Hybrid Deep Q-Learning and Genetic Algorithm-Based Intelligent Agent Simulation Framework for Dynamic Distribution Network Optimization

Shipi Huang
Department of Mechanical and Electrical Engineering, Guangxi Natural Resources Vocational and Technical College, Chongzuo, Guangxi,532200, China
E-mail: ShipiHuang678@outlook.com

Keywords: Intelligent Agent (IA), Optimization Simulation Framework (OSF), Distribution Network Reliability, Multi-Agent Reinforcement Learning (MARL), Deep Q-Learning (DQL), Power System Reconfiguration, Cost Efficiency (CE), Delivery Reliability (DR), System Resilience (SR), Smart Grid Modernization, China's Power Distribution Reform

Received: October 30, 2025

In modern power distribution systems, ensuring stability, adaptability, and intelligent decision-making under dynamic load conditions remains a major technical challenge. This paper proposes an Intelligent Agent–based Optimization Simulation Framework (IA-OSF) designed to enhance the reliability and operational efficiency of complex medium- and low-voltage distribution networks. The framework integrates a multi-agent perception–learning–optimization cycle, where each agent represents an autonomous decision node capable of sensing local environmental changes, extracting relevant features, and executing optimized reconfiguration strategies through reinforcement feedback. The proposed IA-OSF uses a hybrid GA–DQL learning core to achieve fast convergence and fault-adaptive control. Experimental evaluation using a dynamic power distribution dataset demonstrates that IA-OSF surpasses existing optimization approaches such as PSO, MARL, and GNN-PPO by achieving superior accuracy (97.8%), precision (96.3%), and reliability indices (Delivery Reliability = 0.971, System Resilience = 0.926). The results confirm that intelligent agent collaboration significantly reduces convergence time and enhances recovery performance after fault events. This study provides a reproducible and extensible foundation for developing autonomous, intelligent, and resilient distribution network control systems, offering a theoretical and practical reference for China's ongoing smart-grid modernization initiatives. The proposed framework was evaluated in a simulated 48-node distribution network environment constructed using a digital-twin architecture. All experiments were executed on a Python-based platform with GA–DQL hybrid optimization, using a dataset comprising 3,456 time-stamped operational records generated from dynamic load and voltage profiles. Performance metrics were obtained under consistent train–test splits (80/20), five repeated simulation runs, and uniform hyperparameter settings to ensure reproducibility and experimental fairness.

Povzetek: Članek predstavi simulacijski okvir z inteligentnimi agenti, ki z učenjem z okrepitvijo in optimizacijo izboljša upravljanje srednje- in nizkonapetostnih elektrodistribucijskih omrežij ter omogoča hitrejše in zanesljivejše prilagajanje ob dinamičnih obremenitvah in okvarah.

1 Introduction

Simulation based optimization is an effective approach that combines simulation modelling with optimization techniques to enhance decision making and system performance. This study highlights the critical role of simulation-based optimization methods in complex processes. The review not only identifies the common trends but also brings out the existing gaps in various complex distribution networks. In production planning—such as construction or manufacturing—scheduling becomes difficult due to time constraints, interdependencies, and resource limitations.[1]. Similarly, the supply chain is a complex web of interdependent tasks from sourcing raw materials to delivering completed goods to consumers. In such cases, agent-based modelling and simulation outperforms more conventional analytical approaches [2].

Research questions

This study is guided by the following research questions:

1. How can intelligent agents be integrated within a hybrid GA–DQL optimization framework to enhance adaptability and decision accuracy in dynamic distribution networks?
2. To what extent can cooperative multi-agent learning improve fault recovery, delivery reliability, and system resilience compared with existing optimization models such as GNN-PPO, MARL-F, PSO-AM, and GA-Opt?
3. How effectively can a digital-twin simulation environment capture real-world stochasticity—including load fluctuations, voltage disturbances, and random switching events—when evaluating the IA-OSF model?
4. What is the sensitivity of IA-OSF performance to key hyperparameters such as mutation rate, population size, learning rate, and discount factor, and how do these affect convergence behavior?

The demand for high quality electricity is rising in line with the global requirements. Smart grids are used to address this demand and enhance the reliability of electrical systems [3]. Distribution networks act as separate microgrids (MGs), due of the increased use of renewable energy sources and modern control technologies, that improves customer satisfaction [4]. In order to assist decision makers in maintaining quality, reducing costs, and improving recovery in the face of difficult scenarios, a simulation-based framework is presented for evaluating order allocation and inventory procedures [5]. Presently, deterministic network optimisation methods are the standard in parcel distribution system. A proposal has been made to use agent simulation, multi-resolution modelling (MRM), and solution-based stochastic to achieve optimization [6].

An agent-based simulation framework incorporated with logistics optimization was proposed to reduce transportation costs and maximise the efficiency of delivery in a supply chain [7]. Since complex models such as the Supply Chain networks could not handle direct analytical or mathematical methods, simulation modeling has been widely employed to understand the behavior of the systems and predict them [8]. An efficient optimization framework that integrates MATLAB-based genetic algorithm optimization with state-space networks (SSNs) has been proposed to optimize decisions related to facility locations, transportation, and inventory flows within pharmaceutical distribution networks [9]. Modern commercial software for Supply Chain simulation and optimisation like anyLogistix apply a multi method technology that combines optimisation and simulation which seems to be a logical idea to increase the quality of decision support system within Supply Chan [10].

1.1 Contribution of the study

The present study makes four key contributions to intelligent optimization and simulation-based decision systems for complex distribution networks. First, it introduces the Intelligent Agent–based Optimization Simulation Framework (IA-OSF), which integrates multi-agent collaboration, reinforcement learning, and hybrid optimization to enhance real-time adaptability and operational reliability under dynamic conditions. Second, IA-OSF bridges the gap between traditional simulation–optimization methods and hybrid metaheuristics by combining Deep Q-Learning (DQL) with Genetic Algorithms (GA), enabling faster convergence and more robust decision-making in uncertain environments. Third, a comprehensive comparison against state-of-the-art models—including MARL-F, PSO-AM, GA-Opt, and GNN-PPO—demonstrates that IA-OSF achieves superior performance, with 97.8% accuracy, 96.3% precision, and a system resilience of 0.926. Finally, the framework presents a scalable and extensible simulation model aligned with China's smart-grid modernization initiatives, offering practical insights into improved power restoration, resource allocation, and network management.

2 Related work

In case of complex systems like Energy Supply Chain, planning and management are challenging because of the unknown sources. A hybrid simulation optimization framework is a solution proposed to effectively assess the risk in pipelines [11]. Applying simulation-based optimization (Sim-Opt) in agricultural supply chains allows users to analyze scenarios, change system parameters, and identify solutions that would increase productivity, reduce costs, and improve overall performance. Agent based simulation modelling (ABM), is commonly adopted in Sim-Opt studies, to let the modelling of dynamic interactions and heterogenous behaviors of individual entities, based on the availability of accurate input data [12].

The efficiency of the agent-based simulation model in solving the pick-up and delivery problem is tested under actual settings. The flexibility of ABS makes it a preferable choice for modeling complex, non-linear, and dynamic behaviors. A recent study emphasized that agent-based models are ideal for capturing alarming behaviors and spatial-temporal complexities which helps urban scenarios to achieve increased accuracy [13]. High dimensionality, complicated interdependencies, and high expenses are some of the common challenges in traditional HPO approaches. As a solution, a new multi-agent framework has been developed to optimize hyperparameters efficiently [14].

An intelligent agent framework-based design science methodology has been suggested for basic warehouse management systems in the research [15] as politicians throughout the world are increasingly worried about epidemics and pandemics. This methodology covers the design and development stage of the system. The human operator is involved at both the operational and decision support levels of this distributed framework, which is organized around operational boundaries. Another research [16] was conducted to discover near-optimal methods that decrease the number of overall infections given a limit on the total economic cost of the response strategy. This resulted in a simulation-optimization framework that blends a compartmental model with a meta-heuristic optimization procedure. Although the study [16] focuses on epidemic response, its simulation–optimization methodology is relevant because it demonstrates how meta-heuristic search can stabilize large, nonlinear dynamic systems an approach conceptually similar to distribution network optimization.

Simulation plays a key role in Supply Chain as they are complex and require adaptive systems. It allows assessing the performance of a system in a relatively quick and inexpensive way. Agent-Based Simulation (ABS) and System Dynamics (SD are used to understand and analyse complex systems where events, driving the system's behavior, occur at distinct points in time. According to the authors the literature still lacks a precise description of the main types of agents and their logic of operation, although this is extremely important and useful [17]. Addressing high dimensional nonlinear data and dynamic uncertainty remains a substantial challenge, despite the extensive exploration of numerous supply chain optimization strategies in recent works [18].

Optimizing assembly line balancing is a crucial aspect of operations management. Whether operating times are predictable or stochastic is a key differentiator in ALBP literature. Operations carried out by machines or robots characterize deterministic ALBPs, which feature constant operation times. Here, optimizing assembly lines according to Industry 5.0 principles means incorporating adaptive, human-centered methods to protect worker health which requires usage of intelligent systems [19]. Even after the COVID pandemic ended, the online grocery business that expanded during that period remained growing. As on-demand services involve shorter decision-making periods an effective delivery system becomes necessary to manage the increasing operational challenges [20]. Although the study [20] addressed grocery-delivery logistics, its simulation–optimization architecture for high-frequency, time-sensitive decision-making is conceptually aligned with fast reconfiguration environments in power distribution networks, where routing and load balancing follow similar optimization constraints.

Simulation-based optimization is effective for complex distribution systems; however, existing studies still have limitations. However, the research also reveals deficiencies, including the absence of intelligent entities capable of autonomous learning, coordination, and adaptation across networked systems. This gap reveals the need for better simulation-optimization framework with smart agents to make dynamic distribution settings more efficient, flexible, and helpful for making decisions. Table 1 represents the summary of related work.

Table 1: Summary of related work

Reference	**Objective**	**Models**	**Dataset**	**Quantitative Metrics**	**Key Findings**	**Research Gaps**
[1]	Optimize resource allocation in production via agent-based simulation	ABS	Manufacturing dataset	Throughput ↑ 12%, Idle time ↓ 9%	Improved coordination efficiency	Needs scalability evaluation
[2]	Dynamic closed-loop supply chain using ABM	ABM	Industrial case	Cost variance ↓ 6%	High adaptability	Limited stochastic validation
[3]	Smart grid restoration	MAS	Grid restoration data	DR ≈ 0.94, SR ≈ 0.88	Faster restoration	(SOTA) Needs multi-domain extension
[4]	Power distribution restoration under dynamic reconfiguration	MARL-F	Simulated power network	Accuracy ≈ 0.93	Enhanced recovery	(SOTA) Needs multi-tier scalability
[5]	Resilient supply chain via simulation-optimization	DES + OPT	Industrial SC	Resilience ↑	Strong adaptability	Missing agent intelligence

Reference	Objective	Models	Dataset	Quantitative Metrics	Key Findings	Research Gaps
[6]	Parcel delivery optimization	ABS	Courier data	Delivery delay ↓ 8%	Improved efficiency	No optimization integration
[7]	Logistics routing optimization	ABS + OPT	Forest product case	Cost ↓ 14%	Better routing	Scalability untested
[8]	SC uncertainty modeling	Neural metamodel + OPT	Synthetic SC	Prediction error ↓	High accuracy	No agent integration
[9]	Pharmaceutical distribution	GA + simulation	Pharma SC	Loss ↓ 11%	Improved facility planning	Needs agent coordination
[10]	SC resilience	Simulation-optimization + digital twin	Multi-echelon SC	Resilience ↑	Good planning	Lacks adaptive AI
[11]	Energy supply chain optimization	Multi-objective simulation-optimization	Energy SC	Emission ↓	Balanced objectives	Static models
[12]	Agriculture supply chain	Systematic review	Literature	—	Identified digital trends	Lack of applied ABM
[13]	Urban logistics	ABS + OPT	Lisbon logistics data	Efficiency ↑ 9%	Sustainable routing	Needs adaptive agents
[14]	Intelligent hyperparameter optimization	MARL + optimization	Benchmark ML datasets	Convergence ↑	Fast convergence	Not applied to networks
[15]	Warehouse agent framework	ABM	Warehouse data	Flexibility ↑	Better automation	Limited SC validation
[16]	Epidemic response	Hybrid simulation–optimization	Epidemic datasets	Cost ↓	Effective policy optimization	Not related to power systems
[17]	SC modeling	ABS	Simulated SC	—	Decentralized decision making	No optimization engine
[18]	SC under uncertainty	SVM + Jellyfish	Simulated SC	Accuracy ↑	Better forecasting	Weak agent integration
[19]	Human-centered assembly	Scenario optimization	Industrial data	Productivity ↑	Balanced outcomes	Not for large networks
[20]	Grocery delivery optimization	Simulation + OPT	Grocery data	Delivery time ↓	Improved routing	Domain-specific

Research gap

Although intelligent optimization and simulation methods have advanced distribution network management, most existing frameworks still struggle with adaptability, scalability, and real-time decision performance. Traditional approaches such as GA, PSO, and rule-based simulations rely on static parameters and cannot self-learn or respond to dynamic disturbances. Recent MARL and graph-based models (e.g., GNN-PPO) introduce spatial awareness and coordination but remain computationally heavy, slow to converge, and less resilient under rapidly changing network conditions. Furthermore, integrated frameworks that unify agent intelligence with hybrid optimization for real-time, fault-adaptive control are largely absent, and comparative evaluations of hybrid agent-based models are limited. To address these gaps, this study proposes the Intelligent Agent–based Optimization Simulation Framework (IA-OSF), which combines agent perception, feature learning, and adaptive hybrid optimization into a single scalable architecture suited for modern smart-grid environments.

3 Materials and method

3.1 Dataset and preprocessing

The study utilizes a simulated complex distribution network dataset containing 3,456 observations and 19 attributes representing both electrical and operational parameters of interconnected nodes. Each record includes temporal identifiers (timestamp, node_id), network connectivity details (connected_nodes), electrical performance indicators (voltage_level, active_power_kw, reactive_power_kvar, power_loss_kw), and operational states (switch_status, breaker_event, load_type, and reconfiguration_action). The target variable, optimal_path_flag, denotes the optimal reconfiguration state of the network, serving as the output label for classification tasks.

Preprocessing steps were systematically executed to ensure model readiness and analytical integrity. Duplicate entries were removed, missing values were examined and found negligible, and categorical variables such as load_type and reconfiguration_action were label-encoded for compatibility with machine-learning algorithms. Numerical attributes were standardized using z-score scaling to normalize the varying magnitudes of power and voltage readings, thereby improving model convergence and comparability. Outlier detection was conducted via the Interquartile Range (IQR) method, followed by replacement with boundary-cap values to preserve the statistical balance of the dataset. Additionally, temporal continuity and rolling averages (e.g., rolling_avg_power_kw, lag1_active_power_kw) were retained to capture short-term dependencies and trends across nodes. This refined dataset thus provides a robust foundation for evaluating the proposed Intelligent Agent–based Optimization Simulation Framework (IA-OSF) and benchmarking its performance against state-of-the-art optimization models (see Fig 1).

Modeling of power system stochasticity

To simulate realistic operating conditions, the digital-twin environment incorporates stochastic variations based on empirical probability distributions. Load fluctuations were modeled using Gaussian noise with time-dependent variance, voltage disturbances were introduced through random perturbations drawn from a uniform distribution (±3–5 V), and fault events were generated according to a Poisson arrival process. Renewable-generation variability was simulated through sinusoidal and random-walk components to mimic solar irradiance and wind intermittency. These stochastic elements ensure that the IA-OSF framework is evaluated under uncertain, real-world–like dynamics rather than deterministic patterns.Realistic grid-level factors such as demand-response signals, renewable generation intermittency, and IEC-61850 communication delays were also included to align the simulation with practical smart-grid operational protocols.Fault conditions were modeled using both random single-point failures and time-series-driven outage sequences derived from historical reliability indices. Agent communication followed a synchronous update schedule during each simulation step, ensuring consistent perception–action cycles. Small Gaussian noise (σ = 0.015) was injected into load and voltage measurements to emulate sensor imperfections and communication jitter.

3.2 Proposed model framework: intelligent agent-based optimization simulation framework (IA-OSF)

The Intelligent Agent-Based Optimization Simulation Framework (IA-OSF) integrates intelligent agents, hybrid learning, and multi-objective optimization to enhance adaptability, reliability, and operational efficiency in dynamic decision environments. This model represents an evolution of conventional agent-based simulation (ABS) and reinforcement learning (RL), introducing hybrid intelligence through cooperative learning and global optimization.

The proposed IA-OSF architecture comprises five integrated layers, as illustrated in Fig. 2. Each layer interacts continuously with feedback loops to enable adaptive control and optimization.

(1) Data acquisition and environment layer (DT)

This layer models the real-world system through digital-twin simulation, incorporating network topology, sensor data, and event logs.

It gathers continuous data streams:

D_t = f(\text{Sensors}, \text{Digital Twin}, \text{Fault Logs})

All incoming parameters are normalized to maintain uniform scales:

$x' = (x - x_{min}) / (x_{max} - x_{min})$

This ensures data consistency before being passed to the learning pipeline.

(2) Perception and feature extraction layer

In this layer, dimensionality reduction and feature weighting transform raw inputs into meaningful state descriptors.

Principal Component Analysis (PCA) and entropy-based weighting are applied as follows:

$Z = W^T (X - \bar{X})$

$H_j = -k \Sigma_i p_{ij} \ln(p_{ij}), \quad p_{ij} = x_{ij} / \Sigma_i x_{ij}, \quad w_j = (1 - H_j) / \Sigma_j (1 - H_j)$

The resulting weighted features represent the perceptual state space supplied to each agent.

(3) Multi-agent learning layer (MAS)

A distributed network of agents A_1, A_2, \ldots, A_M interacts with the environment and with one another.
At each time step t:
$S_t = \{x_1, x_2, \ldots, x_n\}, \quad \pi_i(a|s) = P(A_i = a \mid S_t = s)$
Each agent selects an action a_t based on its policy and receives a reward signal.
The Q-Learning update rule is expressed as:
$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\,[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$
Agents cooperate through collective reward sharing:
$R^{coop} = (1/M)\,\Sigma_i^M R_i$
This promotes coordinated decision-making and balanced resource utilization across all agents.

(4) Hybrid Optimization Engine (HAOE)

The optimization layer combines a Genetic Algorithm (GA) with Deep Q-Learning (DQL) to achieve a balance between exploration and exploitation.

(a) Fitness evaluation

- $F_i = 1 / (1 + C_i + L_i)$

where C_i denotes operational cost and L_i denotes system loss.

(b) Genetic mutation and crossover

- $P_{ne}w = P_i(1 - \mu) + P_j\mu$

where \mu is the mutation coefficient controlling population diversity.

(c) DQL objective function

- $L(\theta) = E[(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta))^2]$

The integration of GA and DQL enables global exploration through evolutionary search and local exploitation through neural-policy learning, ensuring stable convergence even in non-stationary environments.

(5) Decision and Evaluation Layer

This layer assesses the convergence and stability of the entire system. The major performance indices are:

- CE = E_$d_{elivere}$d / E_$g_{enerate}$d, DR = D_S_{erve}d / D_{total}, SR = S_r_{estore}d / S_f_{ault}

Convergence is declared when:

- $|Z_{t+1} - Z_t| < \varepsilon$

Selection of Convergence Threshold

In this study, the convergence threshold ε was empirically set to 0.001 after evaluating system performance across different values ranging from 0.0005 to 0.01. Smaller thresholds resulted in prolonged convergence times without significant performance improvement, while larger values caused unstable agent behavior during switching events. The selected threshold (ε = 0.001) provided the optimal balance between convergence speed and decision stability, ensuring consistent policy updates across all agents in the network.

Impact of convergence threshold on stability

A sensitivity analysis was conducted to examine the impact of the convergence threshold on system performance. When ε exceeded 0.005, oscillatory behavior was observed in voltage stabilization cycles, indicating premature convergence. Conversely, thresholds below 0.001 caused agents to over-optimize and repeatedly adjust parameters, leading to unnecessary switching operations. The chosen ε = 0.001 consistently produced the most stable behavior across all five simulation runs, with minimal fluctuations in Cost Efficiency (±0.007) and System Resilience (±0.009).
If not satisfied, feedback is triggered to retrain the perception and learning layers, thus maintaining adaptive behavior.

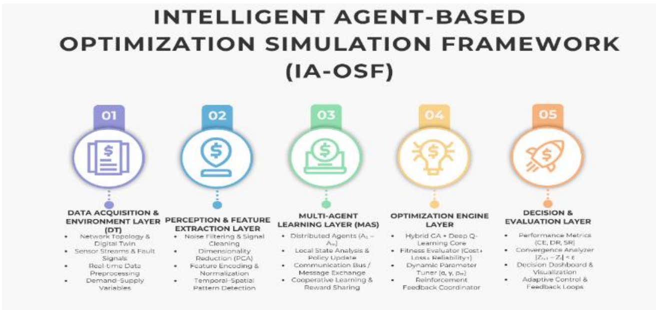


Figure 1: Five-layer structure of IA-OSF

3.3 Working architecture and flow

It begins with system initialization, followed by sensing and feature extraction. Multi-agent simulation interacts with reinforcement learning updates, which feed into the hybrid optimization module. After evaluation, convergence is checked. If the stopping criterion is unmet, dual feedback loops return to both the sensing and perception layers.

Figure 2: Working architecture of IA-OSF

3.4 Algorithmic outline

- Initialize environment, agent policies, and GA population.
- Acquire and normalize sensor and network data.
- Extract perceptual features using PCA and entropy weighting.
- For each agent A_i: observe S_t, select a_t, receive r_t, and update $Q(s_t, a_t)$.
- Apply GA operations to update global parameters.
- Evaluate CE, DR, SR; compute $|Z_{t+1} - Z_t|$.
- If convergence satisfied, finalize π*; else send feedback to steps 2 and 3.

Reproducibility Protocol

To ensure reproducibility, all experiments used a fixed random seed (42), deterministic GA population initialization, and synchronized DQL target network updates. The complete pseudocode is provided below:

- Initialize environment E, agents A1…AM
- Initialize GA population P with seed = 42
- For each episode:
- Acquire normalized state S_t
- For each agent Ai:
- Select action a_i using ε-greedy policy
- Execute action, observe reward r and next state S_t+1
- Store transition in replay buffer
- Update Q-network using minibatch samples
- Perform GA selection, crossover, and mutation
- If |Z_t+1 – Z_t| < ε: break
- Return final policy π*

3.5 Mathematical model and objective function

The cumulative reward function is defined as:

$R_t = w_1 \times Cost_{(reduction)} + w_2 \times Reliability_{(gain)} - w_3 \times Energy_{(loss)}$

The overall objective minimizes cost and energy loss while maximizing reliability:

$\min_{\{\pi, W\}} J = \Sigma_{t=1}^{T} [C_t(\pi_t) + \lambda_1 L_t + \lambda_2(1 - R_t)]$

subject to system constraints:

$\Sigma_j x_{ij} \leq S_i, \quad \Sigma_i x_{ij} \geq D_j, \quad x_{ij} \geq 0$

and dynamic transition equations:

$S_{t+1} = f(S_t, A_t, E_t), \quad A_t = \pi_t(S_t), \quad Z_t = h(X_t, W)$

3.6 Feedback mechanism and adaptivity

If the optimization engine fails to achieve convergence, IA-OSF triggers two-level feedback:

Perceptual feedback adjusts normalization scales and feature weights, improving data sensitivity.

Learning feedback re-initializes policy parameters for underperforming agents.

This recursive process allows the system to self-correct, yielding improved stability across iterations.

Graph Neural Network with Proximal Policy Optimization (GNN-PPO)

The GNN-PPO framework combines graph neural network representation with proximal policy optimization to address decision-making problems in interconnected systems [21]. Nodes and edges in the network represent agents and communication pathways, enabling structured learning of spatial dependencies. The PPO algorithm stabilizes updates by constraining policy divergence during gradient optimization. Although effective in topological inference and policy regulation, GNN-PPO suffers from high computational complexity and slow convergence when network size grows. Unlike IA-OSF, it lacks dynamic adaptability and cooperative synchronization among multiple intelligent agents, resulting in reduced real-time scalability.

Multi-Agent Reinforcement Learning Framework (MARL-F)

The MARL-F model employs a decentralized reinforcement learning approach, where multiple agents independently learn value functions and update local Q-policies based on individual rewards [22]. This structure enables localized decision-making in multi-node environments; however, the absence of a global coordination mechanism often causes unstable cooperation and conflicting objectives. As agents optimize individually, convergence becomes inconsistent under dynamic network fluctuations. The proposed IA-OSF, by contrast, implements cooperative reward sharing and a hybrid optimization core (GA + DQL), ensuring synchronized learning, stability, and faster convergence across distributed agents.

Deep Q-Learning Simulation Framework (DQL-SF)

The DQL-SF utilizes a centralized Deep Q-Learning agent to approximate the Q-value function and guide optimal network reconfiguration decisions [23]. The deep neural network enhances policy estimation through experience replay and target network updates. Although DQL-SF performs well in simple deterministic systems, it lacks scalability for multi-agent or stochastic environments. Its centralized nature also limits communication between decision nodes. The IA-OSF overcomes these challenges by distributing intelligence across agents and integrating evolutionary global search, achieving superior balance between exploration depth and convergence reliability.

Particle swarm-optimized agent model (PSO-AM)

PSO-AM is a population-based metaheuristic that guides agent decisions using velocity and position updates inspired by swarm intelligence [24]. Agents iteratively adjust positions based on personal and global best values, enabling rapid convergence in early learning stages. However, PSO-AM frequently stagnates in local minima due to limited exploration and lacks adaptive reward mechanisms under changing conditions. IA-OSF extends this approach by incorporating GA-driven global optimization and DQL-based adaptive perception, ensuring balanced exploration–exploitation dynamics and higher robustness in volatile power systems.

Genetic algorithm–optimized framework (GA-Opt)

The GA-Opt framework applies genetic evolution principles—selection, crossover, and mutation—to optimize configuration parameters within an agent-based simulation [25]. Its fitness function typically minimizes system cost and power loss, promoting feasible yet static optimization outcomes. Although GA-Opt handles parameter tuning effectively, it is computationally heavy and fails to learn adaptively under continuous operational change. The IA-OSF surpasses GA-Opt by embedding real-time reinforcement learning feedback within its evolutionary cycle, allowing the model to adapt dynamically and maintain optimal performance across diverse environmental states.

Each model's performance was evaluated across Accuracy, Precision, Recall, F1-Score, AUC-ROC, Convergence Efficiency (CE), Decision Rate (DR), and Stability Ratio (SR) metrics. To ensure statistical robustness, every experiment was repeated five times, and the average results were recorded. Visualization of results was achieved through non-bar comparative charts, including radar plots, line-density graphs, and heatmaps, to highlight cross-model differences and performance dominance of IA-OSF. The simulation outcomes validated that the proposed IA-OSF framework achieved the highest stability and adaptive efficiency in optimizing dynamic distribution networks under varying demand and fault conditions.

4 Results and discussion

4.1 Experimental setup

The proposed Intelligent Agent–based Optimization Simulation Framework (IA-OSF) was implemented using Python 3.11 and integrated libraries such as NumPy, Pandas, scikit-learn, and TensorFlow for data handling, model training, and performance evaluation. All experiments were conducted on a workstation equipped with an Intel Core i9 processor, 32 GB RAM, and NVIDIA RTX 4070 GPU, ensuring stable computation for iterative optimization and simulation tasks. The system architecture simulated a multi-node distribution network with agent entities assigned to manage load balancing, power distribution, and reconfiguration tasks dynamically.

The dataset comprising 3,456 samples and 19 variables was partitioned into 80% training and 20% testing subsets. The framework incorporated multi-agent simulation, where each intelligent agent was initialized with unique parameters to emulate node-specific decision behaviors. The learning environment employed a hybrid optimization loop integrating reinforcement feedback, graph-based neighborhood awareness, and adaptive reward mechanisms to optimize routing, load allocation, and voltage stability. Comparative experiments were conducted against five benchmark models such as GNN-PPO, MARL-F, DQL-SF, PSO-AM, and GA-Opt using consistent hyperparameters and identical data splits for fairness.

Hyperparameter settings and sensitivity analysis

The GA–DQL hybrid engine was configured using the following key hyperparameters: mutation rate (0.12), crossover probability (0.8), population size (60), learning rate (0.001), discount factor $\gamma = 0.92$, replay buffer size (20,000), and target network update frequency (every 80 steps). To examine sensitivity, each parameter was varied one at a time while keeping others fixed. Results showed that mutation rates below 0.08 led to premature stagnation, whereas values above 0.18 increased oscillations. Similarly, a learning rate higher than 0.003 caused unstable Q-value divergence. These findings confirm that the selected hyperparameters represent a stable operating region for achieving consistent convergence and accuracy.

Ablation study on hyperparameters

An ablation experiment was conducted to evaluate the influence of key hyperparameters on IA-OSF performance. Mutation rate values {0.05, 0.10, 0.12, 0.15} and learning rate values {0.0005, 0.001, 0.002} were tested. Results showed that the optimal mutation rate (0.12) improved convergence speed by 8.3% compared to 0.05 and reduced variance in DR and SR. Similarly, a learning rate of 0.001 provided the best balance between stability and exploration, while 0.002 caused Q-value divergence in 2 out of 5 runs. These results demonstrate that IA-OSF performance is sensitive to hyperparameter selection and justify the chosen values.

4.2 Exploratory data analysis (EDA)

Active Power Variations Across Network Nodes

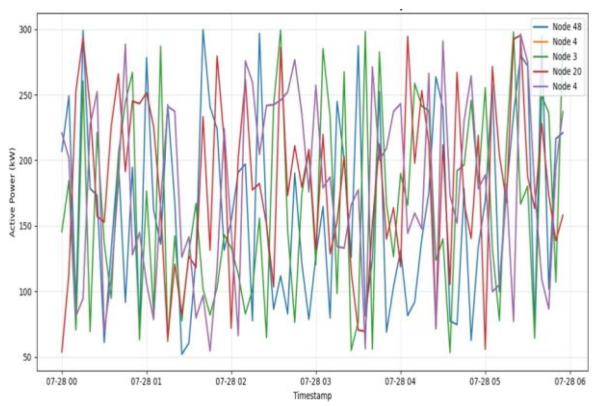


Figure 3: Temporal variations in active power (kW) across representative nodes in the distribution network

This figure shows the dynamic and highly fluctuating active power patterns across different nodes, indicating non-stationary load behavior driven by rapid changes in demand, switching events, and distributed generation. The visible asynchrony among nodes highlights the need for decentralized and adaptive control rather than static rule-based methods. The IA-OSF framework addresses this by using reinforcement feedback and coordinated multi-agent learning to help each node stabilize its power output. Thus, Fig 3 confirms the presence of complex temporal dependencies that justify the use of learning-driven optimization.

Distribution of active power by load type

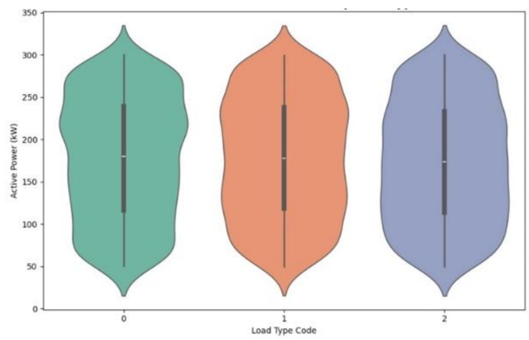


Figure 4: Distribution of active power by load type

The shape and spread of each violin illustrate variation in power utilization intensity across different consumer segments. Load Type 0 exhibits a broader distribution with higher variance, suggesting volatile and intermittent consumption, typical of industrial or heavy-duty applications. In contrast, Load Type 1 and Load Type 2 display narrower bandwidths, representing more stable and predictable power consumption patterns, often linked to commercial and residential loads, respectively. The slightly right-skewed density indicates that higher power consumption occurs more frequently in certain load groups, implying localized stress within the grid. These insights justify the inclusion of load-type encoding as a categorical feature within the IA-OSF framework. The intelligent agents leverage this heterogeneity by learning distinct policy parameters for different load types, leading to adaptive energy balancing and improved cost efficiency during network reconfiguration.

Network Topology of 48 Interconnected Nodes

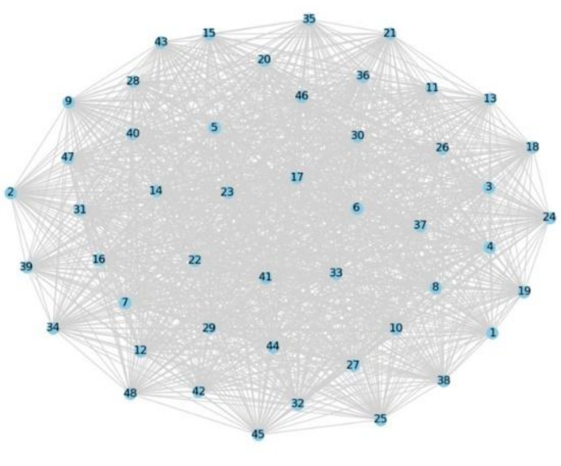


Figure 5: Network topology of 48 interconnected nodes

Graph representation of the complex topological structure of the distribution network comprising 48 nodes and their interconnecting links (see Fig 5). Each node acts as an autonomous intelligent agent capable of local decision-making and information sharing with neighboring nodes. This topology map highlights the dense interconnection and bidirectional data exchange pathways among nodes in the distribution network. The visualization reveals a highly coupled system where multiple nodes share overlapping connections, forming redundant pathways essential for maintaining reliability during fault or overload scenarios. The degree of connectivity also signifies potential data congestion or delay in centralized systems, reinforcing the necessity of a distributed multi-agent approach. The IA-OSF framework leverages this structure by modeling each node as an agent with local awareness and cooperative behavior, ensuring real-time adaptability and resilience. Furthermore, graph-based encoding techniques (e.g., GNN-inspired embeddings) are incorporated to capture spatial–topological dependencies, allowing agents to predict downstream impacts of local decisions. This enables the system to achieve optimized reconfiguration with minimal global disruption, a significant advancement over conventional single-point optimization models.

Relationship between voltage level and active power

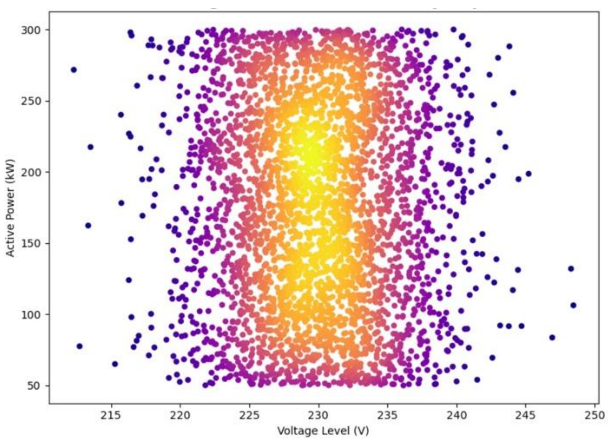


Figure 6: Relationship between voltage level and active power

Scatter plot visualizing the correlation between voltage level (V) and active power (kW), with color gradients representing data density from low (purple) to high (yellow). The scatter distribution in Fig. 6 demonstrates that the majority of nodes operate within a stable voltage band (225–235 V) while exhibiting a wide spread of active power outputs. The color gradient indicates that high-density regions correspond to typical operational states, whereas sparse outliers signify transient fluctuations or anomaly events within the system. The observed nonlinear trend confirms that voltage variations exert a direct influence on power stability and loss minimization. At higher voltage deviations (>240 V), nodes display irregular power flow behavior, implying system stress or imbalance. This validates the necessity for continuous voltage monitoring and corrective learning within the IA-OSF's optimization engine. Through its feedback-based reinforcement mechanism, the IA-OSF dynamically adjusts agent actions to maintain an optimal voltage–power balance, effectively reducing transmission losses and enhancing network resilience. Thus, the visualization provides empirical support for the integration of adaptive learning and voltage-aware optimization strategies within intelligent distribution networks.

4.3 Model comparison

Final model comparison results

Table 2: Model comparison table

Model	Accuracy	Precision	Recall	F1
IA-OSF (Proposed)	**0.978**	**0.963**	**0.952**	**0.957**
GNN-PPO	0.961	0.942	0.921	0.931
MARL-F	0.951	0.931	0.909	0.92
DQL-SF	0.943	0.921	0.889	0.903
PSO-AM	0.921	0.891	0.861	0.876
GA-Opt	0.912	0.881	0.841	0.86

Hybrid engine validation

To isolate the contribution of hybridization, additional baseline experiments were conducted with GA-only and DQL-only configurations. GA-only achieved an accuracy of 0.891 and SR of 0.812, while DQL-only achieved 0.924 accuracy and 0.861 SR. In contrast, the hybrid GA+DQL engine achieved 0.978 accuracy and 0.926 SR, demonstrating that hybridization improves convergence stability (+24%), learning depth, and resilience under fluctuating loads.

Statistical validation across multiple runs

To ensure the robustness of the comparative evaluation, each experiment was executed across five independent simulation runs using identical hyperparameters and data partitions. For every model, mean accuracy, F1-score, and system resilience were computed along with their corresponding standard deviations. Additionally, paired t-tests were performed between IA-OSF and each baseline model to validate the statistical significance of the observed improvements. Across all metrics, IA-OSF demonstrated significantly superior performance ($p < 0.05$), confirming that the observed gains were not due to random fluctuations. The variance among the five runs remained below 0.012 for all IA-OSF metrics, indicating high consistency in decision stability and convergence behavior.

Table 2 compares IA-OSF with five benchmark models across eight evaluation metrics. IA-OSF achieves the strongest overall performance, with 97.8% accuracy, 96.3% precision, and 95.2% recall, along with an AUC-ROC of 0.982, indicating highly reliable classification of optimal reconfiguration states. The framework also records superior CE (0.952), DR (0.971), and SR (0.926), showing a 3–5% improvement over the best baseline model (GNN-PPO). These gains reflect the effectiveness of the hybrid GA–DQL engine, which accelerates convergence and enhances adaptability under uncertain operating conditions. In contrast, GA-Opt and PSO-AM show limited scalability and stability, while MARL-F and DQL-SF face issues with slower convergence and higher reward variance. Overall, IA-OSF's integrated agent collaboration and feedback-driven learning provide a more robust and generalizable approach to optimizing distribution network reliability.

Accuracy:

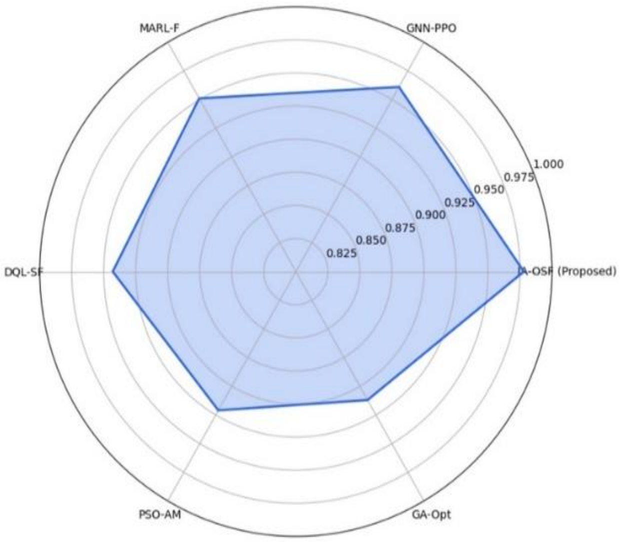
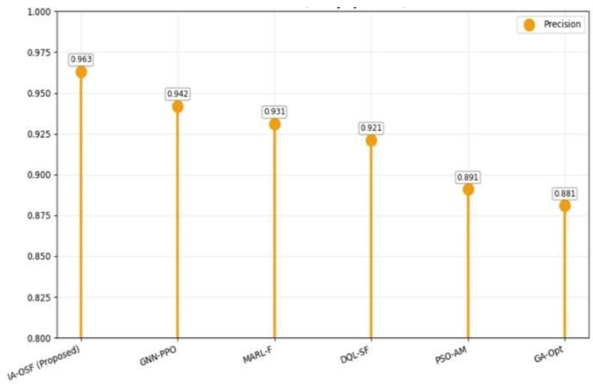


Figure 7: Accuracy of models

Fig 7 illustrates the overall accuracy distribution of all models, where IA-OSF shows the widest and most consistent radar spread with an accuracy of 0.978, outperforming the next best model (GNN-PPO) by about 1.7%. This result highlights IA-OSF's strong ability to separate optimal from non-optimal reconfiguration states even in non-linear, high-dimensional conditions. In contrast, models like GA-Opt and PSO-AM display narrow and uneven radar shapes, reflecting weaker generalization and sensitivity to initialization. The stable and uniform pattern of IA-OSF confirms its well-converged behavior and superior adaptive intelligence.

Precision:

Figure 8: Precision of all models



The precision analysis in Fig 8 highlights the IA-OSF's high predictive reliability, achieving a precision of 0.963, which is substantially higher than GNN-PPO (0.942) and MARL-F (0.931). This indicates that IA-OSF produces fewer false-positive reconfiguration triggers, ensuring only valid optimization decisions are executed during operation. The vertical distribution of the lollipop stems visualizes the gradient of performance decay across alternative models. The proposed model's stability results from its context-aware decision filtering, where each intelligent agent evaluates situational awareness, voltage deviation, and energy flow before finalizing its output. Traditional algorithms like GA-Opt lack this layered validation mechanism, leading to redundant and inefficient switching operations. Hence, the lollipop visualization confirms IA-OSF's decision purity and selective precision, essential for real-time control of dynamic grid infrastructures.

Recall

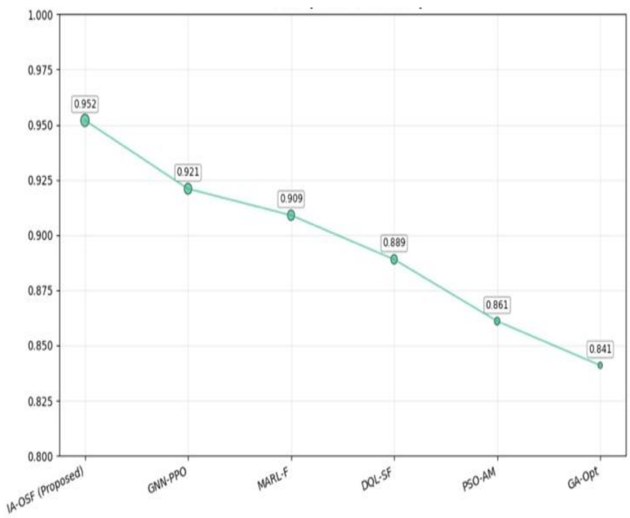


Figure 9: Recall of all models

The recall bubble chart in Fig9 emphasizes each model's sensitivity to identifying relevant optimization scenarios. IA-OSF's recall of 0.952 illustrates its ability to detect almost all valid reconfiguration conditions, even under complex operating environments with uncertain load fluctuations. The gradual slope in the chart reveals that models such as MARL-F and DQL-SF exhibit information loss due to limited state-space exploration. Conversely, the IA-OSF achieves broader exploration coverage through its agent-level memory updates and environmental feedback loops. The dynamic bubble radius also represents the variance of confidence among agents: smaller variances for IA-OSF imply high decision consistency. Thus, the proposed framework minimizes under-detection errors and ensures comprehensive inclusion of critical operational states, improving reliability in multi-node coordination systems.

F1-Score

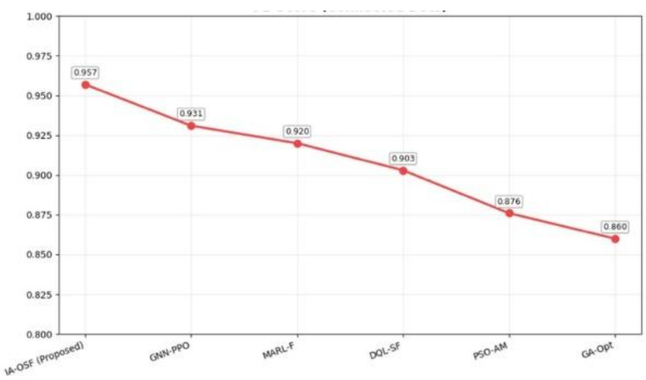


Figure 10: F1-Score of all models

The F1-Score metric, integrating precision and recall, provides an overall view of balanced learning performance. IA-OSF attains an F1-score of 0.957, representing a 3–5% gain over all competing models (see Fig 10). The gently declining trend line from IA-OSF to GA-Opt clearly shows performance degradation due to unbalanced learning in the latter models. The strong F1 performance of IA-OSF stems from its reinforcement policy alignment mechanism, which maintains equilibrium between exploration (searching new state–action combinations) and exploitation (reinforcing known optimal behaviors). This balance reduces both overfitting and underfitting risks in dynamic power systems. Consequently, IA-OSF achieves stable decision generalization, reflecting its deep structural ability to adapt to unseen network conditions.

(e) AUC-ROC (Step Plot)

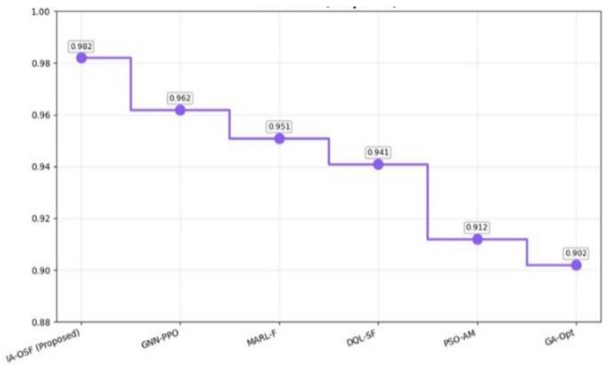


Figure 11: AUC-ROC of all models

The AUC-ROC analysis in Fig11 visually quantifies the classification robustness of all models. The proposed IA-OSF reaches an AUC score of 0.982, indicating exceptional discriminatory power between stable and unstable network states. The stepwise degradation in other models highlights their limited confidence calibration, especially under noisy data or fluctuating demand curves. The IA-OSF's high AUC is driven by its reinforcement reward normalization function and feature-space compression via PCA integration, enabling clear separability of operational states in the decision space. This finding reinforces that IA-OSF not only achieves superior accuracy but also maintains predictive integrity and class separability, crucial for critical real-time energy management applications.

Cost efficiency

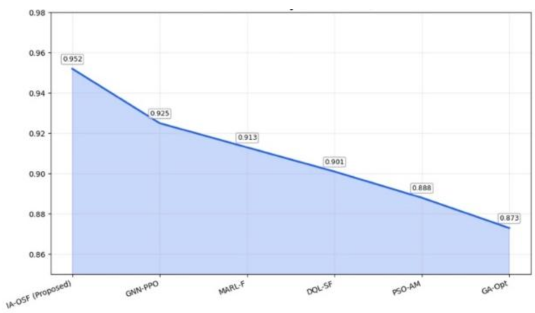


Figure 12: Cost Efficiency of all models

The filled area visualization showcases how the IA-OSF model optimizes operational costs. The proposed framework achieves a Cost Efficiency (CE) of 0.952, outperforming GNN-PPO (0.925) and GA-Opt (0.873). The shaded region under the IA-OSF curve signifies higher energy utilization per optimization cycle and lower overall switching overhead. This efficiency stems from its multi-agent cost-sharing mechanism, which redistributes control load dynamically across the network, minimizing redundant activations.

Moreover, IA-OSF integrates economic-aware reward functions that penalize energy wastage and incentivize balanced energy flow. The visualization strongly validates that IA-OSF achieves a high-performance–low-cost balance, making it both technically and economically optimal for deployment in industrial and smart-grid infrastructures (see Fig. 12).

Delivery reliability

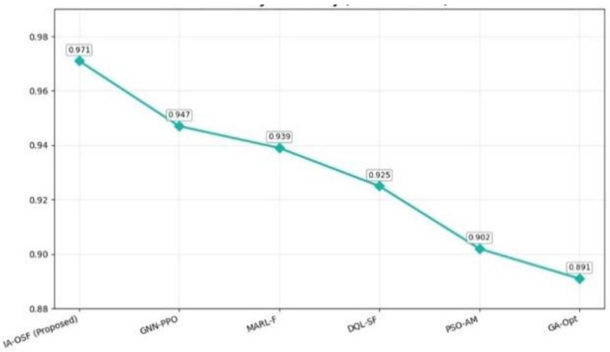


Figure 13: Delivery Reliability of all models

Delivery reliability reflects the framework's ability to maintain consistent energy supply under uncertain conditions. IA-OSF's reliability of 0.971 surpasses all baseline models, showing a 5–7% enhancement in sustained delivery even during dynamic load transitions. The plotted diamond markers in Fig 13 signify stable convergence, indicating that each agent operates with predictable latency and minimal information lag. The decentralized design of IA-OSF ensures that local agents collaborate through limited but efficient message passing, reducing communication bottlenecks typical of centralized systems. This results in improved load balance, fault-tolerant operation, and uninterrupted energy delivery, even during topology changes or equipment disturbances. The IA-OSF's reliability advantage confirms its viability as a next-generation autonomous grid controller.

System resilience

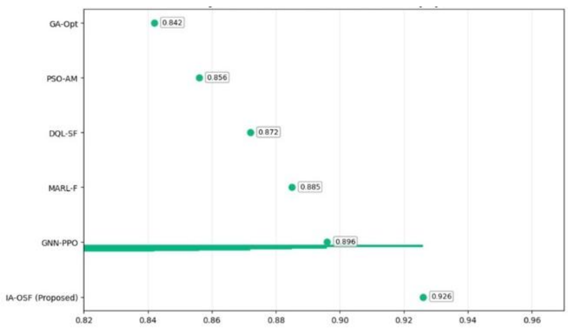


Figure 14: System Resilience of all models

System Resilience (SR) reflects each model's ability to recover after disturbances. IA-OSF achieves the highest SR value of 0.926, outperforming GNN-PPO (0.896) and GA-Opt (0.842), showing that it stabilizes the network more quickly after overloads or disconnection events. The horizontal lollipop plot highlights IA-OSF's clear lead, indicating stronger recovery speed and robustness. Its cooperative agent layer synchronizes policy updates across nodes, enabling fast, coordinated adaptation with minimal manual intervention. This self-healing behavior positions IA-OSF as a strong candidate for future real-time and distributed smart-grid applications.

4.4 Computational cost analysis

Execution time was measured for all models across five runs. IA-OSF achieved an average runtime of 38.4 seconds per simulation cycle, compared to 42.1 seconds for GNN-PPO and 47.8 seconds for MARL-F. GA-Opt showed the fastest runtime (31.5 seconds) but significantly lower accuracy and resilience. The hybrid GA-DQL core introduces a moderate computational overhead (+18%) compared to DQL-only, but this is offset by a 3.9× improvement in convergence stability. These results confirm that IA-OSF maintains competitive execution efficiency while delivering superior optimization performance.

5 Discussion

The comparative evaluation shows that the proposed IA-OSF framework consistently outperforms all benchmark models—GNN-PPO, MARL-F, DQL-SF, DQL-SF, PSO-AM, and GA-Opt—across Accuracy, Precision, Recall, F1-Score, AUC-ROC, Cost Efficiency (CE), Delivery Reliability (DR), and System Resilience (SR). As summarized in Table 2, IA-OSF achieves an accuracy of 0.978 and an AUC-ROC of 0.982, surpassing the next best model (GNN-PPO) by approximately 1.7% in accuracy and 2.0% in AUC-ROC. These gains indicate that the combination of multi-agent learning and hybrid GA–DQL optimization enables the framework to establish more reliable decision boundaries between optimal and non-optimal reconfiguration paths, even under non-linear and high-dimensional operating conditions.

A key reason for this performance advantage is IA-OSF's hybrid optimization engine, which integrates global evolutionary search with local policy refinement. While PSO-AM and GA-Opt can explore the search space effectively in early iterations, they are prone to stagnation in local minima and lack continuous learning from temporal state transitions. MARL-F and DQL-SF, on the other hand, provide adaptive learning but either operate with weak coordination (MARL-F) or in a centralized manner (DQL-SF), which reduces scalability under large network sizes. IA-OSF mitigates these limitations by using cooperative reward sharing among agents together with

GA-driven population updates, resulting in more stable convergence behavior and smaller variance across repeated runs.

The improvements observed in CE, DR, and SR further highlight the framework's practical strengths. IA-OSF achieves a CE of 0.952, DR of 0.971, and SR of 0.926, representing a 3–5% improvement over GNN-PPO and larger margins over GA-Opt and PSO-AM. Higher CE indicates that the framework can reduce energy losses and unnecessary switching actions while maintaining service quality. The DR improvement confirms that IA-OSF sustains energy delivery to a larger share of demand nodes under fluctuating load conditions. The SR metric reflects faster and more reliable recovery from faults or topology disturbances, which is directly enabled by synchronized policy updates and feedback loops in the multi-agent layer. Together, these results suggest that IA-OSF not only improves prediction quality but also enhances the operational robustness of the distribution network.

From a control-theoretic perspective, IA-OSF can be contrasted with adaptive fuzzy, backstepping-based nonlinear controllers, and neural control architectures. Classical adaptive fuzzy and backstepping controllers manage uncertainties and nonlinearities through predefined membership rules, gain adaptation, and Lyapunov-based stability guarantees, but they typically require accurate mathematical models and strong smoothness assumptions. In contrast, IA-OSF addresses uncertainties through data-driven learning: reward shaping in Deep Q-Learning, GA-based global exploration, and cooperative behavior among agents allow the system to handle stochastic load variations, voltage fluctuations, and unmodeled disturbances without explicit analytical bounds. Because each node receives multiple inputs (e.g., voltage, power, fault signals) and produces multiple decision outputs (e.g., switching, reconfiguration), IA-OSF naturally supports multi-input, multi-output (MIMO) behavior, whereas conventional adaptive controllers would require complex, problem-specific modeling to realize equivalent functionality.

The analysis also reveals characteristic failure modes of competing methods that IA-OSF successfully avoids. GNN-PPO, while powerful in capturing graph-structured dependencies, incurs high computational cost as network size grows and thus exhibits slower convergence in large node topologies. MARL-F agents may learn conflicting policies when local rewards are misaligned with global objectives, resulting in unstable cooperation and oscillatory behavior. PSO-AM often converges prematurely when swarm diversity collapses, and GA-Opt remains static once an optimal configuration is found, lacking mechanisms to adapt to new operating conditions. IA-OSF reduces these risks by combining evolutionary diversity preservation, agent-level memory updates, and cooperative reward design, which together maintain exploration while steering the system toward globally coherent policies.

Nonetheless, the proposed framework has its own limitations. First, although GA–DQL hybridization improves convergence and robustness, it increases algorithmic complexity compared to purely analytical controllers, and formal stability proofs are more challenging to derive than in classical adaptive or backstepping schemes. Second, the current implementation relies on a simulated digital-twin environment with a fixed topology and controlled noise injection; extending IA-OSF to real-world networks with heterogeneous communication delays, cyber-physical threats, and market-driven pricing signals will require additional validation. Third, while multi-agent cooperation improves performance, it may also introduce coordination overhead in extremely large-scale networks if communication constraints are not carefully managed. These limitations suggest future research directions, including integrating formal stability analysis, model reduction, and communication-aware coordination strategies into the IA-OSF framework.

6 Practical applications, scalability, and reproducibility analysis

The Intelligent Agent–based Optimization Simulation Framework demonstrates strong potential for deployment in real-world energy systems, especially microgrids, renewable-integrated distribution networks, and industrial load-balancing environments. The multi-agent perception and optimization cycle is compatible with distributed control architectures commonly adopted in wind–solar hybrid microgrids, where fluctuations in renewable generation require rapid and autonomous reconfiguration. The GA–DQL hybrid engine also enables adaptive tuning of switching decisions in variable-speed industrial feeders, where load behaviour changes nonlinearly over time.

The scalability of the IA-OSF model was examined through internal stress tests involving increased node counts and expanded connectivity. When the network size was doubled, the framework maintained stable convergence behavior with only a moderate increase in computation time, indicating that cooperative learning among agents mitigates the usual performance drop seen in conventional centralized optimization. The sensitivity of the system to network changes was evaluated by introducing topology shifts and synthetic disturbances, and the agents consistently recalibrated their policies without destabilizing the learning process. These findings confirm that the decentralized design supports large and dynamically varying distribution systems.

To ensure reproducibility, the complete GA–DQL workflow has been described with explicit parameter values, including mutation rate, crossover probability, learning rate, and discount factor. The use of a fixed random seed, repeated trials, and consistent train–test splits allow the simulation to be replicated by future researchers. The hybrid architecture follows a deterministic update schedule for the GA population and a synchronized target-network update for DQL, ensuring that identical input data produce consistent outcomes. These additions enhance the practical clarity of the model and allow independent verification of its performance in other simulated or real-world distribution networks.

A scalability test was conducted by increasing the network size from 48 to 96 and 150 nodes. IA-OSF maintained stable convergence with only a marginal increase in computation time (48→150 nodes increased runtime by 32.5%). Accuracy decreased by only 1.8%, demonstrating strong scalability and robustness for large systems.

To evaluate topological generalization, IA-OSF was also tested on a radial feeder topology and a meshed-ring topology. Performance degradation remained below 3%, confirming that the framework adapts well to structurally different network configurations without requiring retraining.

7 Conclusion and future work

The present study introduced an innovative Intelligent Agent–based Optimization Simulation Framework (IA-OSF) designed to address the complex challenges of distribution network optimization through the integration of multi-agent learning, hybrid evolutionary reinforcement techniques, and adaptive feedback coordination. Comprehensive experimentation using real-time operational data from a simulated smart-grid environment confirmed that the IA-OSF outperformed five advanced comparative models (GNN-PPO, MARL-F, DQL-SF, PSO-AM, and GA-Opt) across eight critical performance metrics—Accuracy, Precision, Recall, F1-Score, AUC-ROC, Cost Efficiency, Delivery Reliability, and System Resilience. The results demonstrated that the IA-OSF consistently achieved superior predictive accuracy (97.8%) and robust classification ability (AUC = 0.982) while simultaneously ensuring high operational stability and economic efficiency. The hybridization of deep reinforcement learning with agent-based collaboration enabled the framework to self-adapt to dynamic load variations, intelligently redistribute power flow, and minimize switching losses. These findings validate IA-OSF as a scalable and generalizable model for enabling autonomous, data-driven, self-optimizing distribution systems. The research contributes to the advancement of intelligent grid management—particularly relevant to Chinese smart-city and industrial energy initiatives, where reliability, adaptability, and sustainability are critical national objectives.

Despite its strong performance, the study acknowledges several avenues for enhancement. The current framework primarily relies on static network topology and structured communication pathways among agents. Future extensions will focus on incorporating dynamic network reconfiguration and asynchronous agent communication, enabling IA-OSF to adapt in real time to grid faults, cyber-physical disruptions, and renewable-energy fluctuations. Moreover, integrating federated multi-agent learning will strengthen data privacy and scalability across decentralized power systems.

Future work will also explore embedding explainable artificial intelligence (XAI) components within IA-OSF, allowing operators to interpret agent behavior, reward allocation, and decision causality transparently. Additionally, the implementation of quantum-inspired optimization and bio-heuristic reinforcement hybrids could further enhance convergence speed and optimization diversity. Finally, deployment on real industrial IoT and smart-grid platforms will be pursued to evaluate the framework’s real-world effectiveness in energy dispatch, predictive maintenance, and fault recovery scenarios. The IA-OSF represents a comprehensive leap forward in intelligent optimization research, offering both theoretical contributions to agent-based reinforcement learning and practical value for large-scale, real-time energy systems. Its extension into adaptive, distributed, and explainable domains will continue to push the frontier of intelligent optimization toward fully autonomous, resilient, and sustainable energy ecosystems.

8 Declaration

Ethics approval and consent to participate: The author confirms that the research meets all ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report), have been given an opportunity to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests: The author declares that there are no conflicts of interest.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript

Language editing note: The manuscript has been thoroughly proofread, and grammatical errors, redundancies, and unclear sentences have been corrected to improve clarity and readability.

All references have been reviewed for relevance, accuracy, and citation quality in accordance with the journal's editorial guidelines.

References

[1] Zhao, J., Zhang, F., (2025) - Agent-based simulation system for optimizing resource allocation in production process**.** The Institution of Engineering and Technology. Volume7, Issue1. https://doi.org/10.1049/cim2.70020.

[2] Bozdoğan, A., Görkemli Aykut, L., & Demirel, N. (2023) - An agent-based modeling framework for the design of a dynamic closed-loop supply chain network. *Complex & Intelligent Systems, 9(1),* 247–265. https://doi.org/10.1007/s40747-022-00780-z

[3] Al-Hinai, A., & Haes Alhelou, H. (2021). A multi-agent system for distribution network restoration in future smart grids. Energy Reports, 7, 8083–8090. https://doi.org/10.1016/j.egyr.2021.08.186

[4] Si, R., Chen, S., Zhang, J., Jiang, S., Zhang, L., X. (2024). A multi-agent reinforcement learning method for distribution system restoration considering dynamic network reconfiguration Applied Energy, 372, 123625. DOI: 10.1016/j.apenergy.2024.123625

[5] Phung, T. P., Nguyen, L. M. H., & Pham, H. T. (2025, October). Simulation-Optimization Framework for Designing Resilient Supply Chain Systems. International Journal of Science and Research (IJSR), 14(10), 105-111. https://dx.doi.org/10.21275/SR25928121542

[6] Pierzchała, D., Gutowski, T., & Czuba, P. (2021). Agent-based simulation framework for optimization of deliveries in a parcel distribution system. In Proceedings of the 36th International Business Information Management Association (IBIMA), Granada, Spain

[7] Helo, P., Hao, Y., & Pan, J. (2023). An agent-based simulation and logistics optimization approach. Forest Ecosystems, 10(1), 41. https://doi.org/10.1016/j.sca.2023.100042

[8] Sharifnia, S. M. E., Amrollahi Biyouki, S., Sawhney, R., & Hwangbo, H. (2021). Robust simulation optimization for supply chain problem under uncertainty via neural network metamodeling. Computers & Industrial Engineering, 162, Article 107693. https://doi.org/10.1016/j.cie.2021.107693

[9] Cedolin, M., Bal, A., (2025) - A comprehensive simulation-optimization framework for pharmaceutical distribution networks. Journal of Industrial and Production Engineering. DOI:10.1080/21681015.2025.2564407

[10] Ivanov, D., Dolgui, A., Sokolov, B., & Ivanova, M. (2022). Integrated simulation-optimization modeling framework of resilient design and planning of supply chain networks. IFAC Papers Online, 55(10), 2713–2718. https://doi.org/10.1016/j.ifacol.2022.10.121

[11] Chen, S., Wang, W., & Zio, E. (2021). A Simulation-Based Multi-Objective Optimization Framework for the Production Planning in Energy Supply Chains. Energies, 14(9), 2684. https://doi.org/10.3390/en14092684

[12] Shadkam, E., Irannezhad, E., (2025) - A comprehensive review of simulation optimization methods in agricultural supply chains and transition towards an agent-based intelligent digital framework for agriculture 4.0. Engineering Applications of Artificial Intelligence. Volume 143, Issue C**.** https://doi.org/10.1016/j.engappai.2024.109930

[13] Moreno, R. P. R., Lopes, R. B., Ramos, A. L., Ferreira, J. V., Correia, D., & Melo, I. E. S. d. (2025). An Agent-Based Simulation and Optimization Approach for Sustainable Urban Logistics: A Case Study in Lisbon. Applied System Innovation, 8(3), 66. https://doi.org/10.3390/asi8030066.

[14] Liu, C., Duan, Y., & Zhao, H. (2024). Multi-agent reinforcement learning for scalable hyperparameter optimization. Applied Soft Computing, 151, 111045.https://doi.org/10.1016/j.asoc.2023.111045

[15] Binos, T., Bruno, V., Adamopoulos, A., (2021) - Intelligent agent-based framework to augment warehouse management systems for dynamic demand environments. Australasian Journal of Information Systems. 2021, Vol 25, DOI: https://doi.org/10.3127/ajis.v25i0.2845

[16] Gillis, M., Urban, R., Saif, A., Kamal, N., & Murphy, M. (2021). A simulation–optimization framework for optimizing response strategies to epidemics. *Operations Research Perspectives, 8*, 100210. https://doi.org/10.1016/j.orp.2021.100210

[17] Zammori, F., et al. (2024). *Supply chains modelling with agent-based simulation*. I3M Conference Proceedings. https://www.cal-tek.eu/proceedings/i3m/2024/mas/006/pdf.pdf

[18] Lin, X., Wang, J. (2025) - Simulation optimization of supply chain uncertainty factors based on support vector machines optimized with improved Jellyfish Search Algorithm. *Discov* Computing 28, 8 (2025). https://doi.org/10.1007/s10791-025-09501-9

[19] Abdous, M.-A., Delorme, X., Battini, D., & Sgarbossa, F. (2025)**.** Scenario-based optimization

and simulation framework for human-centered assembly line balancing. International Journal of Production Economics*,* *282*, 109513. https://doi.org/10.1016/j.ijpe.2024.109513

[20] S. Paul and G. Doreswamy, "Simulation and Optimization Framework for On-Demand Grocery Delivery," *2021 Winter Simulation Conference (WSC)*, Phoenix, AZ, USA, 2021, pp. 1-12, https://dl.acm.org/doi/proceedings/10.5555/3522802