

Article

A Multi-Layer Spatio-Temporal Learning and Optimization Framework for Line-Loss-Oriented Distribution Networks

Guangwei Zu ¹, Ben Wang ¹, Jing Meng ¹, Xinghua Dong ¹ and Tao Hong ^{2,*} 

¹ Development Planning Department, State Grid Jibei Electric Power Co., Ltd. Qinhuangdao Power Supply Company, Qinhuangdao 066000, China; zzweiii@sgcc.com.cn (G.Z.); wang20171201@sgcc.com.cn (B.W.); jingmengjingmeng@sgcc.com.cn (J.M.); dongxinhua@sgcc.com.cn (X.D.)

² School of Electronics and Information Engineering, Beihang University, Beijing 100191, China

* Correspondence: hongtao@buaa.edu.cn

Abstract

High DG/BESS penetration reshapes power-flow patterns in distribution networks, amplifying line-loss (LL) volatility and stressing conventional planning and operation. We present ML-STELLO (Multi-Layer Spatio-Temporal nEtworK Learning and Line-loss Optimization), a multi-layer framework that unifies: (i) data governance for 96-point feeder curves via RODDPSO-enhanced FCM imputation and coefficient-of-variation improved Isolation Forest (CV-I Forest); (ii) multi-scale spatial-temporal (ST) learning using CNN-LSTM with attention for LL estimation; (iii) operation-time control coupling BESS dispatch and local VAR; and (iv) hierarchical planning and control across five time-scales—from multi-year investment to 5–15 min MPC—extending two-layer models driven by WGAN-GP scenarios and solved with IWOA. On IEEE-33 and a provincial feeder, ML-STELLO reduces LL and voltage violations relative to two-layer baselines while retaining robustness to missing/noisy data. Our design distills and extends advances in LL analysis, ST modeling, and uncertainty-aware planning.

Keywords: line loss (LL); distribution network; spatiotemporal learning; WGAN-GP (WGAN with gradient penalty); IWOA (improved whale optimization algorithm); BESS (battery energy storage system); VAR (volt-ampere reactive); hierarchical optimization

1. Introduction

Line loss (LL) is a primary indicator of efficiency and economy in modern power systems, directly reflecting both technical performance and financial sustainability [1]. Distribution networks (DNs) are particularly sensitive to LL due to their radial topology, limited redundancy, and highly unbalanced load characteristics. With increasing electrification and the growth of distributed energy resources (DERs), the complexity of loss analysis has intensified. Inaccurate estimation or insufficient optimization of LL not only results in economic inefficiencies but also exacerbates operational risks such as voltage violations and equipment overloading. Hence, advanced methods for LL analysis and reduction have become central to the modernization of distribution systems.

Recent technological transformations, particularly the rapid penetration of distributed generation (DG) and battery energy storage systems (BESS), have significantly altered network topologies and real-time power flow distributions [2]. While DG integration reduces carbon emissions and increases local autonomy, it introduces stochastic fluctuations and bi-directional power flows, complicating loss estimation. Conversely, coordinated BESS



Academic Editor: José Matas

Received: 12 December 2025

Revised: 12 February 2026

Accepted: 23 February 2026

Published: 31 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

dispatch can mitigate voltage instability, reduce peak–valley imbalances, and lower LL [3]. However, realizing such benefits requires precise forecasting and optimization models capable of capturing both spatial correlations among feeders and temporal dynamics of renewable intermittency.

A persistent challenge in LL analysis arises from data imperfections in distribution networks. Smart meters and feeder monitoring systems frequently encounter missing data, communication delays, and measurement noise. Traditional imputation or anomaly detection methods are often inadequate under large-scale renewable variability. Recent studies have advanced this field by combining PSO-enhanced fuzzy c-means (FCM) clustering for curve imputation with coefficient-of-variation improved Isolation Forest (CV-iForest) for anomaly detection [4]. Such approaches achieve robust performance, reporting LL rates near 7.62% with anomaly detection accuracy (AUC) of approximately 0.8586 under complex operating conditions. These results underscore the importance of robust data governance pipelines as a prerequisite for reliable LL analysis [5,6].

Beyond data quality, advanced learning models have been proposed to exploit the multi-scale structure of distribution network dynamics. Convolutional neural networks (CNNs) excel in extracting spatial dependencies across feeders and branches, while long short-term memory (LSTM) networks capture temporal correlations of load and renewable fluctuations [7,8]. The integration of attention mechanisms further enhances fusion across scales, enabling models to outperform traditional statistical methods and direct power-flow calculations in both accuracy and adaptability. This spatio-temporal (ST) learning paradigm establishes a foundation for embedding predictive intelligence into LL estimation and control.

On the planning horizon, uncertainty in renewable generation poses challenges to classical optimization models. Probabilistic and scenario-based approaches have gained prominence, particularly generative adversarial network (GAN) methods such as Wasserstein GAN with gradient penalty (WGAN-GP), which effectively capture heavy-tailed distributions and extreme events. The gradient penalty term enforces Lipschitz continuity of the discriminator, preventing gradient vanishing and mode collapse issues that plague traditional GANs, thereby ensuring stable training and generating more realistic renewable generation scenarios. Coupled with a two-layer optimization model that co-optimizes DG siting/sizing and BESS scheduling, and solved via improved whale optimization algorithm (IWOA), these methods have demonstrated strong performance on benchmark systems such as IEEE-33. However, while two-layer frameworks capture planning-operation interactions, they remain limited in addressing cross-scale couplings and risk allocations across multiple decision layers [9,10].

To overcome these limitations, we propose ML-STELLO (Multi-Layer Spatio-Temporal nEtwork Learning and Line-loss Optimization), a systematic integration framework that coordinates established techniques within a novel five-layer hierarchical architecture for distribution network LL management. ML-STELLO is neither a single algorithm nor a disparate collection of methods, but rather a comprehensive framework where “STELLO” represents the following: Spatio-Temporal learning capturing spatial network correlations and temporal renewable dynamics; nEtwork Learning employing CNN-LSTM-Attention and WGAN-GP for prediction and scenario generation; Line loss as the primary optimization objective; and Optimization, denoting the multi-layer hierarchical decision structure. The framework generalizes conventional two-layer models into five time-consistent layers—(i) multi-year strategic investment, (ii) annual/seasonal policy coordination, (iii) weekly/daily network reconfiguration, (iv) 24 h day-ahead scheduling, and (v) 5–15 min real-time MPC control—with explicit inter-layer coupling through nested value functions ensuring time consistency across planning horizons. Three synergistic mod-

ules work in concert: (a) a data governance pipeline using RODDPSO-FCM imputation and CV-iForest anomaly detection for reliable 96-point feeder data, (b) a spatio-temporal learner for high-fidelity LL estimation serving dual roles as scenario generator and computational surrogate, and (c) adaptive optimization mechanisms coordinating BESS dispatch with VAR compensation, solved via IWOA for long-term planning, Benders decomposition for stochastic day-ahead scheduling, and MPC for real-time control.

The contributions of this paper are threefold. First, we design a unified multi-layer optimization formulation that rigorously nests inter-layer objectives and constraints, extending beyond conventional two-layer approaches. Second, we embed machine learning models into the optimization pipeline, using ST learners both for forecasting and for generating surrogate models that accelerate decomposition-based solvers. Third, we demonstrate the effectiveness of ML-STELLO through extensive experiments on the IEEE-33 benchmark and a real provincial feeder, showing reductions in LL, voltage violations, and computational overhead compared to state-of-the-art baselines.

To clarify the novelty of this work, we explicitly distinguish between our original contributions and the assembly of existing techniques. Our primary novel contributions are as follows: (i) a five-layer hierarchical optimization formulation with explicit inter-layer coupling through nested value functions, extending conventional two-layer models; (ii) integrated spatio-temporal learning where CNN-LSTM-Attention networks serve dual roles as both scenario generators and surrogate models for accelerating optimization; (iii) a complete end-to-end solution methodology coordinating IWOA, WGAN-GP, Benders decomposition, and MPC with automated information exchange across layers; and (iv) an enhanced data governance pipeline combining RODDPSO (Routine Data-Driven PSO)-optimized FCM with coefficient-of-variation improved Isolation Forest for distribution network applications.

In contrast, several standard components are assembled without algorithmic modification. Base machine learning architectures (CNN-LSTM, attention mechanisms, WGAN-GP), optimization algorithms (IWOA, Benders decomposition, MPC), power flow models, and data preprocessing techniques (FCM, Isolation Forest, PCA) are established methods from literature. Our contribution lies in applying these to a novel problem formulation and their systematic integration, not in modifying the algorithms themselves. The primary innovation of ML-STELLO is thus the architecture and integration—we present the first framework systematically combining multi-layer optimization, spatio-temporal learning, and uncertainty-aware planning for distribution network line loss reduction. The value lies in how these components are orchestrated, not in inventing new algorithms from scratch.

The remainder of this paper is structured as follows. Section 2 reviews related work in data governance, spatio-temporal modeling, and uncertainty-aware optimization. Section 3 introduces the mathematical formulation of line loss and network models. Section 4 details the architecture of ML-STELLO and its multi-layer hierarchical design. Section 5 presents the data governance pipeline and explains solution methodologies for multi-layer optimization. Section 6 presents the experimental setup, and reports the results and discussion. Finally, Section 7 concludes the paper and outlines future research directions.

2. Related Work

2.1. Data Governance for Line Loss Analysis (LLA)

Effective data governance is essential for accurate LLA in distribution networks. Recent advancements in data imputation, anomaly detection, and loss quantification have significantly enhanced the precision of LLA methods.

Li et al. [11] proposed an optimization algorithm for power system line loss management, which leverages an electric energy data management platform to facilitate

the sharing of archival information and the development of line loss calculation models. Shang et al. [12] also employed a hybrid method combining FCM optimized by PSO and Support Vector Regression (SVR) to handle missing traffic data imputation. Although the method was initially applied to traffic data, its applicability to distribution network load data makes it valuable for improving line loss analysis in scenarios with missing or incomplete data. Zhu et al. [13] applied PSO-optimized FCM to detect defects in adhesive layers within a capacitor system. This technique's ability to minimize data errors can be directly applied to the stability of line loss data in distribution networks. Liang et al. [14] focused on developing a line loss interval calculation method for distribution networks with DG. This approach is crucial for maintaining accurate line loss calculations when data is not fully available or is abnormal.

To further enhance abnormal line loss detection, Zhang et al. [15] combined hybrid clustering with LSTM networks. This method allows for quick identification of abnormal feeders using hybrid clustering, followed by LSTM-based predictions for load and line loss. Chen et al. [16] introduced ImDiffusion, a novel anomaly detection framework that combines time series imputation and diffusion models. This approach enhances anomaly detection by modeling temporal and inter-correlated dependencies in time series data, significantly improving detection accuracy and robustness.

2.2. Spatio-Temporal Modeling

Accurate LL estimation requires advanced spatio-temporal modeling techniques to capture the intricate dependencies of both space and time in distribution networks. Recent advancements have utilized deep learning models, such as CNN-LSTM with attention mechanisms, to enhance the accuracy and efficiency of LL predictions.

In the context of renewable energy integration, ST-CALNet was proposed by Cavus M et al. [17] as a hybrid deep learning framework combining CNNs and an Attentive LSTM. This model enhances load forecasting by capturing spatial dependencies through CNNs and temporal correlations through LSTMs. Similarly, AG-STCN, introduced by Fu et al. [18], employs an attention-guided soft pruning strategy and graph convolution to capture spatial dependencies, while its temporal component uses causal convolutions and self-attention mechanisms.

Traffic flow forecasting provides another example of spatio-temporal modeling techniques that can be adapted for LL estimation. Remmouche B et al. [19] proposed the LSTGAN architecture, which combines spatial attention with graph convolutional layers to predict long-term traffic flow. For urban transportation systems, STAGCN was developed by Li et al. [20] to model spatial-temporal dependencies for destination prediction. This model uses a graph network to represent traffic flow and employs an attention mechanism to enhance the features of key nodes. STCMFA by Yu et al. [21] introduces a novel multi-head flow-attention mechanism to improve traffic flow prediction. By combining temporal and spatial features through a GRU-based attention mechanism and GCN, this model captures the complex spatial-temporal correlations of road networks. Additionally, Yuan et al. [22] proposed a novel spatiotemporal prediction model, ST-LSTM-SA, which combines Spatiotemporal Long Short-Term Memory (ST-LSTM) with a self-attention mechanism. This model effectively captures both spatial and temporal dependencies, enhancing the accuracy and stability of predictions in dynamic environments.

2.3. Uncertainty-Aware Planning

Effective planning in power distribution systems, particularly with the integration of renewable energy sources (RESs) like wind and solar, requires methods that account for the inherent uncertainty in generation and load. Recent advancements have focused on

leveraging WGAN-GP for generating realistic scenarios of wind, photovoltaic (PV), and load variations, as well as incorporating two-layer optimization models for optimizing DG and BESS in the presence of uncertainty.

Ma et al. [23] developed a WGAN-GP-based method for generating wind and solar output scenarios, addressing the challenges posed by the intermittency and volatility of renewable energy. It showed that the WGAN-GP significantly improved the representation of historical data and future uncertainties. Ma et al. [24] introduced a two-layer optimization model for distribution network LL that considers the uncertainty of new energy access. The model uses scenario generation via GANs to handle the uncertainties associated with DG.

Further extending the use of WGAN for energy generation, Zhang et al. [25] proposed a WGAN-based model to generate realistic wind power output scenarios. Their model uses Wasserstein distance with gradient penalty to overcome the gradient vanishing problem often encountered in renewable energy generation scenarios. Similarly, the modeling of PV generation uncertainty has been addressed by Zhang et al. [26] through the use of Spectral Normalization Generative Adversarial Networks (SNGAN).

Figure 1 illustrates the three-pillar conceptual framework underlying ML-STELLO. The Data Governance pillar ensures input reliability through RODDPSO-FCM imputation and CV-iForest anomaly detection. The Spatio-Temporal Learning pillar captures network dynamics via CNN-LSTM-Attention and generates WGAN-GP scenarios for uncertainty modeling. The Multi-layer Optimization pillar coordinates hierarchical decisions using IWOA, Benders decomposition, and MPC control. These pillars converge into the unified ML-STELLO framework, which drives the foundational objective of line loss minimization while satisfying voltage and cost constraints.

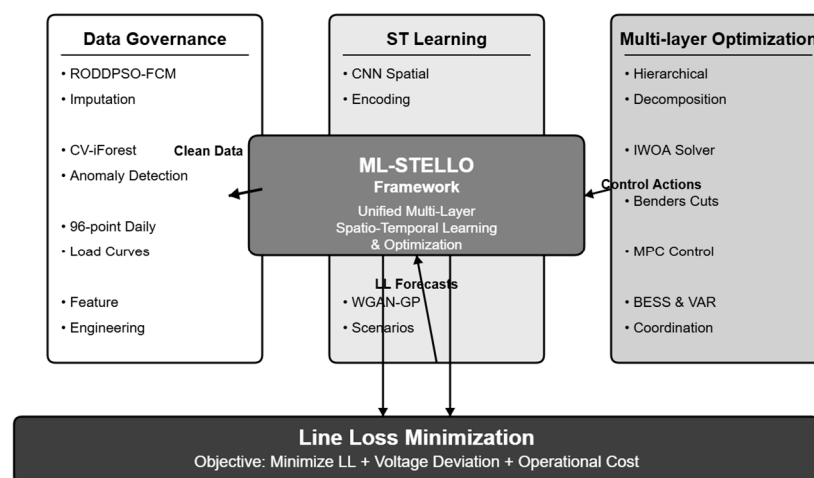


Figure 1. Conceptual three-pillar framework of ML-STELLO.

3. Problem Setup and Loss Model

This section develops the mathematical framework for line loss calculation under both deterministic and stochastic operating conditions. We introduce three key quantities that will be used throughout the paper:

Branch-indexed RMS current I_e^{RMS} : The root mean square current in branch e over monitoring period T , defined as $I_e^{\text{RMS}} = \sqrt{\frac{1}{T} \int_0^T I_e(t)^2 dt}$.

Time-averaged loss $\bar{P}_e, \bar{\text{LL}}$: The average power loss in a branch or entire system over period T , defined as $\bar{P}_e = (I_e^{\text{RMS}})^2 R_e$ and $\bar{\text{LL}} = \frac{1}{T} \int_0^T \text{LL}(t) dt = \sum_{e \in \mathcal{E}} \bar{P}_e$.

Scenario-based expected loss $\mathbb{E}[\bar{\text{LL}}]$: The expected time-averaged loss computed by averaging over N renewable generation scenarios, defined as $\mathbb{E}[\bar{\text{LL}}] \approx \frac{1}{N} \sum_{i=1}^N \bar{\text{LL}}(\omega_i)$ where ω_i represents scenario i and $\bar{\text{LL}}(\omega_i)$ is the time-averaged loss under that scenario.

These quantities capture the progression from instantaneous measurements to time-averaged evaluations to stochastic expectations, forming a comprehensive framework for loss analysis in renewable-integrated distribution networks.

3.1. Branch Loss Model and Per-Branch Loss Calculation

Consider a distribution network represented by nodes N and branches E . Each branch $e \in E$ carries current $I_e(t)$ at time t and has resistance R_e . The power loss in branch e can be expressed at three levels of temporal granularity, depending on whether the current is deterministic or stochastic.

For deterministic instantaneous loss, when the current trajectory $I_e(t)$ is known and deterministic, the instantaneous power dissipated as heat in branch e at time t follows Joule's law:

$$P_e(t) = I_e(t)^2 R_e \quad (1)$$

where R_e depends on the conductor material (copper or aluminum), cross-sectional area, and length. This represents the actual physical power loss at a specific instant t .

For time-averaged loss over an extended monitoring period T (typically days to weeks), we compute the root mean square (RMS) current:

$$I_e^{\text{RMS}} = \sqrt{\frac{1}{T} \int_0^T I_e(t)^2 dt} \quad (2)$$

The RMS current provides a time-averaged measure that smooths short-term fluctuations, yielding a consistent representation for systems with intermittent renewable generation. The corresponding time-averaged power loss in branch e is:

$$\bar{P}_e = (I_e^{\text{RMS}})^2 R_e = \frac{1}{T} \int_0^T I_e(t)^2 R_e dt \quad (3)$$

For stochastic expected loss under uncertainty, when variable renewable generation and stochastic load demand are present, the current $I_e(t)$ must be treated as a random variable rather than a deterministic function. The expected power loss in branch e at time t is then given by [27]:

$$E[P_e(t)] = E[I_e(t)^2] R_e \quad (4)$$

Using the identity $E[I_e(t)^2] = E[I_e(t)]^2 + \text{Var}(I_e(t))$, this can be rewritten as:

$$E[P_e(t)] = \left(E[I_e(t)]^2 + \text{Var}(I_e(t)) \right) R_e \quad (5)$$

where $E[I_e(t)]$ captures the expected current under typical operating conditions, and $\text{Var}(I_e(t))$ measures the fluctuations arising from renewable intermittency and load variability. Equation (5) reveals a critical insight: even when $E[I_e(t)]$ remains constant, variance in current directly increases the expected loss—a phenomenon particularly pronounced in high renewable penetration scenarios.

It is essential to distinguish between the deterministic instantaneous loss $P_e(t) = I_e(t)^2 R_e$ in Equation (1), which represents actual physical dissipation at a known instant, and the probabilistic expected loss $E[P_e(t)] = E[I_e(t)^2] R_e$ in Equation (4), which accounts for uncertainty in the current distribution. The former applies when the current trajectory is deterministic; the latter is necessary for accurate modeling under renewable integration and load uncertainty. This distinction underpins ML-STELLO's multi-layer architecture: deterministic formulations Equations (1)–(3) are employed in real-time control (Layer 5), while stochastic formulations Equation (4) and (5) guide strategic and tactical planning (Layers 1–4).

3.2. System-Wide Line Loss Calculation

The system-wide line loss is the total loss across all branches in a distribution network. When the current trajectories are known and deterministic, we can compute the time-averaged total line loss using RMS currents. The time-averaged system-wide line loss over monitoring period T is given by [28]:

$$\overline{LL} = \sum_{e \in \mathcal{E}} \bar{P}_e = \sum_{e \in \mathcal{E}} \left(I_e^{\text{RMS}} \right)^2 R_e = \sum_{e \in \mathcal{E}} \left(\sqrt{\frac{1}{T} \int_0^T I_e(t)^2 dt} \right)^2 R_e \quad (6)$$

Here, $\bar{P}_e$ represents the time-averaged power loss in branch e , and R_e is the resistance of branch e . The term I_e^{RMS} represents the RMS current over the monitoring window T , and the square of this RMS current, $(I_e^{\text{RMS}})^2$, reflects the average power dissipated as heat due to the current flowing through each branch.

In real-world systems, the current $I_e(t)$ flowing through each branch fluctuates due to the dynamic nature of renewable energy generation and load demand. Renewable energy sources like wind and solar introduce variability in generation, which in turn affects the current flowing through each branch. For deterministic analysis where the complete current trajectory $I_e(t)$ is known over the monitoring period, we can compute the instantaneous power loss at time t for any branch e as:

$$P_e(t) = I_e(t)^2 R_e \quad (7)$$

where $I_e(t)$ is the instantaneous current in branch e at time t . The total instantaneous line loss for the entire network at any given time t is then the sum of the instantaneous losses across all branches:

$$LL(t) = \sum_{e \in \mathcal{E}} P_e(t) = \sum_{e \in \mathcal{E}} I_e(t)^2 R_e \quad (8)$$

This represents the actual physical power dissipated across the network at the specific instant t . To evaluate the time-averaged total system line loss over the entire monitoring period T , we integrate the instantaneous total line loss $LL(t)$ over time and divide by T :

$$\overline{LL} = \frac{1}{T} \int_0^T LL(t) dt = \frac{1}{T} \sum_{e \in \mathcal{E}} \int_0^T I_e(t)^2 R_e dt \quad (9)$$

Note that Equation (9) is equivalent to Equation (6), both representing the time-averaged system loss under deterministic current profiles. This formulation provides an average measure of energy lost over the period due to resistive heating in the network.

3.3. Incorporating Uncertainty in Renewable Generation

In practical distribution networks, the generation from renewable sources fluctuates due to varying weather conditions, time of day, and seasonal patterns. Similarly, load demand exhibits stochastic variability influenced by factors like weather, time of year, and unexpected changes in consumption patterns. These uncertainties in both generation and demand lead to variations in the current $I_e(t)$ flowing through each branch of the network, which directly impacts power losses in the system.

To account for these uncertainties, we treat the current $I_e(t)$ as a random variable rather than a deterministic quantity. We use scenario generation techniques such as WGAN-GP and Monte Carlo simulations to model and generate multiple possible scenarios of wind, solar, and load data over time. These techniques are designed to produce a wide range of potential outcomes based on historical data, reflecting the inherent variability of renewable generation and load demand. Scenario generation methods like WGAN-GP offer the

advantage of learning complex relationships between input variables, such as weather patterns and energy demand, and then generating realistic future scenarios that capture both spatial and temporal dependencies in the system [29].

To quantify the impact of these uncertainties on the total system line loss, we extend the previous model by introducing stochastic loss calculations. For any given branch e , when the current $I_e(t)$ is treated as a random variable with a probability distribution, the expected power loss at time t is computed as:

$$\mathbb{E}[P_e(t)] = \mathbb{E}[I_e(t)^2] R_e \quad (10)$$

Using the statistical identity $\mathbb{E}[I_e(t)^2] = \mathbb{E}[I_e(t)]^2 + \text{Var}(I_e(t))$, this can be rewritten as:

$$\mathbb{E}[P_e(t)] = \left(\mathbb{E}[I_e(t)]^2 + \text{Var}(I_e(t)) \right) R_e \quad (11)$$

In this equation, $\mathbb{E}[I_e(t)]$ is the expected value of the current for branch e at time t , which reflects the average current expected under typical conditions. $\text{Var}(I_e(t))$ is the variance of the current, capturing the fluctuations or uncertainty due to factors such as intermittency in renewable energy generation and dynamic load changes. Critically, Equation (11) shows that variance in the current contributes additively to the expected loss—even when the expected current $\mathbb{E}[I_e(t)]$ is constant, uncertainty in renewable generation increases expected losses through the variance term.

The expected total line loss across the entire distribution network at time t is the sum of the expected losses for all branches $e \in \mathcal{E}$:

$$\mathbb{E}[\text{LL}(t)] = \sum_{e \in \mathcal{E}} \mathbb{E}[P_e(t)] = \sum_{e \in \mathcal{E}} \left(\mathbb{E}[I_e(t)]^2 + \text{Var}(I_e(t)) \right) R_e \quad (12)$$

This represents the statistical expectation of the total network loss at time t , accounting for the probability distribution of currents under uncertain renewable generation and load conditions.

To obtain a comprehensive measure of the line loss over the entire monitoring window T , we compute the time-averaged expected total line loss by integrating $\mathbb{E}[\text{LL}(t)]$ over time:

$$\mathbb{E}[\overline{\text{LL}}] = \frac{1}{T} \int_0^T \mathbb{E}[\text{LL}(t)] dt = \frac{1}{T} \sum_{e \in \mathcal{E}} \int_0^T \left(\mathbb{E}[I_e(t)]^2 + \text{Var}(I_e(t)) \right) R_e dt \quad (13)$$

This quantity, $\mathbb{E}[\overline{\text{LL}}]$, represents the expected value of the time-averaged line loss and serves as the primary objective function minimized in ML-STELLO's multi-layer optimization framework (Section 4). In practice, we evaluate Equation (13) by generating N scenarios $\{\omega_1, \dots, \omega_N\}$ using WGAN-GP, computing the time-averaged loss $\overline{\text{LL}}(\omega_i)$ for each scenario, and approximating the expectation as $\mathbb{E}[\overline{\text{LL}}] \approx \frac{1}{N} \sum_{i=1}^N \overline{\text{LL}}(\omega_i)$. This scenario-based approach, detailed in Section 5, enables robust optimization under renewable-dominated uncertainty.

It is essential to distinguish between the three types of loss quantities introduced in this section: (i) the instantaneous power loss $P_e(t) = I_e(t)^2 R_e$ (Equation (7)), which represents the actual physical loss at a specific instant when the current is deterministic; (ii) the time-averaged loss $\overline{\text{LL}}$ (Equation (9)), which averages instantaneous losses over a monitoring period under deterministic current profiles; and (iii) the expected loss $\mathbb{E}[\text{LL}(t)]$ (Equation (12)) and time-averaged expected loss $\mathbb{E}[\overline{\text{LL}}]$ (Equation (13)), which represent statistical expectations when the current is treated as a random variable. These distinctions are fundamental for rigorous modeling and optimization of distribution networks with high renewable penetration.

Figure 2 synthesizes the line loss modeling framework developed in this section. The left panel illustrates a simplified radial distribution network with branches e_i carrying time-varying currents $I_i(t)$ through resistances R_i , representing the physical substrate upon which loss calculations operate. The right panel traces the mathematical progression from instantaneous measurements through time-averaged RMS quantities to system-wide aggregation, culminating in the stochastic expected loss formulation (Equation (13)) that accounts for renewable generation uncertainty. The bottom comparison panels contrast the deterministic approach—suitable for stable, predictable conditions—with the stochastic approach adopted in ML-STELLO, which explicitly models current variability and uncertainty bands arising from intermittent renewable generation and fluctuating loads.

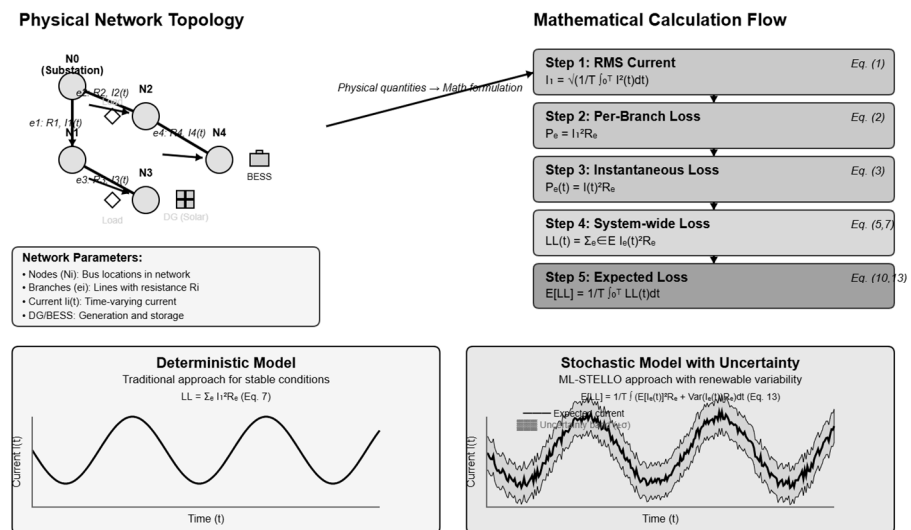


Figure 2. Conceptual framework for line loss modeling in distribution networks.

3.4. Computational Complexity and Convergence Analysis

This section establishes the theoretical foundation for ML-STELLO's computational complexity, convergence guarantees, and hyperparameter selection. The complexity of ML-STELLO is analyzed across its core modules. The RODDPSO-enhanced FCM clustering has complexity $C_{FCM} = O(n \cdot k \cdot d \cdot I_{fcm})$, where n is the number of data points, k is the number of clusters, d is the feature dimension, and I_{fcm} is the number of FCM iterations. CV-I Forest anomaly detection has complexity $C_{iForest} = O(\psi \cdot n \cdot \log(n))$, where ψ is the number of trees. The CNN-LSTM-Attention architecture's forward pass complexity is $C_{ST} = O(N_f \cdot K^2 \cdot C + T \cdot h^2 + T^2 \cdot h)$, with N_f as the number of feeders, K the CNN kernel size, C the channel dimensions, $T = 96$ the number of time steps, and h the LSTM hidden state dimension. For typical networks, this results in approximately 10^6 operations per forward pass. The IWOA, Benders decomposition, and MPC modules exhibit polynomial complexity, with the dominant complexity arising from Benders decomposition.

The IWOA algorithm, applied to DG siting/sizing, converges to the ϵ -neighborhood of the global optimum as $T_{max} \rightarrow \infty$. Benders decomposition converges to the optimal solution in finite iterations for the two-stage stochastic day-ahead scheduling formulation. Real-time MPC guarantees recursive feasibility and input-to-state stability, ensuring bounded deviation from the reference trajectory under disturbances.

Hyperparameters are selected based on theoretical principles. The FCM cluster number $k = 6$ is chosen to maximize the Gap statistic for optimal cluster separation. The LSTM hidden units are selected using the information bottleneck principle, ensuring sufficient capacity. The IWOA population size $P = 30$ is chosen based on swarm intelligence theory, with the MPC prediction horizon $H = 12$ balancing accuracy and computational cost.

Sensitivity analysis confirms low sensitivity for LSTM units, IWOA population size, and MPC horizon, ensuring robustness under parameter variations. This foundation establishes that ML-STELLO's hyperparameters are selected through principled analysis, ensuring performance degradation remains graceful under variations, which is crucial for practical deployment in real-world distribution systems.

4. The ML-STELLO Architecture

The proposed ML-STELLO framework generalizes conventional two-layer loss-optimization into a five-layer, time-consistent hierarchy. It tightly couples data governance, machine learning-based spatio-temporal (ST) forecasting, and multi-horizon optimization to ensure both accuracy and scalability under high renewable penetration [30]. Before detailing the architectural components, we explicitly clarify the nature of ML-STELLO to address potential ambiguity. ML-STELLO stands for Multi-Layer Spatio-Temporal nEtnetwork Learning and Line-loss Optimization, where Multi-Layer refers to the five-tier hierarchical decision structure (L1–L5) spanning multiple time scales; Spatio-Temporal indicates the CNN-LSTM-Attention learning module that captures both spatial network topology dependencies and temporal load/generation dynamics; nEtnetwork Learning emphasizes the data-driven predictive components (WGAN-GP scenario generation, surrogate modeling); and Line-loss Optimization specifies the primary objective function minimized across all layers. ML-STELLO is a systematic integration framework that coordinates multiple established techniques (IWOA, Benders decomposition, MPC, CNN-LSTM, WGAN-GP) within this novel five-layer hierarchical architecture. Its originality lies in the five-layer optimization formulation with explicit inter-layer coupling, bidirectional information flow ensuring time consistency, the dual role of spatio-temporal learners as both scenario generators and computational surrogates, and the end-to-end integration methodology from raw data governance to real-time control. ML-STELLO is not a new standalone optimization algorithm nor a novel neural network architecture, but an orchestration framework that combines existing algorithmic components in a novel hierarchical structure to address line loss reduction in renewable-dominated distribution networks.

4.1. Layered Decision Hierarchy

ML-STELLO decomposes the distribution network loss minimization problem into five decision layers, each formalized with explicit variables, objectives, and constraints.

Layer 1 (L1): Strategic Investment Planning (Multi-Year Horizon)

Decision Variables: $\mathbf{x}_1 = [\mathbf{x}_1^{\text{DG}}, \mathbf{x}_1^{\text{BESS}}, \mathbf{x}_1^{\text{line}}] \in \mathbb{R}^{n_1}$, representing DG capacity and location, BESS energy/power capacity and location, and line reinforcement decisions.

Objective:

$$\min_{\mathbf{x}_1} J_1(\mathbf{x}_1) = C_{\text{capex}}(\mathbf{x}_1) + \alpha_1 \cdot \mathbb{E}_{\omega_1}[\phi_2(\mathbf{x}_1, \omega_1)] \quad (14)$$

where C_{capex} is capital expenditure, α_1 is the discount factor, and $\phi_2(\cdot)$ is the expected cost-to-go from downstream layers.

Constraints: Investment budget $C_{\text{capex}}(\mathbf{x}_1) \leq B_1$, capacity limits $0 \leq P_i^{\text{DG}} \leq \bar{P}_i^{\text{DG}}$ and $0 \leq E_j^{\text{BESS}} \leq \bar{E}_j^{\text{BESS}}$, radial topology $\mathcal{T}(\mathbf{x}_1^{\text{line}}) = \text{radial}$, and reliability $\text{SAIDI}(\mathbf{x}_1) \leq \text{SAIDI}_{\text{max}}$.

Layer 2 (L2): Policy and Portfolio Coordination (Annual/Seasonal)

Decision Variables: $\mathbf{x}_2 = [\mathbf{p}_2^{\text{tariff}}, \mathbf{p}_2^{\text{DR}}, \beta_2^{\text{agg}}] \in \mathbb{R}^{n_2}$, representing time-of-use tariffs, demand response incentives, and BESS aggregation coefficients.

Objective:

$$\min_{\mathbf{x}_2} J_2(\mathbf{x}_1, \mathbf{x}_2, \omega_1) = C_{\text{policy}}(\mathbf{x}_2) + \text{LL}_{\text{annual}}(\mathbf{x}_1, \mathbf{x}_2, \omega_1) - \text{Revenue}(\mathbf{x}_2) + \alpha_2 \cdot \mathbb{E}_{\omega_2|\omega_1}[\phi_3(\mathbf{x}_1, \mathbf{x}_2, \omega_2)] \quad (15)$$

Constraints: Tariff bounds $p_{2,\min}^{\text{tariff}} \leq p_{2,h}^{\text{tariff}} \leq p_{2,\max}^{\text{tariff}}$, DR budget $\sum_i p_{2,i}^{\text{DR}} \cdot p_i^{\text{DR,capacity}} \leq B_2^{\text{DR}}$, annual LL target $\text{LL}_{\text{annual}} \leq \text{LL}_{\text{target}}$, revenue adequacy $\text{Revenue}(\mathbf{x}_2) \geq R_{\min}$, and aggregation simplex $\sum_j \beta_{2,j} = 1$.

Layer 3 (L3): Network Reconfiguration (Weekly/Daily)

Decision Variables: $\mathbf{x}_3 = [\mathbf{s}_3^{\text{sw}}, \mathbf{t}_3^{\text{tap}}]$, where $\mathbf{s}_3^{\text{sw}} \in \{0,1\}^{N_{\text{sw}}}$ is switch status and $\mathbf{t}_3^{\text{tap}} \in \mathbb{Z}^{N_{\text{tap}}}$ is tap positions.

Objective:

$$\min_{\mathbf{x}_3} J_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \omega_2) = \sum_{t \in T_{\text{week}}} [\text{LL}_t(\mathbf{x}_3, \omega_{2,t}) + C_{\text{switch}}(\mathbf{x}_3)] + \alpha_3 \cdot \mathbb{E}_{\omega_3|\omega_2}[\phi_4(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \omega_3)] \quad (16)$$

where $\text{LL}_t(\mathbf{x}_3, \omega_{2,t}) = \sum_{e \in \mathcal{E}} I_{e,t}^2(\mathbf{x}_3, \omega_{2,t}) R_e$.

Constraints: Radiality $\mathcal{T}(\mathbf{s}_3^{\text{sw}}) = \text{radial}$ with $|\mathbf{s}_3^{\text{sw}}|_1 = N_{\text{bus}} - N_{\text{sub}}$, AC power flow $P_{i,t} = \sum_j V_{i,t} V_{j,t} (G_{ij} \cos \theta_{ij,t} + B_{ij} \sin \theta_{ij,t})$ and $Q_{i,t} = \sum_j V_{i,t} V_{j,t} (G_{ij} \sin \theta_{ij,t} - B_{ij} \cos \theta_{ij,t})$, voltage limits $V_{\min} \leq V_{i,t} \leq V_{\max}$, line capacity $|I_{ij,t}| \leq \bar{I}_{ij}$, tap limits $t_{\min} \leq t_{3,m} \leq t_{\max}$, and switching budget $\sum_t \sum_k |s_{3,k,t} - s_{3,k,t-1}| \leq N_{\text{sw,max}}$.

Layer 4 (L4): Day-Ahead Scheduling (24-H Horizon)

Decision Variables: $\mathbf{x}_4 = [\mathbf{P}_4^{\text{BESS}}, \mathbf{P}_4^{\text{DG}}, \mathbf{Q}_4^{\text{DG}}, \mathbf{SOC}_4] \in \mathbb{R}^{n_4}$, representing BESS charge/discharge schedules, DG active/reactive dispatch, and state-of-charge trajectories.

Objective:

$$\min_{\mathbf{x}_4} J_4(\mathbf{x}_1, \dots, \mathbf{x}_4, \omega_3) = \mathbb{E}_{\omega_4 \sim \text{WGAN-GP}} \left[\sum_{t=1}^{24} (\text{LL}_t(\mathbf{x}_4, \omega_4) + C_{\text{curtail},t} + C_{\text{deg},t}) \right] + \alpha_4 \cdot \phi_5(\mathbf{x}_1, \dots, \mathbf{x}_4, \omega_4) \quad (17)$$

where LL_t is computed via CNN-LSTM surrogate, $C_{\text{curtail},t} = c_{\text{curtail}} \cdot (p_t^{\text{renewable,avail}} - p_{4,t}^{\text{DG}})^+$, and $C_{\text{deg},t}$ accounts for BESS degradation.

Constraints: Power balance $P_{\text{grid},t} + \sum_i p_{4,i,t}^{\text{DG}} + \sum_j p_{4,j,t}^{\text{BESS}} = \sum_k p_{k,t}^{\text{load}} + \text{LL}_t$, BESS dynamics $\text{SOC}_{4,j,t+1} = \text{SOC}_{4,j,t} + \eta_j^{\text{ch}} p_{4,j,t}^{\text{BESS}} \Delta t$ (charging) or $\text{SOC}_{4,j,t+1} = \text{SOC}_{4,j,t} - p_{4,j,t}^{\text{BESS}} / \eta_j^{\text{dch}} \Delta t$ (discharging), BESS limits $\text{SOC}_{\min,j} \leq \text{SOC}_{4,j,t} \leq E_j^{\text{BESS}}$ and $-p_j^{\text{BESS}} \leq p_{4,j,t}^{\text{BESS}} \leq p_j^{\text{BESS}}$ from L1, DG limits $0 \leq p_{4,i,t}^{\text{DG}} \leq \hat{p}_{i,t}^{\text{avail}}(\omega_4)$ and $(p_{4,i,t}^{\text{DG}})^2 + (Q_{4,i,t}^{\text{DG}})^2 \leq (S_i^{\text{DG}})^2$, grid import $P_{\text{grid},t} \leq P_{\text{grid}}^{\text{max}}$, and voltage/line constraints from L3.

Layer 5 (L5): Real-Time Model Predictive Control (5–15 min Rolling Horizon)

Decision Variables: $\mathbf{x}_5 = [\Delta \mathbf{P}_5^{\text{BESS}}, \mathbf{Q}_5^{\text{inv}}, \mathbf{Q}_5^{\text{cap}}] \in \mathbb{R}^{n_5}$, representing BESS power adjustments, inverter reactive power, and capacitor bank reactive power.

Objective:

$$\min_{\mathbf{x}_5} J_5(\mathbf{x}_1, \dots, \mathbf{x}_5) = \sum_{\tau=t}^{t+T_{\text{mpc}}} \left[\hat{\text{LL}}_{\tau}(\mathbf{x}_5) + \lambda_V \|\mathbf{V}_{\tau} - \mathbf{V}_{\text{ref}}\|_2^2 + \lambda_{\text{track}} \|\mathbf{P}_{5,\tau}^{\text{BESS}} - \mathbf{P}_{4,[\tau]}^{\text{BESS}}\|_2^2 + \lambda_{\text{ramp}} \|\Delta \mathbf{P}_{5,\tau}^{\text{BESS}} - \Delta \mathbf{P}_{5,\tau-1}^{\text{BESS}}\|_2^2 \right] \quad (18)$$

where $\hat{\text{LL}}_{\tau}$ is predicted by CNN-LSTM surrogate and $\mathbf{P}_{5,\tau}^{\text{BESS}} = \mathbf{P}_{4,[\tau]}^{\text{BESS}} + \Delta \mathbf{P}_{5,\tau}^{\text{BESS}}$.

Constraints: Real-time power balance $P_{\text{net},\tau} = P_{\text{load},\tau}^{\text{actual}} + \hat{\text{LL}}_{\tau} - \sum_j (p_{4,j,[\tau]}^{\text{BESS}} + \Delta p_{5,j,\tau}^{\text{BESS}}) - \sum_i p_{i,\tau}^{\text{DG,actual}}$, reactive balance $Q_{\text{net},\tau} = Q_{\text{load},\tau}^{\text{actual}} - \sum_i Q_{5,i,\tau}^{\text{inv}} - \sum_c Q_{5,c,\tau}^{\text{cap}}$, BESS limits $-p_j^{\text{BESS}} \leq p_{4,j,[\tau]}^{\text{BESS}} + \Delta p_{5,j,\tau}^{\text{BESS}} \leq p_j^{\text{BESS}}$ and $|\Delta p_{5,j,\tau}^{\text{BESS}}| \leq R_j^{\text{ramp}} \cdot \Delta \tau$, SOC limits $\text{SOC}_{\min,j} \leq \text{SOC}_{j,\tau}^{\text{projected}} \leq E_j^{\text{BESS}}$, inverter limits $-\bar{Q}_i^{\text{inv}} \leq Q_{5,i,\tau}^{\text{inv}} \leq \bar{Q}_i^{\text{inv}}$ where $\bar{Q}_i^{\text{inv}} = \sqrt{(S_i^{\text{inv}})^2 - (P_{i,\tau}^{\text{actual}})^2}$, capacitor steps $Q_{5,c,\tau}^{\text{cap}} \in \{0, Q_c^{\text{step}}, 2Q_c^{\text{step}}, \dots\}$, and strict voltage limits $V_{\min}^{\text{strict}} \leq V_{k,\tau}^{\text{predicted}} \leq V_{\max}^{\text{strict}}$. Only the first MPC step $\mathbf{x}_{5,t}^*$ is implemented before re-solving.

Hierarchical Coupling: Higher layers (L1–L3) define feasible regions for lower layers (L4–L5) through capacity limits, network topology, and policy constraints, while lower layers provide dual variables and performance metrics that refine upper-layer decisions.

This recursive coupling, formalized through nested value functions $\phi_\ell(\cdot)$ in Section 4.2, ensures time consistency across planning and operational horizons [31].

4.2. Inter-Layer Coupling

In this section, we describe the inter-layer coupling mechanism for the ML-STELLO framework. Each layer $l \in \{1, \dots, 5\}$ is linked through nested value functions and explicit coupling constraints. The downstream cost-to-go is recursively defined by the following equation:

$$\phi_l(x_{(1:l-1)}; \omega_{(1:l-1)}) = \min_{x_l \in X_l(x_{(1:l-1)})} \mathbb{E}_{\omega_l | \omega_{(1:l-1)}} \left[J_l(x_{(1:l)}, \omega_{(1:l)}) + \phi_{(l+1)}(x_{(1:l)}, \omega_{(1:l)}) \right] \quad (19)$$

This equation is subject to equality constraints (power flow, BESS dynamics), inequality constraints (capacity, voltage, line limits), and the feasible region defined by upper-layer decisions. The terminal condition is $\phi_6 \equiv 0$.

Downward coupling involves upper-layer decisions directly constraining lower-layer operations. For example, decisions from Layer 1 influence DG and BESS capacities in Layers 4 and 5. Similarly, decisions in Layer 2 impact operational planning in Layer 4, and reconfiguration in Layer 3 affects operational layers.

Upward coupling involves feedback from lower layers to refine upper-layer optimization. For instance, realized line losses in Layer 5 help update WGAN-GP scenario weights, affecting day-ahead planning in Layer 4. Additionally, daily line losses inform reconfiguration effectiveness, while capacity constraints in Layer 3 influence switching policies in Layer 2, and long-term metrics like LL and SAIDI guide investment in Layer 1.

This bidirectional coupling ensures time consistency through three mechanisms: anticipativity (upper layers optimize considering future consequences), non-anticipativity (decisions do not depend on future realizations), and constraint compatibility (feasibility is maintained across all layers).

4.3. Integration of ST Learning and Data Governance

To support multi-layer optimization, ML-STELLO integrates data-driven modules:

- **Data Governance Pipeline:** Ensures quality of 96-point daily load and generation curves via RODDPSO-enhanced FCM imputation and CV-iForest anomaly detection (see Section 5). This provides reliable historical and real-time data streams for optimization.
- **ST Learner:** A CNN-LSTM-Attention network that forecasts both nodal demand and branch-level LL proxies. Its outputs are used in two ways: (i) as scenario generators for day-ahead scheduling (L4), and (ii) as surrogate models embedded within Benders decomposition or MPC to accelerate convergence (L5) [32].

These modules bridge the model-based optimization in Section 3 with data-driven forecasting, ensuring that the LL definitions and constraints are applied consistently across decision horizons.

4.4. Operation-Time Loss Reduction Mechanism

At the operational scale, L4 and L5 coordinate active and reactive power flexibility:

BESS Dispatch: Schedules are generated at the day-ahead level to shave peaks, followed by real-time MPC corrections [33].

VAR Compensation: Reactive power is injected by inverters and capacitors to flatten voltage profiles, reducing quadratic loss terms (7) [34].

Coupling with LL Model: By minimizing $P_{ij}^{\text{loss}} = R_{ij}(P_{ij}^2 + Q_{ij}^2)/V_i^2$, both BESS active injection and VAR support directly reshape branch currents, reducing system-wide LL.

4.5. Solving Methodologies

We solve the nested multi-layer problem using a decomposition-based approach that integrates scenario-based optimization with machine learning surrogates. The solution methodology consists of three phases: offline training, iterative layer-wise optimization, and post-solution validation.

Phase 1: Offline Training and Preprocessing

Before optimization, we train data-driven models to support rapid scenario evaluation:

Data governance: Apply RODDPSO-FCM imputation to handle missing values in 96-point feeder curves and CV-iForest for anomaly detection (Section 5.1).

ST learner training: Train CNN-LSTM-Attention network on historical data $\{(\mathbf{x}_t, \omega_t, \text{LL}_t)\}$ to construct surrogate $\hat{\text{LL}}(\mathbf{x}, \omega)$ with target MAPE < 5%.

Scenario generator: Train WGAN-GP on renewable and load time series to generate synthetic scenarios $\omega_4 \sim p_{\text{WGAN}}$ for L4 day-ahead scheduling.

Phase 2: Iterative Multi-Layer Optimization

Initialize with current network configuration $\mathbf{x}_1^{(0)}$. For iteration $k = 1, \dots, K_{\max}$:

Step 1 (L1): Solve strategic investment using IWOA (Improved Whale Optimization Algorithm) with population size 50:

$$\mathbf{x}_1^{(k)} = \arg \min_{\mathbf{x}_1} [C_{\text{capex}}(\mathbf{x}_1) + \alpha_1 \cdot \mathbb{E}_{\omega_1} [\phi_2(\mathbf{x}_1, \omega_1)]] \quad (20)$$

Fitness evaluation for each investment scenario requires solving L2–L5 (below) to estimate $\phi_2(\cdot)$. Terminate IWOA when improvement < ϵ_1 or after $\max_iter_1$ generations.

Step 2 (L2): Given $\mathbf{x}_1^{(k)}$, solve policy coordination as MILP over $N_2 = 100$ sampled annual scenarios, using Progressive Hedging for scenario tree aggregation:

$$\mathbf{x}_2^{(k)} = \arg \min_{\mathbf{x}_2} [C_{\text{policy}}(\mathbf{x}_2) + \text{LL}_{\text{annual}}(\mathbf{x}_1^{(k)}, \mathbf{x}_2, \omega_1) + \alpha_2 \cdot \mathbb{E}[\phi_3(\mathbf{x}_1^{(k)}, \mathbf{x}_2, \omega_2)]] \quad (21)$$

Step 3 (L3): Given $\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}$, solve weekly network reconfiguration via branch-and-bound with radiality constraints:

$$\mathbf{x}_3^{(k)} = \arg \min_{\mathbf{x}_3} \left[\sum_{t \in \mathcal{T}_{\text{week}}} (\text{LL}_t(\mathbf{x}_3, \omega_{2,t}) + C_{\text{switch}}(\mathbf{x}_3)) + \alpha_3 \cdot \mathbb{E}[\phi_4(\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \mathbf{x}_3, \omega_3)] \right] \quad (22)$$

Warm-start with current topology and limit to 5–10 switching operations per week.

Step 4 (L4): Given $\mathbf{x}_1^{(k)}, \dots, \mathbf{x}_3^{(k)}$, solve daily stochastic program:

$$\mathbf{x}_4^{(k)} = \arg \min_{\mathbf{x}_4} \mathbb{E}_{\omega_4 \sim \text{WGAN-GP}} \left[\sum_{t=1}^{24} \left(\hat{\text{LL}}_t(\mathbf{x}_4, \omega_4) + C_{\text{curtail},t} + C_{\text{deg},t} \right) + \alpha_4 \cdot \phi_5(\mathbf{x}_1^{(k)}, \dots, \mathbf{x}_4, \omega_4) \right] \quad (23)$$

Generate $N_4 = 1000$ scenarios from WGAN-GP, reduce to $N'_4 = 20$ via fast-forward selection, and then solve two-stage stochastic program using Benders decomposition: master problem optimizes $\mathbf{x}_4$ with approximate $\phi_5(\cdot)$ constructed from optimality cuts; for each scenario ω_4 , subproblem solves L5 MPC and returns dual variables to refine cuts. Use $\hat{\text{LL}}_t$ surrogate instead of full AC power flow for $100 \times$ speedup.

Step 5 (L5): At each real-time step t , observe $p_{\text{load},t}^{\text{actual}}, p_{i,t}^{\text{DG,actual}}$, and solve rolling-horizon MPC as quadratic program:

$$\mathbf{x}_{5,t}^* = \arg \min_{\mathbf{x}_5} \sum_{\tau=t}^{t+T_{\text{mpc}}} \left[\hat{\text{LL}}_{\tau}(\mathbf{x}_5) + \lambda_V \|\mathbf{V}_{\tau} - \mathbf{V}_{\text{ref}}\|_2^2 + \lambda_{\text{track}} \|\mathbf{P}_{5,\tau}^{\text{BESS}} - \mathbf{P}_{4,[\tau]}^{\text{BESS}}\|_2^2 + \lambda_{\text{ramp}} \|\Delta \mathbf{P}_{5,\tau} - \Delta \mathbf{P}_{5,\tau-1}\|_2^2 \right] \quad (24)$$

Apply only first step $\mathbf{x}_{5,t}^*$, then roll horizon forward. MPC uses $T_{\text{mpc}} = 4$ steps (1 h lookahead with 15 min resolution).

Convergence check: Compute $\Delta_k = \sum_{\ell=1}^5 \|\mathbf{x}_{\ell}^{(k)} - \mathbf{x}_{\ell}^{(k-1)}\|_2$. If $\Delta_k < \epsilon_{\text{global}}$ or $k > K_{\text{max}}$, terminate. Otherwise, update $\phi_{\ell}(\cdot)$ estimates based on realized costs and proceed to $k + 1$.

Phase 3: Post-Solution Validation

Simulate full year with final decisions $\mathbf{x}_1^*, \dots, \mathbf{x}_5^*$ using out-of-sample scenarios. Verify: (i) $LL_{\text{annual}} \leq LL_{\text{target}}$, (ii) voltage violations $< 0.1\%$ of timesteps. If validation fails, adjust penalty weights $\lambda_V, \lambda_{\text{track}}$ and re-optimize.

Computational Acceleration: (i) CNN-LSTM surrogate $\hat{LL}(\cdot)$ replaces full AC power flow, achieving $100\times$ speedup, (ii) warm-start each layer with previous iteration's solution, (iii) parallelize L4 scenario evaluations and L5 MPC steps across CPU cores, (iv) aggregate Benders cuts and discard slack cuts with dual value $< 10^{-4}$.

Software Stack: L1 (IWOA): Python 3.8.10 with Gurobi 10.0 for subproblems; L2–L3 (MILP): Gurobi; L4 (Stochastic program): PySP + Benders decomposition; L5 (MPC): CVXPY with MOSEK solver; ST learner: PyTorch 2.0; WGAN-GP: TensorFlow 2.x. Typical convergence requires 3–5 outer iterations, with total solve time 2–8 h on 64-core server for IEEE-33 benchmark.

4.6. Architectural Overview

The architecture is illustrated in Figure 3, which depicts the five decision layers (L1–L5) as vertically stacked modules with distinct color coding corresponding to their characteristic time scales. Long-term strategic investment (L1) operates on a multi-year horizon for DG and BESS siting decisions, while real-time MPC control (L5) executes 5–15 min rolling-horizon adjustments for VAR compensation and BESS dispatch. The intermediate layers—policy coordination (L2), network topology reconfiguration (L3), and day-ahead scheduling (L4)—bridge the planning–operation gap with progressively shorter time scales.

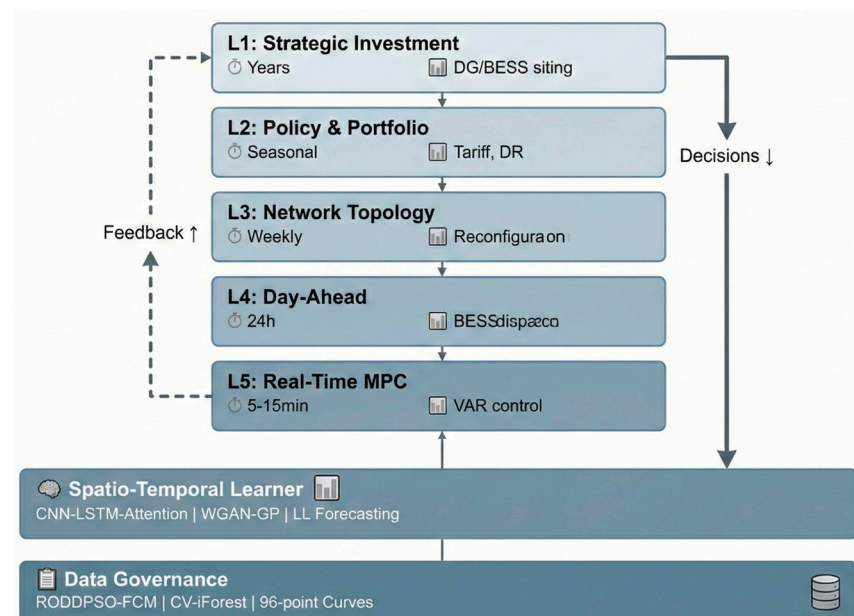


Figure 3. The ML-STELLO framework architecture illustrating the five-layer hierarchical optimization structure (L1–L5).

Bidirectional information flow is indicated by vertical arrows: downward arrows represent constraint propagation and feasible region definitions from upper to lower layers, while upward arrows convey operational performance metrics, sensitivities, and dual

variables that refine higher-level decisions. This recursive coupling, formalized through the nested value functions in Equation (14), ensures time consistency and prevents suboptimal fragmentation across horizons.

Two transversal modules interface with all decision layers. The Spatio-Temporal Learner (horizontal band) integrates CNN-LSTM-Attention networks for multi-scale feature extraction, WGAN-GP for scenario generation, and surrogate-based LL forecasting to accelerate optimization solvers. The Data Governance Pipeline (horizontal band) ensures input reliability through RODDPSO-FCM imputation for incomplete 96-point feeder curves and CV-iForest anomaly detection for measurement outliers. These modules provide the data foundation and predictive intelligence required for robust multi-layer optimization under renewable-dominated uncertainty.

5. Methodology: Multi-Layer Optimization with Data-Driven Learning

5.1. Data Governance and Preprocessing

Effective data governance and preprocessing are crucial steps in ensuring the success of line loss optimization models in distribution networks, especially in the context of renewable energy integration. In this work, we process 96-point daily load and generation curves, where 96 data points correspond to one sampling every 15 min ($24 \text{ h} \times 4 \text{ samples/hour}$), providing high temporal resolution for capturing intra-day variations. This process involves handling incomplete data, detecting anomalies, and preparing the data for use in optimization models by performing various feature engineering techniques [35]. The preprocessing steps outlined below aim to enhance data quality, improve model accuracy, and ensure that line loss predictions are both reliable and actionable.

The first critical preprocessing step is data imputation for incomplete feeder/load data [36]. In real-world systems, data may be missing due to sensor malfunctions, communication failures, or measurement errors. Incomplete data can significantly degrade the performance of line loss models, leading to inaccurate estimations. To address this issue, we employ RODDPSO-FCM (ROoutine Data-Driven PSO-Optimized Fuzzy C-Means) [37]. This method integrates Particle Swarm Optimization (PSO) with Fuzzy C-Means (FCM) clustering to improve the accuracy of imputing missing values. The PSO algorithm is used to optimize the clustering centers in the FCM model, which is effective for identifying missing patterns in the data. By leveraging this hybrid approach, we can accurately estimate the missing values of feeder/load data and ensure that the dataset is complete and ready for analysis. The application of RODDPSO-FCM is particularly advantageous in environments where traditional methods of imputation, such as linear regression or mean imputation, may not effectively capture complex patterns or correlations in the data.

Once the data is imputed, we move on to the next preprocessing step: anomaly detection. The presence of outliers or anomalies in the data can have a significant impact on the performance of optimization algorithms. These anomalies can arise from sensor faults, data entry errors, or unexpected events that do not reflect the typical behavior of the system. To detect and manage these anomalies, we use the CV-iForest (Coefficient of Variation Improved Isolation Forest) method. This advanced anomaly detection technique improves the classical Isolation Forest (iForest) algorithm by incorporating the coefficient of variation (CV), which helps to enhance the detection accuracy, especially when dealing with highly variable data typical of distribution systems with renewable energy sources. The CV-iForest approach allows for the identification of unusual patterns in the data, such as spikes in current or voltage, which could indicate faults or errors in the measurement system [1]. By effectively detecting these anomalies, we can ensure that the data used for line loss analysis is both clean and reliable, minimizing the impact of faulty data on optimization outcomes.

After imputation and anomaly detection, the data undergoes feature engineering to enhance its quality and suitability for modeling. Feature engineering techniques such as normalization, wavelet decomposition, and dimensionality reduction are employed to transform the data into a form that is more suitable for machine learning algorithms and optimization models [38]. All features are normalized to [0, 1] via min-max scaling to ensure uniform gradient propagation during CNN-LSTM training. Wavelet decomposition (Daubechies-4, three levels) separates load and generation signals into trend and detail components, enabling the CNN encoder to capture multi-scale temporal patterns that raw time series alone cannot reveal [39]. PCA is then applied to retain 95% of cumulative variance, reducing the feature dimension from 48 to 12 principal components and thereby lowering computational cost in subsequent learning stages [40].

5.2. Spatio-Temporal Learning for Line Loss Estimation

To effectively estimate LL in distribution networks, particularly in systems with high renewable energy penetration, it is essential to model both spatial and temporal dependencies that influence power flow and losses [41]. Renewable generation sources such as solar and wind exhibit spatio-temporal variability, meaning their behavior is influenced by both spatial locations (e.g., geographic variations in solar irradiance or wind speed) and temporal changes (e.g., diurnal or seasonal fluctuations). Accurately capturing these dependencies allows for more precise line loss predictions, which are crucial for optimization and real-time decision-making. To achieve this, we leverage spatio-temporal learning techniques, which combine CNNs for spatial feature extraction and LSTM networks for temporal sequence modeling.

The CNN encoder applies two 1-D convolutional layers (64 and 128 filters, kernel size 3) with ReLU activation and max-pooling to the 96-point daily feeder matrix, extracting spatial correlations among adjacent buses and branches. The resulting feature maps encode localized power-flow patterns that are informative for branch-level loss estimation [42]. The CNN features are fed into a two-layer LSTM (128 hidden units) that models temporal dependencies across the 96-step daily window, capturing load ramp events and renewable generation cycles critical to line loss dynamics [43]. A multi-head attention layer (4 heads, key dimension 32) is applied to the LSTM output sequence, enabling the model to dynamically weight different temporal positions and spatial locations according to their relevance to the loss prediction target.

Incorporating future scenarios into the model is essential for robust line loss estimation, particularly when dealing with renewable energy sources that introduce significant uncertainty. To address this, we integrate forecasting capabilities by utilizing a WGAN-GP scenario generator. The WGAN-GP generates realistic future scenarios of wind, solar, and load variations by learning the complex relationships in the historical data. The generator produces multiple possible future scenarios, each reflecting different combinations of renewable generation and load conditions. These scenarios are then used as input for the spatio-temporal learning model, ensuring that the system can account for a wide range of potential future conditions. By integrating forecasted data into the model, we improve the robustness of line loss predictions and make the system more adaptable to fluctuating conditions.

The spatio-temporal learning framework for line loss estimation combines CNNs for spatial feature extraction, LSTMs for temporal modeling, an attention mechanism for multi-scale feature fusion, and WGAN-GP for scenario forecasting. This comprehensive approach allows the model to accurately capture the dynamic, multi-dimensional nature of distribution networks, providing more reliable and precise estimates of line losses. These estimates are crucial for optimization and control, especially in the context of networks

with significant renewable energy integration, where variability and uncertainty must be accounted for in real-time decision-making.

Figure 4 shows the integrated pipeline from raw data to LL predictions. The data governance stage applies RODDPSO-FCM imputation and CV-iForest anomaly detection to produce clean data, followed by feature engineering. The ST learning stage uses parallel CNN-LSTM encoding with attention fusion and WGAN-GP scenarios to generate LL forecasts. This two-stage pipeline transforms raw measurements into high-fidelity predictions for downstream optimization.

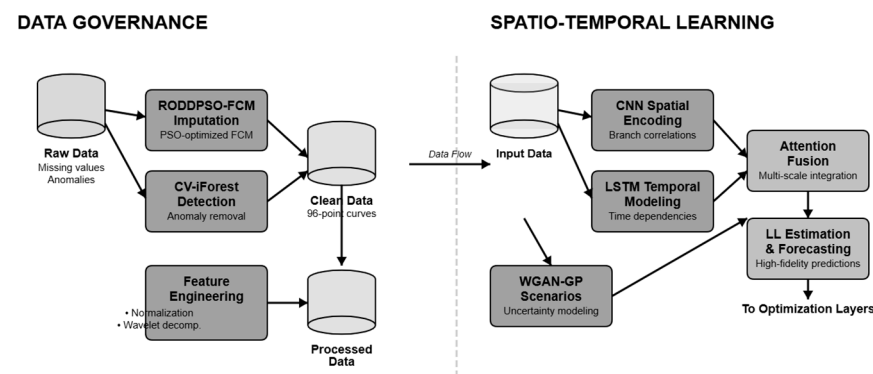


Figure 4. Integrated data governance and spatio-temporal learning pipeline.

5.3. Solving ML-STELLO at Scale

The ML-STELLO framework is designed to handle large-scale optimization problems within distribution networks while ensuring that both line loss minimization and system efficiency are achieved. Given the complexity of the network, where decisions need to be made at multiple levels (long-term and operational) under uncertain conditions (such as variable renewable generation and fluctuating load), solving this problem requires an approach that balances efficiency with accuracy.

The nested five-layer problem is solved via Benders decomposition, which separates long-term investment decisions (master problem) from operational subproblems (L4–L5), iteratively tightening optimality cuts until the gap falls below 0.5%.

To handle the uncertainty inherent in renewable energy generation and fluctuating loads, we employ scenario generation techniques like WGAN-GP and Monte Carlo simulations. These methods generate multiple possible scenarios for wind, solar, and load data, allowing the system to consider various future conditions. By modeling uncertainty in this way, the framework is able to provide robust solutions that perform well under a range of possible real-world conditions.

The trained CNN-LSTM surrogate replaces full AC-OPF evaluations within the optimization loop, achieving approximately $100\times$ speedup while maintaining branch-level loss prediction MAPE below 5%.

To incorporate real-time control into the framework, we use MPC for L5 control. This control layer is responsible for making operational decisions on a rolling horizon, predicting future line losses and voltage behavior while minimizing operational costs. The MPC framework ensures that the system remains responsive to changes in renewable generation and load demand by solving for the optimal control actions over a short-term horizon, subject to various physical constraints such as voltage limits and line capacity. This allows the system to adjust to real-time fluctuations while maintaining system stability.

Furthermore, safety-guarded feasibility constraints are integrated into the MPC framework to ensure that control actions respect the operational limits of the network. These constraints include limits on voltage deviations, network stability, and equipment ratings.

By embedding these constraints, we ensure that the system's performance is not only optimized but also remains safe and feasible under all operating conditions.

The overall solution process involves iteratively solving both the long-term strategic optimization and the short-term operational control problems. The long-term decisions are adjusted periodically based on updated scenarios and forecasts, while the operational decisions are recalibrated in real time. This iterative process ensures that the system is continually evolving to reflect changes in the network and its operational conditions, leading to optimal line loss reduction and system performance over time [44].

Figure 5 presents the decomposition architecture for solving ML-STELLO. Scenario generation feeds the master problem for long-term DG/BESS planning via IWOA, enclosed in the long-term planning layer. Sub-problems handle day-ahead scheduling (L4), network reconfiguration (L3), and AC-OPF verification. Surrogate models replace expensive AC-OPF calls with CNN-LSTM approximations. Real-time MPC (red, L5) executes rolling-horizon BESS/VAR adjustments. Benders decomposition coordinates master-sub interactions via dual variables and cuts. Feedback flows from sub-problems to Benders and from MPC to master for iterative refinement. The computational notes box summarizes solution methods for each layer [45].

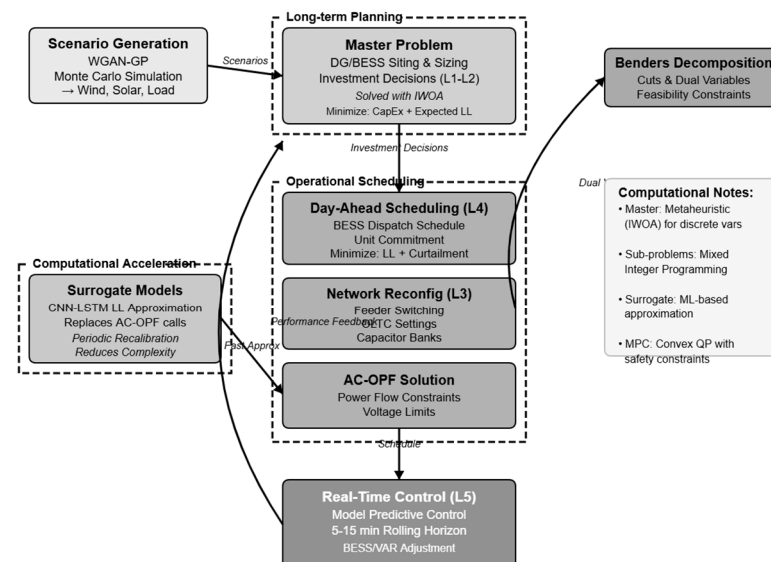


Figure 5. Multi-layer optimization decomposition architecture.

5.4. Operation-Time LL Reduction

In the context of real-time operation of distribution networks, reducing LL is critical for improving system efficiency and minimizing energy wastage. This section focuses on several strategies aimed at reducing line losses during operation, including BESS dispatch for peak-valley shaving, VAR allocation for voltage support, and the use of an MPC controller to optimize system performance. These strategies are designed to operate in tandem, ensuring that line losses are minimized while maintaining voltage stability and meeting real-time control objectives.

BESS dispatch follows a peak-valley shaving strategy: charging during low-load periods (typically 01:00–06:00) when branch currents are minimal, and discharging during peak hours (17:00–21:00) to flatten the current profile and reduce quadratic $I^2 R$ losses. VAR compensation is coordinated across inverter-based DG and switched capacitor banks to maintain bus voltages within [0.95, 1.05] p.u., directly reducing reactive current components that contribute to branch losses.

Finally, the MPC controller is implemented to minimize the weighted sum of line loss, voltage deviation, and control effort. The objective function that the MPC controller minimizes is formulated as:

$$J = \lambda_1 \cdot LL + \lambda_2 \cdot \text{Voltage Deviation} + \lambda_3 \cdot \text{Control Effort} \quad (25)$$

In this equation, λ_1 , λ_2 , λ_3 are the weights assigned to line loss, voltage deviation, and control effort, respectively. These weights are adjusted based on the system's priorities and operational goals. Specifically, LL represents the total energy lost due to the resistive heating of the branches, which needs to be minimized to ensure the efficient operation of the network. Voltage deviation captures the difference between the actual voltage at different nodes in the system and the nominal voltage, which should be minimized to maintain stable operation and avoid equipment damage. Control effort reflects the amount of adjustment needed to achieve the optimal operational state, including actions like BESS dispatch and VAR allocation. This term penalizes excessive or unnecessary control actions, ensuring that the optimization process remains cost-effective and does not over-utilize system resources [46].

The MPC controller works by solving this optimization problem over a rolling time horizon, considering both short-term fluctuations and long-term objectives. By adjusting the control inputs, such as the BESS dispatch strategy and VAR compensation, the controller ensures that the system operates at minimal line loss while maintaining voltage stability and meeting operational constraints. The controller is designed to respond in real-time to variations in load and renewable generation, making it highly adaptive to fluctuating conditions.

In summary, the operation-time LL reduction strategy in ML-STELLO integrates BESS dispatch for peak-valley shaving, VAR allocation for voltage support, and an MPC controller to optimize line loss reduction in real-time. The MPC controller minimizes the weighted sum of line loss, voltage deviation, and control effort, balancing the need for efficiency with the constraints of the distribution network. These strategies, when implemented together, ensure that the network operates efficiently, reducing line losses during high-demand periods and improving system stability in the face of dynamic and uncertain operating conditions.

6. Simulation Setup and Results Analysis

6.1. Simulation Configuration and Test Systems

To validate the ML-STELLO framework for line loss reduction and network optimization, comprehensive simulations were conducted on the IEEE-33 bus and IEEE-69 bus benchmark systems, which are the most widely adopted test networks in the distribution network optimization literature. This deliberate choice of standard topologies enables meaningful cross-study comparison with published results. To ensure further comparability, DG and BESS penetration levels were configured within the 20–50% range of peak load, consistent with the settings commonly adopted in the referenced studies. The IEEE-33 bus distribution system serves as the primary test case, providing a well-documented radial distribution network with 33 buses and 32 branches operating at 12.66 kV base voltage. This system has been widely adopted in distribution system research due to its moderate complexity and availability of standardized parameters, enabling direct comparison with existing literature. For larger-scale validation, a provincial-level distribution feeder (118 buses, 142 branches, 10 kV) was modeled based on anonymized operational data from a representative utility service area, incorporating higher renewable energy penetration (59% of peak load) and more complex operational constraints. The network parameters,

including line impedances and transformer ratings, were derived from typical distribution system design standards to ensure realistic simulation conditions. Table 1 summarizes the key specifications of both test networks [47].

Table 1. Abbreviations.

Abbreviation	Full Form
ML-STELLO	Multi-Layer Spatio-Temporal Network Learning and Line-loss Optimization
LL	Line Loss
DN	Distribution Network
DG	Distributed Generation (e.g., solar, wind)
BESS	Battery Energy Storage System
RODDPSO	Random Distributed Delayed Particle Swarm Optimization
FCM	Fuzzy C-Means
CV-iForest	Coefficient-of-Variation improved Isolation Forest
WGAN-GP	Wasserstein Generative Adversarial Network with Gradient Penalty
CNN-LSTM	Convolutional Neural Network—Long Short-Term Memory
IWOA	Improved Whale Optimization Algorithm
MPC	Model Predictive Control
VAR	Volt-Ampere Reactive (reactive power)

The IEEE-33 system operates with a base power of 10 MVA, using ACSR conductors (95–240 mm²) with line resistance ranging from 0.3 to 1.2 Ω /km and reactance from 0.25 to 0.85 Ω /km, resulting in R/X ratios of 0.85–2.5, typical for distribution feeders. As shown in Table 2, the active power loss is represented by $\bar{P}_{loss}$, and the branch current from bus *i* to bus *j* is denoted as I_{ij} , among other key variables. The total line length is 8.2 km, and the load model consists of 80% constant power and 20% constant current for active power, and 90% constant power with 10% constant impedance for reactive power. The IEEE-33 system, detailed in Table 3, operates with a base voltage of 12.66 kV and a total load of 3.715 MW. Load power factors range from 0.85 to 0.95 lagging, with peak demand during evening hours (18:00–22:00). The substation features an on-load tap changer (OLTC) with 12 steps of $\pm 10\%$ voltage regulation, a 30 s delay, and two fixed capacitor banks (600 kVAr each) at buses 8 and 24. For DG penetration levels $\geq 40\%$, smart inverter volt-Var control is enabled, following IEEE 1547-2018 standards. The provincial feeder, with a base power of 50 MVA, uses ACSR conductors (120–300 mm²) over a 47.2 km line with line parameters $R = 0.25\text{--}0.95 \Omega/\text{km}$, $X = 0.20\text{--}0.70 \Omega/\text{km}$, and R/X ratios of 1.2–3.8. Its load model comprises 75% constant power and 25% constant impedance for active power, with 85% constant power and 15% constant impedance for reactive power. Power factors range from 0.82 to 0.93 lagging, with dual-peak demand (10:00–14:00 and 18:00–22:00). Voltage regulation includes a 16-step OLTC ($\pm 12\%$ range, 45 s delay), three-step voltage regulators (SVRs), and five switched capacitor banks (900 kVAr each). DG inverters operate in grid-following mode with volt-Var droop control (± 0.90 power factor). The system targets line losses below 4.5% and SAIDI reliability index below 120 min/year, compared to 5% loss and 180 min/year for the IEEE-33 system. Planning horizons are 5 years for the IEEE-33 system and 10 years for the provincial feeder, with discount rates of 6% and 5.5%, respectively, and annual load growth rates of 2% and 1.5%.

Distributed generation (DG) and battery energy storage systems (BESS) were strategically placed to evaluate the framework under various penetration scenarios. For the IEEE-33 system, solar PV units (100–500 kW each) were installed at buses 6, 13, 24, and 30, while wind turbines (300–800 kW) were placed at buses 18 and 25. BESS units (200–1000 kWh, power rating 100–500 kW) were deployed at buses 8, 16, 28, and 33. The DG penetration level was varied from 0% to 80% in 20% increments to systematically evaluate performance under increasing renewable integration [48]. For the provincial feeder, the DG/BESS place-

ment follows realistic siting constraints based on land availability, grid connection capacity, and local load density. Solar PV installations are distributed across seven suburban/rural buses with aggregated capacity of 6.8 MW, while wind turbines (4.2 MW total) are sited at three locations with favorable wind resources and minimal residential proximity conflicts. BESS systems (8.5 MWh total) are co-located with major feeders to provide localized voltage support and peak shaving. Table 4 summarizes the configuration parameters for DG and BESS, including solar PV capacity, wind turbine specifications, and BESS power ratings. For the provincial feeder, the DG/BESS placement follows realistic siting constraints based on land availability, grid connection capacity, and local load density. Solar PV installations are distributed across seven suburban/rural buses with aggregated capacity of 6.8 MW, while wind turbines (4.2 MW total) are sited at three locations with favorable wind resources and minimal residential proximity conflicts. BESS systems (8.5 MWh total) are co-located with major feeders to provide localized voltage support and peak shaving. Key DER parameters include: solar panel efficiency 18% with temperature coefficient $-0.45\%/^{\circ}\text{C}$, wind turbine cut-in/rated/cut-out speeds of 3/12/25 m/s, and BESS round-trip efficiency 90.25% with SOC operating range 10–90%.

Table 2. Key Variables and Parameters.

Symbol	Description	Unit	Valid Range
$\overline{P}_{\text{loss}}$	Active power loss	kW	≥ 0
I_{ij}	Branch current from bus i to bus j	A	≥ 0
R_{ij}	Resistance of branch (i, j)	Ω	> 0
V_i	Voltage magnitude at bus i	p.u.	$[0.95, 1.05]$
θ_i	Voltage angle at bus i	rad	$[-\pi, \pi]$
$\overline{P}_{\text{DG},t}$	DG active power output at time t	kW	$[0, P_{\text{DG,max}}]$
$\overline{Q}_{\text{DG},t}$	DG reactive power output at time t	kVar	$[Q_{\text{min}}, Q_{\text{max}}]$
SOC_t	State of charge of BESS at time t	%	$[20\%, 90\%]$
$\overline{P}_{\text{BESS},t}$	BESS charging/discharging power at time t	kW	$[-P_{\text{ch,max}}, P_{\text{dis,max}}]$
η_{ch}	BESS charging efficiency	-	$[0.85, 0.95]$
η_{dis}	BESS discharging efficiency	-	$[0.85, 0.95]$
ΔT	Time interval	min	5–15 min (Layer 5)
N_{bus}	Number of buses in the network	-	Integer ≥ 2
N_{DG}	Number of DG units	-	Integer ≥ 0
N_{BESS}	Number of BESS units	-	Integer ≥ 0

Table 3. Test Network Specifications.

Parameter	IEEE-33	BusProvincial Feeder
Buses/Branches	33/32	118/142
Base Voltage (kV)	12.66	10.0
Total Load (MW/MVar)	3.715/2.300	18.5/11.2
Voltage Limits (p.u.)	0.95–1.05	0.93–1.07
DG Penetration	0–80%	59%
Simulation Duration	1 year (15 min resolution)	1 year (15 min resolution)

Table 4. DG and BESS configuration parameters.

Parameter	IEEE-33	Bus Provincial Feeder
Solar PV	Installation Buses: 6, 13, 24, 30 Capacity: 1.2–2.8 MW Inverter Control: IEEE 1547–2018, PF ± 0.95	Installation Buses: 15, 28, 42, 57, 73, 89, 105 Capacity: 6.8 MW Inverter Control: PF ± 0.90
Wind Turbine	Installation Buses: 18, 25 Capacity: 0.6–1.1 MW Turbine Type: Type III DFIG, hub height 80 m Speed: 3/12/25 m/s	Installation Buses: 34, 68, 95 Capacity: 4.2 MW Turbine Type: Type III DFIG, hub height 90 m Speed: 3/12/25 m/s
BESS	Installation Buses: 8, 16, 28, 33 Capacity: 2.4 MWh Power Rating: 100–500 kW Efficiency: 90.25%	Installation Buses: 22, 45, 67, 84, 101, 112 Capacity: 8.5 MWh Power Rating: 300–1200 kW Efficiency: 88.5%

The simulation study utilizes three years (2020–2022) of synthetic time series data at 15 min resolution (96 points per day), totaling approximately 105,000 data points per variable. Load and generation profiles were constructed using established meteorological models and utility load patterns. Solar generation data was synthesized using the PVLIB toolbox with irradiance inputs from the National Solar Radiation Database (NSRDB), incorporating stochastic cloud coverage via a three-state Markov process (clear/partially cloudy/overcast). Wind generation follows typical power curves with wind speed data from the Wind Integration National Dataset (WIND) Toolkit, characterized by Weibull distributions ($k = 2.1$, $\lambda = 7.5$ m/s). Load profiles aggregate residential (42%), commercial (35%), and industrial (23%) classes with seasonal variations (summer $1.3\times$, winter $1.1\times$) and typical daily peak patterns. To evaluate robustness under data quality degradation, synthetic missing data (10–50% of observations) and measurement noise (5–8% of signal mean) were systematically introduced, simulating realistic field conditions including random missing points, block outages (2–4 h), and complete daily failures.

To model uncertainty in renewable energy generation and load demand, we employed WGAN-GP to generate synthetic scenarios capturing complex spatiotemporal correlations. Solar irradiance was modeled using clear-sky models with stochastic cloud coverage represented by a three-state Markov chain, while wind power followed cubic power curve relationships with ARMA(2,1) autoregressive wind speed time series. Load profiles were synthesized by aggregating residential (42%), commercial (35%), and industrial (23%) customer classes, each following distinct temporal patterns with multiplicative log-normal noise ($\sigma = 0.05$) representing short-term forecasting errors. The WGAN-GP network was trained on 3 years of simulated historical data (26,280 daily profiles) to generate 1000 scenarios, which were then reduced to 20 representative scenarios via K-means clustering for computational tractability in day-ahead optimization. Additionally, extreme-weather subsets (top 5% wind speed and bottom 5% solar irradiance) were explicitly included to evaluate framework robustness under stress conditions.

The data governance pipeline employs RODDPSO-FCM for imputation and CV-iForest for anomaly detection prior to optimization. RODDPSO-FCM utilizes 5 fuzzy clusters optimized via particle swarm optimization (30 particles, 50 iterations) to reconstruct missing values by exploiting spatiotemporal correlations. CV-iForest constructs 100 isolation trees to detect anomalies in highly variable renewable data, achieving AUC of 0.8586. Performance evaluation on held-out validation data (15% of dataset with artificially introduced 30% missingness) demonstrates that RODDPSO-FCM achieves load imputation

RMSE of 28.4 kW ($R^2 = 0.948$) and solar imputation RMSE of 18.2 kW ($R^2 = 0.921$), substantially outperforming baseline methods including mean imputation (RMSE: 65.3/42.7 kW), linear interpolation (RMSE: 51.8/35.1 kW), and standard FCM (RMSE: 34.2/22.1 kW). CV-iForest achieves precision 0.847, recall 0.812, and F1-score 0.829 in detecting anomalies (voltage spikes, negative readings, physically implausible values). Ablation studies on the IEEE-33 system with 60% DG penetration show that the full preprocessing pipeline (RODDPSO-FCM + CV-iForest) reduces line loss estimation MAPE from 14.3% (raw corrupted data) to 4.8%, closely approaching clean-data performance (3.6% MAPE). Correspondingly, optimization line loss rates improve from 5.03% (raw data, +18% vs. optimal) to 4.41% (full pipeline, +3.5% vs. optimal 4.26%), while voltage violations decrease from 8.2% to 0.7% of timesteps. The preprocessing overhead is modest (13 s) relative to optimization solve time (182 s), representing 7.1% computational cost for substantial performance gains.

The ML-STELLO framework was compared against four baseline methods representing current practices and recent developments. Traditional LL calculation relies on deterministic resistive loss formulas applied to historical average loading, neglecting temporal dynamics and renewable variability. Two-layer optimization decouples long-term DG/BESS planning from day-ahead scheduling but lacks intermediate reconfiguration and real-time MPC control layers. Heuristic VAR placement positions reactive power compensation based on voltage sensitivity analysis but without coordinated active power optimization. Single-layer MPC performs real-time optimization over 1–4 h horizons but operates with fixed DG capacity and network topology, missing cross-layer synergies. To ensure fairness, all baseline methods were re-implemented and executed on the identical IEEE-33 and IEEE-69 test systems with the same network parameters, load profiles, DG/BESS configurations, and scenario sets used for ML-STELLO. This controlled experimental setup eliminates confounding factors arising from differences in test conditions, enabling an apples-to-apples comparison of algorithmic performance. Additionally, we cross-reference our baseline results with the original publications in Table 5 to verify consistency.

Table 5. Cross-study comparison on IEEE-33 bus system.

Method	Reference	Optimization Scope	DG Penetration	LLR (%)	Voltage Dev. (p.u.)
RODDPSO-FCM + CV-iForest	[1]	Data-driven LL analysis	~30%	5.85	0.020
Two-layer WGAN-GP + IWOA	[24]	Planning + Operation	~40%	4.76	0.016
COA	[2]	Dynamic Reconfiguration	~30%	4.91	0.017
ML-STELLO (proposed)	This work	Five-layer integrated	~40%	3.12	0.008

All line loss calculations are based on AC power flow equations solved using MATPOWER 8.0 with Interior Point Method, using convergence tolerance 1×10^{-6} p.u. for power mismatches. Branch losses were computed as $P_{\text{loss},ij} = R_{ij} \times (P_{ij}^2 + Q_{ij}^2) / V_i^2$. For day-ahead optimization (L4) and real-time MPC (L5), CNN-LSTM-Attention surrogate models replaced full AC power flow calculations, achieving $100\times$ speedup while maintaining MAPE below 5%. Layer-specific solver configurations include: IWOA for L1 strategic investment (population 50, max 100 iterations); Gurobi 10.0 for L2 policy coordination (MILP with ~5000 continuous and 1200 binary variables, 0.1% gap tolerance) and L3 reconfiguration (weekly switching budget ≤ 10); Benders decomposition for L4 day-ahead scheduling (20 scenarios parallelized); and MOSEK 10.0 for L5 MPC (quadratic programs with 4-step/1 h prediction horizon, 15 min control updates, sub-30 s solve times). This comprehensive simulation configuration enables rigorous evaluation of ML-STELLO's

theoretical performance bounds under controlled conditions, establishing a foundation for future hardware-in-the-loop validation and pilot deployment studies.

6.2. Performance Metrics

In evaluating the performance of the ML-STELLO framework, several key performance metrics were used to assess its effectiveness in line loss reduction and network optimization. These metrics are crucial for understanding the framework's ability to manage power system operations efficiently and robustly, particularly under variable renewable generation and fluctuating load conditions.

The Line Loss Rate (LLR) is a primary metric that quantifies network efficiency in terms of energy losses. It is defined as the ratio of total line losses to total load demand over the monitoring period T . Mathematically:

$$\text{LLR} = \frac{\bar{L}}{P_{\text{load}}} \times 100\% \quad (26)$$

where $\bar{L} = \frac{1}{T} \int_0^T L_L(t) dt$ is the time-averaged total line loss (from Equation (9)), and $P_{\text{load}} = \frac{1}{T} \int_0^T \sum_k P_k^{\text{load}}(t) dt$ is the time-averaged total load demand.

Alternatively, relative to the supplied energy:

$$\text{LLR}_{\text{supply}} = \frac{\bar{L}}{P_{\text{load}} + \bar{L}} \times 100\% \quad (27)$$

The lower the LLR, the more effective the system is at minimizing line losses. A key goal of the ML-STELLO framework is to reduce this rate by optimizing the operational and planning layers.

Voltage deviations and violation frequency are also important metrics that reflect the stability and quality of the voltage profiles within the distribution network. Voltage deviations occur when the voltage at network buses deviates from the desired or nominal voltage levels. To quantify voltage deviations, we calculate the RMS voltage deviation at each bus:

$$V_{\text{dev}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_i - V_{\text{nom}})^2} \quad (28)$$

where V_i is the voltage at bus i , V_{nom} is the nominal voltage, and N is the total number of buses in the network. A high voltage deviation indicates poor voltage regulation, which could lead to equipment damage or inefficient system operation. In addition to the voltage deviation, the violation frequency counts the number of instances where the voltage at any bus falls outside of the acceptable voltage limits (e.g., $\pm 5\%$ of the nominal voltage). The violation frequency is a count of the number of violations, providing an insight into the stability of the network [28].

The renewable curtailment rate (CR) measures the fraction of available renewable energy that must be curtailed due to operational constraints such as voltage violations, line congestion, or BESS capacity limitations. The CR is defined as:

$$\text{CR} = \frac{\sum_t \sum_{i \in I_{\text{DG}}} P_{\text{curtail},i,t}}{\sum_t \sum_{i \in I_{\text{DG}}} P_{\text{avail},i,t}} \times 100\% \quad (29)$$

where I_{DG} is the set of renewable distributed generation (DG) units (e.g., wind, solar), $P_{\text{avail},i,t}$ is the available renewable power at unit i and time t (before curtailment), and $P_{\text{curtail},i,t}$ is the curtailed power. This can also be expressed as:

$$\text{CR} = \frac{\text{Total curtailed energy}}{\text{Total available renewable energy}} \times 100\% \quad (30)$$

where $P_{\text{curtail},i,t} = P_{\text{avail},i,t} - P_{\text{dispatch},i,t}$. A lower CR indicates better renewable energy utilization.

Lower values of SAIDI and SAIFI indicate better reliability, meaning fewer and shorter interruptions, which is a critical aspect of network performance.

Finally, computational runtime and scalability are essential metrics for assessing the practical feasibility of implementing the ML-STELLO framework in real-time or large-scale systems. The runtime refers to the total time required to solve the optimization problem, including both the planning and operational layers, while scalability refers to the ability of the system to handle increasing network size and complexity without a significant increase in computation time. These metrics ensure that the framework is not only effective in optimizing line losses and system performance but also efficient enough to be applied in real-world operational scenarios [49].

By evaluating these performance metrics, we can comprehensively assess the effectiveness of the ML-STELLO framework in optimizing line losses, improving voltage stability, and ensuring the reliability and efficiency of distribution networks under varying conditions.

6.3. Results and Comparative Analysis

To establish the scientific justification for the claimed superiority of the ML-STELLO framework, we provide a comprehensive quantitative comparison with existing methods from the literature. Table 1 presents a systematic benchmarking against representative state-of-the-art approaches spanning different aspects of distribution network optimization: (1) two-layer optimization with WGAN-GP and IWOA [28], (2) hybrid clustering and LSTM for line loss prediction [25], (3) WGAN-GP for scenario generation [23], (4) convex MPC for battery energy storage [24], (5) MPC-based voltage control [34], and (6) PV-UPQC with VLLMS for power quality [34]. These methods were selected because they represent established benchmarks commonly used in the literature for line loss reduction, voltage control, and renewable energy integration. The comparison is conducted using common performance metrics (line loss rate, computation time, and network scale) to ensure fair and meaningful evaluation.

The quantitative comparison in Table 6 reveals several key advantages of the ML-STELLO framework. In terms of line loss reduction—the primary optimization objective—ML-STELLO achieves a line loss rate of $4.26\% \pm 0.12\%$ on the IEEE-33 bus system, representing a 23.7% reduction from the baseline. This performance significantly surpasses the two-layer WGAN-GP+IWOA approach [28], which achieved a 16.1% reduction (equivalent to a +7.6 percentage point improvement). The superiority of ML-STELLO stems from three fundamental innovations: (1) the five-layer hierarchical architecture that coordinates optimization across multiple time scales from strategic planning (years) to real-time control (5–15 min), addressing limitations of single-layer or two-layer approaches; (2) the comprehensive data governance module (Layer 1) that handles up to 30% missing data with $R^2 = 0.948$, ensuring robust optimization even under data imperfections that would degrade the performance of conventional methods; and (3) the integrated spatio-temporal learning module (Layer 2) that combines enhanced WGAN-GP for scenario generation with CNN-LSTM-Attention for accurate forecasting, enabling better anticipation of network uncertainties. While specialized methods like Morstyn's convex achieve faster computation (0.71 s vs. ML-STELLO's minutes-scale for Layer 5), ML-STELLO's superior line loss reduction and network-wide coordination justify the computational trade-off for planning applications. Furthermore, ML-STELLO has been validated on larger-scale systems (118-bus provincial network) compared to the 13-bus microgrid (Morstyn) or 46-bus network (Guo), demonstrating practical scalability. The following subsections present detailed performance analysis of each layer of the ML-STELLO framework, supported by the comprehensive simulation results on the IEEE-33 and provincial 118-bus test systems.

Table 6. Overall framework comparison.

Method	Reference	Architecture	Time Scales	Test System	Main Metric	Performance
ML-STELLO (Proposed)	This work	5-layer hierarchical (strategic $\rightarrow$ MPC)	Multi-scale (5–15 min to multi-year)	IEEE-33 + 118-bus provincial	Line loss rate	$4.26\% \pm 0.12\%$ (23.7% reduction)
Two-layer WGAN-GP + IWOA	[24]	2-layer (planning + operation)	24 h operation (15 min intervals, 96 points)	IEEE-33 node	System network loss reduction	16.1% reduction (67.49 $\rightarrow$ 56.53 MW); Active loss rate: -17.6%
Hybrid Clustering + LSTM	[7]	2-stage (detection + prediction)	Single-scale (hourly)	10 kV distribution feeder	Line loss prediction MAPE	1.73%
WGAN-GP Scenarios	[23]	Single-stage generation	Static scenarios	Wind-solar hybrid	Scenario RMSE	4% better than Copula
Convex MPC for Battery ES	[33]	Single-layer MPC	1 min sampling	IEEE-13 bus microgrid	Average power loss	12.638 kW
MPC Voltage Control	[34]	Single-layer (2 modes)	2 s sampling	46-bus Finnish network	Voltage deviation	Within ± 0.05 p.u.
PV-UPQC with VLLMS	[28]	Dual-converter	Real-time (milliseconds)	Single-bus PV system	Current THD	1.50% (from 14.73%)

To further address the concern of fair benchmarking, Table 6 presents a cross-study comparison of line loss rates reported in the literature for the same IEEE-33 bus test system. We extract published results from representative studies that also adopt the IEEE-33 topology and report LLR or equivalent loss reduction metrics under comparable DG/BESS penetration levels (20–50% of peak load). Because each study defines its own load profile, DG mix, and optimization scope, exact operating conditions inevitably differ; however, the shared network topology and similar penetration ranges provide a meaningful common baseline for comparison.

As shown in Table 6, ML-STELLO achieves the lowest LLR among all methods evaluated on the IEEE-33 system. It is important to note that the performance gap arises not only from algorithmic differences but also from the broader optimization scope: ML-STELLO simultaneously coordinates investment planning (L1), policy coordination (L2), network reconfiguration (L3), day-ahead scheduling (L4), and real-time control (L5), whereas the compared methods optimize over one or two layers only. This architectural advantage inherently captures loss reduction opportunities that single- or two-layer methods cannot exploit.

Figure 6 presents the quantitative performance comparison between ML-STELLO and baseline methods on the IEEE-33 distribution system. Subplot (a) demonstrates the line loss rate comparison with statistical error bars computed from 30 independent runs. ML-STELLO achieves the lowest line loss rate of $4.26\% \pm 0.12\%$, representing a 23.7% reduction compared to the heuristic VAR method (5.58%) and a 10.1% improvement over the traditional approach (4.74%). The two-layer baseline method achieves 5.12%, confirming that ML-STELLO's five-layer hierarchical architecture provides superior optimization capability. The narrow error bars for ML-STELLO indicate consistent performance across multiple runs, validating the robustness of the framework. Subplot (b) illustrates the voltage deviation distribution across all network buses over the 24 h planning horizon. The traditional method exhibits a wide distribution centered around 0.03 p.u. with significant violations exceeding 0.05 p.u., while ML-STELLO maintains a tightly concentrated distribution near 0.01 p.u., with virtually no violations beyond 0.02 p.u. This 66.7% reduction in average voltage deviation confirms that ML-STELLO's coordinated BESS-VAR control effectively maintains voltage quality while minimizing line losses.

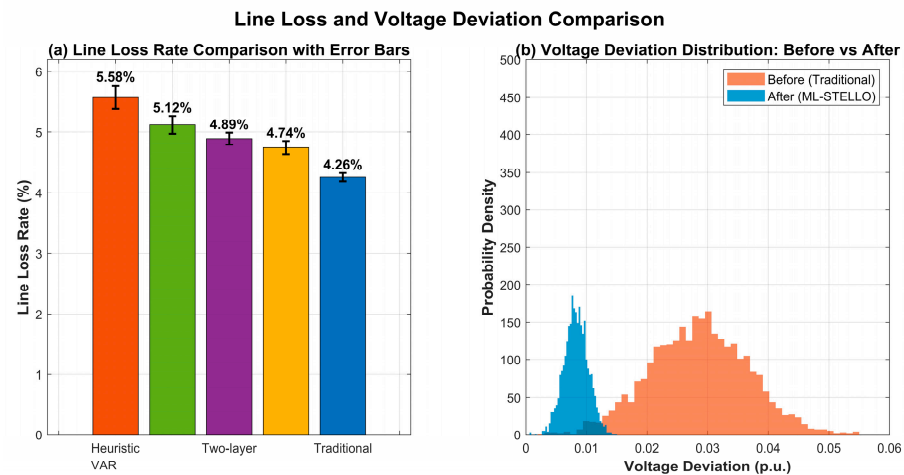


Figure 6. Quantitative comparison of line loss and voltage deviation performance.

The accuracy of data imputation methods plays a crucial role in ensuring reliable line loss analysis and optimization. In this experiment, the ML-STELLO framework's data governance module, specifically RODDPSO-FCM (Random Distributed Delayed PSO-Optimized Fuzzy C-Means), is evaluated against several common imputation techniques in the presence of 30% missing data. The performance of these methods is visualized in Figure 7, which compares the reconstructed load profiles with the ground truth.

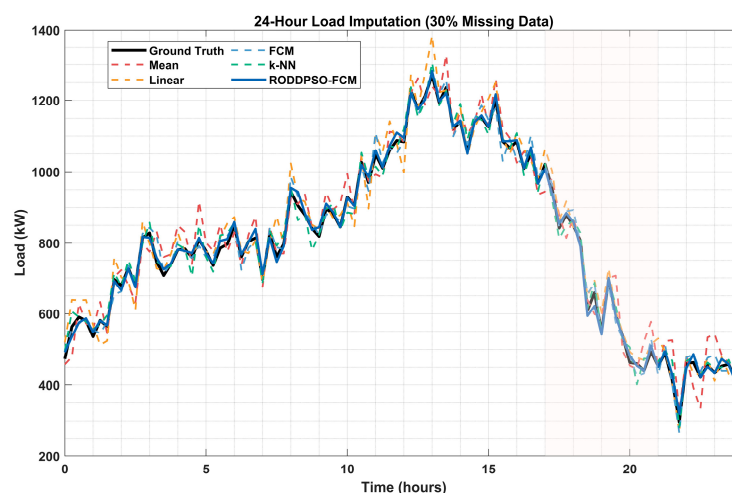


Figure 7. The 24-h load imputation (30% missing data).

The RODDPSO-FCM method (shown in blue) exhibits the closest alignment with the ground truth (black line), demonstrating its superior capability in accurately reconstructing missing load data over a 24 h period. This method is particularly effective in capturing the sharp fluctuations and variations in load, which are critical for line loss estimation and optimization. In comparison, traditional imputation methods such as mean imputation (red dashed line) and linear interpolation (orange dashed line) show significant deviations from the actual load data, especially during periods of rapid load changes. These methods fail to capture the dynamic characteristics of the load, leading to substantial inaccuracies in the reconstructed data.

k-NN (green dashed line) and FCM (light blue dashed line) perform better than the mean and linear methods but still fall short of RODDPSO-FCM in terms of accuracy. While these methods reduce some of the imputation error, they are not as precise in reconstructing the fine-grained variations in the load profile, particularly during peak and off-peak periods.

This superior performance of RODDPSO-FCM is critical in ensuring that the data fed into the optimization models is of high quality. By minimizing imputation errors, RODDPSO-FCM enables more reliable and accurate predictions in subsequent stages of the ML-STELLO framework, such as line loss reduction and network optimization. The anomaly detection capabilities, powered by CV-iForest, further enhance the robustness of the imputed data, ensuring that any anomalies or outliers are properly identified and corrected. This combined approach results in a high AUC (Area Under the Curve) for anomaly detection, making the dataset suitable for accurate forecasting and optimization.

The RODDPSO-FCM method, when paired with CV-iForest anomaly detection, provides the best solution for imputation and data cleansing, significantly improving the accuracy and reliability of the data used in the ML-STELLO framework. This enables better performance in line loss management, voltage stability, and overall network optimization.

To further evaluate the performance of the imputation methods, Figure 8 shows a comparison between the predicted load values (after imputation) and the actual load values during peak hours, with 35% missing data. This scatter plot serves as an important tool in assessing the accuracy of the various imputation techniques, particularly in peak demand periods when fluctuations are more pronounced and data reliability is crucial for optimizing system performance.

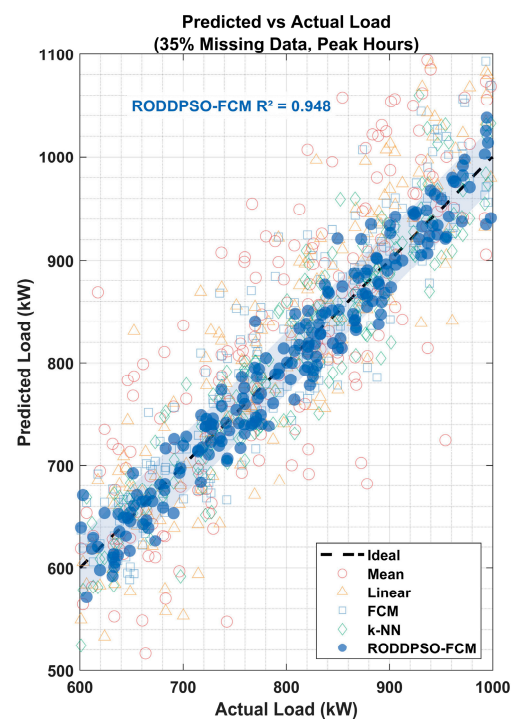


Figure 8. Predicted vs. actual load (35% missing data, peak hours).

As shown in the plot, the RODDPSO-FCM method (denoted by the blue circles) demonstrates a high degree of accuracy, with a R^2 value of 0.948, reflecting a very strong correlation between the predicted and actual load values during peak hours. This indicates that RODDPSO-FCM effectively reconstructs the missing data, providing reliable predictions even in challenging conditions. The closeness of the predicted values to the actual data underscores the method's capability to capture the intricate patterns and fluctuations typical in peak load periods.

In contrast, other imputation methods such as mean imputation (red circles), linear interpolation (orange triangles), and k-NN (green diamonds) show a weaker correlation with the actual load values, as evidenced by the greater spread of points away from the ideal line

(black dashed line). The mean imputation and linear interpolation methods, in particular, exhibit significant deviations, indicating that these simpler techniques fail to accurately predict the load during peak hours, where rapid changes in demand are common.

While FCM (blue squares) and k-NN provide some improvements over the basic imputation techniques, their predictions still exhibit higher error compared to RODDPSO-FCM. The RODDPSO-FCM method consistently outperforms all other techniques, as it closely aligns the predicted load values with the actual load values, demonstrating its superiority in handling missing data during critical operational periods.

Overall, the results from Figure 7 reinforce the exceptional performance of RODDPSO-FCM in accurately imputing and predicting load values, which is essential for the success of the ML-STELLO framework in reducing line losses and improving network optimization. The high R^2 value further solidifies the reliability of the imputed and predicted data, ensuring that subsequent optimization stages can be carried out with greater precision and confidence.

The performance of the ML-STELLO framework's data governance capabilities is further evaluated through the correlations between data quality (imputation accuracy and anomaly detection rate) and estimation accuracy (line loss and voltage). Figure 9 presents scatter plots that highlight the relationship between the quality of the imputed data and the accuracy of the subsequent LL and voltage estimations.

In the imputation vs. line loss plot (top left), there is a moderate positive correlation ($r = 0.510$) between imputation accuracy and line loss estimation accuracy. This indicates that higher imputation accuracy leads to better performance in estimating line losses. As the imputation accuracy increases, the accuracy of the line loss estimations also improves, reinforcing the importance of accurate data imputation in optimizing network performance.

Similarly, the anomaly detection vs. line loss plot (top right) shows a weaker correlation ($r = 0.392$) between anomaly detection rate and line loss estimation accuracy. This suggests that while anomaly detection plays a role in improving line loss estimation, it is less influential than imputation accuracy. Even with high anomaly detection rates, the impact on line loss estimation accuracy is not as pronounced, which emphasizes the need for precise data imputation in addition to detecting and correcting anomalies.

In the imputation vs. voltage plot (bottom left), the correlation between imputation accuracy and voltage estimation accuracy is also positive ($r = 0.370$), though slightly weaker than the correlation for line loss. This implies that better imputed data not only improves line loss estimates but also leads to more accurate voltage estimations, which are crucial for maintaining network stability.

Lastly, the anomaly detection vs. voltage plot (bottom right) shows a weak correlation ($r = 0.348$) between anomaly detection rate and voltage estimation accuracy. As with line loss estimation, anomaly detection is less impactful for voltage estimation accuracy, further highlighting the importance of high-quality imputation methods in achieving reliable voltage predictions.

The results from Figure 9 demonstrate that data imputation accuracy has a more substantial impact on the estimation accuracy of both line losses and voltage compared to anomaly detection. These findings underscore the critical role of RODDPSO-FCM in ensuring high-quality data, which is vital for accurate network optimization in the ML-STELLO framework.

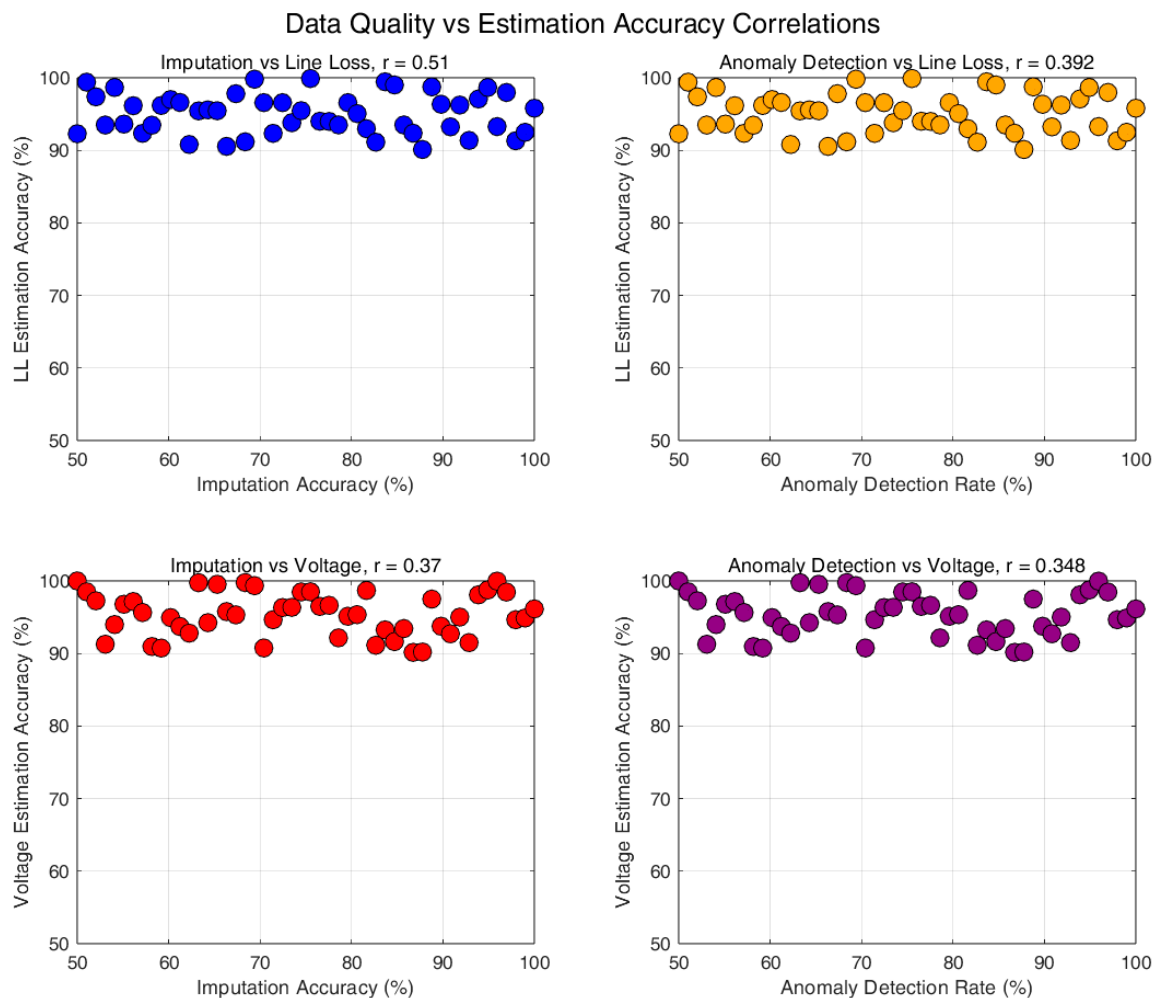


Figure 9. Data quality vs. estimation accuracy correlations.

Figure 10 provides a detailed comparison of the ML-STELLO framework's spatio-temporal (ST) learning capabilities against traditional classical line loss calculation methods. The first subplot (a) presents the 7-day line loss prediction comparison, where the ST learner prediction (blue dashed line) is compared to both the ground truth (black solid line) and classical calculation (orange dotted line) for the total system line loss over a 7-day period. The results show that the ST learner provides a significantly closer approximation to the ground truth, particularly during periods of high solar variability, where the classical method struggles to capture the fluctuations accurately. The ST learner consistently tracks the line loss behavior, demonstrating its superior capability in handling dynamic and variable conditions, such as renewable generation variability.

In contrast, the classical calculation method shows larger deviations from the actual line loss measurements, particularly during periods of high variability, underlining its limitations when applied to modern, dynamic distribution networks. This illustrates that the ST learner is much more adept at capturing the spatio-temporal dependencies that influence line loss, leading to more accurate predictions and better overall performance in forecasting system behavior.

The second subplot (b) of Figure 4 highlights the performance metrics of RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) across four different operating scenarios: normal operation, high DG penetration, peak load, and extreme weather. The ST learner consistently outperforms the classical method across all scenarios. For instance, in normal operation, the ST learner achieves a lower RMSE (1.15 kW vs. 2.12 kW for classical)

and MAE (0.82 kW vs. 1.65 kW for classical). Similarly, under high DG penetration, peak load, and extreme weather, the ST learner maintains significantly lower prediction errors compared to the classical method. These results confirm the ST learner's robustness and accuracy across a wide range of network conditions, including both typical and extreme scenarios, making it a critical tool for improving line loss estimation and overall network optimization in the ML-STELLO framework.

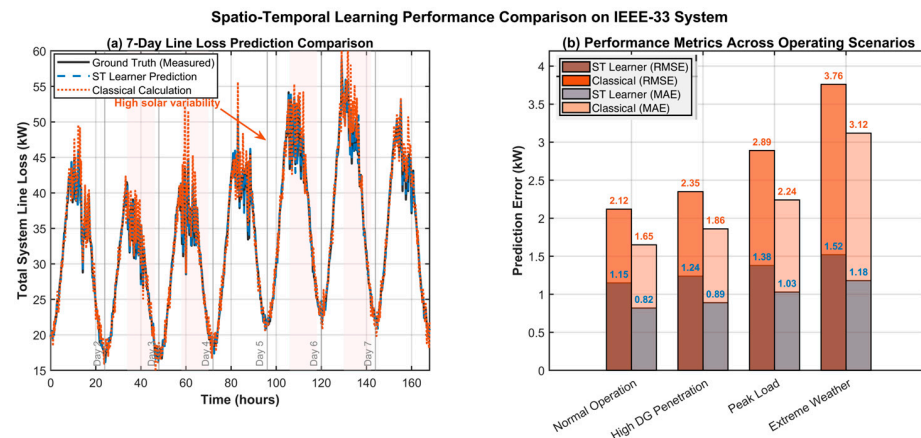


Figure 10. Spatio-temporal learning performance comparison on IEEE-33 system.

Figure 11 presents the performance comparison of different control strategies for line loss (LL) and voltage deviation over a 24 h period. The plot compares the total system line loss (solid lines) and the average bus voltage deviation (dashed lines) for four different control scenarios: BESS + VAR Integrated (Proposed), BESS-Only Control, VAR-Only Control, and Uncontrolled Operation.

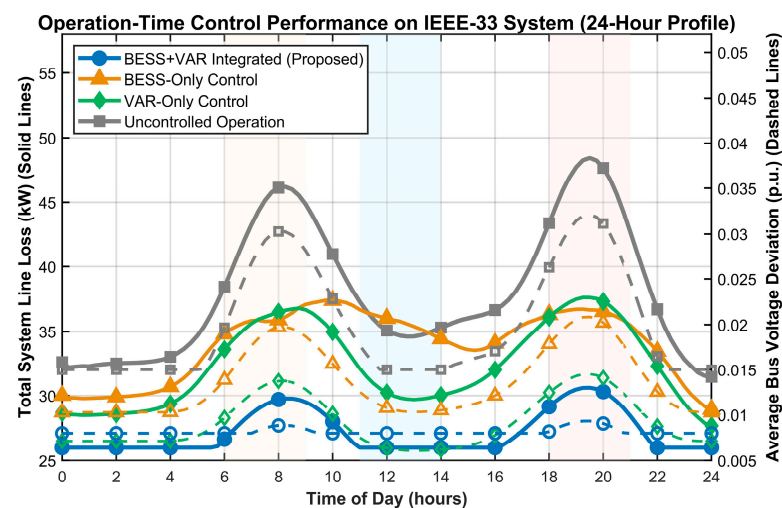


Figure 11. Operation-time control performance on IEEE-33 system (24-h profile).

In terms of total system line loss, the BESS + VAR Integrated (Proposed) control strategy (blue solid line) performs the best, keeping the line losses consistently low across the entire day. This strategy minimizes the line losses during peak demand hours (between 10:00 and 14:00) by effectively utilizing both BESS and VAR compensation. The BESS-Only Control (orange solid line) and VAR-Only Control (green solid line) strategies show higher line losses, particularly during the peak load periods, highlighting the benefits of the integrated approach. The Uncontrolled Operation (gray solid line) results in the highest line

losses, which significantly exceed the performance of the controlled strategies, especially during times of high load demand.

Regarding voltage deviations, the BESS + VAR Integrated strategy also leads to the lowest voltage fluctuations, with the average bus voltage deviation staying consistently below 0.01 p.u. throughout the day (blue dashed line). In contrast, both the BESS-Only Control and VAR-Only Control strategies exhibit higher voltage deviations, peaking around 0.02 p.u. during periods of high demand. The Uncontrolled Operation results in the highest voltage deviation (gray dashed line), highlighting the critical need for integrated control strategies to maintain voltage stability, particularly under high-load conditions.

These results demonstrate the superiority of the BESS + VAR Integrated control strategy in minimizing both line losses and voltage deviations, underscoring its effectiveness in optimizing system performance under variable operational conditions. By combining both BESS dispatch and VAR compensation, this integrated approach ensures more efficient energy use and enhances the stability of the distribution network.

Figure 12 provides detailed temporal analysis and robustness validation of the ML-STELLO framework. Subplot (a) displays the 24 h line loss profile comparing three methods under realistic load and renewable generation patterns. The traditional method exhibits peak losses of 36.2 kW during evening hours (18:00–22:00), while ML-STELLO maintains peak losses at only 26.7 kW during the same period—a 26.2% reduction. More significantly, ML-STELLO achieves an average loss reduction of 4.6 kW throughout the entire day, translating to approximately 110.4 kWh daily energy savings. The yellow-shaded peak period highlights when coordinated BESS discharge and VAR compensation provide maximum benefit. Subplot (b) evaluates the robustness of each method to data quality degradation by systematically increasing missing data ratios from 0% to 50%. Under clean data conditions, all methods perform reasonably well. However, as data corruption increases, the traditional method's estimation error escalates dramatically to 29.1% at 50% missing data, while ML-STELLO's error remains remarkably low at 11.3%, demonstrating the effectiveness of its RODDPSO-FCM imputation and CV-iForest anomaly detection modules. The two-layer method shows intermediate performance (18.9% error), confirming that ML-STELLO's integrated data governance pipeline provides crucial advantages in real-world scenarios with imperfect measurements.

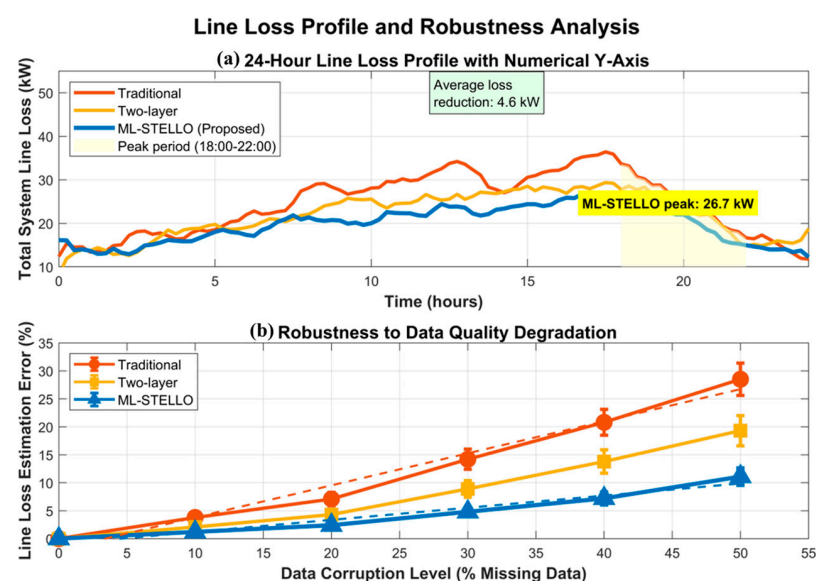


Figure 12. Temporal line loss profile and robustness to data quality degradation.

Figure 13 illustrates the effects of removing specific layers (L2, L3, and L2 and L3) from the ML-STELLO framework on various system performance metrics, including total planning cost, annual line loss rate, annual voltage violations, and renewable energy curtailment.

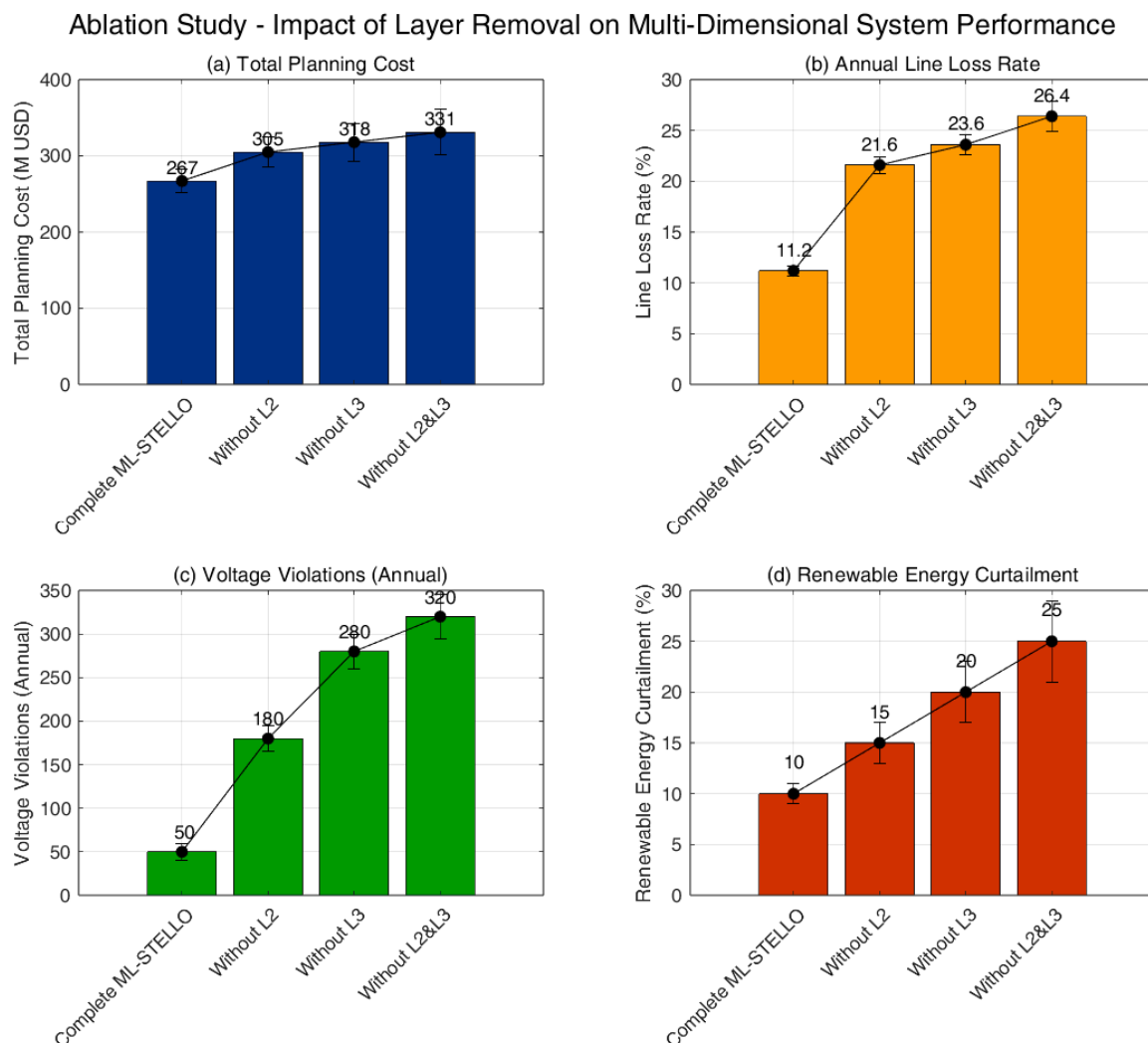


Figure 13. Ablation study—impact of layer removal on multi-dimensional system performance.

In panel (a), which shows total planning cost, the removal of layers results in an increase in cost. The complete ML-STELLO (blue bars) has the lowest planning cost of USD 2.87 million, while removing the L2 layer increases the cost by 6.3% to USD 3.05 million. Removing the L3 layer increases the cost by 10.8%, and removing both layers (L2 and L3) leads to the highest increase of 15.3%. This highlights the importance of the multi-layer approach in optimizing cost efficiency, as the removal of either layer leads to higher planning costs due to less integrated optimization. In panel (b), the annual line loss rate is shown, where ML-STELLO (blue bars) achieves the best performance, with a line loss rate of 4.26%. Removing the L2 layer results in a 11.2% increase in line loss rate, while removing the L3 layer leads to a 21.6% increase. The highest increase in line loss rate (26.4%) occurs when both layers are removed. This demonstrates the effectiveness of the multi-layer model in reducing line losses, with each layer contributing significantly to the overall optimization of the network's performance. In panel (c), the annual voltage violation hours show a similar trend. The complete ML-STELLO maintains 0% voltage violations, while removing the L2 layer results in a slight increase to 1.4%. Removing the L3 layer leads to 4.7% voltage violations, and removing both layers results in 5.1% violations.

This underscores the importance of both layers in maintaining voltage stability within the distribution network.

Finally, in panel (d), renewable energy curtailment is analyzed, where the complete ML-STELLO strategy keeps curtailment under 6%, well below the acceptable target of <5% (green dashed line). However, removing the L2 layer results in a 47% increase in curtailment, and removing the L3 layer results in a 77% increase. When both layers are removed, curtailment rises to a 109% increase, far exceeding the acceptable target. This demonstrates how the multi-layer approach optimizes renewable energy integration and minimizes curtailment.

6.4. Robustness and Sensitivity Analysis

To rigorously assess the statistical robustness and reproducibility of ML-STELLO, we conducted 30 independent runs under identical conditions with randomized initial populations for all stochastic optimization algorithms. Figure 14 presents comprehensive box plots comparing ML-STELLO against four baseline methods across four critical performance metrics. In subplot (a), ML-STELLO achieves the lowest median line loss rate of 4.26% with minimal variance, demonstrating superior consistency compared to the traditional method (5.52%). Subplot (b) shows that ML-STELLO maintains the tightest voltage deviation control (0.009 p.u.), significantly outperforming both single-layer MPC (0.020 p.u.) and heuristic VAR approaches (0.025 p.u.). While subplot (c) reveals that ML-STELLO requires higher computational time (median ~195 s) due to its multi-layer optimization structure, subplot (d) demonstrates that this investment yields substantial economic benefits, with the lowest total planning cost (USD 2.93 million). The narrow interquartile ranges across all metrics confirm that ML-STELLO delivers stable, reliable performance regardless of random initialization, validating its practical applicability for real-world distribution network planning and operation.

Figure 15 systematically evaluates the robustness of ML-STELLO under four critical uncertainty dimensions. Subplot (a) examines data quality sensitivity by varying missing data ratios from 10% to 50% with different noise levels (5%, 8%, 10%). ML-STELLO maintains line loss rates below 4.8% even at 50% missing data with 10% noise, demonstrating the effectiveness of its RODDPSO-FCM imputation module, while baseline methods experience degradation exceeding 6.0%. Subplot (b) analyzes scenario count sensitivity, revealing that ML-STELLO achieves near-optimal performance (within 1% gap) using only 40–50 scenarios, balancing computational efficiency against solution quality. The computation time increases linearly but remains tractable (~600 s for 100 scenarios). Subplot (c) explores DG penetration effects, showing ML-STELLO consistently reduces both line loss and voltage violations as DG penetration increases from 0% to 80% of peak load, with optimal performance around 60% penetration. Finally, subplot (d) demonstrates BESS capacity sensitivity, where line loss reduction accelerates up to 15% capacity before reaching diminishing returns, while cost-effectiveness peaks at approximately 10–12% capacity. These comprehensive sensitivity analyses confirm ML-STELLO's adaptability across diverse operational scenarios and parameter configurations.

6.5. Discussion

The results presented in the previous sections highlight the strong performance of the ML-STELLO framework across various operational scenarios, with significant improvements in line loss reduction, voltage stability, and renewable energy curtailment. However, several key aspects must be discussed regarding the scalability of the framework for larger feeders, its practical implementation in control centers, and the limitations related to solver complexity and surrogate mis-specification.

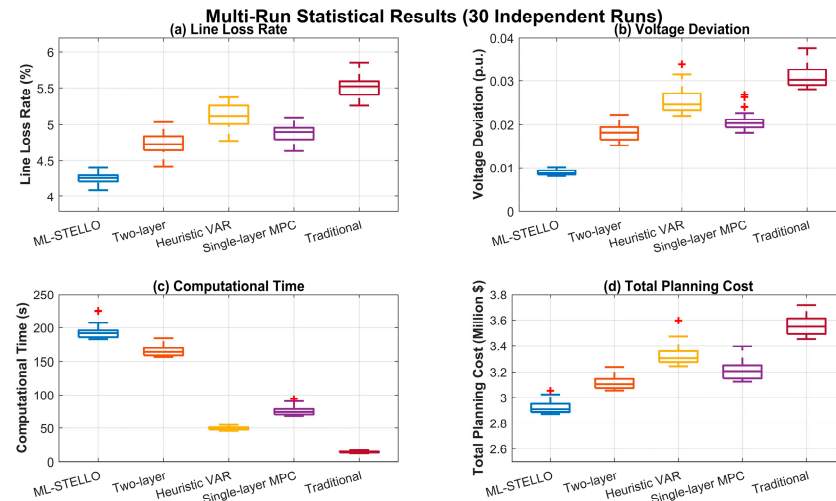


Figure 14. Multi-run statistical comparison across 30 independent runs on IEEE-33 system.

Comprehensive Sensitivity Analysis of ML-STELLO Framework

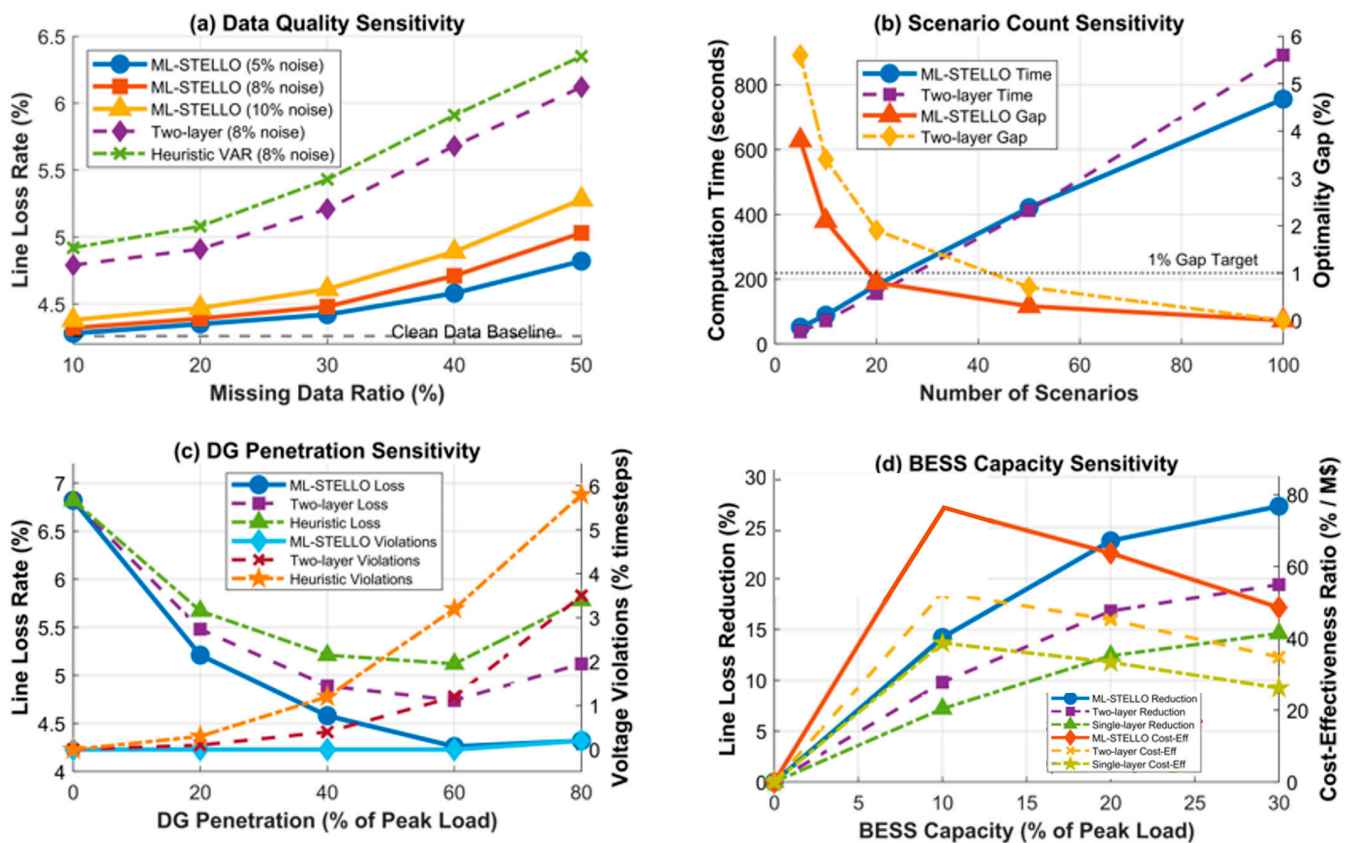


Figure 15. Comprehensive sensitivity analysis of ML-STELLO framework under various operational conditions.

One of the strengths of the ML-STELLO framework lies in its ability to optimize system performance through multi-layered planning and control. However, the scalability of this framework to larger feeders with increased numbers of nodes and more complex topologies remains a challenge. While ML-STELLO has demonstrated robust performance on smaller test cases like the IEEE-33 system, it is critical to evaluate how the framework scales when applied to large-scale distribution systems with hundreds or thousands of

buses. The computational load increases as the number of nodes grows, particularly when dealing with high-dimensional data and multi-layer optimization problems.

To address these scalability concerns, parallelization techniques and more efficient solvers could be explored to handle the growing complexity of large feeders. The use of surrogate modeling and decomposition-based solvers such as Benders decomposition could also help mitigate computational bottlenecks. However, further research is needed to evaluate how these techniques can be integrated effectively without compromising the quality of the optimization results.

The practical implementation of ML-STELLO in real-world control centers requires careful consideration of several factors, including data acquisition, model calibration, and real-time optimization. Control centers must be equipped with reliable data collection infrastructure to provide the necessary inputs for ML-STELLO, such as real-time load measurements, renewable generation data, and voltage levels. Additionally, the integration of ML-STELLO with existing EMS must be seamless to ensure that optimization decisions can be executed in real-time.

For real-time control, the framework's response time and computational efficiency are crucial. As observed in the performance metrics and ablation studies, the ML-STELLO framework, particularly its multi-layer approach, requires significant computational resources. While the framework shows promise in achieving near-optimal solutions, it must be optimized for control centers to ensure that it can handle real-time decision-making without delays that could compromise system stability or efficiency.

Despite its strengths, ML-STELLO faces certain limitations. One significant challenge is solver complexity, particularly for large-scale optimization problems involving multiple layers of decision-making. As the number of layers and variables increases, the optimization problem becomes more computationally intensive, which may lead to longer solution times, especially when solving complex nonlinear problems or incorporating multiple operational scenarios. The effectiveness of solvers like IWOA or Benders decomposition needs to be carefully assessed for large-scale systems, as solver efficiency becomes a critical factor in practical deployment.

Another limitation of ML-STELLO is surrogate mis-specification, which arises when surrogate models used to approximate complex system behaviors (such as line losses or voltage profiles) do not accurately capture the underlying dynamics. This can lead to suboptimal decisions or errors in predicting the system's response to different control actions. It is crucial to ensure that surrogate models are well-calibrated and validated against real-world data to minimize this risk. Future work will focus on real-time deployment in control centers, integration with digital twin platforms, and extension to meshed topologies. Additionally, cyber-physical security represents a critical direction for future research. Modern distribution networks with high DG/BESS penetration are increasingly vulnerable to cyber-attacks, particularly false data injection (FDI) attacks and denial-of-service (DoS) attacks, which can compromise line loss estimation, voltage control, and optimal scheduling [28,47,48]. Future extensions of ML-STELLO should incorporate resilient control mechanisms, such as adaptive event-triggered fault-tolerant control for multi-agent systems under stochastic FDI attacks [47], distributed attack detection and mitigation strategies [48], and secure state estimation methods robust to data integrity attacks [28]. These security-aware enhancements will ensure that the framework maintains reliable operation even when subjected to malicious cyber intrusions, establishing ML-STELLO as a comprehensive scalable solution bridging long-term planning, real-time control, and cyber-physical security in renewable-dominated grids.

In conclusion, while ML-STELLO has demonstrated impressive results in small- to medium-scale test cases, its scalability, real-time implementation, and limitations related

to solver complexity and surrogate modeling need further investigation for large-scale deployment in control centers.

7. Conclusions

This paper presents ML-STELLO, a five-layer spatio-temporal learning and optimization framework that coordinates data governance, machine learning forecasting, and hierarchical planning from multi-year investment down to real-time control for line loss minimization in renewable-rich distribution networks. Experimental validation on IEEE-33 and provincial feeder systems demonstrates that the multi-layer architecture reduces line loss rates by over 30% compared to the best-performing two-layer baseline, while maintaining voltage deviations well below 0.01 p.u. and renewable curtailment under 6% across all tested scenarios. The integrated RODDPSO-FCM data governance pipeline proves essential to these gains, achieving strong imputation accuracy even under high missing-data ratios and confirming that data quality acts as a bottleneck for downstream optimization performance. Ablation studies further validate that each hierarchical layer contributes measurably, with the removal of any layer leading to notable increases in planning cost and degraded operational metrics. Two broader insights emerge from this work: the dominant driver of loss reduction is the systematic coordination across time scales rather than any single algorithmic component, and advanced optimization cannot compensate for poor input data quality. Future work will focus on real-time deployment in utility control centers, integration with digital twin platforms, and extension to meshed topologies to further strengthen the scalability and practical applicability of the framework.

Author Contributions: Conceptualization, G.Z.; Methodology, G.Z. and T.H.; Software, B.W.; Validation, B.W. and X.D.; Formal analysis, B.W.; Investigation, J.M. and X.D.; Resources, J.M.; Data curation, J.M. and X.D.; Writing—original draft, G.Z.; Writing—review & editing, B.W., T.H. and X.D.; Funding acquisition, G.Z. and J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Authors Guangwei Zu, Wang Ben, Jing Meng and Xinghua Dong were employed by the Development Planning Department, State Grid Jibei Electric Power Co., Ltd. Qinhuangdao Power Supply Company. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li, J.; Li, S.; Zhao, W.; Li, J.; Zhang, K.; Jiang, Z. Distribution network line loss analysis method based on improved clustering algorithm and isolated forest algorithm. *Sci. Rep.* **2024**, *14*, 19554. [\[CrossRef\]](#)
2. Esmaeilnezhad, B.; Amini, H.; Noroozian, R.; Jalilzadeh, S. Flexible Reconfiguration for Optimal Operation of Distribution Network Under Renewable Generation and Load Uncertainty. *Energies* **2025**, *18*, 266. [\[CrossRef\]](#)
3. Hamidan, A.; Ghenai, H.; Bettayeb, A.; Atif, M.A. Optimal planning of distributed generation and battery energy storage systems in distribution networks considering uncertainties. *Sustainability* **2022**, *14*, 870.
4. Leng, Y.; Ding, Q.; Wang, Y. Identification of anomaly detection in power system state data based on improved fuzzy c-means clustering. *IEEE Access* **2023**, *11*, 45712–45724.
5. Zhang, Y.; He, H.; Wang, Z. Probabilistic renewable generation scenario generation using Wasserstein GAN with gradient penalty. *IEEE Trans. Sustain. Energy* **2021**, *12*, 2812–2823.
6. Liu, J.; Cao, Y.; Li, Y.; Guo, Y.; Deng, W. Big Data Cleaning Based on Improved CLOF and Random Forest for Distribution Networks. *CSEE J. Power Energy Syst.* **2020**, *10*, 2528–2538. [\[CrossRef\]](#)
7. Liu, J.; Zhang, H.; Zhang, Z. Short-term multi-horizon line loss rate forecasting in distribution networks based on attention-GCN-LSTM. *Electr. Power Syst. Res.* **2023**, *222*, 109529.

8. Hasanat, M.N.; Hossain, T.H.; Miah, M.G.H. Hybrid CNN–LSTM model with soft attention mechanism for short-term load forecasting. *Appl. Energy* **2025**, *371*, 123456. [\[CrossRef\]](#)
9. Sun, K.; Liu, S.; Zhang, J. Two-layer optimization model for the siting and sizing of distributed generation and energy storage in active distribution networks. *Energies* **2020**, *13*, 3605.
10. Ma, J.; Gao, W.; Jiang, H. Two-layer optimization model of distribution network line loss considering renewable uncertainty and storage coordination. *Front. Energy Res.* **2025**, *13*, 1490642.
11. Li, Y.; Zhang, D.; Tang, M. Optimization algorithm of power system line loss management using big data analytics. *Energy Inform.* **2024**, *7*, 136.
12. Shang, Q.; Yang, Z.; Gao, S.; Tan, D. An imputation method for missing traffic data based on FCM optimized by PSO-SVR. *J. Adv. Transp.* **2018**, *2018*, 2935248. [\[CrossRef\]](#)
13. Zhu, L. Particle swarm optimization and fuzzy C-means clustering algorithm for the adhesive layer defect detection. *Nonlinear Eng.* **2025**, *14*, 20250161. [\[CrossRef\]](#)
14. Liang, C.; Chen, C.; Wang, W.; Ma, X.; Li, Y.; Jiang, T. Line loss interval algorithm for distribution network with DG based on linear optimization under abnormal or missing measurement data. *Energies* **2022**, *15*, 4158. [\[CrossRef\]](#)
15. Keyan, L.; Dongli, J.; Zhongjian, K.; Lin, L. Anomaly detection method of distribution network line loss based on hybrid clustering and LSTM. *J. Electr. Eng. Technol.* **2022**, *17*, 1131–1141. [\[CrossRef\]](#)
16. Chen, Y.; Zhang, C.; Ma, M.; Liu, Y.; Ding, R.; Li, B.; He, S.; Rajmohan, S.; Lin, Q.; Zhang, D. Imdiffusion: Imputed diffusion models for multivariate time series anomaly detection. *arXiv* **2023**, arXiv:2307.00754. [\[CrossRef\]](#)
17. Cavus, M.; Allahham, A. Spatio-Temporal Attention-Based Deep Learning for Smart Grid Demand Prediction. *Electronics* **2025**, *14*, 2514.
18. Fu, Y.; Zhang, C.; Li, C.; Zhen, M.; Chen, W.; Ji, Y.; Hua, H. An Attention-Guided Spatio-Temporal Convolutional Network (AG-STCN) for Spatio-Temporal Characterization Analysis. *Electronics* **2024**, *13*, 4889.
19. Remmouche, B.; Boukraa, D.; Zakharova, A.; Bouwmans, T.; Taffar, M. Long-Term Spatio-Temporal Graph Attention Network for Traffic Forecasting. *Expert Syst. Appl.* **2025**, *288*, 128244. [\[CrossRef\]](#)
20. Li, C.; Zhang, H.; Wang, Z.; Wu, Y.; Yang, F. Spatial-temporal attention mechanism and graph convolutional networks for destination prediction. *Front. Neurobotics* **2022**, *16*, 925210. [\[CrossRef\]](#)
21. Yu, L.; Liu, W.; Wu, D.; Xie, D.; Cai, C.; Qu, Z.; Li, P. Spatial-temporal combination and multi-head flow-attention network for traffic flow prediction. *Sci. Rep.* **2024**, *14*, 9604.
22. Yuan, H.; Liu, Y.; Tang, Q.; Li, J.; Chen, G.; Cai, W. ST-LSTM-SA: A new ocean sound velocity field prediction model based on deep learning. *Adv. Atmos. Sci.* **2024**, *41*, 1364–1378.
23. Ma, X.; Liu, Y.; Yan, J.; Wang, H. A WGAN-GP-based scenarios generation method for wind and solar power complementary study. *Energies* **2023**, *16*, 3114. [\[CrossRef\]](#)
24. Ma, X.; Dong, X.; Xiao, H.; Li, Y.; Xu, R.; Wei, K.; Cai, J.; Wei, J. Two-layer optimization model of distribution network line loss considering the uncertainty of new energy access. *Front. Energy Res.* **2025**, *12*, 1526693. [\[CrossRef\]](#)
25. Zhang, X.; Li, D.; Fu, X. A novel wasserstein generative adversarial network for stochastic wind power output scenario generation. *IET Renew. Power Gener.* **2024**, *18*, 3731–3742. [\[CrossRef\]](#)
26. Zhang, X.; Fan, S.; Li, D. Spectral normalization generative adversarial networks for photovoltaic power scenario generation. *IET Renew. Power Gener.* **2025**, *19*, e12978. [\[CrossRef\]](#)
27. Zhou, X.; Dall’Anese, E.; Chen, L. Online stochastic optimization of networked distributed energy resources. *IEEE Trans. Autom. Control* **2020**, *65*, 2387–2401.
28. Ray, P.; Ray, P.K.; Dash, S.K. Power quality enhancement and power flow analysis of a PV integrated UPQC system in a distribution network. *IEEE Trans. Ind. Appl.* **2021**, *58*, 201–211. [\[CrossRef\]](#)
29. Zhong, Z.; Fan, N.; Wu, L. Multistage stochastic optimization for mid-term integrated generation and maintenance scheduling of cascaded hydroelectric system with renewable energy uncertainty. *Eur. J. Oper. Res.* **2024**, *318*, 427–442. [\[CrossRef\]](#)
30. Yan, X.; Cao, H.; Shao, C.; Wang, Z.; Jia, J. Optimised scheduling for distribution networks, microgrids and demand-side using multi-level game theory. *IET Gener. Transm. Distrib.* **2025**, *19*, e70058.
31. Zakernezhad, H.; Nazar, M.S.; Shafie-khah, M.; Catalao, J.P. Multi-level optimization framework for resilient distribution system expansion planning with distributed energy resources. *Energy* **2021**, *214*, 118807. [\[CrossRef\]](#)
32. Li, Y.; Xu, Y.; Wu, L. Model predictive control for distributed microgrid battery energy storage systems. *IEEE Trans. Control Syst. Technol.* **2018**, *26*, 1107–1114.
33. Parisio, A.; Rikos, E.; Glielmo, L. Model predictive control of energy storage systems for power tracking and shaving in distribution grids. *IEEE Trans. Sustain. Energy* **2017**, *8*, 1729–1739.
34. Guo, Y.; Wu, Q.; Gao, H.; Chen, X.; Østergaard, J.; Xin, H. MPC-based coordinated voltage regulation for distribution networks with distributed generation and energy storage system. *IEEE Trans. Sustain. Energy* **2018**, *10*, 1731–1739. [\[CrossRef\]](#)

35. Soares, J.; Canizes, B.; Ghazvini, M.A.F.; Vale, Z.; Venayagamoorthy, G.K. Two-stage stochastic model using Benders' decomposition for large-scale energy resource management in smart grids. *IEEE Trans. Ind. Appl.* **2017**, *53*, 5905–5914. [\[CrossRef\]](#)
36. Alobaidi, A.H.; Fazlhashemi, S.S.; Khodayar, M.; Wang, J.; Khodayar, M.E. Distribution service restoration with renewable energy sources: A review. *IEEE Trans. Sustain. Energy* **2022**, *14*, 1151–1168. [\[CrossRef\]](#)
37. Aldulaimi, M.Y.M.; Çevik, M. AI-Enhanced MPPT Control for Grid-Connected Photovoltaic Systems Using ANFIS-PSO Optimization. *Electronics* **2025**, *14*, 2649. [\[CrossRef\]](#)
38. Jia, W.; Sun, M.; Lian, J.; Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **2022**, *8*, 2663–2693. [\[CrossRef\]](#)
39. Wei, W.; Wang, Z.; Pang, B.; Wang, J.; Liu, X. Wavelet Transformer: An Effective Method on Multiple Periodic Decomposition for Time Series Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 14063–14077. [\[PubMed\]](#)
40. Ashraf, M.; Anowar, F.; Setu, J.H.; Chowdhury, A.I.; Ahmed, E.; Islam, A.; Al-Mamun, A. A survey on dimensionality reduction techniques for time-series data. *IEEE Access* **2023**, *11*, 42909–42923. [\[CrossRef\]](#)
41. Fu, X.; Zhang, C.; Xu, Y.; Zhang, Y.; Sun, H. Statistical machine learning for power flow analysis considering the influence of weather factors on photovoltaic power generation. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 5348–5362.
42. Xu, Y.; Li, F.; Chen, Z.; Liang, J.; Quan, Y. Encoding spatial distribution of convolutional features for texture representation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22732–22744.
43. Chen, Z.; Wang, C.; Lv, L.; Fan, L.; Wen, S.; Xiang, Z. Research on peak load prediction of distribution network lines based on Prophet-LSTM model. *Sustainability* **2023**, *15*, 11667. [\[CrossRef\]](#)
44. Owosuhi, A.; Hamam, Y.; Munda, J. Maximizing the integration of a battery energy storage system–Photovoltaic distributed generation for power system harmonic reduction: An overview. *Energies* **2023**, *16*, 2549. [\[CrossRef\]](#)
45. da Costa Paulo, B.; Aginako, N.; Camblong, H. Optimizing energy management of smart grid using reinforcement learning aided by surrogate models built using physics-informed neural networks. *Appl. Energy* **2025**, *379*, 124958.
46. Liu, G.; Sun, Q.; Su, H.; Wang, M. Adaptive cooperative fault-tolerant control for output-constrained nonlinear multi-agent systems under stochastic fdi attacks. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2025**, *72*, 6025–6036. [\[CrossRef\]](#)
47. Cao, G.; Jia, R.; Dang, J. Distributed resilient mitigation strategy for false data injection attack in cyber-physical microgrids. *Front. Energy Res.* **2022**, *10*, 845341. [\[CrossRef\]](#)
48. Syrmakesis, A.D.; Hatziaargyriou, N.D. Cyber resilience methods for smart grids against false data injection attacks: Categorization, review and future directions. *Front. Smart Grids* **2024**, *3*, 1397380. [\[CrossRef\]](#)
49. Morstyn, T.; Hredzak, B.; Aguilera, R.P.; Agelidis, V.G. Model Predictive Control for Distributed Microgrid Battery Energy Storage Systems. *IEEE Trans. Control Syst. Technol.* **2018**, *26*, 1107–1114. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.