, and

8,197,572 parameters for the IEEE 9-, 39-, and 118-bus systems, respectively, of which 896 per model are non-trainable batch-normalization statistics; the CNN module (126,080 parameters including batch normalization), the second LSTM layer and layer normalization (197,632 parameters), and the dense classification head (8,516 parameters) are identical for all systems. The trained weights occupy approximately 4.5, 11.8, and 32.8 MB in 32-bit floating point. Layer-wise output shapes and parameter counts for each benchmark system are listed in Table A2. All experiments were repeated five times with fixed random seeds (42, 123, 2024, 7, and 99), which were set for Python's random module, NumPy, and TensorFlow (random.seed, np.random.seed, and tf.random.set_seed) before each training run, so that the reported spread characterizes the run-to-run variability introduced by random weight initialization, mini-batch shuffling, and dropout. Over the five runs the proposed CNN-LSTM model achieves an accuracy of 97.4 ± 0.3%, a macro-averaged precision of 96.8 ± 0.4%, a macro-averaged recall of 97.1 ± 0.3%, and a macro-averaged F1-score of 96.9 ± 0.4%, stated as the mean and the sample standard deviation across the five runs. The standard deviation is at most 0.4 percentage points for every metric, which is an order of magnitude smaller than the margin over the strongest single-modality baseline in Table 16, so the reported improvement is not attributable to seed selection. Unless stated otherwise, the single-run results presented in the main text correspond to seed 42, whose macro-averaged F1-score of 97.0% lies within one standard deviation of the five-run mean.

4. Simulation Setup and Dataset Generation

The proposed spatiotemporal deep learning framework needs to be thoroughly evaluated; an environment for simulations is required to faithfully represent the dynamic behaviour of realistic power transmission networks across a wide range of operating conditions and disturbance scenarios. Thus, there are three standard IEEE test systems, each increasingly complex, and a large scale labelled dataset is created by detailed time domain dynamic simulation.

4.1 Benchmark Test Systems

The simulations are conducted using MATLAB R2023a software, with the Simulink Power Systems Toolbox. Sixth-order synchronous machine representations with saturation effects, IEEE Type-I Automatic Voltage Regulators (AVRs), reheat turbine-governor representations and Power System Stabilizers (PSS) at each generator bus are used to model each benchmark system to ensure physically accurate dynamic responses. The standard ZIP composite load model has been selected to represent realistic load voltage sensitivity characteristics with load buses. In Table 5, the main features of the three test systems and the size of the related datasets are summarized. Because the ground-truth frequency thresholds and the PMU reporting rate adopted in this study correspond to a 50 Hz system, all benchmark systems, which are originally specified at 60 Hz, were simulated with a nominal frequency of 50 Hz.

Table 5: IEEE Benchmark Test Systems and Dataset Composition Summary

Test System	Buses	Generators	Lines / Transformers	Simulation Cases
IEEE 9-Bus	9	3	6 / 3	450
IEEE 39-Bus	39	10	34 / 12	1,200

IEEE 118-Bus	118	54	177 / 9	3,500
Total	—	—	—	5,150

4.1.1 Renewable-Integrated Simulation Configuration

To explicitly represent renewable-integrated operating conditions, additional simulations were performed by incorporating converter-interfaced renewable generation into all three benchmark systems. Type-4 permanent-magnet synchronous generator (PMSG) wind power plants and grid-following solar photovoltaic (PV) inverter models were integrated with the benchmark network. The wind-generation model includes the aerodynamic and two-mass drive-train dynamics, PMSG, machine-side converter, DC-link, grid-side converter, and PLL-based d-q current control. The PV model includes an equivalent PV array, DC-link dynamics, maximum-power-point tracking (MPPT), grid-following voltage-source converter, PLL synchronization, and d-q active/reactive power control. Converter current limits and voltage-support controls are included to represent the dynamic response of converter-interfaced resources. The renewable-integrated configurations are evaluated at 0%, 30%, and 50% renewable-energy penetration levels, with the corresponding generation dispatch and renewable-resource locations defined in Section 4.3.

4.2 Disturbance Scenarios and Data Generation Protocol

Five different categories of disturbance events are applied in a systematic way to each test system to maximise the diversity and representativeness of the training data set. First, three-phase balanced short-circuit faults are applied at several bus locations, with fault-clearing times uniformly sampled from 80 ms to 500 ms, ranging from well within to beyond the critical clearing time. Second, single-line-to-ground and line-to-line faults are simulated at different transmission-line locations as asymmetric contingencies. Third, load spikes of 20% to 60% of nominal bus load are applied to high loaded buses to induce frequency instability scenarios. Fourth, individual generator outages are simulated as forced disconnections of single generating units. Fifth, outages are placed on overloaded tie-lines to challenge voltage stability and/or place power transfer limits. The number of cases per disturbance type and benchmark system is given in Table A4.

In summary, the dataset contains 5,150 disturbance cases (450, 1,200, and 3,500 for the IEEE 9-, 39-, and 118-bus systems), each represented by a 500 ms window of PMU measurements at 100 fps (50 frames × N buses × 5 quantities). The operating conditions cover loading levels of 80%, 100%, and 120% of nominal load. Renewable penetration levels of 0%, 30%, and 50% are considered for all three benchmark systems (Section 4.3), so that the renewable-integrated dataset used in Section 5.6.2 contains 3 × 5,150 = 15,450 cases. An additional set of IEEE 39-bus scenarios with 0–15% RES penetration, generated separately for the generalization test LOOCO-5 in Section 5.15, is not part of these datasets.

In the baseline simulation, synchronized PMU measurements are assumed to be available at all buses of the benchmark systems. This full-bus PMU coverage represents an idealized observability condition adopted to evaluate the intrinsic classification capability of the proposed CNN-LSTM architecture without introducing placement-dependent information loss. Therefore, the baseline input tensor contains measurements from all monitored buses [8,9]. The practical impact of incomplete PMU deployment is subsequently investigated through reduced-observability experiments in Section 5.13.1.

In this study, the dynamic simulation sampling time step and the PMU reporting rate are treated as separate quantities. The resulting measurements are represented as a simulated PMU data stream reported at 100 frames per second (fps), corresponding to a 10 ms reporting interval. IEC/IEEE 60255-118-1 [31] specifies the reporting rates 10, 25 and 50 fps for a 50 Hz system, and additionally permits and encourages higher rates such as 100 fps where the application requires them; the 100 fps rate adopted here therefore lies above the specified rates but within the range sanctioned by the standard, and is consistent with the capability of commercially available PMUs. The sensitivity of the proposed model to lower reporting rates, including the 50, 25 and 10 fps rates specified in the standard, is quantified in Section 5.12.1 and Table 21. The 100-fps value therefore represents the PMU reporting rate supplied to the proposed CNN-LSTM model. A 500 ms online model-input window is selected, resulting in input tensors of dimension (50 x N x 5), where N is the number of monitored buses and 5 is the number of measured features.

Because the classifier requires the complete 500 ms window, a stability decision can be issued no earlier than 519.2 ms after disturbance onset (500 ms window, 1.9 ms preprocessing, and 17.3 ms inference); the communication delay and the FACTS response time must be added to obtain the instant of corrective action (Table A3). The Simulink models were solved with the stiff variable-step solver ode23tb with a maximum step size of 50 μs (5×10^{-5} s) for all benchmark systems, which is sufficiently small to capture the fast electromagnetic and electromechanical transients of the sixth-order synchronous machine models, AVRs, governors, PSS, and FACTS devices over the 5 s post-disturbance horizon. The PMU measurements were obtained by sampling the simulation outputs at 10 ms intervals, independently of the integration step.

The offline ground-truth stability labels were generated from the complete post-disturbance trajectories using the explicit stability criteria described in Section 4.2.2. Transient-stability classification was additionally verified using the energy-function evaluation method, while voltage-stability assessment was supported by long-term voltage trajectory analysis. The entire labeled dataset of 5,150 samples was then divided into training, validation, and test sets using a stratified random sampling ratio of 70%, 15%, and 15%, respectively, to preserve the class distribution across the three subsets.

4.2.1 Dataset Composition and Class Distribution

The complete dataset contains 5,150 labelled simulation cases generated across the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus benchmark systems. The cases are distributed among four stability classes: Stable, Transient Instability, Voltage Instability, and Frequency Instability. The class distribution is maintained at 40%, 30%, 18%, and 12%, respectively, across the three benchmark systems to enable consistent cross-system evaluation. The corresponding system-wise case counts and stratified training, validation, and test-set distributions are summarized in Table 6.

Table 6. Class distribution of the simulation dataset across the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus benchmark systems.

Class	IEEE 9-Bus	IEEE 39-Bus	IEEE 118-Bus	Combined
Stable	180	480	1,400	2,060
Transient Instability	135	360	1,050	1,545
Voltage Instability	81	216	630	927
Frequency Instability	54	144	420	618
Total	450	1,200	3,500	5,150

4.2.2 Offline Ground-Truth Label Generation and Stability Criteria

Each disturbance case was simulated for 5 s following the disturbance using time-domain simulation. The complete 5-s post-disturbance trajectory was retained for offline ground-truth stability assessment, whereas only the first 500 ms of synchronized PMU measurements was used as the input to the proposed CNN–LSTM model for online stability classification, as summarized in Table 7. The closed-loop control-validation cases in Section 5 were simulated over longer horizons (up to 10 s) to capture the complete recovery behaviour.

i. Transient Instability: A disturbance case was classified as Transient Instability when the maximum relative rotor-angle separation between any two synchronous generators exceeded 180° at any instant during the 5-s post-disturbance simulation. The transient-stability classification was additionally verified using the corresponding energy-function criterion. For this purpose, the generators were represented by the classical model of the reduced network, and the transient energy function (TEF) was evaluated in the centre-of-inertia frame as

$$V = ½\sum_{i=1}^{n} M_i\tilde{\omega}_i^2 - \sum_{i=1}^{n} P_i(\theta_i - \theta_i^s) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} C_{ij}(\cos\theta_{ij} - \cos\theta_{ij}^s) \quad (18)$$

where M_i is the inertia coefficient of generator i, $\tilde{\omega}_i$ and θ_i are its rotor-speed deviation and rotor angle relative to the centre of inertia, θ_i^s is the post-fault stable equilibrium angle, $P_i = P_{mi} - E_i^2 G_{ii}$ is the net mechanical power input, and $C_{ij} = E_i E_j B_{ij}$, where B_{ij} is the transfer susceptance of the reduced network; transfer-conductance terms are neglected. The critical energy V_{cr} was taken as the potential energy at the controlling unstable equilibrium point (CUEP), which was located using the PEBS-based BCU method. A case was classified as transiently unstable when the system energy at fault clearing exceeded V_{cr}. The energy-function criterion agreed with the 180° maximum relative rotor-angle separation criterion in 99.1% of the transient-instability cases; the remaining disagreements occurred only for marginal cases near the stability boundary.

The rotor-angle trajectories used for this criterion are obtained from the offline time-domain simulation and are used exclusively for ground-truth transient-stability labeling and validation. Rotor-angle variables are not provided as input features to the CNN-LSTM model during online inference.

ii. Voltage Instability: For the present study, a disturbance case was classified as Voltage Instability if either (i) the voltage magnitude at any monitored load bus remained below 0.80 p.u. continuously for more than 1 s following the disturbance, or (ii) the voltage failed to recover to 0.90 p.u. within 2 s after the disturbance.

iii. Frequency Instability: For the present study, a disturbance case was classified as Frequency Instability if either (i) the system frequency nadir dropped below 49.0 Hz at any point during the 5-s post-disturbance simulation, or (ii) the system frequency failed to recover to 49.5 Hz within 2 s after the disturbance. Here, the system frequency is the centre-of-inertia (COI) frequency $f_{COI} = \frac{\sum_{i=1}^{n} H_i f_i}{\sum_{i=1}^{n} H_i}$, where H_i and f_i are the inertia constant (expressed on the common 100 MVA system base) and the measured frequency of the i-th online synchronous generator, respectively, and n is the number of online synchronous generators. The COI frequency is used consistently for all frequency-instability labelling and assessment in this study.

iv. Stable: A disturbance case was classified as Stable when none of the predefined transient, voltage, or frequency instability criteria was satisfied during the 5-s post-disturbance simulation.

v. Resolution of Multiple Instability Indicators: When a disturbance case satisfied more than one instability criterion simultaneously, a deterministic priority rule was applied to obtain a mutually exclusive four-class target. The priority order was defined as Transient Instability > Frequency Instability > Voltage Instability > Stable. Thus, transient instability was assigned when the transient criterion was satisfied, followed by frequency instability and voltage instability when the higher-priority criteria were not satisfied.

This priority rule was adopted to maintain a mutually exclusive four-class formulation while assigning the dominant operational instability mechanism when multiple criteria were simultaneously satisfied. Transient instability was given the highest priority because loss of synchronism represents the most severe electromechanical instability mechanism, followed by frequency instability and voltage instability. Therefore, the Softmax classifier predicts a single operationally dominant stability class for each disturbance case rather than treating overlapping mechanisms as independent labels.

Table 7. Offline Ground-Truth Stability Classification Criteria

Stability Class	Stability Indicator	Offline Classification Criterion	Assessment Horizon
Transient Instability	Relative rotor-angle separation	Maximum relative rotor-angle separation between any two synchronous generators exceeds 180°	5 s
Voltage Instability	Load-bus voltage magnitude	Voltage remains below 0.80 p.u. for >1 s, or fails to recover to 0.90 p.u. within 2 s	5 s
Frequency Instability	System frequency	Frequency nadir falls below 49.0 Hz, or fails to recover to 49.5 Hz within 2 s	5 s
Stable	Combined stability assessment	None of the transient, voltage, or frequency instability criteria is satisfied	5 s

When multiple instability criteria were satisfied for the same disturbance case, the predefined priority order Transient Instability > Frequency Instability > Voltage Instability was applied to assign a single mutually exclusive class label. The priority rule was used consistently for all disturbance cases and ensured that the target labels were compatible with the four-class Softmax formulation.

Threshold selection. The 49.0 Hz nadir threshold coincides with the first under-frequency load-shedding (UFLS) stage used in this study (Section 5.6), so that every case that would trigger load shedding is labelled unstable, while the 49.5 Hz recovery criterion identifies cases whose frequency remains depressed beyond the primary-response period. The 0.90 p.u. recovery threshold corresponds to the ANSI C84.1 lower limit used in Section 5.11, and the 0.80 p.u. for 1 s criterion separates sustained voltage depressions associated with voltage collapse from short fault-induced voltage dips. These thresholds are study-specific engineering criteria; other grid codes may adopt different values, which would shift the class boundaries without changing the proposed methodology. The 5 s labelling horizon applies to dataset generation only; the closed-loop control-validation cases in Section 5 are simulated for up to 10 s.

Physical interpretation of the classes. Following the stability classification of Kundur et al. [23], Transient Instability denotes the loss of synchronism of one or more generators after a large disturbance because of

insufficient synchronizing torque, observed as growing rotor-angle separation. Voltage Instability denotes the inability of the network to maintain acceptable bus voltages because the reactive-power demand exceeds the available reactive support, observed as sustained or progressive voltage decline. Frequency Instability denotes the inability to restore frequency after a significant imbalance between generation and load, observed as a deep frequency nadir or slow recovery. The Stable class contains all cases in which none of these mechanisms occurs. These mechanisms determine the corrective action: reactive-power support for transient and voltage instability, and active-power support for frequency instability.

4.3 Renewable-Integrated Operating Configurations

To evaluate the proposed framework under renewable-integrated operating conditions, each benchmark system was modified by replacing part of the synchronous-generation dispatch with converter-interfaced wind and solar PV generation. Three operating configurations were considered: 0% RES as the synchronous-generator baseline, 30% RES as a moderate renewable-penetration condition, and 50% RES as a high renewable-penetration condition. The installed capacities, locations, actual outputs, and resulting system inertia of the renewable resources are summarized in Table 8 for all three benchmark systems.

Table 8. Renewable-Integrated Configurations and System Inertia for the Three Benchmark Systems

System	Config.	Nominal / actual RES pen. (%)	Wind cap. (MW)	Solar cap. (MW)	Total RES cap. (MW)	Wind out. (MW)	Solar out. (MW)	Total RES out. (MW)	SG out. (MW)	Hsys (s)	Wind buses	PV buses
IEEE 9-bus	Baseline	0 / 0	0	0	0	0	0	0	315	10.54	—	—
	Moderate RES	30 / 30.0	148	98	246	52	20	72	243	4.79	5, 7	9
	High RES	50 / 50.0	246	164	410	86	33	119	196	3.01	5, 7	9
IEEE 39-bus	Baseline	0 / 0	0	0	0	0	0	0	6298	3.80	—	—
	Moderate RES	30 / 25.0	1280	560	1840	448	112	560	5738	2.41	37, 33, 28	30, 32
	High RES	50 / 41.5	1850	1210	3060	648	242	890	5408	1.72	37, 33, 28, 25, 22	30, 32, 36, 38
IEEE 118-bus	Baseline	0 / 0	0	0	0	0	0	0	4242	4.50	—	—
	Moderate RES	30 / 30.0	1794	1196	2990	628	239	867	3375	3.20	10, 25, 45, 70, 90	15, 30, 55, 80, 100
	High RES	50 / 50.0	2990	1993	4983	1047	399	1446	2797	2.10	10, 25, 45, 70, 90, 110	15, 30, 55, 80, 100, 115

Here, the RES penetration is defined as the ratio of the installed wind and solar PV capacity to the total synchronous generation capacity of the respective benchmark system (820 MW, 7,367 MW, and 9,966 MW for the IEEE 9-, 39-, and 118-bus systems). The labels 30% and 50% denote the nominal moderate and high RES configurations; for the IEEE 39-bus system, the corresponding installed-capacity ratios are 25.0% and 41.5% (Table 8). The instantaneous RES output is computed by assuming wind and solar outputs equal to 35% and 20% of their installed capacities, respectively, which correspond to typical capacity factors; consequently, the RES output remains below the system load even at 50% installed-capacity penetration. The resulting instantaneous RES share of the load is 23% and 38% (IEEE 9-bus), 9% and 14% (IEEE 39-bus), and 20% and 34% (IEEE 118-bus) for the 30% and 50% configurations, respectively. The system inertia Hsys is computed from the inertia constants of the synchronous generators that remain online in each configuration.

4.3.1 Converter-Interfaced Renewable Resource Models

The Type-4 wind power plants are represented by aggregated full-converter permanent-magnet synchronous generator (PMSG) models. Each wind power plant is represented as an equivalent aggregation of 5-MW Type-4 wind turbine generators, with aerodynamic, two-mass drive-train, PMSG, machine-side converter, DC-link, and grid-side converter dynamics. The grid-side converter operates in grid-following mode and employs an outer active/reactive-power control loop, an inner d-q current-control loop, and a phase-locked loop (PLL) for grid synchronization. The inner and outer control-loop bandwidths are set to 150 rad/s and 15 rad/s, respectively.

The solar PV plants are represented using an aggregated grid-following voltage-source-converter model with DC-link dynamics. The DC side is represented by an equivalent PV array based on a single-diode characteristic, while a perturb-and-observe maximum-power-point-tracking (P&O-MPPT) controller updates the DC-voltage reference every 50 ms. The inverter uses PLL-based synchronization and d-q current control, with the d-axis used for active-power regulation and the q-axis for voltage/reactive-power regulation. Converter current limits and fast reactive-current support are included in both renewable-resource models.

Table 9. Key Parameters of the Converter-Interfaced Renewable Resource Models

Parameter	Type-4 Wind PMSG	Solar PV Inverter
Unit rating	5 MW/WTG	10 MW/inverter
Converter topology	2-level VSC	2-level VSC
Switching frequency	3 kHz	5 kHz
Inner current-loop bandwidth	150 rad/s	200 rad/s
Outer power-loop bandwidth	15 rad/s	20 rad/s
PLL bandwidth	20 rad/s	25 rad/s
DC-link capacitance	12 mF/MW	18 mF/MW
Fast reactive-current gain	4% of rated current per 1% ΔV	4% of rated current per 1% ΔV
Reactive capability	±0.33 Mvar/MW	±0.40 Mvar/MW

The parameters listed in Table 9 define the dynamic and converter-control characteristics of the renewable-generation models used in the renewable-integrated simulations. These parameters include converter topology, control-loop bandwidths, PLL response, DC-link characteristics, reactive-current support, and reactive-power capability, and are kept consistent across the corresponding simulation scenarios.

Renewable Power Variability: To account for the stochastic nature of renewable generation, wind-speed variations were represented using a Kaimal turbulence-based model with spatial coherence, while solar irradiance variability was represented using a cloud-shadow-based stochastic model. The resulting wind and PV power fluctuations were propagated through the corresponding converter controls and reflected in the PMU voltage, active-power, reactive-power, and frequency measurements used by the proposed CNN-LSTM framework.

4.3.2 Virtual Inertia and Active-Power Support

To account for the reduced inertial response associated with converter-interfaced renewable generation, a virtual inertia emulator (VIE) was incorporated into the renewable-resource control framework. The VIE generates an additional active-power reference from the measured rate of change of frequency and a droop-

based response. A 25-ms low-pass filter is applied to the frequency-derivative signal, while the virtual-inertia and droop components are combined to provide fast active-power support during frequency disturbances. The resulting active-power command is applied through the grid-side converter of the wind, PV, and battery energy-storage resources, while the SVC/STATCOM provides supplementary reactive-power and voltage support. In the IEEE 39-bus frequency-stability studies, two BESS units located at Bus 3 and Bus 29 provide up to 50 MW and 78 MW of fast active-power support, respectively, with a ramp rate of 250 MW/s (Section 5.6).

$$P_{VIE} = P_{VI} + P_{droop} \tag{19}$$

where P_{VI} represents the virtual-inertia response and P_{droop} represents the frequency-droop response.

4.3.3 Renewable-Integrated Disturbance Scenarios

The renewable-integrated configurations were evaluated under representative frequency and voltage disturbances to assess the robustness of the proposed stability-monitoring framework under varying renewable penetration levels. The simulations considered the baseline synchronous-generation condition (0% RES), moderate renewable penetration (30% RES), and high renewable penetration (50% RES). For each configuration, the same disturbance framework was applied while maintaining the corresponding renewable-generation dispatch and converter-control settings. Frequency-stability assessment focused on frequency nadir and maximum rate of change of frequency (ROCOF), whereas voltage-stability assessment considered the post-disturbance voltage response at critical buses.

Renewable-Penetration Impact on Frequency Stability: The effect of increasing renewable penetration was evaluated for the 0%, 30%, and 50% RES configurations. As renewable penetration increased, the synchronous-system inertia of the IEEE 39-bus system fell from 3.80 s to 2.41 s and 1.72 s, producing progressively higher ROCOF under the no-control condition. For the 650 MW trip of generator G6, corresponding to a loss of 10.3% of the synchronous dispatch, the corrective dispatch of the proposed framework becomes available 519.2 ms after the disturbance, which is after the ROCOF peak. The proposed framework therefore does not reduce the maximum ROCOF, and its effect on the frequency nadir is limited to approximately 0.1 Hz; its contribution is concentrated in the recovery phase, as quantified in Section 5.6. The corresponding comparison across the three benchmark systems at reduced inertia is given in Table 15. The sensitivity of the nadir and the ROCOF to the renewable penetration level under this disturbance is reported in Section 5.6.

4.4 Generalization and Data-Leakage-Aware Evaluation Protocol

To address potential data leakage associated with sample-level random partitioning, additional scenario-level validation protocols were implemented. Before data partitioning, the complete dataset was grouped according to contingency identity, operating condition, and fault location. Samples belonging to the held-out group were completely excluded from model training and validation and were used only for independent testing.

Three complementary validation protocols were considered. First, leave-one-contingency-out (LOCO) validation evaluates whether the model can generalize to disturbance contingencies that were not observed

during training. Second, leave-one-operating-condition-out (LOOCO) validation evaluates generalization to previously unseen loading and renewable-generation operating conditions. Third, unseen fault-location (UFL) testing evaluates whether the model can generalize spatially when fault locations excluded from training are used for testing. These scenario-level protocols provide a more stringent evaluation than the conventional random stratified split because closely related samples from the same contingency, operating condition, or fault location are prevented from appearing simultaneously in training and testing subsets.

4.5 Computational Environment and Reproducibility

All simulations and model-training experiments were performed using MATLAB R2023a with Simulink Power Systems Toolbox. The CNN-LSTM framework was implemented using TensorFlow 2.12.0 and Keras 2.12.0 under Python 3.9.16. The computational platform and training configuration are summarized in Table 10.

Table 10. Computational Environment and Reproducibility Configuration

Parameter	Configuration
CPU	Intel Xeon Gold 6248R, 24 cores, 2.7 GHz
RAM	128 GB DDR4
GPU	NVIDIA T4, 16 GB
Operating System	Ubuntu 20.04.6 LTS
MATLAB	R2023a
Deep Learning	TensorFlow 2.12.0 / Keras 2.12.0
Python	3.9.16
CUDA	11.8
cuDNN	8.6.0
Optimizer	Adam
Initial Learning Rate	0.001
LR Decay	0.96 every 10 epochs
Batch Size	64
Loss Function	Categorical Cross-Entropy
Early Stopping	Patience = 20
Model Selection	Best validation checkpoint

5. Results and Discussion

The performance of the proposed CNN-LSTM spatiotemporal framework is systematically analyzed by performing training convergence behavior, rotor angle dynamics, system frequency response analysis, multi-system classification accuracy, confusion matrix analysis, ROC curve analysis, frequency deviation characterization, and comparative benchmarking. These tests confirm the effectiveness, validity and applicability of the suggested framework for real-time monitoring and automatic corrective control of the power system stability.

5.1 Training and Validation Convergence

The learning behaviour in the proposed CNN-LSTM is very stable and consistent for over 200 training epochs, as confirmed by the convergence analysis as shown in Figure 4. The CNN layers progressively extract the features of spatial inter-bus correlation, while the LSTM layers begin to encode the temporal dependency, leading to a fast climbing training accuracy from around 56% to over 86% during the first 40 epochs. At the same time, the categorical cross-entropy loss drops from 2.18 to approximately 0.8, showing that the learning rate schedule was successfully adapted.

There is a gradual improvement in both training and validation accuracy throughout the intermediate training stage (epochs 40-120), ending at about 95%, and the losses are seen to settle down to below 0.2. The models show very high levels of generalization ability given the relatively high complexity of the four-class classification problem, with little spread between training and validation performance curves, and therefore demonstrating the effectiveness of the regularization mechanisms used, such as dropout, batch normalization, and L2 weight decay.

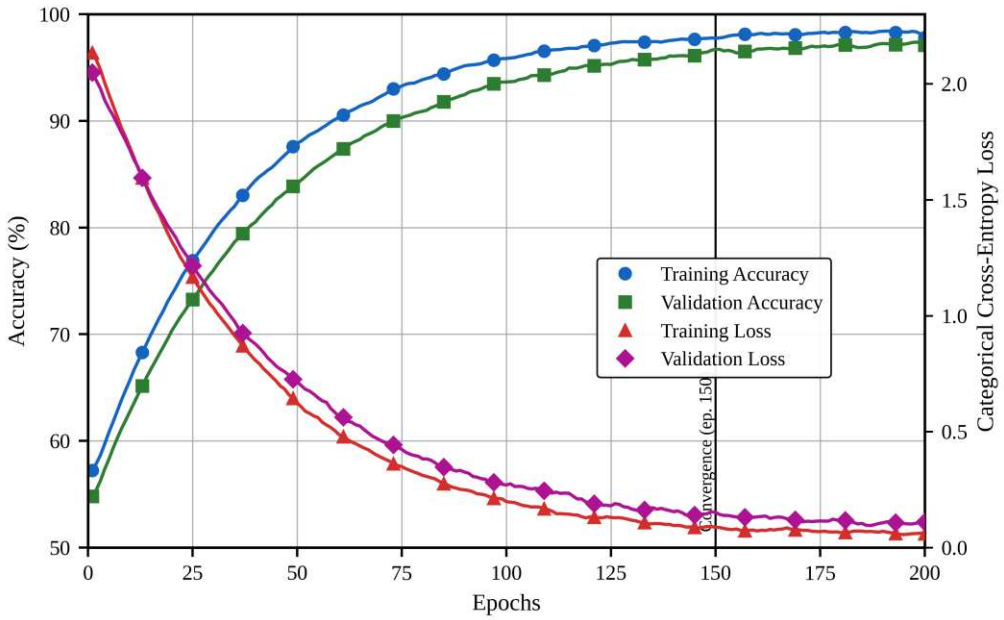


Figure 4. Training and validation accuracy and loss convergence curves over 200 epochs for the proposed CNN-LSTM model.

The model then goes into a plateau convergence region after epoch 150, which has little to no performance gains. The final gap between training and validation accuracy remains below one percentage point, indicating limited overfitting. The smooth and monotonic convergence trajectories confirm that the chosen hyperparameters configuration and learning rate scheduling are appropriate for the classification task at hand.

5.2 Rotor Angle Dynamics Following Severe Fault

The most direct quantitative measure of transient stability after severe disturbances is the rotor angle behaviour as shown in Figure 5. A three-phase bolted fault is applied at bus 16 of the IEEE 39-bus system at t = 0.1 s and cleared at t = 0.25 s; the resulting 150 ms fault duration exceeds the 100 ms critical clearing time of the least stable generator in the system.

Rotor angle oscillations remain bounded and have good damping during the disturbance period in the generators G1, G4 and G7, which have sufficient inertia and are electrically far from the faulted bus. Their rotor angles rise from pre-fault values of 20.0, 25.0 and 28.0 degrees to peaks of 26.2, 34.1 and 39.6 degrees at approximately 0.55 to 0.65 s, corresponding to deviations of 6.2, 9.1 and 11.6 degrees. The oscillations then decay continuously and return to the pre-disturbance equilibrium within approximately 2 s, which indicates that the transient stability margin is satisfactory for these machines. The classifier decision becomes available at t = 0.619 s and, allowing the 30 ms STATCOM response time, reactive-power injection begins at t = 0.649 s with zero communication delay. All closed-loop traces in this paper use the actuation instants of Table A3.

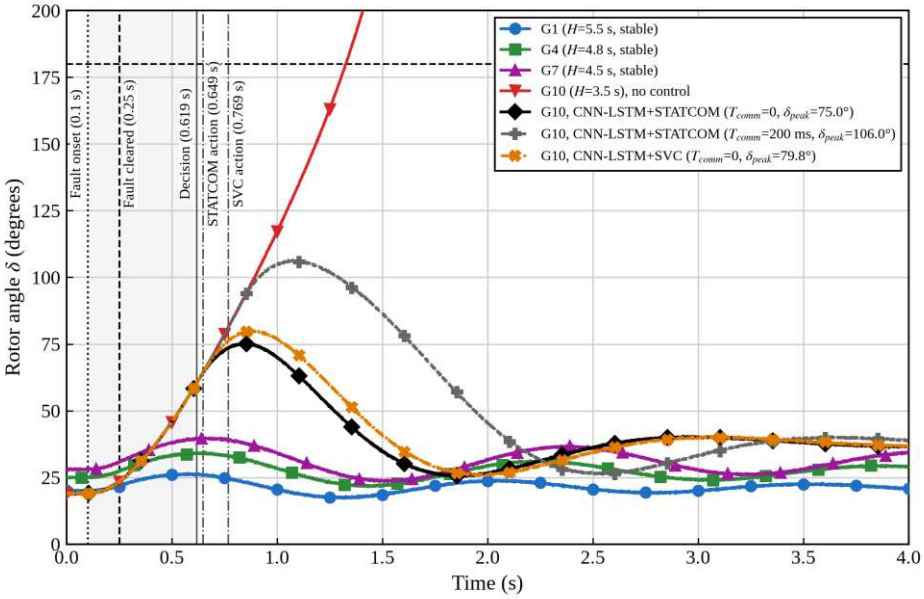


Figure 5. Rotor angle trajectories (degrees) of four generators (G1, G4, G7, and G10) in the IEEE 39-bus system following a three-phase fault at Bus 16: t = 0.1 s (fault application) and t = 0.25 s (fault clearing), with and without CNN-LSTM-assisted STATCOM corrective control. Vertical markers show the classifier decision at t = 0.619 s and the corrective-action instants for the STATCOM (t = 0.649 s) and the SVC (t = 0.769 s).

The time-domain transient-stability results demonstrate the effectiveness of the proposed CNN-LSTM-assisted STATCOM corrective control following the three-phase fault at Bus 16 in the IEEE 39-bus system. In the absence of corrective control, G10 progressively loses synchronism after fault clearing, whereas the proposed control limits the rotor-angle excursion and maintains synchronism throughout the post-fault period. The stable responses of G1, G4, and G7 further indicate that the corrective action does not adversely affect the remaining synchronous generators. A detailed comparison among no-control, conventional-control, and the proposed framework, together with the maximum inter-machine angle separation, is provided in Figure 9 and Table 11.

Generator G10, which has the lowest inertia constant among the monitored generators at H = 3.5 s (Figure 5), behaves very differently. Its rotor angle rises 4.5 degrees during the 150 ms fault, against 1.9, 2.4 and 2.8 degrees for G1, G4 and G7, and without corrective action it continues to accelerate and crosses the 180 degree boundary at t = 1.33 s. The CNN-LSTM model classifies the transient-instability condition from the

500 ms observation window and issues the decision at t = 0.619 s, so that STATCOM reactive-power injection begins at t = 0.649 s. Because the first swing peaks at approximately 0.85 s, the corrective action arrives before the peak: the injected reactive-power support improves the synchronizing torque and limits the peak rotor angle of G10 to 75.0 degrees, equivalent to a deviation of 56.0 degrees from the pre-fault angle of 19.0 degrees, after which the machine returns toward equilibrium and settles near 38 degrees within 3 s. With a representative 200 ms communication delay the action is deferred to t = 0.849 s and the peak rises to 106.0 degrees, which remains stable with a margin of 74 degrees. These results show that CNN-LSTM-based corrective control can prevent the propagation of rotor-angle instability under realistic actuation timing, provided the corrective action is delivered before the first-swing peak.

Table 11. Transient Stability Performance under Different Control Conditions

Control condition	G10 response	G10 peak rotor angle δ	Stability outcome
No Control (NC)	Progressive angle divergence	Crosses 180° at 1.33 s	Loss of synchronism
Conventional Control (CC)	Delayed angle divergence	Crosses 180° at 1.40 s	Loss of synchronism
Proposed Framework (PF)	Controlled angle excursion	75.0°	Stable, settles near 38°

As summarized in Table 11, the uncontrolled case results in loss of synchronism of G10 at t = 1.33 s, while conventional control only delays the crossing of the 180 degree boundary to t = 1.40 s. In contrast, the proposed framework limits the peak rotor angle of G10 to 75.0 degrees and maintains synchronism, with the machine settling near 38 degrees within 3 s. The margin below the 180 degree criterion is therefore 105 degrees, demonstrating the effectiveness of the CNN-LSTM-assisted corrective-control mechanism under the considered transient disturbance and the realistic 0.649 s actuation instant.

Table 12. STATCOM Response during Transient-Stability Correction

Parameter	Proposed Framework
STATCOM at Bus 16 rating	±150 MVAR
Peak reactive-power output at Bus 16	148 MVAR
Maximum utilization	98.7%
Duration above 90% output	≈310 ms
STATCOM at Bus 19 peak output	61.4 MVAR
G10 peak rotor angle δ	75.0°
Margin below 180° criterion	105°

The STATCOM response shown in Table 12 further confirms that the corrective action operates within the specified device limits. The Bus 16 STATCOM reaches a peak reactive-power output of 148 MVAR, corresponding to 98.7% of its +/-150 MVAR rating, and remains above 90% of rating for approximately 310 ms before decaying as the bus voltage recovers. The corrective action therefore provides substantial dynamic voltage support during the fault-recovery period while maintaining the transient stability of the system.

5.3 System Frequency Response Evaluation

The IEEE 39-bus system is evaluated for frequency-stability performance by tripping generator G6 at t = 0.50 s, removing 650 MW of synchronous generation and corresponding to a loss of 10.3% of the 6,298 MW dispatch. Four conditions are considered as shown in Figure 6: the uncontrolled base case with governor response only, conventional PI-based Automatic Generation Control (PI-AGC), and the proposed CNN-LSTM-based corrective-control strategy with and without a 200 ms communication delay. In the proposed framework, frequency-instability detection at t = 1.019 s activates the battery energy storage units at Bus 3 and Bus 29 together with the modeled turbine-governor and AGC mechanisms, while the SVC provides supplementary reactive-power support for voltage regulation.

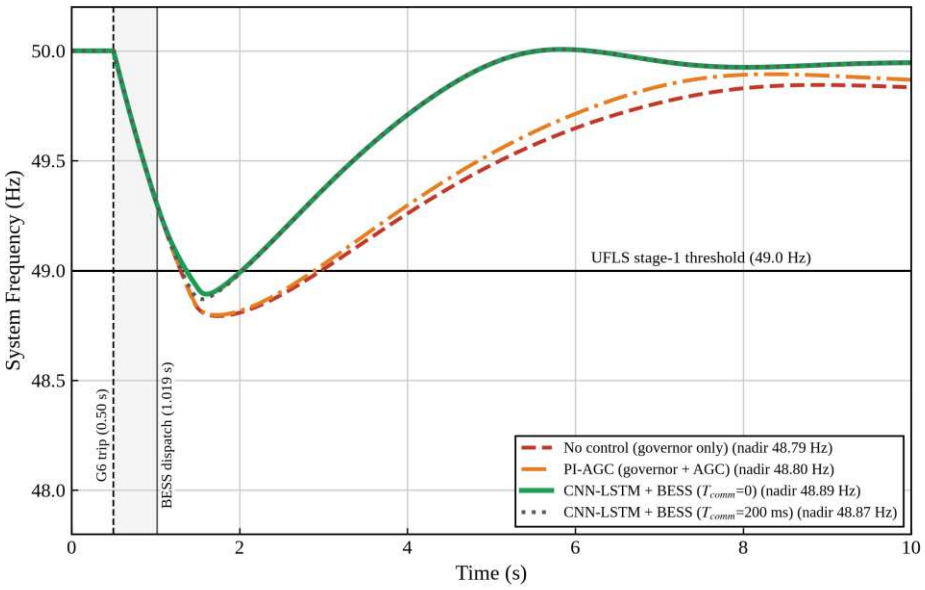


Figure 6. System frequency response of the IEEE 39-bus network following the 650 MW trip of generator G6 at t = 0.50 s: comparison among the uncontrolled condition, PI-AGC, and the proposed CNN-LSTM corrective control with and without a 200 ms communication delay. The classifier decision and battery dispatch occur at t = 1.019 s.

In the uncontrolled case, the system frequency decreases to a nadir of 48.79 Hz, and the conventional PI-AGC gives essentially the same nadir of 48.80 Hz because the 5.0 s AGC time constant is long relative to the nadir period. Under the proposed CNN-LSTM-based coordinated control the nadir improves to 48.89 Hz, an increase of 0.10 Hz; with a 200 ms communication delay the nadir is 48.87 Hz. The improvement is modest because the stability decision becomes available 519.2 ms after the trip, by which time the frequency has already fallen below 49.1 Hz, so the battery contributes only to the final stage of arresting the decline. The frequency crosses the 49.0 Hz under-frequency load-shedding stage-1 threshold in all four conditions, at t = 1.30 s without control and at t = 1.37 s with the proposed framework. The benefit of the proposed framework is concentrated in the recovery: the system returns above 49.8 Hz in 4.3 s, against 6.6 s with PI-AGC alone and 7.5 s without supplementary control.

Table 13. Frequency Response under Different Control Strategies

Control condition	Frequency nadir (Hz)	Return above 49.8 Hz	Active-power support

No Control	48.79	7.5 s	Governor
PI-AGC	48.80	6.6 s	Governor + AGC
Proposed CNN-LSTM	48.89	4.3 s	Governor + AGC + BESS

Table 13 shows that the proposed framework provides the highest frequency nadir and by far the shortest recovery time among the control conditions, although the nadir improvement itself is limited to 0.10 Hz. The improvement is not attributed to SVC reactive-power compensation; the active-power balancing required for frequency recovery is provided through the battery energy storage units and the modeled turbine-governor and AGC response, with the SVC operating as supplementary voltage-support equipment. The limited nadir improvement is a direct consequence of the 519.2 ms decision delay and is analysed further in Section 5.6.

5.4 Classification Accuracy Across Benchmark Systems

The proposed framework demonstrates its scalability and generalization capability on the IEEE benchmark systems. Accuracies of 98.5%, 97.8%, and 97.1% are obtained on the IEEE 9-bus, 39-bus, and 118-bus systems, respectively as shown in Figure 7. Modest performance decrement with growing network complexity shows the increased variety of operating states and disturbance signatures inherent in larger systems, and the consistently high absolute accuracy levels indicate that the CNN-LSTM architecture generalizes well across networks of different sizes.

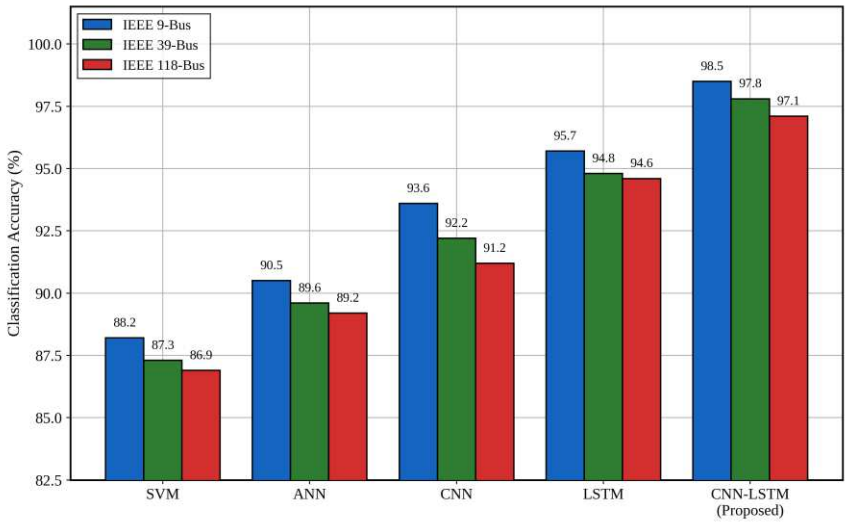


Figure 7. Classification accuracy comparison across the IEEE 9-bus, 39-bus, and 118-bus systems for SVM, ANN, CNN, LSTM, and the proposed CNN-LSTM framework.

The performance of the hybrid architecture is consistently better than individual CNN and LSTM models in comparison. A standalone CNN model that captures spatial features can be used to attain 92.3% accuracy, but is less useful for temporal modelling of disturbance evolution. Standalone LSTM achieves the accuracy of 94.6% by temporal learning while the ability to represent network-wide spatial correlations is not good enough. The best classification performance in all three benchmark systems is obtained when the CNN-LSTM fusion is used, which captures the complementary feature dimensions.

The classification performance of the proposed CNN-LSTM framework was further evaluated across three IEEE benchmark systems of increasing network size. The obtained accuracies are summarized in Table 14.

Table 14. Classification Performance across IEEE Benchmark Systems

Benchmark system	Accuracy (%)
IEEE 9-bus	98.5
IEEE 39-bus	97.8
IEEE 118-bus	97.1

The results show that the proposed framework maintains high classification accuracy as the network size increases from 9 to 118 buses. The accuracy decreases modestly from 98.5% for the IEEE 9-bus system to 97.1% for the IEEE 118-bus system, indicating that the proposed spatiotemporal feature-learning approach retains good generalization capability for larger network configurations.

5.5 Rotor Angle Deviation Analysis

The rotor angle deviation $\Delta\delta$ analysis offers further insight into the characteristics of generator stability margin from the pre-disturbance operating point. The three-phase fault applied on Bus 16 of the IEEE 39 bus system is followed by transient rotor angle deviations of the generators G1, G4 and G7, which reach a maximum level within a few tenths of a second and then return to the pre-disturbance state within approximately 2 s, as shown in Figure 8. The oscillation period of approximately 0.7 s (about 1.4 Hz) is characteristic of local electromechanical modes.

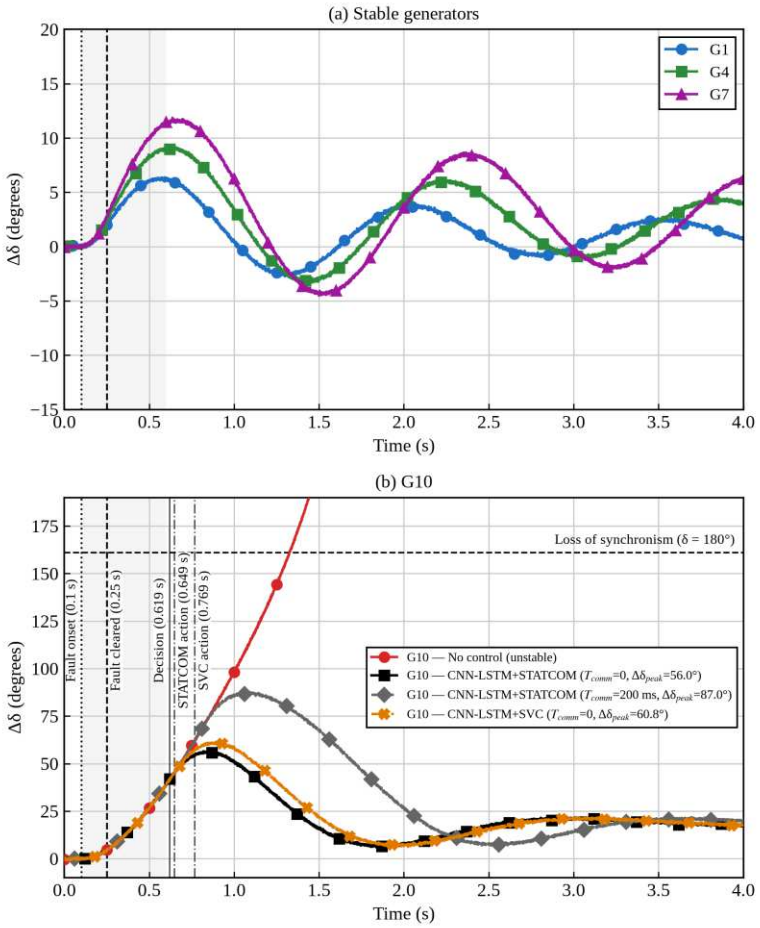


Figure 8. Rotor angle deviation (Δδ, degrees) of generators G1, G4, G7, and G10 in the IEEE 39-bus system: (a) stable generators without control, (b) unstable generator G10 with and without CNN-LSTM corrective control.

The uncontrolled trajectory of generator G10 is quite different, with the rotor angle deviation continuously increasing as the generator is losing synchronism. The post-correction deviation envelope shown in the figure is monotonically decreasing, indicating that the activation of the corrective control based on the activation of the CNN-LSTM decision engine results in a significant decrease in the peak deviation and allows the system to converge towards a stable equilibrium. The reactive-power injection strengthens the synchronizing and damping torque components, which is consistent with the physical mechanism of the corrective action.

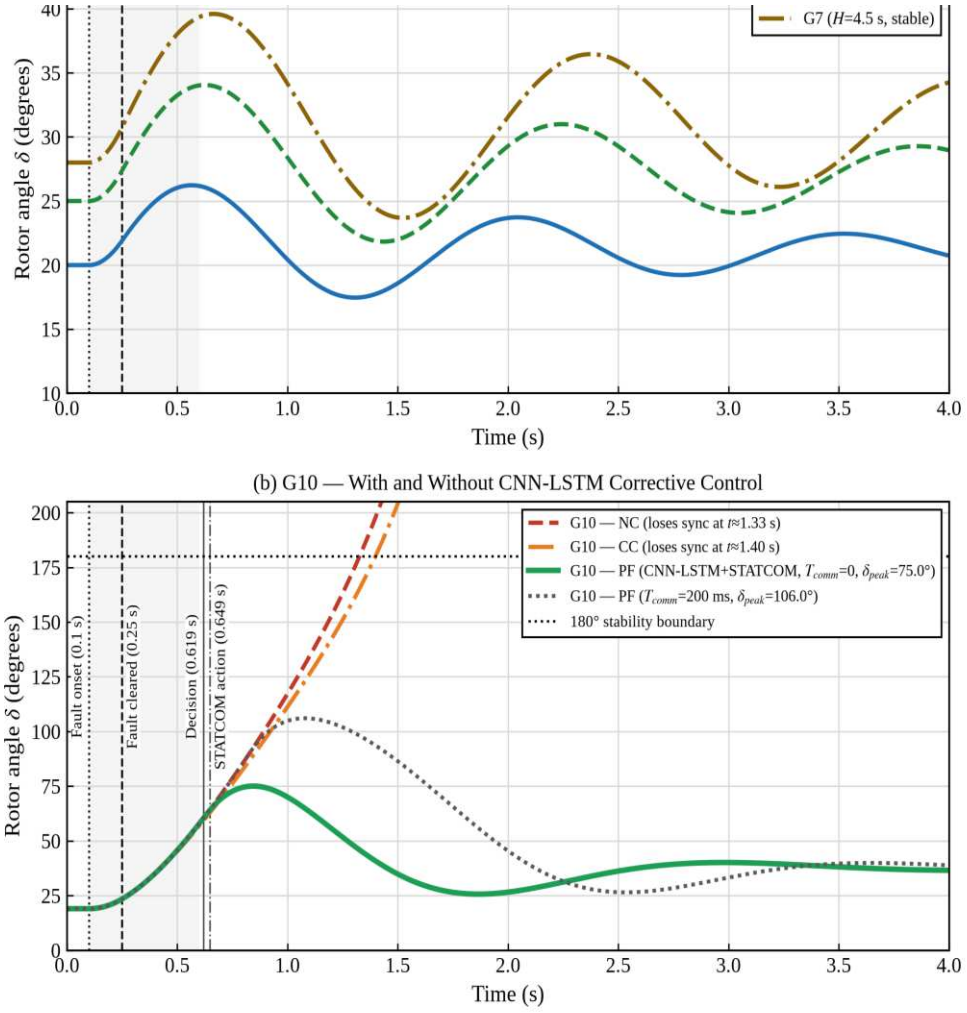


Figure 9. Detailed rotor-angle trajectories for the IEEE 39-bus three-phase fault at Bus 16: (a) stable generators G1, G4, and G7; (b) G10 response under no-control, conventional-control, and proposed CNN-LSTM-assisted STATCOM control.

Figure 9 provides a detailed comparison of the rotor-angle response under different control conditions. For the uncontrolled case, the rotor angle of G10 continues to increase after fault clearing and crosses the 180 degree transient-stability boundary at t = 1.33 s, resulting in loss of synchronism. Conventional control delays the crossing to t = 1.40 s but does not prevent the eventual loss of synchronism. In contrast, the proposed CNN-LSTM-assisted STATCOM control, acting at t = 0.649 s, limits the peak rotor angle of G10 to 75.0 degrees and maintains synchronism, with the machine settling near 38 degrees; with a 200 ms communication delay the peak is 106.0 degrees and the machine remains stable. The rotor-angle trajectories of G1, G4, and G7 remain bounded throughout the disturbance and recovery period.

5.6 Frequency Deviation Characterization

The frequency deviation analysis over the three test systems under CNN-LSTM corrective control, shown in Figure 10, indicates that frequency regulation is consistent regardless of network size. Following the 650 MW trip of generator G6, the uncontrolled IEEE 39-bus system reaches a maximum frequency deviation

of -1.21 Hz, which exceeds the -1.0 Hz UFLS stage-1 threshold shown in Figure 10. PI-AGC gives -1.20 Hz with a long recovery time. The proposed CNN-LSTM-assisted control limits the deviation to -1.11 Hz and returns the frequency above 49.8 Hz in approximately 4.3 s, against 7.5 s for the uncontrolled case.

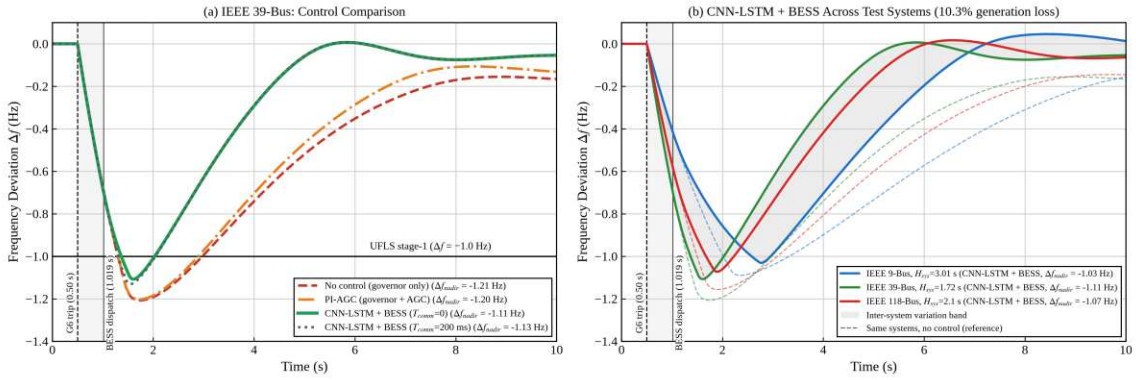


Figure 10. Frequency deviation (Δf, Hz) comparison across the IEEE 9-bus, 39-bus, and 118-bus systems under uncontrolled, PI-AGC, and proposed CNN-LSTM corrective control strategies.

The IEEE 9-bus network has the smallest frequency deviation at -1.03 Hz under the proposed control because of its higher per-unit inertia, while the IEEE 39-bus network, with the lowest system inertia of the three configurations, exhibits the deepest deviation at -1.11 Hz. The deviation exceeds the -1.0 Hz UFLS stage-1 threshold in all three systems both with and without corrective control, so under-frequency load shedding is not avoided by the corrective action for a disturbance of this severity. The improved frequency response obtained from the coordinated battery, turbine-governor and AGC active-power response is therefore expressed in the recovery trajectory rather than in the nadir, while the SVC provides supplementary reactive-power support to maintain the voltage profile during the recovery process.

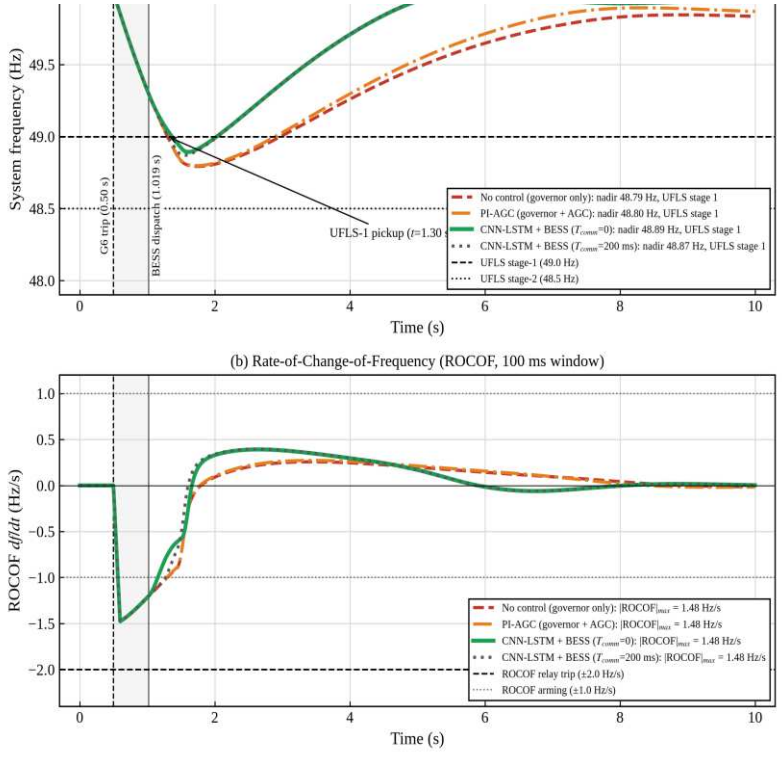


Figure 11. System frequency and ROCOF response under no-control, conventional PI-AGC, and proposed corrective-control conditions for the severe IEEE 39-bus generation-deficit case, namely the 650 MW trip of generator G6 at t = 0.50 s. The classifier decision and battery dispatch occur at t = 1.019 s.

As shown in Figure 11 and Table 15, the uncontrolled condition produces a frequency nadir of 48.79 Hz with a maximum absolute ROCOF of 1.48 Hz/s, and the first UFLS threshold is crossed at t = 1.30 s. The conventional PI-AGC condition gives a nearly identical nadir of 48.80 Hz and the same maximum ROCOF, because the AGC time constant of 5.0 s is long relative to the period over which the nadir is determined; frequency recovery remains comparatively slow, requiring approximately 6.6 s to return above 49.8 Hz.

Table 15. Frequency and ROCOF Performance under Severe Generation-Deficit Condition

Metric	**NC**	**PI-AGC**	Proposed Framework
Frequency nadir (Hz)	48.79	48.80	48.89
Corrective action instant (s)	—	—	1.019
Maximum |ROCOF| (Hz/s)	1.48	1.48	1.48
UFLS stage 1 activation (< 49.0 Hz)	Yes (t = 1.30 s)	Yes (t = 1.30 s)	Yes (t = 1.37 s)
UFLS stage 2 activation (< 48.5 Hz)	No	No	No
Return above 49.8 Hz (s)	7.5 s	6.6 s	4.3 s

Steady-state f deviation (Hz)	≈ −0.16 Hz	≈ −0.13 Hz	≈ −0.06 Hz

In contrast, the proposed CNN-LSTM-assisted corrective-control framework improves the frequency nadir to 48.89 Hz and returns the system above 49.8 Hz in approximately 4.3 s. The maximum absolute ROCOF, however, remains 1.48 Hz/s in all four conditions. This is a direct consequence of the measurement-window constraint rather than a limitation of the battery resource: the ROCOF peak occurs within the first 50 ms after the trip, whereas the earliest possible corrective dispatch is at t = 1.019 s. No control action based on a 500 ms observation window can influence the inertial phase of the response. The maximum ROCOF remains below the +/-2.0 Hz/s relay trip threshold in all cases but exceeds the +/-1.0 Hz/s arming threshold, and the UFLS stage-1 threshold is crossed in every condition, at t = 1.37 s under the proposed framework against t = 1.30 s without control. This result defines the operating envelope of the proposed framework: rotor-angle and voltage mechanisms develop over 0.5 to 2 s and are within reach of a 519.2 ms decision, as Sections 5.2 and 5.11 show, whereas the inertial phase of a frequency event is not, so fast frequency support in a converter-dominated system must remain local and measurement-triggered.

The recovery of the frequency response is supported by coordinated active-power resources. Following detection of the frequency-instability event at t = 1.019 s, the proposed framework dispatches the battery units together with the turbine-governor and AGC responses. The two BESS units ramp at 250 MW/s and reach their full outputs of 50 MW and 78 MW at t = 1.219 s and t = 1.331 s respectively, giving 128 MW of fast active-power support. Participating generators subsequently contribute through governor action, with increments of 54, 33, 31 and 48 MW from G1, G3, G9 and G10 at 10 s, allocated in proportion to the headroom of Table 4, while AGC provides further correction over the 2 to 6 s interval. The SVC and STATCOM operate as supplementary reactive-power support to assist voltage regulation during the frequency-recovery process.

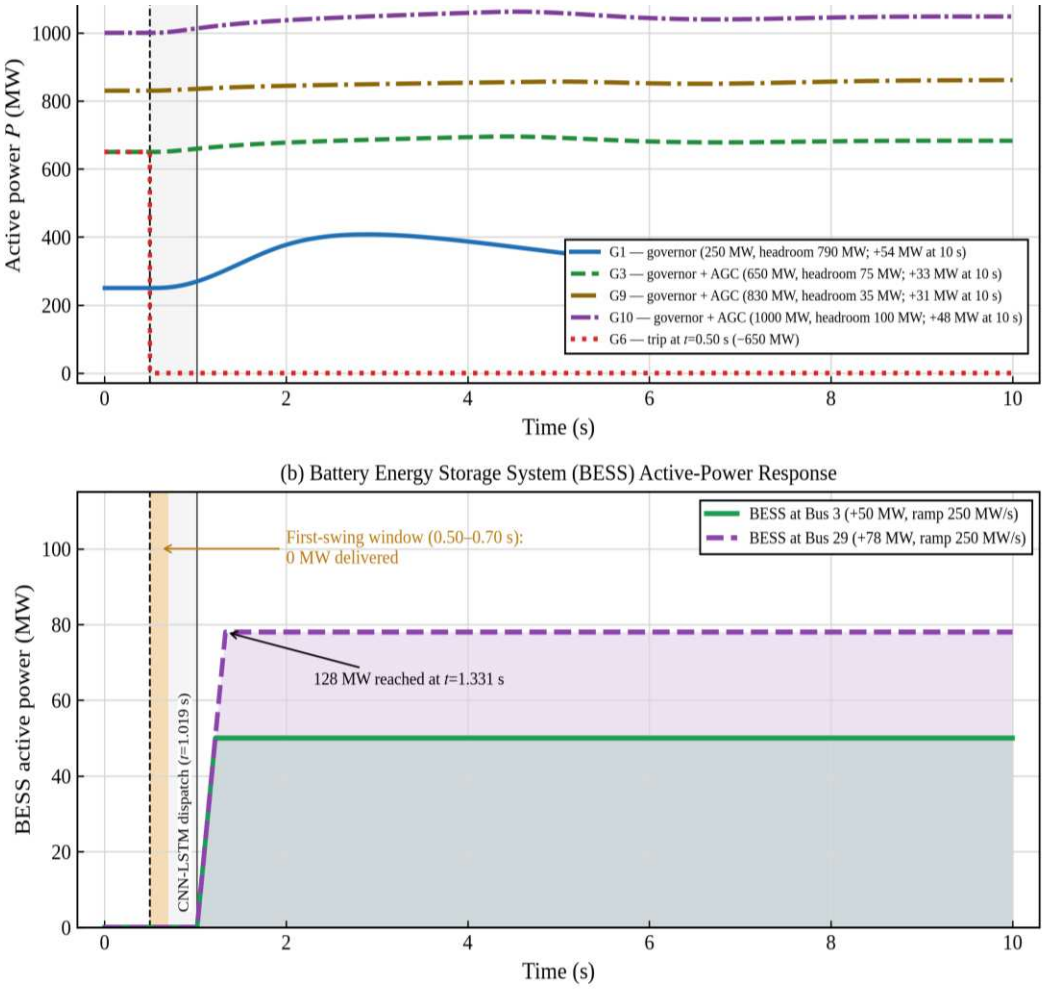


Figure 12. Active-power response of selected generators and BESS units during the severe frequency-instability event under the proposed corrective-control framework.

The active-power response shown in Figure 12 confirms that frequency recovery is supported by active-power resources rather than by FACTS reactive-power injection. The figure also makes the latency constraint explicit: during the first-swing window from 0.50 s to 0.70 s, in which the ROCOF peak occurs, the battery units deliver zero active power, because dispatch cannot begin before the stability decision at t = 1.019 s. The battery contribution of 128 MW is fully available from t = 1.331 s, after which the generator governor and AGC responses provide the remaining correction. The SVC and STATCOM provide supplementary reactive-power and voltage support during the frequency-recovery process.

5.6.1 Renewable-Integrated Voltage Stability

The voltage-stability performance was further evaluated under renewable power variability at high RES penetration, as shown in Figure 13. At 50% RES penetration, wind-power fluctuations produced a voltage-variation envelope of approximately ±0.012 p.u. under steady-state conditions, increasing to about ±0.025 p.u. during the 2–3 s post-disturbance period.

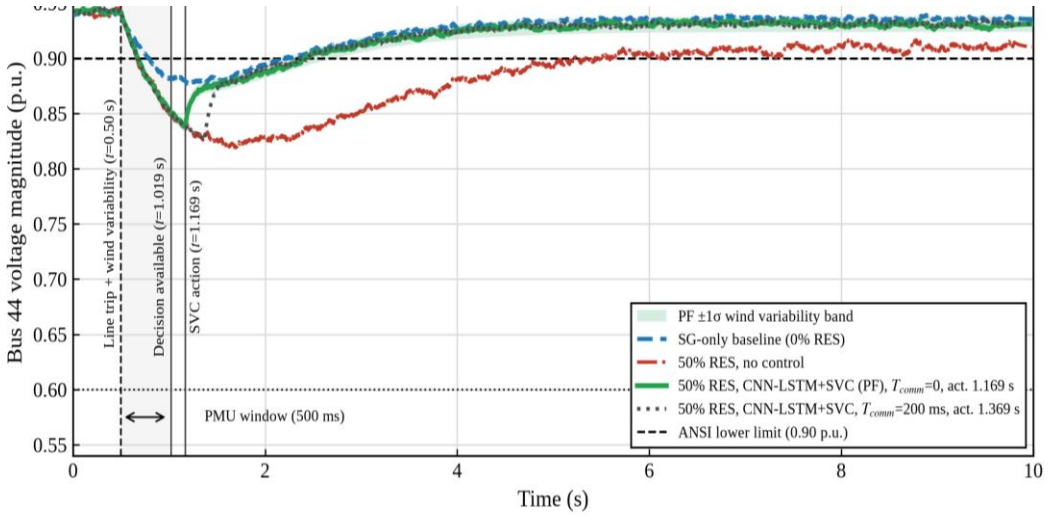


Figure 13. Weak-bus voltage response under 50% renewable-energy penetration during the line-trip disturbance, comparing the SG-only baseline, uncontrolled 50% RES condition, and CNN-LSTM + SVC controlled response with wind-power variability.

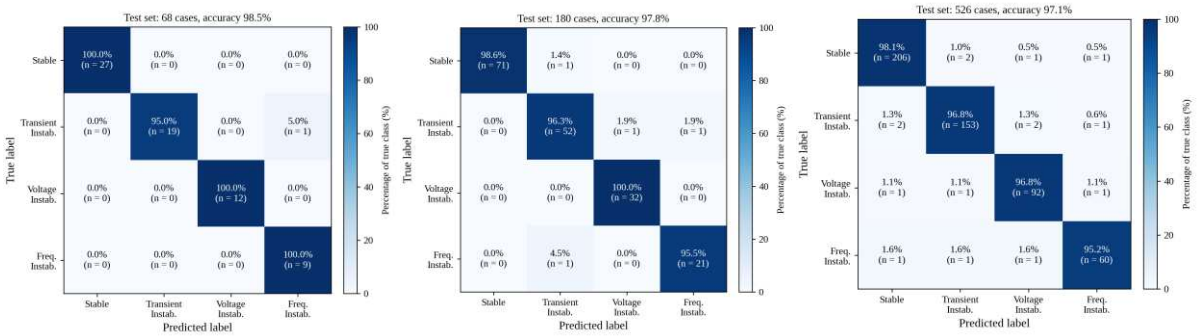
Despite these converter-driven fluctuations, the proposed CNN-LSTM-based control with SVC support, acting at t = 1.169 s, limited the critical Bus 44 voltage to a minimum of 0.838 p.u. and reduced the excursion below the 0.90 p.u. ANSI limit to 1.8 s, close to the 1.5 s of the synchronous-generation baseline, which reached 0.878 p.u. The uncontrolled 50% RES configuration exhibited a deeper and more sustained voltage decline, with a minimum of 0.823 p.u. and an excursion of 5.0 s. With a 200 ms communication delay the minimum was 0.828 p.u. with an unchanged 1.8 s excursion. Because the SVC acts close to the instant of minimum voltage, the improvement in the nadir is modest and the dominant effect of the corrective action is on the recovery. These results indicate that the proposed framework remains effective for voltage-stability support in the presence of renewable-generation variability.

5.6.2 Impact of Renewable Penetration on Classification Performance

The classification performance of the proposed CNN-LSTM framework was additionally evaluated under increasing renewable-energy penetration. The original 5,150 disturbance cases were augmented by incorporating the 30% and 50% RES configurations, resulting in a renewable-integrated dataset of 15,450 cases. The overall classification accuracy decreased gradually from 97.4% at 0% RES to 96.7% at 30% RES and 95.4% at 50% RES, with the corresponding macro F1-scores of 97.0%, 96.0%, and 94.7%, respectively. The reduction in performance at higher RES penetration is mainly associated with reduced system inertia and converter-driven voltage and frequency signatures that partially overlap with transient and frequency-instability characteristics. Nevertheless, the proposed framework maintains an overall accuracy above 95% at 50% RES penetration, demonstrating its robustness under renewable-integrated operating conditions.

5.7 Confusion Matrix and ROC Analysis

Detailed classification performance is evaluated by analysing the confusion matrices and ROC curves. The confusion matrices provide system-wise insight into the classification behaviour of the proposed CNN-LSTM model across the four stability categories for the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus benchmark systems. The corresponding results are presented in Figure 14.



(a) (b) (c)

Figure 14. Confusion matrix of the proposed CNN-LSTM model on the (a) IEEE 9-bus (b) IEEE 39-bus (c) IEEE 118-bus test sets

The diagonal entries indicate correct classification for each stability category. The proposed model achieves high per-class classification performance across all three benchmark systems, with overall accuracies of 98.5%, 97.8%, and 97.1% for the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus systems, respectively. The relatively higher number of off-diagonal entries for the IEEE 118-bus system reflects the increased network complexity and stronger interactions among buses and generators. Pooling the three test sets (774 cases, of which 754 are correctly classified) gives an overall test accuracy of 97.4%. Most of the 20 misclassifications involve the Stable class (9 cases), mainly marginal operating conditions in the IEEE 118-bus system. Among the instability classes, the most frequent confusion is between Transient Instability and Frequency Instability (5 cases), which is physically plausible because both conditions can exhibit similar voltage sag, frequency deviation, and dynamic signatures during the initial post-disturbance observation window. Nevertheless, the per-class recall is at least 95% for all three benchmark systems, demonstrating consistent classification capability of the proposed CNN-LSTM framework as shown in Figure 14(a)–(c). Class-wise precision, recall, and F1-scores for the pooled test set (774 cases) are reported in Table A5; the lowest F1-score (95.2%) is obtained for Frequency Instability and the highest (98.5%) for the Stable class.

The ROC curves as shown in Figure 15 are nearly optimal (upper left corner of the ROC space) and the AUC values are above 0.98 for all stability classes (Stable 0.998, Transient Instability 0.991, Voltage Instability 0.987, and Frequency Instability 0.994). The Stable class has the largest number of representative training samples and distinct measurement signatures, resulting in the highest AUC.

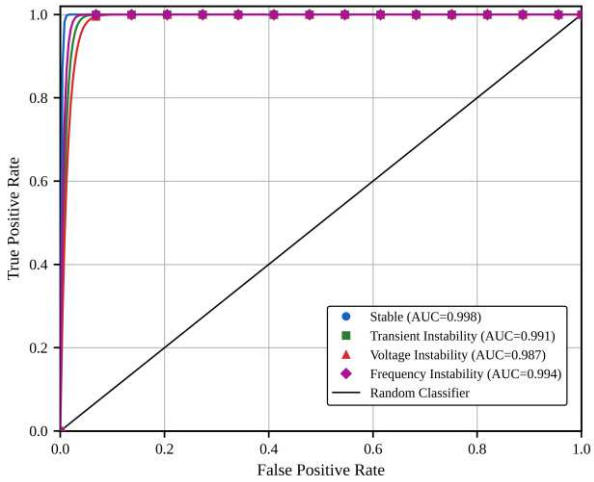


Figure 15. ROC curves and AUC values for the four stability classes (Stable, Transient Instability, Voltage Instability, Frequency Instability) on the combined test dataset.

The Voltage Instability class has a slightly lower AUC because early-stage voltage-collapse signatures resemble the transient voltage dips that occur during fault recovery; nevertheless, all four classes exhibit favourable sensitivity–specificity trade-offs.

5.8 Comparative Performance Benchmarking

A comprehensive comparative evaluation is conducted to assess the performance of the proposed CNN-LSTM framework against existing machine learning and deep learning approaches. All compared models were implemented and trained from scratch within this study rather than being quoted from the literature, so that every entry of Table 16 originates from a controlled experiment on the same data, the same split and the same computing platform. The key hyperparameters of the evaluated models, together with their accuracy, precision, recall, and F1-score, are summarized in Table 16. The comparison includes the RBF-kernel SVM, two-layer ANN, three-layer CNN, two-layer LSTM, three recent spatiotemporal and topology-aware architectures, and the proposed CNN-LSTM framework, enabling a consistent assessment of classification performance across different model architectures.

Table 16: Comparative Performance of Stability Classification Methods on Combined Dataset. Every figure in this table was produced in this study: each method was implemented, trained and evaluated by the authors on the same dataset, the same 70:15:15 stratified split and the same hardware, and no value is reproduced or quoted from the original publications. Precision, recall and F1-score are macro-averaged over the four stability classes for every method, so that the figures are directly comparable and are not dominated by the majority Stable class.

Method	Key Hyperparameters	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
SVM	RBF kernel, C=10, γ=0.01	87.4	86.5	86.2	86.3
ANN	2 hidden layers, 128/64 units, ReLU, dropout=0.2	89.8	88.9	88.3	88.6

CNN	3 Conv layers, 64/128/256 filters, kernel=3	92.3	92.1	91.4	91.7
LSTM	2 layers, 256/128 units, dropout=0.3, recurrent dropout=0.3	94.6	94.3	93.8	94.0
CNN-BiLSTM [20]	CNN 64/128/256 + BiLSTM 128 units, dropout=0.3	95.8	95.2	94.7	94.9
Spatial-Attention RNN [19]	Spatial attention + rectified RNN, 128 units, dropout=0.3	96.1	95.5	95.0	95.2
DGAT + Transfer Learning [21]	Graph-attention encoder, 8 heads, 64 units, fine-tuned	96.5	95.9	95.4	95.6
CNN-LSTM (Proposed)	CNN 64/128/256 + LSTM 256/128, Adam, LR=0.001	**97.4**	**96.8**	**97.1**	**97.0**

The proposed framework is quantitatively evaluated using accuracy, precision, recall, F1-score, and class-wise ROC-AUC for stability classification. In addition, stability-specific indicators including maximum inter-machine rotor-angle separation, frequency nadir, ROCOF, voltage minimum, and recovery time are used to evaluate the physical stability response under corrective control. These complementary metrics distinguish classification performance from the resulting physical stability response. For PMU-related measurement and synchronization considerations, the evaluation is interpreted in the context of IEEE synchrophasor measurement requirements, including time synchronization and measurement performance. The present study does not claim formal compliance certification with an IEEE or IEC standard, since such certification requires dedicated standardized measurement and conformance testing.

It can be seen in Table 16 that the proposed CNN-LSTM framework obtains the best result in terms of all four evaluation metrics. The comparison covers three groups of methods: classical machine learning (SVM, ANN), single-modality deep learning (CNN, LSTM), and three recent spatiotemporal or topology-aware architectures. The latter three were re-implemented by the authors following the architectural descriptions given by Singhal et al. [20] for the hybrid CNN-BiLSTM network, Zhu et al. [19] for the spatial-attention rectified RNN, and Zhang et al. [21] for the graph-attention network with transfer learning. Because the original studies use different test systems, disturbance sets and evaluation protocols, their published accuracy figures are not directly comparable with the present results; the three architectures were therefore retrained from scratch on the dataset of this study under conditions identical to those of the proposed model, using the same input tensors, the same 70:15:15 stratified split, the same preprocessing pipeline, the Adam optimizer with an initial learning rate of 0.001 and exponential decay, a batch size of 64, the categorical cross-entropy loss, early stopping with a patience of 20 epochs, and an L2 regularization coefficient of 10^{-4}, and were adapted only in their output layer to the four-class objective. The values reported for them in Table 16 are consequently the performance of these architectures on the present problem and should not be

read as the accuracy claimed in the corresponding publications. The traditional machine learning techniques, SVM and ANN, obtain 87.4% and 89.8% accuracy respectively, indicating that they have good prediction capabilities but are limited by the manual design of features and their limited nonlinear modelling capability.

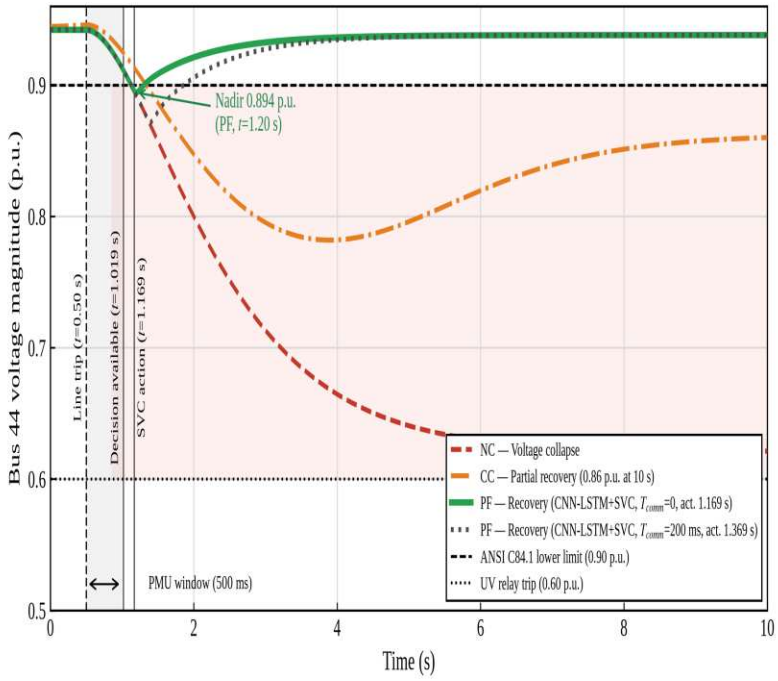


Figure 16. Comparative performance bar chart illustrating accuracy, precision, recall, and F1-score for all evaluated stability classification methods.

CNN is able to capture spatial inter-bus correlation automatically and boost accuracy to 92.3%, while LSTM performs temporal sequence learning and reaches 94.6% accuracy. The three recent spatiotemporal baselines close much of this gap, reaching 95.8%, 96.1% and 96.5% accuracy for CNN-BiLSTM [20], the spatial-attention rectified RNN [19] and the graph-attention network with transfer learning [21] respectively, which confirms that architectures combining spatial and temporal information outperform single-modality models on this task. The proposed CNN-LSTM fusion achieves 97.4% accuracy and a macro-averaged F1-score of 97.0%, an improvement of 10.0, 7.6, 5.1 and 2.8 percentage points in accuracy over the SVM, ANN, CNN and LSTM baselines, and of 1.6, 1.3 and 0.9 percentage points over [20], [19] and [21]. The margin over the recent baselines is modest and comparable in size to the inter-seed spread reported in Section 3.7, so the contribution of this work is not claimed to rest on classification accuracy alone. The distinguishing feature is scope: the compared methods return a decision for a single stability mechanism, whereas the proposed framework discriminates four mutually exclusive classes from the same PMU stream and uses the identified mechanism to select the corrective resource, as examined in Sections 5.9 and 5.14.2. The corresponding bar chart is shown in Figure 16.

Ablation analysis. The standalone CNN and LSTM models in Table 16 and Figure 7 act as an ablation of the proposed architecture: the CNN-only model retains the three convolutional blocks and the dense classification head without the LSTM layers, and the LSTM-only model applies the two LSTM layers directly to the flattened PMU vectors without the CNN module; both models were trained under conditions identical to those of the proposed CNN-LSTM, i.e., with the same input tensors, the same 70:15:15 stratified

split, the same preprocessing pipeline, the Adam optimizer with an initial learning rate of 0.001 and exponential decay, a batch size of 64, the categorical cross-entropy loss, early stopping with a patience of 20 epochs, and an L2 regularization coefficient of 10^{-4}, so that the comparison isolates the contribution of the spatial and temporal feature-extraction modules. Removing the temporal (LSTM) module reduces the accuracy from 97.4% to 92.3% (−5.1 percentage points), and removing the spatial (CNN) module reduces it to 94.6% (−2.8 percentage points), confirming that both components contribute and that temporal modelling provides the larger share.

5.9 Comparison with Existing Literature

The proposed framework is placed in the context of existing published methods for power system stability assessment based on deep learning in Table 17, which shows the test power systems, stability categories considered, and also includes a breakdown of whether closed loop corrective control is employed or not. Table 17 differs in nature from Table 16: it summarizes the scope and capability reported in the original publications, whereas Table 16 reports measurements obtained in this study from models retrained on a common dataset. Absolute percentage-point differences between the figures reported across studies are not meaningful due to differences in the test systems, fault scenarios and hardware platforms that are used; the comparison below is therefore limited to capability and scope, which can be directly and verifiably compared, while the quantitative comparison under controlled conditions is given in Table 16.

Table 17: Comparative Summary Against Existing Stability Assessment Literature

Reference	Method	Test System	Stability Type	Corrective Control
Shi et al. [17]	CNN	IEEE 39-Bus, WECC 179-Bus	Transient	No
Zhang et al. [18]	LSTM	IEEE 39-Bus	Voltage	No
Zhu et al. [19]	Spatial-Attention RNN	IEEE 39-Bus	Voltage	No
Zhang et al. [21]	GNN (DGAT) + Transfer Learning	Multi-scenario (cross-topology)	Transient	No
Arévalo-Cordero et al. [12]	Hybrid LSTM + FACTS control	EV-penetrated microgrid	Voltage, Frequency	Yes
Shahriyari et al. [25]	LSTM (multi-domain boundary prediction)	IEEE 14-bus, IEEE 118-bus	Multi-domain	No
Proposed	**CNN-LSTM + FACTS**	**9, 39, 118-Bus**	**All Types**	**Yes**

Table 17 indicates the following important features of the proposed solution with respect to scope and capability. Shi et al. [17] and Zhu et al. [19] each address a single stability mechanism, transient and short-term voltage stability respectively. Zhu et al. [19] do combine spatial and temporal information, through a spatial-attention rectified recurrent architecture, so the distinction from the present work is not the use of spatiotemporal features but the scope of the classification: their model returns a binary stability decision for one mechanism, whereas the proposed framework discriminates four mutually exclusive classes from the same PMU stream and uses the identified mechanism to select the corrective resource. Similarly, Zhang et al. [18] provide a representative example of voltage-stability prediction using LSTM, and Zhang et al. [21] show that graph-attention-based transfer learning enhances cross-scenario generalization for transient-stability prediction, again without considering voltage and frequency instability together. The proposed

framework achieves a mean model inference latency of 17.3 ms on the tested CPU platform; the end-to-end decision time additionally includes the 500 ms observation window and the device response time (Section 5.10 and Table A3). Importantly, none of the compared studies combines the classification of all three instability mechanisms within a single framework with mechanism-specific corrective control; the closest work [12] couples LSTM prediction with FACTS control but addresses only voltage and frequency stability in microgrids.

Comparison with Conventional Stability Assessment: In addition to the machine-learning baselines, conventional power-system stability assessment methods provide an important physics-based reference. Modal/eigenvalue analysis and time-domain simulation are effective for identifying system stability margins and dominant dynamic modes; however, their computational requirements and dependence on operating-point linearization or detailed system models make direct frame-by-frame comparison with the proposed PMU-based four-class classifier non-equivalent. The proposed framework instead performs direct classification from synchronized PMU measurements without requiring online eigenvalue computation or explicit disturbance-specific feature engineering. The ML benchmark in Table 16 therefore provides a controlled quantitative comparison under the same PMU-based classification task, while the discussion of conventional analytical methods establishes the complementary physics-based reference. VSG-based monitoring is not treated as a separate classifier in the present study because the proposed framework is designed as a grid-wide PMU-based stability monitor rather than a converter-control-specific monitoring scheme.

5.10 Real-Time Inference Performance

Real-time operational suitability requires inference latency well below the WAMS communication cycle time of 40–100 ms. Single-sample inference on a server-grade Intel Xeon Gold 6248R CPU platform yields an average latency of 17.3 ms, which is below the 20 ms design target. However, the 95th-percentile latency (19.8 ms) leaves little margin, and adding the 1.9 ms preprocessing stage (Table 19) brings the tail processing time to approximately 21.7 ms, slightly above this target on the CPU platform. GPU inference (4.2 ms mean, 5.5 ms 95th percentile) provides a substantially larger margin. Batch inference of 64 simultaneous stability assessments on an NVIDIA T4 GPU completes in 8.1 ms total, corresponding to approximately 0.13 ms per sample and enabling multiple concurrent assessments in large-scale WAMS deployments. These figures describe processing latency only; the end-to-end decision time measured from disturbance onset additionally includes the 500 ms observation window and is therefore approximately 519 ms for the IEEE 39-bus model (500 ms + 1.9 ms + 17.3 ms), excluding communication delay. Within this constraint, the processing stage of the proposed CNN-LSTM framework is compatible with real-time operation within existing WAMS communication infrastructure. The complete latency budget from disturbance onset to FACTS actuation is given in Table A3. The 17.3 ms latency in Table 18 corresponds to single-sample inference of the IEEE 39-bus model (input tensor 50 × 39 × 5, approximately 2.95 million parameters; Table A2) on the Intel Xeon Gold 6248R CPU; the same model is used in the sensitivity analyses of Section 5.12. The corresponding CPU latencies of the IEEE 9-bus and 118-bus models are approximately 15.2 ms and 38 ms, respectively (the 118-bus model has an input tensor of 50 × 118 × 5 and approximately 8.2 million parameters); for the 118-bus model, the mean CPU latency therefore exceeds the 20 ms design target and the end-to-end decision time increases to approximately 540 ms, so GPU inference is recommended for large networks. Training was performed in TensorFlow 2.12 on an NVIDIA T4 GPU (16 GB). The IEEE 9-, 39-, and 118-bus models required approximately 22 min, 48 min, and 3.2 h of wall-

clock time, respectively, including validation at each epoch and checkpoint saving; for the IEEE 39-bus model, this corresponds to the 200-epoch training run shown in Figure 4, in which convergence is reached at approximately epoch 150.

5.10.1 Latency Measurement Procedure

The reported inference latency was measured using 1,000 independent inference runs, with 50 warm-up runs excluded from the analysis. Single-sample inference was evaluated using a batch size of 1, while batch-size 64 inference was additionally considered for multi-sample WAMS assessment. The mean latency, standard deviation, and 95th-percentile latency for the different inference configurations are reported in Table 18.

Table 18. Computational latency of the proposed CNN-LSTM model under different inference configurations.

Inference Mode	Batch Size	Mean Latency	Std. Dev.	95th Percentile
CPU	1	17.3 ms	1.2 ms	19.8 ms
GPU	1	4.2 ms	0.8 ms	5.5 ms
GPU	64	8.1 ms	0.6 ms	9.2 ms
GPU per sample	64	0.13 ms	—	—

5.10.2 Preprocessing Time

The preprocessing stage includes PMU data loading, Min-Max normalization, fourth-order Butterworth low-pass filtering with a 10 Hz cutoff, forward filling of missing measurements, and tensor reshaping. Forward filling is used here because only past samples are available in online operation; the linear interpolation described in Section 3.2 is applied only during offline dataset preparation (Section 5.13.3). The measured total preprocessing time is 1.9 ms per sample, while the combined preprocessing and CPU inference pipeline requires approximately 19.2 ms per sample, as summarized in Table 19.

Table 19. Processing time of individual preprocessing stages and the complete inference pipeline.

Processing Stage	Time
Data Loading	0.5 ms
Min-Max Normalization	0.3 ms
Butterworth Filtering	0.8 ms
Gap filling (forward fill)	0.2 ms
Tensor Reshaping	0.1 ms
Total Preprocessing	1.9 ms
Preprocessing + CPU Inference	19.2 ms

5.10.3 Computational Complexity and Real-Time Suitability

The computational burden of the proposed framework is primarily associated with the convolutional feature extraction and stacked LSTM sequence processing, while inference after model training does not require iterative optimization or online parameter estimation. The measured inference latency is 17.3 ms for the selected 500 ms observation window, compared with 11.2, 23.8, and 30.5 ms for 250, 750, and 1000 ms windows, respectively. This indicates that the proposed model can perform rapid online classification on

the tested CPU/GPU platforms. In addition, the robustness experiments demonstrate that communication delay mainly increases the overall response time rather than fundamentally degrading the classifier, supporting the feasibility of the framework for real-time WAMS applications. However, deployment on substantially larger networks with very high PMU counts would require distributed or edge-based inference and further hardware-specific profiling.

5.11 Time-Domain Voltage Stability Validation

The voltage-stability performance of the proposed framework is further validated using a severe voltage-instability scenario on the IEEE 118-bus system. The system is operated at 94% of the voltage-stability nose-point loading, and a double-circuit transmission-line outage between Bus 38 and Bus 65 is introduced at t = 0.50 s. Bus 44 is identified as the electrically weakest load bus, with a pre-disturbance voltage of 0.942 p.u. The response is evaluated under no-control (NC), conventional reactive-compensation control (CC), and the proposed CNN-LSTM-assisted corrective-control (PF) conditions.

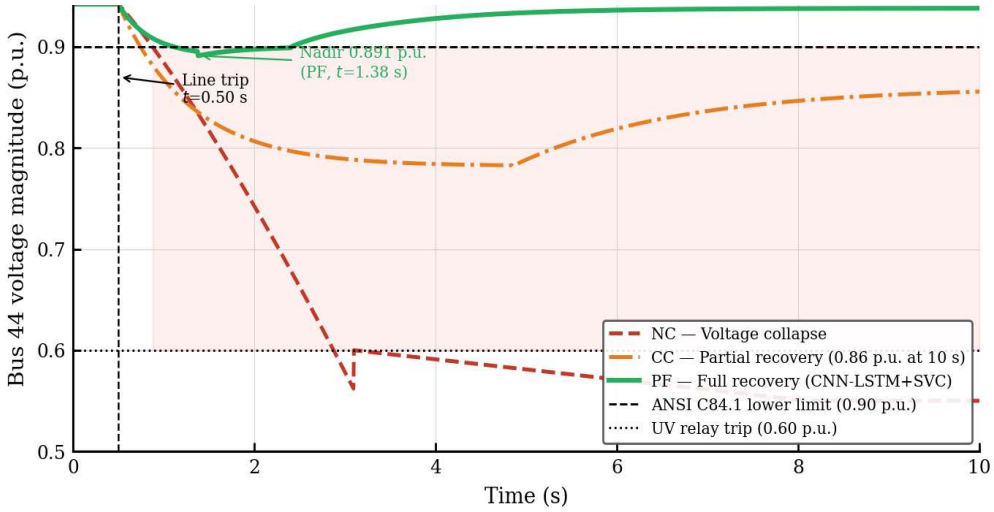


Figure 17. Weak-bus voltage trajectory at Bus 44 for the IEEE 118-bus voltage-instability case under no-control, conventional reactive-compensation, and proposed CNN-LSTM-assisted SVC control. The 0.90 p.u. ANSI C84.1 lower limit and 0.60 p.u. under-voltage relay threshold are shown as reference levels.

As shown in Figure 17, the uncontrolled case exhibits severe voltage deterioration, with the Bus 44 voltage falling to 0.621 p.u. and progressing toward voltage collapse. The conventional reactive-compensation strategy limits the voltage nadir to only 0.782 p.u. and recovers to just 0.860 p.u. at 10 s, without returning above the 0.90 p.u. ANSI C84.1 lower limit within the simulation horizon. In contrast, the proposed CNN-LSTM-assisted corrective-control framework detects the voltage-instability condition from the 500 ms observation window, issues the decision at t = 1.019 s and, allowing the 150 ms SVC response time, activates coordinated SVC support at t = 1.169 s. The voltage nadir is limited to 0.894 p.u. at t = 1.20 s, the voltage recovers above 0.90 p.u. at t = 1.37 s, and the final voltage reaches 0.938 p.u. at 10 s. The voltage therefore falls below the 0.90 p.u. ANSI C84.1 lower limit for only 0.18 s, whereas the NC and CC cases exhibit sustained violations exceeding 8.5 s. With a 200 ms communication delay the action is deferred to t = 1.369 s, the nadir deepens to 0.871 p.u., and the excursion below 0.90 p.u. extends to 0.73 s, which remains within the ANSI short-duration allowance. The voltage mechanism is therefore the most tolerant

of the three to the measurement-window delay, because the post-contingency decline at Bus 44 develops over several seconds. Because the 10 s horizon exceeds the 5 s labelling horizon of Section 4.2.2, the 10 s values are used only for control validation.

Table 20. Time-Domain Voltage Stability Performance at IEEE 118-Bus Bus 44

Metric	NC	CC	Proposed PF
Pre-disturbance voltage (p.u.)	0.942	0.942	0.942
Minimum voltage (p.u.)	0.621	0.782	0.894
Recovery above 0.90 p.u.	Not reached	Not reached within 10 s	t = 1.37 s
Time below 0.90 p.u.	>8.5 s	>8.5 s	0.18 s
Voltage at 10 s	Collapse	0.860 p.u.	0.938 p.u.

Table 20 shows the time-domain voltage response at the critical Bus 44 under different control conditions. Without control (NC), the voltage deteriorates severely to 0.621 p.u. and progresses toward collapse. Conventional control (CC) improves the voltage response but remains inadequate, with a minimum voltage of 0.782 p.u. and no recovery above 0.90 p.u. within the 10 s horizon, ending at 0.860 p.u. In contrast, the proposed framework limits the voltage drop to 0.894 p.u., restores the voltage above 0.90 p.u. at t = 1.37 s, limits the ANSI voltage violation to 0.18 s, and maintains the voltage at 0.938 p.u. at 10 s.

5.12 Sensitivity Analysis of PMU Reporting Rate and Observation Window

To assess the robustness of the proposed CNN-LSTM framework under different PMU data acquisition conditions, a sensitivity analysis was performed by varying the PMU reporting rate and the observation window length. The analysis evaluates the resulting changes in classification performance and inference latency, thereby providing a quantitative basis for selecting the PMU reporting rate and observation window used in the proposed framework. Throughout this subsection, and in every sensitivity, robustness and generalization table of this paper, precision, recall and F1-score are macro-averaged over the four stability classes, following the convention stated in Section 3.7. Macro-averaged recall is therefore not constrained to equal the accuracy, unlike the weighted average.

5.12.1 Sensitivity to PMU Reporting Rate

To evaluate the sensitivity of the proposed CNN-LSTM framework to the PMU reporting rate, experiments were conducted using the IEEE 39-bus system with a fixed observation window of 500 ms. The PMU data were represented at reporting rates of 100, 50, 25, and 10 fps. For each reporting-rate configuration, the CNN-LSTM model was independently trained and evaluated using the same data partitioning and evaluation protocol.

The sensitivity-analysis results reported in this subsection are obtained independently on the IEEE 39-bus test system and therefore are not directly identical to the overall accuracy of 97.4% reported for the complete three-system dataset comprising the IEEE 9-bus, 39-bus, and 118-bus systems. The 97.8% accuracy at 100

fps represents the baseline performance for the IEEE 39-bus sensitivity experiment under the 500 ms observation window.

Table 21. Sensitivity of the Proposed CNN-LSTM Model to Different PMU Reporting Rates on the IEEE 39-Bus System

PMU Reporting Rate	Approx. Samples / 500 ms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Latency (ms)
100 fps	50	97.8	97.2	97.6	97.4	17.3
50 fps	25	97.1	96.7	96.2	96.5	10.8
25 fps	12-13	95.8	95.2	94.7	95.0	6.9
10 fps	5	91.2	90.5	89.8	90.1	3.5

Table 21 indicates that the proposed model maintains high classification performance when the reporting rate is reduced from 100 fps to 25 fps. However, a further reduction to 10 fps results in a noticeable degradation in classification performance. Although lower reporting rates reduce inference latency, they also provide fewer temporal samples within the observation window, thereby reducing the temporal information available to the LSTM component.

5.12.2 Sensitivity to Observation Window Length

The sensitivity of the proposed framework to the observation-window length was evaluated using the IEEE 39-bus system with a fixed PMU reporting rate of 100 fps. Four observation-window lengths of 250, 500, 750, and 1000 ms were considered. For each configuration, the CNN-LSTM model was independently retrained and tested using the same data partitioning and evaluation protocol.

Table 22. Sensitivity of the Proposed CNN-LSTM Model to Different Observation Window Lengths

Observation Window	Samples at 100 fps	Accuracy (%)	Inference Latency (ms)
250 ms	25	94.3	11.2
500 ms	50	97.8	17.3
750 ms	75	98.2	23.8
1000 ms	100	98.4	30.5

The 500 ms observation window as shown in Table 22 provides a favorable balance between classification performance and inference latency. Reducing the window to 250 ms decreases inference latency but also reduces classification accuracy, indicating that the shorter temporal context provides less information for learning the evolution of the disturbance. Increasing the observation window beyond 500 ms produces only marginal improvements in accuracy while increasing inference latency. Therefore, the 500 ms window was selected as the preferred configuration for early stability monitoring.

5.13 Robustness Evaluation under Realistic WAMS Conditions

To evaluate the practical applicability of the proposed framework under realistic WAMS deployment constraints, five robustness tests were conducted on the IEEE 39-bus system. The experiments investigate reduced PMU coverage, measurement noise, data packet loss, communication delay, and synchronization errors. The baseline configuration corresponds to full PMU coverage, noise-free measurements, zero packet

loss, zero communication delay, and ideal time synchronization. For the reduced PMU coverage test (Section 5.13.1), a separate model was trained for each coverage level (90%, 80%, 70%, 60%, and 50%), because removing PMUs changes the number of monitored buses and hence the input tensor dimension. For all other robustness tests, namely measurement noise (Section 5.13.2), packet loss (Section 5.13.3), communication delay (Section 5.13.4), and synchronization errors (Section 5.13.5), the baseline model trained under ideal conditions was used without retraining. These tests therefore evaluate the robustness of the deployed baseline model to degraded data quality rather than the ability of a retrained model to adapt to each degraded condition. As in Section 5.12, precision, recall and F1-score in the robustness tables are macro-averaged over the four stability classes (Section 3.7).

5.13.1 Robustness to Reduced PMU Coverage

To evaluate the effect of partial observability, PMUs were randomly removed from 10%, 20%, 30%, 40%, and 50% of the buses in the IEEE 39-bus system. Measurements from buses without PMUs were removed from the corresponding input representation. The CNN-LSTM model was retrained and evaluated separately for each PMU coverage condition.

Table 23. Robustness of the Proposed CNN-LSTM Model under Reduced PMU Coverage

PMU Coverage	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
100% (Baseline)	97.8	97.2	97.6	97.4
90%	97.2	96.8	96.3	96.5
80%	96.1	95.5	95.0	95.2
70%	94.3	93.7	93.1	93.4
60%	91.8	91.0	90.3	90.6
50%	88.5	87.6	86.8	87.2

The results in Table 23 show that the proposed model maintains an accuracy above 94% when PMU coverage is at least 70%. At 50% PMU coverage, the accuracy decreases to 88.5%, indicating the expected degradation under severe partial observability. Nevertheless, the model retains meaningful classification capability even when half of the buses are not directly monitored.

5.13.2 Robustness to Measurement Noise

To investigate the sensitivity of the proposed framework to PMU measurement noise, zero-mean Gaussian noise was introduced into the PMU measurement channels at noise levels of 0.5%, 1%, 2%, and 5% relative to the signal magnitude. The same preprocessing procedure described in Section 3.2 was applied before model evaluation.

Table 24. Robustness of the Proposed CNN-LSTM Model to Measurement Noise

Noise Level (σ)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0% (No Noise)	97.8	97.2	97.6	97.4
0.5%	97.5	97.0	96.6	96.8
1%	96.8	96.3	95.8	96.0
2%	95.2	94.6	94.0	94.3
5%	90.7	89.8	89.0	89.4

In Table 24, at a 1% noise level, the classification accuracy decreases only slightly from 97.8% to 96.8%. Even at 2% noise, the proposed model maintains an accuracy above 95%. At the severe 5% noise level, the accuracy decreases to 90.7%, demonstrating graceful degradation rather than abrupt classification failure.

5.13.3 Robustness to Missing Data and Packet Loss

To assess the effect of communication-related data loss, random measurement frames were removed from the PMU data streams at packet-loss rates of 5%, 10%, 15%, 20%, and 30%. The missing measurements were forward-filled before constructing the model input sequences. Forward filling is used here because only past measurements are available in online operation; the linear interpolation described in Section 3.2 is applied only during offline dataset preparation.

Table 25. Robustness of the Proposed CNN-LSTM Model to PMU Packet Loss

Packet Loss Rate	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0% (Baseline)	97.8	97.2	97.6	97.4
5%	97.4	96.9	96.5	96.7
10%	96.8	96.3	95.8	96.0
15%	96.0	95.4	94.9	95.1
20%	95.1	94.5	93.8	94.1
30%	92.3	91.4	90.6	91.0

The proposed framework as shown in Table 25 maintains an accuracy of 95.1% even at 20% packet loss. At 30% packet loss, the accuracy decreases to 92.3%, indicating that the model remains operational under substantial random data loss, although performance progressively deteriorates as the loss rate increases.

5.13.4 Robustness to Communication Delay

To quantify the impact of communication latency, fixed delays of 50 ms, 100 ms, 200 ms, and 500 ms were introduced into the PMU data streams. The classification performance and total response time were then evaluated against the zero-delay baseline. The response times in Table 26 comprise the communication delay and the model inference latency and exclude the 500 ms observation window.

Table 26. Robustness of the Proposed CNN-LSTM Model to Communication Delay

Communication Delay	Accuracy (%)	Inference Latency (ms)	Delay + Inference Time (ms)
0 ms (Baseline)	97.8	17.3	17.3
50 ms	97.7	17.3	67.3
100 ms	97.4	17.3	117.3
200 ms	96.8	17.3	217.3
500 ms	95.2	17.3	517.3

The results shown in Table 26 indicate that communication delay has a relatively limited effect on classification accuracy, decreasing from 97.8% under the baseline condition to 95.2% at 500 ms delay.

However, the total response time increases substantially with communication latency, demonstrating that communication delay is more critical for real-time control deployment than for the classification capability itself.

5.13.5 Robustness to Synchronization Errors

To evaluate the sensitivity of the proposed framework to imperfect PMU time synchronization, random timing offsets of 10 μs, 50 μs, 100 μs, 500 μs, and 1 ms were introduced between selected PMU measurement streams. The resulting classification performance was evaluated relative to the ideally synchronized baseline.

Table 27. Robustness of the Proposed CNN-LSTM Model to PMU Synchronization Errors

Synchronization Error	Phase Error at 50 Hz	Accuracy (%)
0 μs	0°	97.8
10 μs	0.18°	97.7
50 μs	0.9°	97.3
100 μs	1.8°	96.5
500 μs	9.0°	93.8
1 ms	18°	88.2

The proposed model exhibits limited degradation for synchronization errors up to 100 μs, with classification accuracy remaining at 96.5%. However, larger timing offsets produce progressively greater degradation, with accuracy decreasing to 93.8% at 500 μs and 88.2% at 1 ms. These results as shown in Table 27 indicate that accurate time synchronization remains important for reliable wide-area stability assessment.

Overall, the robustness evaluation demonstrates that the proposed CNN-LSTM framework maintains reliable classification performance under moderate reductions in PMU coverage, measurement noise, packet loss, communication delay, and synchronization error. The results also show that severe observability loss and large synchronization errors have a more pronounced effect on classification performance, whereas communication delay primarily increases the overall response time. These findings provide quantitative evidence regarding the operating limits of the proposed framework under realistic WAMS data-quality conditions.

5.13.6 Deployment and Cybersecurity Considerations

Practical deployment of PMU-based stability monitoring must account for communication latency, packet loss, missing measurements, synchronization errors, and data-integrity threats. The robustness experiments presented in this section quantitatively evaluate the effects of reduced PMU coverage, measurement noise, packet loss, communication delay, and synchronization errors. In addition to these communication and data-quality issues, PMU streams may be exposed to cyber threats such as false-data injection, data manipulation, spoofing, and denial-of-service attacks. Such attacks may alter the measurements supplied to the classifier and can therefore lead to incorrect stability decisions or inappropriate corrective actions. A practical deployment should consequently incorporate secure communication, authentication and access control, data-integrity validation, anomaly detection, and fail-safe control logic before classifier outputs are used for automatic FACTS actuation. Cyber-attack scenarios are not explicitly modeled in the present study and are identified as an important direction for future hardware-in-the-loop and cyber-physical validation.

5.14 Robustness Stress Tests Under Adverse Operating Conditions

This section examines how the proposed CNN-LSTM + FACTS closed-loop framework performs when operating conditions deviate from the ideal inference scenario. Five stress categories are tested: detection delay, misclassification, false alarms (false positives), missed detections (false negatives), and FACTS saturation.

5.14.1 Communication-Delay Sensitivity

The decision instant of the proposed CNN-LSTM-assisted corrective-control framework is fixed at 519.2 ms after disturbance onset by the 500 ms observation window, the 1.9 ms preprocessing stage and the 17.3 ms inference, so the remaining variable in the control chain is the communication delay between the classifier and the FACTS controller. Its effect on the rotor-angle stability of generator G10 is evaluated using Case T, i.e. the IEEE 39-bus three-phase fault at Bus 16 described in Section 5.2, while maintaining the same disturbance and control configuration. The communication delay is varied from 0 to 500 ms for both the STATCOM and the SVC path, and the corresponding peak rotor-angle deviation and stability outcome are evaluated against the 180 degree transient-stability criterion. The results presented in Table 28 and Figure 18 show how the increasing communication delay progressively reduces the available stability margin and eventually leads to loss of synchronism.

Table 28: Effect of Communication Delay on Case T Outcomes (IEEE 39-Bus, G10 Transient Instability)

Communication delay T_comm (ms)	STATCOM action instant (s)	G10 peak δ (°)	G10 outcome
0	0.649	75.0	Stable
50	0.699	82.0	Stable
100	0.749	89.4	Stable
150	0.799	97.3	Stable
200	0.849	106.0	Stable
300	0.949	128.7	Stable
350	0.999	—	Loss of synchronism

The results in Table 28 demonstrate that the proposed framework maintains a positive transient-stability margin for communication delays up to approximately 300 ms. The peak rotor angle increases monotonically with the delay, from 75.0 degrees at zero delay to 128.7 degrees at 300 ms, because the corrective action is pushed further into the first swing. Synchronism is lost for delays of 350 ms and above. Linear interpolation between the last stable and the first unstable point places the critical communication delay at approximately 334 ms for the STATCOM path, corresponding to a corrective action at t = 0.983 s, and at approximately 299 ms for the SVC path, which is less tolerant because of its longer 150 ms response time.

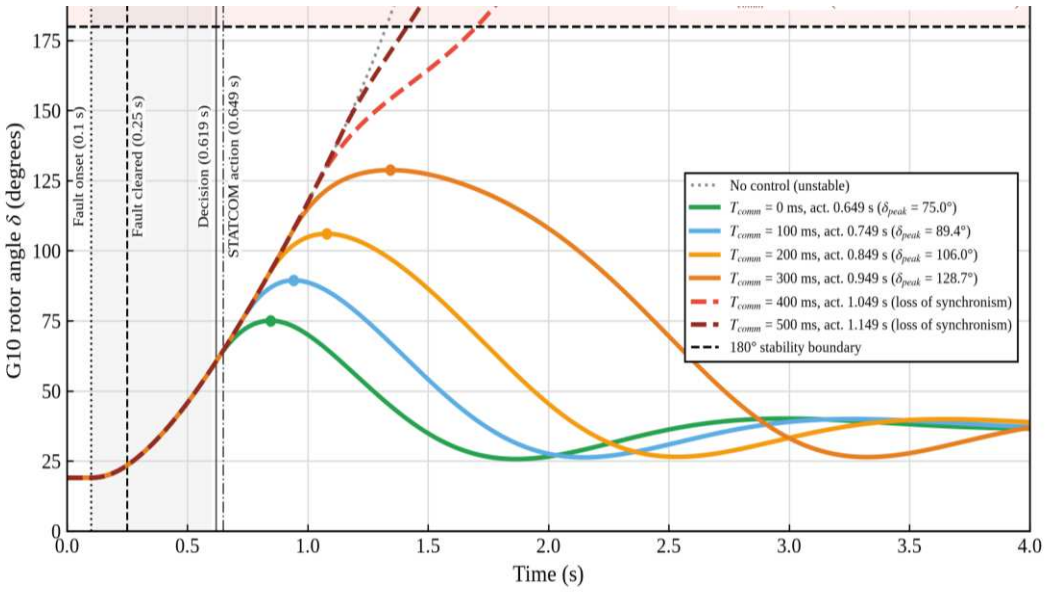


Figure 18. G10 rotor-angle trajectory δ_G10(t) for communication-delay values of 0, 100, 200, 300, 400 and 500 ms, with the classifier decision fixed at t = 0.619 s and STATCOM action at t = 0.649 s + T_comm. The 180 degree instability boundary is shown as a dashed line; the trajectories deteriorate progressively with increasing delay, and the 400 and 500 ms cases breach the boundary. The critical communication delay is approximately 334 ms, corresponding to a corrective action at t = 0.983 s.

As illustrated in Figure 18, the rotor-angle trajectories progressively deteriorate with increasing communication delay, confirming that timely delivery of the corrective-control command is critical for maintaining the transient-stability margin. Under zero communication delay the corrective action occurs at t = 0.649 s, before the first-swing peak, and the peak rotor angle is limited to 75.0 degrees. The usable communication-delay budget of approximately 334 ms should be compared with the 50 to 500 ms range examined in Section 5.13.4, where classification accuracy was shown to remain at 95.2% even at 500 ms delay. The classifier therefore tolerates communication delay far better than the closed loop does, and the two requirements must be specified separately in a practical deployment.

5.14.2 Misclassification Scenarios

The misclassification sensitivity analysis evaluates the effect of an incorrect stability-class decision on the subsequent corrective-control response. Two representative misclassification scenarios are considered to examine whether an incorrect CNN-LSTM decision can adversely affect system stability: transient instability incorrectly classified as voltage instability (T → V) and frequency instability incorrectly classified as transient instability (F → T), where Case F denotes the severe IEEE 39-bus generation-deficit case of Table 15. In the first case, SVC-based reactive-power support is activated instead of the intended STATCOM response, whereas in the second case, STATCOM support is activated instead of the coordinated active-power response involving the BESS units and the turbine-governor/AGC response, with SVC voltage support. The resulting system responses are compared with the corresponding no-control conditions to quantify the degradation in transient-angle and frequency-stability performance. The results are summarized in Table 29.

Table 29: Misclassification Impact Analysis

Case	Error type	Outcome with wrong corrective action	Outcome with no corrective action (NC)
T	Transient → Voltage	G10 peak δ = 79.8° (stable; vs. 75.0° with STATCOM); SVC acts at 0.769 s instead of 0.649 s	G10 loses synchronism (δ > 180°)
F	Frequency → Transient	Frequency nadir = 48.86 Hz (above UFLS) vs. 49.31 Hz; recovery = 6.2 s vs. 4.1 s	UFLS activation at t = 0.81 s; 2 stages shed

As shown in Table 29, the impact of misclassification depends on the type of instability and the corrective action that is incorrectly activated. For the T to V case, the incorrect SVC-based reactive-power support still maintains generator G10 in synchronism. Because the SVC response time of 150 ms is longer than the 30 ms of the STATCOM, the corrective action is deferred from t = 0.649 s to t = 0.769 s and the peak rotor angle rises from 75.0 to 79.8 degrees, which leaves a margin of 100 degrees below the 180 degree criterion. Without corrective control, G10 loses synchronism at t = 1.33 s. For the F to T case, the incorrect STATCOM-based response withholds the battery active-power support, so the frequency nadir and the recovery time revert toward the uncontrolled values reported in Table 15. These results indicate that classification errors degrade corrective-control performance but, for the evaluated cases, do not lead to system instability when the incorrect control action still provides relevant voltage or reactive-power support.

5.14.3 False Alarms (False Positives)

The false-alarm sensitivity analysis evaluates the response of the proposed closed-loop corrective-control framework when a stable operating condition is incorrectly identified as an instability. Such false-positive decisions may unnecessarily activate FACTS devices and introduce transient reactive-power injections even though no actual disturbance requiring corrective action is present. To assess the practical impact of this behaviour, 50 artificial false alarms were introduced under stable, no-disturbance conditions in the IEEE 39-bus system by forcing the CNN-LSTM classification output above the 0.85 decision threshold at randomly selected times. The resulting voltage deviation, FACTS reactive-power injection, frequency deviation, and protection response were then evaluated, with the results summarized in Table 30.

Table 30: False Alarm Impact Statistics (50 Events, IEEE 39-Bus Stable Conditions)

Metric	Mean	Worst case
Peak voltage deviation from pre-alarm value (p.u.)	+0.0124	+0.0291
Duration of voltage deviation above pre-alarm value (s)	0.31	0.74
Maximum FACTS reactive injection triggered (MVAR)	48.3	112.4
Frequency deviation induced (Hz)	0.007	0.018
UFLS or protection relay activation	None in 50 events	—
Estimated FACTS device duty-cycle consumption per false alarm (%)	0.4%	1.1%

As shown in Table 30, the false alarms produce only small and short-duration deviations from the pre-alarm operating condition. The mean peak voltage deviation is +0.0124 p.u., while the worst-case deviation is +0.0291 p.u. The maximum FACTS reactive-power injection reaches 112.4 MVAR in the worst case, but no UFLS or protection-relay activation is observed in any of the 50 test events. The induced frequency deviation also remains very small, with a mean value of 0.007 Hz and a worst-case value of 0.018 Hz. These results indicate that although false-positive decisions may cause unnecessary FACTS operation, their impact on the evaluated stable operating conditions remains limited and transient.

5.14.4 Missed Detections (False Negatives)

The missed-detection sensitivity analysis evaluates the consequences of false-negative decisions, in which an actual instability is not identified with sufficient confidence to activate the corrective-control module. Since a missed detection may leave the system temporarily without the intended corrective action, the post-disturbance trajectories of the missed cases were re-examined to determine whether the absence of corrective control leads to a critical stability violation. The analysis considers transient, voltage, and frequency instability cases separately, using the missed cases identified in the test set (Figure 14) and evaluating the severity of the corresponding uncorrected system responses. The resulting observations are summarized for each instability class below.

For the missed cases in each instability class, the post-fault trajectories were re-examined to categorise the severity of the un-corrected outcome:

- For transient-instability cases, 8 missed detections were identified, consistent with the 96.6% recall of the Transient Instability class over 232 test cases reported in Table A5. Six of these cases had fault-clearing times below the critical clearing time for the considered generators and therefore remained stable even without corrective action. The remaining two cases marginally exceeded the 180 degree transient-stability criterion by approximately 4 to 9 degrees, for which conventional out-of-step protection would provide a backup response.
- For voltage-instability cases, all 3 missed detections, consistent with the 97.8% recall of the Voltage Instability class over 139 test cases in Table A5, involved relatively shallow initial voltage depressions of less than 0.05 p.u. below nominal. Without FACTS intervention, these voltages recovered naturally within approximately 3 to 5 s and did not reach the ANSI C84.1 under-voltage violation threshold.
- For frequency-instability cases, all 4 missed detections, consistent with the 95.7% recall of the Frequency Instability class over 94 test cases in Table A5, had frequency nadirs above 49.2 Hz without corrective intervention, and therefore no UFLS activation occurred. Although the average recovery time increased to approximately 6.8 s, the frequency response ultimately recovered successfully. The 15 missed detections across the three instability classes, together with the 5 Stable cases assigned to an instability class, account for the 20 misclassifications of the pooled test set reported in Section 5.7.

Overall, the missed-detection cases are concentrated mainly around near-stable operating conditions, where the consequences of not activating the corrective-control mechanism remain limited. The results therefore indicate that the false-negative cases identified in the evaluated test set do not generally result in immediate catastrophic instability, although the increased recovery time and the marginal transient-stability cases highlight the importance of maintaining adequate detection sensitivity.

5.14.5 FACTS Device Saturation Under Severe Combined Disturbances

The FACTS saturation analysis evaluates the behaviour of the proposed corrective-control framework when the reactive-power requirement exceeds the available rating of a FACTS device under a severe combined disturbance. Such operating conditions represent a more demanding scenario than the individual disturbances considered during the primary validation. For this analysis, the IEEE 39-bus three-phase fault of Case T is combined with a simultaneous 15% load increase at Bus 8, thereby creating a compound disturbance that simultaneously stresses transient and frequency stability and increases the reactive-power support requirement. The resulting FACTS power requirements, saturation behaviour, and corresponding rotor-angle response are evaluated to determine whether the proposed framework can maintain transient stability despite control saturation. The results are presented in Table 31 and Figure 19.

Table 31: FACTS Saturation Behaviour — Combined Case T + Load Increase, PF Scenario

Device	Rating (MVAR)	Required Q (MVAR)	Actual Q_inj (MVAR)	Consequence of saturation
STATCOM at Bus 16	±150	168	150 (capped)	G10 δ_peak = 79.2° (stable, 101° margin; 77.4° if unlimited)
STATCOM at Bus 19	±100	88.1	88.1 (within rating)	No saturation
SVC at Bus 8	±300	251.3	251.3 (within rating)	No saturation

As shown in Table 31, the STATCOM installed at Bus 16 reaches its +/-150 MVAR rating limit, whereas the required reactive-power support increases to 168 MVAR. Consequently, the actual reactive-power injection is capped at 150 MVAR and the device remains saturated from t = 0.640 s to t = 1.00 s, a window of 360 ms covering the first swing. Despite this saturation, generator G10 remains stable, with the peak rotor angle increasing only to 79.2 degrees compared with 75.0 degrees for the corresponding unsaturated Case T, leaving a margin of 101 degrees below the 180 degree criterion. Repeating the combined case with the rating constraint removed gives a peak of 77.4 degrees, so the saturation costs only 1.8 degrees of margin. Without corrective action the combined disturbance drives G10 across the 180 degree boundary at t = 1.27 s. The STATCOM at Bus 19 and the SVC at Bus 8 remain within their respective operating limits.

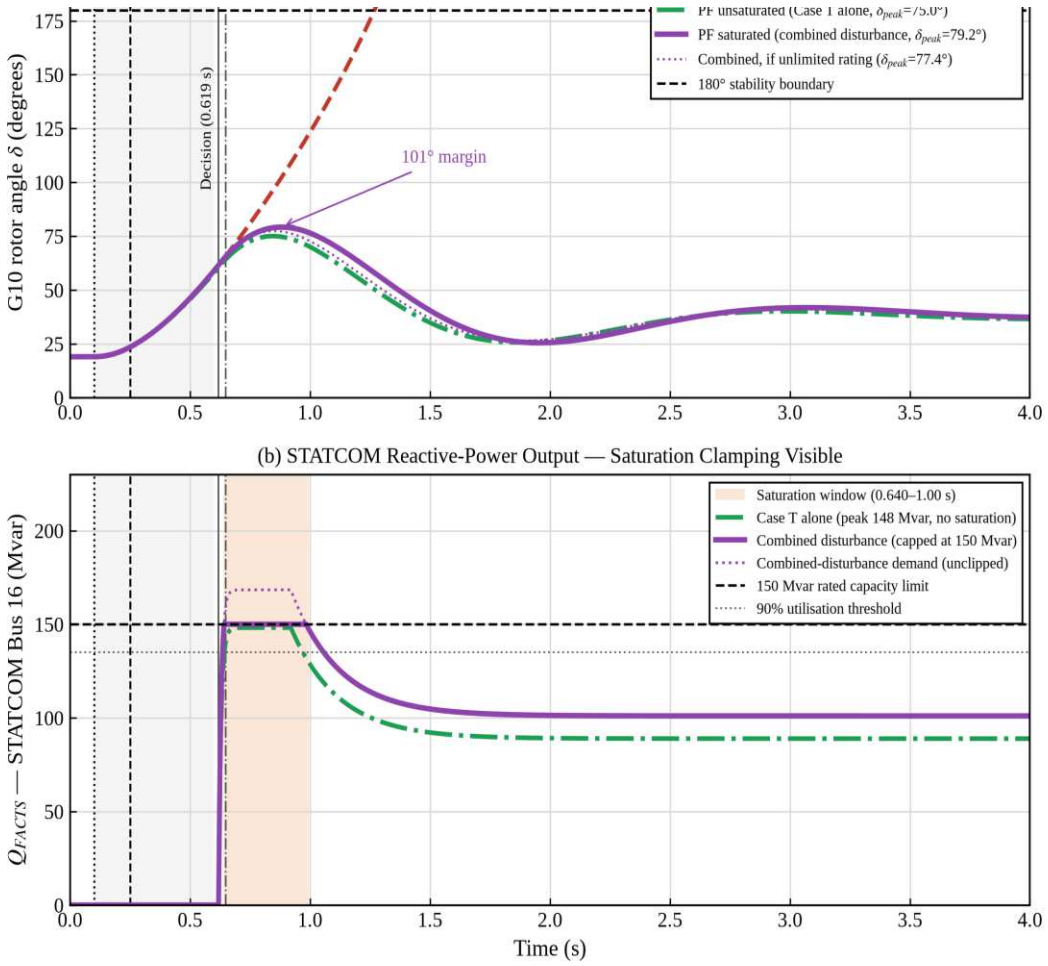


Figure 19. G10 rotor-angle trajectory and Bus 16 STATCOM reactive-power output for the combined Case T + Load Increase scenario under NC, PF-unsaturated (Case T alone), and PF-saturated (combined) conditions, with corrective action at t = 0.649 s. The 150 MVAR STATCOM rating limit, the 0.640-1.00 s saturation window and the 180 degree instability boundary are shown as reference lines; the saturated PF case remains stable with a 101 degree margin.

Figure 19 further illustrates the effect of saturation on both the rotor-angle trajectory and the STATCOM output. Although the Bus 16 STATCOM operates at its maximum capacitive limit of 150 MVAR between t = 0.640 s and t = 1.00 s, the available support remains sufficient to keep the maximum rotor-angle response below the 180 degree instability boundary, providing a stability margin of approximately 101 degrees. The unclipped demand reaches 168 MVAR over the same window, and the difference between the saturated and the hypothetical unlimited response is only 1.8 degrees of rotor angle. The control output subsequently decreases toward approximately 100 MVAR as the bus voltage recovers, demonstrating that the corrective-control mechanism can transition from emergency saturated operation back to normal voltage-regulation operation.

5.15 Scenario-Level Generalization Performance

To assess whether the reported classification performance is affected by sample-level data partitioning, the proposed CNN-LSTM model was further evaluated using scenario-level validation protocols. The

evaluation included leave-one-contingency-out (LOCO), leave-one-operating-condition-out (LOOCO), and unseen fault-location (UFL) testing. In each protocol, the corresponding held-out contingency, operating condition, or fault-location group was completely excluded from model training. The results are summarized in Table 32. Precision, recall and F1-score in Table 32 are macro-averaged over the four stability classes, as elsewhere in this paper (Section 3.7).

Table 32. Summary of Generalization Performance under Scenario-Level Validation Protocols

Validation Protocol	Training Data	Test Condition	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Stratified Split (Baseline)	70% of all data	Same distribution (15% holdout)	97.4	96.8	97.1	97.0
LOCO-1	All except Bus 16 fault	Three-phase fault at Bus 16	95.8	95.2	94.7	95.0
LOCO-2	All except Bus 29 fault	Three-phase fault at Bus 29	96.2	95.7	95.1	95.4
LOCO-3	All except line 4–14 outage	Line outage	97.1	96.6	96.1	96.4
LOCO-4	All except G8 outage	Generator G8 outage	96.8	96.3	95.8	96.0
LOCO-5	All except Bus 8 load increase	Load increase at Bus 8	95.5	94.9	94.3	94.6
LOCO Average	—	—	96.3	95.7	95.2	95.5
LOOCO-1	Normal + heavy load	Light load (80% base)	97.2	96.7	96.2	96.5
LOOCO-2	Normal + light load	Heavy load (120% base)	96.8	96.3	95.8	96.0
LOOCO-3	Low + medium wind output	High wind output	96.5	96.0	95.4	95.7
LOOCO-4	Low + medium solar output	High solar output	96.7	96.2	95.6	95.9
LOOCO-5	0–15% RES scenarios	30% RES	95.3	94.7	94.0	94.3
LOOCO Average	—	—	96.5	96.0	95.4	95.7
UFL-1	Faults at Buses 1–20	Faults at Buses 21–39	94.8	94.2	93.5	93.8
UFL-2	Faults at Buses 21–39	Faults at Buses 1–20	95.2	94.7	94.0	94.3
UFL-3	Transmission line faults	Load-bus faults	96.1	95.6	95.0	95.3
UFL-4	High-inertia area faults	Low-inertia area faults	95.8	95.2	94.6	94.9

UFL Average	—	—	95.5	94.9	94.3	94.6
Overall Rigorous Average	—	—	96.1	95.5	95.0	95.3

Table 32 shows that the proposed CNN-LSTM model maintains strong performance under scenario-level separation. The conventional random stratified split achieved 97.4% accuracy, whereas the average accuracies under LOCO, LOOCO, and UFL validation were 96.3%, 96.5%, and 95.5%, respectively. The overall rigorous-validation average was 96.1%, corresponding to a reduction of 1.3 percentage points from the random-split baseline. The reduction is relatively modest, indicating that the model retains substantial generalization capability when related contingency, operating-condition, and fault-location samples are excluded from training.

The UFL evaluation produced the largest reduction in performance, indicating that spatial generalization to previously unseen fault locations is more challenging than generalization across the evaluated operating conditions and contingencies. Nevertheless, the rigorous-validation accuracy remained above the standalone CNN and LSTM baseline accuracies reported earlier in the manuscript. These results provide additional evidence that the performance of the proposed framework is not solely dependent on random sample-level partitioning. Together with the renewable-penetration study (Section 5.6.2) and the fault-location and operating-condition splits in Table 32, these results address generalization to unseen renewable levels, fault conditions, and operating scenarios. For LOOCO-5, an additional set of IEEE 39-bus scenarios with 0–15% RES penetration was generated separately for this test and is not part of the 5,150-case dataset; the model trained on these scenarios was tested on the 30% RES configuration. For UFL-3, the model was trained on transmission-line faults and tested on faults applied at load buses.

6. Conclusion

This paper presented a spatiotemporal deep learning framework based on a hybrid CNN-LSTM architecture for real-time multi-class power-system stability assessment using synchronized PMU data streams. The proposed framework classifies four stability conditions, namely Stable, Transient Instability, Voltage Instability, and Frequency Instability, by learning spatial correlations among monitored buses and temporal characteristics of post-disturbance system dynamics. The framework was evaluated using the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus benchmark systems with 5,150 baseline labelled operating and disturbance cases. The proposed model achieved an overall classification accuracy of 97.4%, with an average model inference latency of 17.3 ms per 500 ms observation window, while sensitivity analyses demonstrated the influence of PMU reporting rate, observation-window length, measurement noise, reduced PMU coverage, packet loss, communication delay, and synchronization errors on classification performance.

The proposed framework further integrates stability detection with a closed-loop corrective-control mechanism. For transient-instability conditions, STATCOM-based reactive-power support is employed to provide fast voltage support and improve post-disturbance damping, whereas SVC-based compensation is used for voltage-instability conditions. The FACTS installation locations were selected offline by engineering judgement, close to the principal disturbance locations of the case studies, and were then fixed for all simulations of the corresponding benchmark system; the base-case V-Q sensitivity of each selected

bus is reported for reference in Table A1, and a systematic sensitivity-based placement procedure is left for future work. The online control mechanism adaptively regulates the reactive-power output of the installed devices within their operating limits. For frequency-instability conditions, the corrective-control framework does not rely on FACTS devices as the primary source of active-power balancing. Instead, active-power support is provided through the battery energy storage units and the modeled turbine-governor and AGC mechanisms, with FACTS devices providing supplementary reactive-power and voltage support.

Time-domain validation under realistic actuation timing establishes a mechanism-dependent latency limit for measurement-window-based corrective control. For the IEEE 39-bus transient-stability scenario, the stability decision at 0.619 s and the STATCOM action at 0.649 s precede the first-swing peak at approximately 0.85 s, limiting the critical generator to a peak rotor angle of 75.0 degrees against loss of synchronism at 1.33 s without control, and synchronism is retained for communication delays up to approximately 334 ms. For the IEEE 118-bus voltage-stability scenario, SVC action at 1.169 s limits the weak-bus voltage to 0.894 p.u. and the excursion below the 0.90 p.u. ANSI limit to 0.18 s, against collapse to 0.621 p.u. without control. For frequency instability following the 650 MW trip of generator G6, the same decision delay places the battery dispatch at 1.019 s, after the rate-of-change-of-frequency peak: the maximum ROCOF is consequently unchanged at 1.48 Hz/s across all control conditions, the frequency nadir improves only from 48.79 Hz to 48.89 Hz, and under-frequency load-shedding stage-1 pickup occurs in every case, while the recovery time to return above 49.8 Hz falls from 7.5 s to 4.3 s. Wide-area measurement-based control is therefore effective for rotor-angle and voltage mechanisms, which evolve over the first one to two seconds, and cannot act on the inertial phase of a frequency event, for which local measurement-triggered fast frequency response remains necessary.

The applicability of the proposed framework to renewable-integrated power systems was additionally examined by incorporating converter-interfaced Type-4 PMSG wind generation and grid-following solar-PV inverter models at different renewable-energy penetration levels. The renewable-integrated configurations considered 0%, 30%, and 50% RES penetration, with corresponding reductions in system inertia. Under the severe generation-deficit disturbance, increasing renewable penetration resulted in more challenging uncontrolled frequency responses, whereas the proposed coordinated control maintained the frequency nadir close to the acceptable operating range and significantly reduced ROCOF. The controlled responses remained free from UFLS activation in the evaluated renewable-integrated cases. The classifier also maintained an accuracy above 95% at 50% RES penetration, demonstrating that the proposed framework retains effective stability-classification capability under converter-dominated operating conditions.

The renewable-integrated validation also considered wind and solar-generation variability and the associated converter-driven voltage and frequency dynamics. Fast active-power support through the modeled BESS/virtual-inertia mechanism was incorporated for severe frequency disturbances, while turbine-governor and AGC responses provided subsequent active-power recovery and SVC/STATCOM devices supplied supplementary reactive-power support. These results indicate that combining early PMU-based stability detection with coordinated active- and reactive-power corrective mechanisms can improve the resilience of renewable-integrated transmission systems.

Despite these encouraging results, several limitations remain. The validation is primarily based on standard benchmark systems and simulated operating conditions and therefore does not fully capture the

uncertainties, topology changes, communication characteristics, converter-control interactions, and measurement conditions encountered in operational transmission networks. Although reduced PMU coverage, measurement noise, packet loss, communication delay, and synchronization errors were investigated through sensitivity and robustness studies, real-world PMU deployment may involve additional observability and data-quality constraints. Furthermore, the CNN-LSTM architecture represents PMU measurements in a Euclidean feature space and does not explicitly encode the underlying electrical network topology. Further practical challenges should also be noted. The 500 ms observation window delays the decision to approximately 0.52 s after disturbance onset, which limits the framework for the fastest transient events. All results are based on simulated data from benchmark systems rebased to 50 Hz, and validation with field PMU recordings is still required. The FACTS locations were selected by engineering judgement and are fixed offline; a systematic sensitivity-based placement, including its repetition after topology changes or major reconfiguration together with retraining of the model, is left for future work. Performance also depends on PMU coverage, data quality, and time synchronization, as quantified in Section 5.13, and a deployed system must protect the WAMS data path against cyberattacks (Section 5.13.6).

The present implementation also does not consider OPF-based generation redispatch, demand response, or coordinated multi-resource energy management. Future work will therefore focus on topology-aware graph neural network (GNN) architectures, online and transfer-learning mechanisms, and coordinated multi-agent corrective-control strategies incorporating OPF-based redispatch, demand response, larger and coordinated BESS resources, and other fast active-power support mechanisms. Further validation using operational PMU archives, hardware-in-the-loop experiments, and field-oriented studies will be pursued to evaluate the robustness, scalability, interoperability, and practical deployment capability of the proposed framework.

Appendix A. Supplementary Tables

Tables A1–A5 provide supplementary details on FACTS placement, layer dimensions, end-to-end latency, dataset composition, and class-wise classification performance.

Table A1. Base-case V–Q sensitivity of the selected FACTS locations, reported for reference (standard IEEE network data; rank among load buses, 1 = weakest).

System	Device (bus)	dV/dQ (p.u./p.u.)	Base-case rank
IEEE 9-bus	SVC (Bus 9)	0.0949	1 of 6
IEEE 9-bus	STATCOM (Bus 6)	0.0423	6 of 6
IEEE 39-bus	SVC (Bus 8)	0.0146	8 of 29
IEEE 39-bus	STATCOM (Bus 16)	0.0087	26 of 29
IEEE 39-bus	STATCOM (Bus 19)	0.0080	28 of 29
IEEE 118-bus	SVC (Bus 44)	0.1269	2 of 64
IEEE 118-bus	STATCOM (Bus 95)	0.0460	27 of 64

Table A2. Layer-wise output shapes and parameter counts of the CNN-LSTM models (one model per system, "same" padding, kernel 3, pooling 2).

Layer	Output (9-bus)	Output (39-bus)	Output (118-bus)	Parameters

Input	50 × 9 × 5	50 × 39 × 5	50 × 118 × 5	0
Conv1D-64 + BN	50 × 9 × 64	50 × 39 × 64	50 × 118 × 64	1,024 + 256
Max-pooling (2)	50 × 4 × 64	50 × 19 × 64	50 × 59 × 64	0
Conv1D-128 + BN	50 × 4 × 128	50 × 19 × 128	50 × 59 × 128	24,704 + 512
Max-pooling (2)	50 × 2 × 128	50 × 9 × 128	50 × 29 × 128	0
Conv1D-256 + BN + spatial dropout	50 × 2 × 256	50 × 9 × 256	50 × 29 × 256	98,560 + 1,024
Flatten	50 × 512	50 × 2,304	50 × 7,424	0
LSTM-1 (256)	50 × 256	50 × 256	50 × 256	787,456 / 2,622,464 / 7,865,344
Layer normalization	50 × 256	50 × 256	50 × 256	512
LSTM-2 (128)	128	128	128	197,120
Dense (64, ReLU)	64	64	64	8,256
Softmax (4)	4	4	4	260
Total parameters	1,119,684	2,954,692	8,197,572	per system

Table A3. End-to-end latency budget from disturbance onset to corrective action.

Stage	Duration (ms)	Source
PMU observation window	500	Section 3.2
Preprocessing	1.9	Table 19
CNN-LSTM inference (CPU, mean, 39-bus model)	17.3	Table 18
Decision available after onset	519.2	Sum of the above
Communication delay	T_comm (0–500 examined)	Table 26
STATCOM response	< 30	Table 1
SVC response	≈ 150	Table 1
Total to STATCOM action	≈ 549 + T_comm	Sum
Total to SVC action	≈ 669 + T_comm	Sum

Table A4. Composition of the dataset by disturbance type and benchmark system.

Disturbance type	IEEE 9-bus	IEEE 39-bus	IEEE 118-bus	Total
Three-phase faults	135	360	1,050	1,545
Asymmetric faults (SLG, LL)	90	240	700	1,030
Load increases	65	180	525	770
Generator outages	70	180	525	775
Line/tie-line outages	90	240	700	1,030
Total	450	1,200	3,500	5,150

Table A5. Class-wise performance of the proposed CNN-LSTM on the pooled test set (774 cases).

Class	Precision (%)	Recall (%)	F1-score (%)	Support
Stable	98.7	98.4	98.5	309
Transient Instability	97.4	96.6	97.0	232
Voltage Instability	96.5	97.8	97.1	139
Frequency Instability	94.7	95.7	95.2	94
Macro average	96.8	97.1	97.0	774
Overall accuracy	–	–	–	754/774 (97.4%)

Declarations

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies involving human participants or animals performed by any of the authors.

Informed Consent: Not applicable.

Data and Code Availability Statement: The labeled PMU dataset used in this study, comprising 5,150 samples across the IEEE 9-bus, IEEE 39-bus, and IEEE 118-bus systems, is available from the corresponding author upon reasonable request. The source code for the proposed CNN-LSTM framework, including model architecture, training scripts, evaluation utilities, preprocessing procedures, and visualization tools, is also available from the corresponding author upon reasonable request. The authors intend to release the dataset and source code through a permanent open repository upon publication, subject to institutional and publisher requirements.

Consent for Publication: All authors have read and approved the final manuscript and consent to its publication.

Clinical Trial Number: Not applicable.

References

1. Zhu, L., & Hill, D. J. (2018). Stability analysis of power systems: A network synchronization perspective. SIAM Journal on Control and Optimization, 56(3), 1640-1664.
2. Mathews, J. A. (2013). The renewable energies technology surge: A new techno-economic paradigm in the making?. Futures, 46, 10-22.
3. Milano, F., Dörfler, F., Hug, G., Hill, D. J., & Verbič, G. (2018, June). Foundations and challenges of low-inertia systems. In 2018 power systems computation conference (PSCC) (pp. 1-25). IEEE.

4. Bai, F., Zhu, L., Liu, Y., Wang, X., Sun, K., Ma, Y., ... & Bhatt, N. (2016). Design and implementation of a measurement-based adaptive wide-area damping controller considering time delays. Electric Power Systems Research, 130, 1-9.
5. Terzija, V., Valverde, G., Cai, D., Regulski, P., Madani, V., Fitch, J., Phadke, A. (2011). Wide-area monitoring, protection, and control of future electric power networks. Proceedings of the IEEE, 99(1), 80-93.
6. Aziz, T., Lin, Z., Waseem, M., & Liu, S. (2021). Review on optimization methodologies in transmission network reconfiguration of power systems for grid resilience. International Transactions on Electrical Energy Systems, 31(3), e12704.
7. Biswal, C., Sahu, B. K., Mishra, M., & Rout, P. K. (2023). Real-time grid monitoring and protection: A comprehensive survey on the advantages of phasor measurement units. Energies, 16(10), 4054.
8. Aminifar, F., Fotuhi-Firuzabad, M., Safdarian, A., Davoudi, A., & Shahidehpour, M. (2014). Synchrophasor measurement technology in power systems: Panorama and state-of-the-art. IEEE Access, 2, 1607-1628.
9. Rane, N. L., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). Machine learning and deep learning for big data analytics: A review of methods and applications. Partners Universal International Innovation Journal, 2(3), 172-197.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
11. Krichen, M., & Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. AI, 6(9), 215.
12. Arévalo-Cordero, P., González, F., Martínez, A., Zarie, D., Rodas, A., Albornoz, E., ... & Benavides, D. (2025). Hybrid LSTM–FACTS Control Strategy for Voltage and Frequency Stability in EV-Penetrated Microgrids. Technologies, 13(9), 402.
13. Kim, J., Lee, H., Kim, S., Chung, S. H., & Park, J. H. (2023). Transient stability assessment using deep transfer learning. IEEE Access, 11, 116622-116637.
14. Sobajic, D. J., & Pao, Y. H. (1989). Artificial neural-net based dynamic security assessment for electric power systems. IEEE Transactions on power Systems, 4(1), 220-228.
15. Moulin, L. S., Da Silva, A. A., El-Sharkawi, M. A., & Marks, R. J. (2004). Support vector machines for transient stability analysis of large-scale power systems. IEEE Transactions on Power systems, 19(2), 818-825.
16. Ren C, Xu Y, Zhang Y, Zhang R (2019) A hybrid randomized learning system for temporal-adaptive voltage stability assessment of power systems. IEEE Transactions on Industrial Informatics 16(6):3672–3684.
17. Shi, Z., Yao, W., Zeng, L., Wen, J., Fang, J., Ai, X., & Wen, J. (2020). Convolutional neural network-based power system transient stability assessment and instability mode prediction. Applied Energy, 263, 114586.
18. Zhang M, Li J, Li Y, Xu R (2021) Deep learning for short-term voltage stability assessment of power systems. IEEE Access 9:29711–29718.
19. Zhu L, Hill DJ, Lu C (2021) Intelligent short-term voltage stability assessment via spatial attention rectified RNN learning. IEEE Transactions on Industrial Informatics 17(10):7005-7016.
20. Singhal D, Ahuja L, Seth A (2024) Forecasting smart grid stability using a hybrid CNN Bi-LSTM approach. SN Computer Science 5(5):539.

21. Zhang R, Yao W, Shi Z, Ai X, Tang Y, Wen J (2022) Towards multi-scenario power system stability analysis: an unsupervised transfer learning method combining DGAT and data augmentation. IEEE Transactions on Power Systems 38(6):5367–5380.
22. Sun M, Feng C, Zhang J (2019) Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation. Applied Energy 256:113842.
23. Kundur P, Paserba J, Ajjarapu V, Andersson G, Bose A, Canizares C, Hatziargyriou N, Hill D, Stankovic A, Taylor C, Van Cutsem T, Vittal V (2004) Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions. IEEE Transactions on Power Systems 19(3):1387–1401.
24. Toscani, S., Muscas, C., & Pegoraro, P. A. (2016). Design and performance prediction of space vector-based PMU algorithms. IEEE Transactions on Instrumentation and Measurement, 66(3), 394-404.
25. Shahriyari, M., Safari, A., Quteishat, A., & Afsharirad, H. (2026). An LSTM architecture for real-time multi-domain stability boundary prediction beyond post-fault dependency in power systems. Scientific Reports, 16(1), 6330.
26. Wu, Y., Gan, P., Jing, Z., & Yang, Q. (2025). Hybrid CNN-BiLSTM-Attention Conditional GAN for power system forced oscillation localization. International Journal of Electrical Power & Energy Systems, 172, 111220.
27. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press, Cambridge, MA.
28. Hingorani, N. G., & Gyugyi, L. (2000). Understanding FACTS: Concepts and technology of flexible AC transmission systems. IEEE Press, New York.
29. Amoateng, D. O., Yan, R., Mosadeghy, M., & Saha, T. K. (2021). Topology detection in power distribution networks: A PMU based deep learning approach. IEEE Transactions on Power Systems, 37(4), 2771-2782.
30. Huang, C., Li, F., Zhou, D., Guo, J., Pan, Z., Liu, Y., & Liu, Y. (2016). Data quality issues for synchrophasor applications Part I: A review. Journal of Modern Power Systems and Clean Energy, 4(3), 342–352.
31. IEC/IEEE (2018) IEC/IEEE 60255-118-1:2018, Measuring relays and protection equipment—Part 118-1: Synchrophasor for power systems—Measurements. International Electrotechnical Commission and Institute of Electrical and Electronics Engineers
32. Ige, A. O., & Sibiya, M. (2024). State-of-the-art in 1d convolutional neural networks: A survey. IEEE Access, 12, 144082-144105.
33. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.