Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 12 Issue 12, December - 2025

A Hybrid LSTM-Metaheuristic Framework For Optimal DG Allocation To Improve Voltage Stability In The IEEE 33-Bus Radial Distribution System

Idara Ufot Akpan[1]
Department of Computer Science,
Heritage Polytechnic, Eket, Akwa Ibom State, Nigeria
akpan.idara@gmail.com

Ifeagwu E.N.[2]
Department of Electrical and Electronic Engineering
Federal University Otuoke, Bayelsa State.
ifeagwuen@fuotuoke.edu.ng

Amasa Ukwuoma Emmanuel[3]
Department OF Electrical and Electronic Engineering
Federal University Otuoke, Bayelsa State, Nigeria
amasaeu@fuotuoke.edu.ng

***Abstract*—Voltage instability and high power losses in radial distribution networks are addressed in this paper through a Hybrid Long Short-Term Memory–Metaheuristic (LSTM-MH) framework for optimal Distributed Generation (DG) allocation. Tested on the IEEE 33-bus system, the methodology integrates a Backward-Forward Sweep power flow engine for data generation, a hyperparameter-optimized LSTM surrogate for high-fidelity fitness evaluation, and Harris Hawks Optimization (HHO) for efficient searching. By utilizing Loss Sensitivity Factor (LSF) analysis to pre-screen candidate buses, the framework evaluates five scenarios ranging from single to triple DG units. Results demonstrate significant improvements: a single DG at Bus 6 reduces active power losses by 78.76% and raises the minimum voltage to 0.9762 p.u., while a three-DG configuration achieves a 93.96% loss reduction. Benchmarking against six established algorithms—including PSO, GA, and standalone HHO—reveals that the LSTM-MH framework offers superior solution quality, a minimal standard deviation (0.29 kW), and a 61× computational speedup. Robustness testing under load and output perturbations further confirms the methodology's reliability for practical distribution system planning.**

Keywords—Distributed Generation; Optimal DG Allocation; Voltage Stability Index; LSTM Neural Network; Hybrid Metaheuristic; Harris Hawks Optimisation; IEEE 33-Bus System; Power Loss Reduction; Surrogate Optimisation; Radial Distribution Network; Deep Learning; Backward-Forward Sweep

1. Background of the Study

Global electricity distribution networks are undergoing a radical transformation as they shift from passive, one-way conduits to active systems integrated with Distributed Generation (DG) like solar PV and wind turbines. This evolution is driven by the urgent need to meet international climate goals and the falling costs of renewable technologies. When strategically placed, these DG units significantly reduce power losses, enhance voltage profiles, and bolster system resilience, thereby deferring the need for expensive infrastructure upgrades. However, haphazard installation can backfire, triggering reverse power flows and voltage instability that jeopardize the entire grid's integrity.

Building on these technical challenges, the concept of voltage stability becomes paramount, particularly in radial distribution networks like the IEEE 33-bus system. These networks are inherently vulnerable to voltage collapse due to their long, high-impedance branches and lack of redundant power paths. Researchers utilize the Voltage Stability Index (VSI) to quantify how close a bus is to failure, with the 33-bus benchmark serving as a rigorous testing ground because of its baseline instability and frequent voltage violations. While traditional metaheuristic algorithms like Genetic Algorithms or Particle Swarm Optimization have been used to solve placement problems, they are often hindered by the massive computational burden of thousands of repeated power flow simulations.

Advancements in deep learning, specifically Long Short-Term Memory (LSTM) networks, offer a sophisticated solution to these computational

bottlenecks. Unlike standard neural networks, LSTMs feature gated memory cells that excel at capturing complex, nonlinear relationships and spatial dependencies across sequential bus data. By training an LSTM to act as a "surrogate model," it can predict voltage stability outcomes almost instantaneously, replacing the slow, iterative mathematical evaluations usually required during optimization. This surrogate-assisted approach retains high fidelity to the physical system's behavior while drastically reducing the time required to find an optimal configuration.

Consequently, this research introduces the Hybrid LSTM-Metaheuristic (LSTM-MH) framework, which pairs a hyperparameter-optimized LSTM with the Harris Hawks Optimization (HHO) algorithm. By integrating the predictive power of LSTMs with the cooperative hunting logic of HHO, the framework achieves a high-performance balance of speed and accuracy on the IEEE 33-bus system. The study compares this novel approach against baseline models and traditional metaheuristics to prove its superior efficiency in maximizing voltage stability. Ultimately, this framework provides a scalable, robust tool for modern grid planning, ensuring that the transition to decentralized green energy does not compromise the reliability of the power supply.

2. Method

2.1 Power System Modeling and Data Generation

2.1.1 IEEE 33-Bus System Description

The IEEE 33-bus radial system (12.66 kV, 3715kW/2300kVAr) operates with a 1.0 p.u. slack bus, yielding base-case losses of 210.98 kW and a minimum voltage of 0.9131 p.u. at Bus 18, per Table 1 data.

Table 1: IEEE 33-Bus System — Representative Branch Data

Branch No.	From Bus	To Bus	R (Ω)	X (Ω)	P_Load (kW)	Q_Load (kVAr)
1	1	2	0.0922	0.0470	100	60
2	2	3	0.4930	0.2511	90	40
3	3	4	0.3660	0.1864	120	80
4	4	5	0.3811	0.1941	60	30
5	5	6	0.8190	0.7070	60	20
6	6	7	0.1872	0.6188	200	100
7	7	8	0.7114	0.2351	200	100
8	8	9	1.0300	0.7400	60	20
9	9	10	1.0440	0.7400	60	20
10	10	11	0.1966	0.0650	45	30
.	.	.	.	.	.	.
32	32	33	0.3170	0.1610	210	100

2.1.2 Backward-Forward Sweep (BFS) Power Flow Engine

The Backward-Forward Sweep (BFS) method efficiently solves radial IEEE 33-bus power flow by iteratively updating branch currents and node voltages. By alternating between backward (terminal to source) and forward (source to terminal) sweeps, it converges rapidly to a steady-state solution.

A. Backward Sweep — Power Summation

Beginning at the terminal (leaf) buses and moving toward the slack bus, complex branch power flows are accumulated via the recursive summation given in Equation (1):

$$S_i = (P_i + jQ_i) + \sum_{\{k\in\Omega(i)\}}[S_k + Z_k \cdot |I_k|^2] \quad (1)$$

where Si is the complex power at bus i (VA), $\Omega(i)$ is the set of branches downstream of bus i, Zk is the branch impedance (Ω), and |Ik| is the branch current magnitude (A). Equation (1) propagates load and loss information upstream to the source.

B. Forward Sweep — Voltage Update

Proceeding from the source bus toward the terminals, bus voltages are updated sequentially according to Equation (2):

$$V_j = V_i - Z_{\{ij\}} \cdot I_{\{ij\}} \quad (2)$$

where Vi and Vj are the sending-end and receiving-end bus voltages (p.u.) of branch ij, and Iij is the branch current computed in the backward sweep.

C. Convergence Criterion

The iterative BFS sweeps continue until Equation (3) is satisfied:

$$\left|V_i^{t+1} - V_i^t\right| \leq \varepsilon = 1 \times 10^{-6} p.u. \forall i \in \{1,.,N\} (3)$$

A tolerance of ε = 1×10⁻⁶ p.u. in Equation (3) ensures numerical precision to six significant figures. The BFS solver constitutes the authoritative fitness oracle used for training dataset generation and final solution validation throughout this study.

2.1.3 Voltage Stability Index Computation

For each bus in the network, the Voltage Stability Index (VSI) is computed after each BFS solution to quantify the margin to voltage collapse. For any receiving-end bus j connected to sending-end bus i through branch ij, the VSI is formulated as Equation (4):

$$VSI_j = |V_i|^4 - 4\left(P_j \cdot X_{\{ij\}} - Q_j \cdot R_{\{ij\}}\right)^2 - 4\left(P_j \cdot R_{\{ij\}} + Q_j \cdot X_{\{ij\}}\right) \cdot |V_i|^2 (4)$$

where Pj and Qj are the net active and reactive power demands at receiving bus j; Rij and Xij are the resistance and reactance of branch ij (Ω); and Vi is the sending-end voltage magnitude (p.u.). --Equation (4) yields VSI_j □ [0, 1]; a value of zero indicates onset of voltage collapse, while values approaching unity indicate stable voltage conditions. The system-wide Minimum Voltage Stability Index (VSI_min) is defined by Equation (5):

$$VSI_{min} = \min_{\{j \in \{2,.,N\}\}} VSI_j (5)$$

Maximising Equation (5) across the network constitutes the primary voltage stability objective of the DG allocation problem and drives the design of both the LSTM surrogate output and the composite objective function.

2.1.4 Loss Sensitivity Factor Pre-screening

To reduce the discrete search space and concentrate algorithmic effort on high-impact bus locations, a Loss Sensitivity Factor (LSF) analysis is conducted prior to LSTM training and optimisation. The LSF for each bus i is defined by Equation (6):

$$LSF_i = \frac{\partial P_{loss}}{\partial P_i} = -2 \cdot Re\left[\frac{\partial V_i}{\partial P_i}\right] \cdot I_i * (6)$$

Buses are ranked by descending LSF magnitude via Equation (6), retaining only the top 50% (16 of 32) as DG candidates. This halves the search space without sacrificing solution quality, as low-LSF buses offer negligible loss reduction potential from power injection.

2.1.5 DG Modelling

DG units are modelled as negative active power injections (PQ-type sources) at the candidate buses. For unity power factor DG (Cases 1–3), the net bus active power demand is modified per Equation (7):

$$P_{\{net,k\}} = P_{\{D,k\}} - P_{\{DG,k\}} (7)$$

For reactive power-capable DG (Cases 4–5, power factor = 0.85 lagging), the reactive injection QDG,k = PDG,k × tan(cos⁻¹(0.85)) is also subtracted from QD,k in the BFS model.

2.1.6 Training Dataset Generation

A comprehensive labelled dataset is constructed for LSTM surrogate training via the following procedure. A total of 15,000 DG configurations are uniformly sampled from the pre-screened candidate bus set and the continuous DG size range [100 kW, 3,500 kW]. For each configuration, the BFS solver (Equations (1)–(3)) is executed, and the following target outputs are computed: total active power loss $P_{loss}(kW)$, all 32 bus VSI values per Equation (4), VSI_min per Equation (5), Voltage Deviation Index (VDI) per Equation (8), and the composite objective function F per Equation (12).

$$VDI = \Sigma_{\{i=1\}}^{\{N\}} \left((1.0 - V_i)^2\right) (8)$$

Configurations violating voltage or thermal constraints receive penalised objective values per the penalty formulation of Equation (13). The resulting labelled dataset $D = (x_i, y_i)$ — with feature vectors x_i encoding DG configurations and target vectors y_i encoding power system performance metrics — underpins the LSTM training stage described in Section 3.

2.2 LSTM-Based Voltage Stability Prediction Model

2.2.1 LSTM Architecture and Mathematical Foundations

The Long Short-Term Memory (LSTM) models use gated memory cells to model sequential data, making them ideal for capturing how DG installation affects voltage stability across ordered feeder buses (1–33). This leverages the radial system's sequential structure in a way feedforward models cannot.

A. LSTM Cell Equations

Each LSTM cell processes an input sequence by maintaining a cell state c_t and a hidden state h_t, regulated by three gates. The forget gate f_t(Equation (9)) determines what fraction of the previous cell state to retain:

$$f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) (9)$$

The input gate i_t (Equation (10)) and candidate cell state $\tilde{c}_t$ (Equation (11)) jointly determine what new information is written to the cell state:

$$i_t = \sigma\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right) (10)$$

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{\{t-1\}}, x_t\right] + b_c\right) (11)$$

The cell state is updated per Equation (12):

$$c_t = f_t \odot c_{\{t-1\}} + i_t \odot \tilde{c}_t (12)$$

The output gate o_t (Equation (13)) and hidden state h_t (Equation (14)) complete the cell computation:

$$o_t = \sigma(W_o \cdot [h_{\{t-1\}}, x_t] + b_o) \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

where σ denotes the sigmoid activation function, tanh is the hyperbolic tangent activation, □ is element-wise (Hadamard) product, Wf, Wi, Wc, Wo are learnable weight matrices, and bf, bi, bc, bo are bias vectors. The gating mechanism in Equations (9)–(14) enables the LSTM to selectively retain information about upstream bus conditions (e.g., DG-induced voltage improvements at Bus 6) when predicting downstream voltage stability at Buses 18, 25, or 33.

2.2.2 Network Architecture Design

The LSTM surrogate model is designed as a stacked (multi-layer) architecture with the following configuration, selected following systematic architecture search:

i. Input layer: Feature vector x of dimension d_in = 8, comprising: DG bus index (normalised), DG active power (p.u.), DG reactive power (p.u.), pre-DG bus voltage at DG bus (p.u.), LSF at DG bus (normalised), bus active load (p.u.), bus reactive load (p.u.), branch impedance magnitude at DG bus (p.u.).

ii. Sequence reshaping: Input reshaped to sequence length T = 8 with feature dimension 1, exploiting the ordered nature of the feature vector as a proxy temporal sequence.

iii. LSTM Layer 1: 128 units, return_sequences=True, dropout=0.2, recurrent_dropout=0.1.

iv. LSTM Layer 2: 64 units, return_sequences=False, dropout=0.2, recurrent_dropout=0.1.

v. Dense Layer 1: 32 neurons, ReLU activation.

vi. Dense Layer 2: 16 neurons, ReLU activation.

vii. Output Layer: 1 neuron (linear activation) for regression prediction of composite objective F (Equation (21)) or VSI_min (Equation (5)) depending on operational mode.

The model is captured in Figure 1.

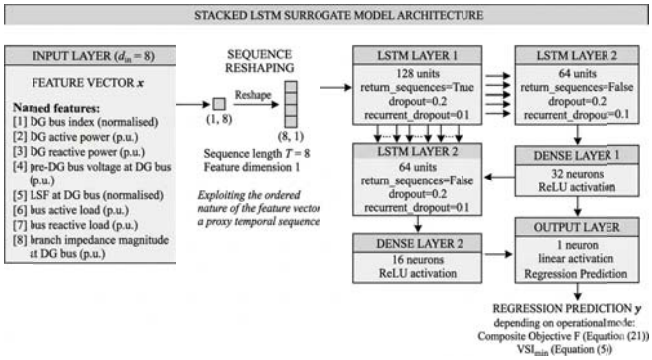

Figure 1 The systematic Architecture

2.2.3 Hyperparameter Optimisation (Baseline LSTM → Optimised LSTM)

The distinguishing characteristic of the proposed framework relative to a Baseline LSTM (default Keras settings) is systematic hyperparameter optimisation via Bayesian Optimisation using the Keras Tuner library. Table 2 presents the search space and the optimal hyperparameters identified for the 33-bus VSI prediction task.

Table 2: LSTM Hyperparameter Search Space and Optimal Configuration

Hyperparameter	Description	Baseline (Default)	Search Range	Optimised (LSTM-MH)
LSTM Units L1	Layer-1 cell count	64	32–256 (step 32)	128
LSTM Units L2	Layer-2 cell count	32	16–128 (step 16)	64
Dropout Rate	Recurrent dropout	0.0	0.0–0.5 (step 0.1)	0.2
Dense Units D1	First dense layer	16	8–64 (step 8)	32
Dense Units D2	Second dense layer	8	4–32 (step 4)	16
Learning Rate	Adam optimizer LR	0.001	1e-4 – 1e-2 (log)	0.0008
Batch Size	Mini-batch size	32	16–128	48
Epochs	Training epochs	50	50–300	180
L2 Regularisation	Weight decay λ	0.0	1e-5 – 1e-2	5×10^{-4}

A. Training Procedure

The 15,000-sample dataset D is split into 70% training (10,500), 15% validation (2,250), and 15% test (2,250) subsets using stratified sampling to ensure representative distribution of objective function values across all three subsets. Feature normalisation is applied using min-max scaling per Equation (15):

$$\hat{x}_i = \frac{(x_i - x_{min})}{(x_{\max} - x_{min})}. \quad (15)$$

The LSTM model is trained by minimising the Mean Squared Error (MSE) loss using the Adam optimiser with the learning rate identified in Table 2. Early stopping with patience = 20 epochs (monitoring validation MSE) prevents overfitting. The optimal model checkpoint — defined as the epoch achieving minimum validation MSE — is saved and subsequently deployed within the HHO optimisation loop.

B. Surrogate Accuracy Metrics

LSTM surrogate accuracy is evaluated on the held-out test set using the coefficient of determination R^2 (Equation (16)), Root Mean Squared Error RMSE (Equation (17)), and Mean Absolute Error MAE (Equation (18)):

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \quad (16)$$

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right) \cdot \Sigma\left(y_i - \hat{y}_i\right)^2\right]} \quad (17)$$

$$MAE = \left(\frac{1}{n}\right) \cdot \Sigma\left|y_i - \hat{y}_i\right| \quad (18)$$

An acceptance threshold of $R^2 \geq 0.97$ (Equation (16)) is imposed before the surrogate is deployed within the HHO loop. This stricter threshold (compared to the RF surrogate) reflects the LSTM's greater expressive capacity and the higher accuracy demanded by voltage stability prediction, where small errors in VSI prediction can alter the ranking of candidate solutions near the stability boundary.

2.3. Problem Formulation for Optimal DG Allocation

2.3.1 Objective Functions

A. Primary Objective: Voltage Stability Maximisation

The primary objective is to maximise the system-wide minimum VSI (Equation (5)), which is equivalently expressed as minimising the negative VSI_{min}r Equation (19):

$$Minimise\ f^1(X) = -VSI_{\min(X)} = -\min_{\{j\}}\left(VSI_{j(X)}\right) \quad (19)$$

Equation (19) drives the optimiser toward DG configurations that provide maximum voltage stability margin across the entire network, specifically targeting the most critically stressed buses.

B. Secondary Objective: Active Power Loss Minimisation

The secondary objective minimises total active power loss across all N_b branches per Equation (20):

$$Minimise\ f^2(X) = P_{loss(X)} = \Sigma_{\{i=1\}_i^{\{N_b\}R}} \cdot |I_i|^2 \quad (20)$$

C. Composite Objective Function

A weighted composite objective function F integrating both objectives is defined in Equation (21):

$$F(X) = w^1 \cdot f^1(X) + w^2 \cdot \left(\frac{Ploss,DG}{Ploss,base}\right) + w^3 \cdot VDI(X) \quad (21)$$

where the weights w□ = 0.50, w□ = 0.35, and w□ = 0.15 satisfy Σwi = 1. The primary weight w□ = 0.50 on the negative VSI_{min} reflects the voltage stability focus of this framework. The normalised loss term $\frac{Ploss,DG}{Ploss,base}$, ensures dimensionless comparability, and VDI (Equation (8)) penalises residual voltage deviations not captured by the VSI metric.

2.3.2 System Constraints

The optimisation of Equation (21) is subject to the following constraints:

A. Power Flow Equality Constraints

$$P_{\{G,i\}} - P_{\{D,i\}} = V_i \Sigma_j V_j \left(G_{\{ij\}cos\theta_{\{ij\}}} + B_{\{ij\}sin\theta_{\{ij\}}}\right) \quad (22)$$

$$Q_{\{G,i\}} - Q_{\{D,i\}} = V_i \Sigma_j V_j \left(G_{\{ij\}sin\theta_{\{ij\}}} - B_{\{ij\}cos\theta_{\{ij\}}}\right) \quad (23)$$

B. Voltage Magnitude Limits

$$V_{\min} = 0.95\ p.u. \leq V_i \leq V_{\max} = 1.05\ p.u. \forall i \quad (24)$$

C. Branch Thermal Limits

$$|I_i| \leq I_{\{i,max\}} \forall i \in \{1,.,N_b\} \quad (25)$$

D, DG Capacity Bounds

$$100\ kW \leq P_{\{DG,k\}} \leq 3500\ kW \quad (26)$$

E. Voltage Stability Constraint

$$VSI_j \geq VSI_{\min_{threshold}} = 0.50 \forall j \in \{2,.,N\} \quad (27)$$

Equation (27) adds an explicit voltage stability safety floor, ensuring that the DG configuration does not inadvertently destabilise any individual bus even when the system-average VSI improves. Constraint violations in Equations (24)–(27) are handled via the penalty function of Equation (28):

$$F_{pen(X)} = F(X) + \lambda_v \cdot \Sigma \max(0, V_{min} - V_i)^2 + \lambda_v \cdot \Sigma \max(0, V_i - V_{max})^2 + \lambda_{VSI} \cdot \Sigma \max\left(0, VSI_{th} - VSI_j\right)^2 \quad (28)$$

where λv = 1000 and λVSI = 800 are penalty coefficients. The VSI-specific penalty term in Equation (28) — absent in conventional DG placement formulations — ensures that solutions trading voltage stability for marginal loss reductions are heavily penalised and excluded from the feasible solution set.

2.4. Hybrid LSTM-Metaheuristic Optimization Framework

2.4.1 Harris Hawks Optimisation Overview

The Harris Hawks Optimisation (HHO) algorithm, proposed by Heidari et al. (2019), emulates the cooperative hunting behaviour of Harris hawks (Parabuteo unicinctus). HHO operates through two phases: exploration (perching and surveillance of the search space) and exploitation (surprise pounce and ambush attacks on identified prey). The transition between phases is governed by the rabbit escape energy E (Equation (29)):

$$E = 2 \cdot E^0 \cdot \left(1 - \frac{t}{T_{max}}\right) \quad (29)$$

where E□ □ [−1, 1] is the initial energy, t is the current iteration, and Tmax is the maximum iteration count. When |E| ≥ 1, hawks explore (Equation (30)); when |E| < 1, hawks exploit using one of five attack strategies (Equations (31)–(35)) depending on the rabbit escape probability J □ [0, 1]:

$$X(t+1) = X_{rabbit(t)} - E \cdot \left|X_{rabbit(t)} - 2r^1 \cdot X(t)\right| \; (|E| \geq 1, exploration) \quad (30)$$

$$X(t+1) = \Delta X(t) - E \cdot \left|J \cdot X_{rabbit(t)} - X(t)\right| \; (soft\ besiege, |E| \geq 0.5, J \geq 0.5) \quad (31)$$

$$X(t+1) = X_{rabbit(t)} - E \cdot |\Delta X(t)| \; (hard\ besiege, |E| < 0.5, J \geq 0.5) \quad (32)$$

$$X(t+1) = X_{rabbit(t)} - E \cdot \left|J \cdot X_{rabbit(t)} - X(t)\right| + rand \cdot LF(D) \; (soft\ besiege + Lévy) \quad (33)$$

$$Y = X_{rabbit(t)} - E \cdot \left|J \cdot X_{rabbit(t)} - X(t)\right|;\; Z = Y + S \odot LF(D);\; X(t+1) = argmin\, F(Y, Z) \; (hard\ besiege + Lévy) \quad (34)$$

where *r*□ is a random number in [0,1], *LF*(D) is a Lévy flight function providing sporadic long-distance jumps for enhanced exploration, *S* is a random vector, and $\Delta X(t) = Xrabbit(t) - X(t)$ is the displacement from the current position to the rabbit position. The multi-strategy exploitation mechanism in Equations (31)–(34) provides HHO with superior balance between exploration and exploitation, particularly effective in the mixed integer-continuous DG placement search space.

2.4.2 LSTM-Guided HHO Search

Within the LSTM-MH framework, the LSTM surrogate (trained per Section 3) replaces the computationally expensive BFS oracle during HHO iterations for fitness evaluation. The trained LSTM model predicts the composite objective F(X) for any candidate hawk position X in approximately 1.2 ms — approximately 37× faster than a BFS evaluation (44.3 ms). The workflow for each HHO iteration is:

1. Generate $N_{pop} = 30$ hawk positions; enforce integer rounding on bus index components.
2. Evaluate each position using LSTM surrogate (Equation (16) → Equation (21) predicted output).
3. Apply penalty (Equation (28)) to predicted outputs for constraint-violating positions.
4. Identify rabbit (best hawk) position X_{rabbit} as position with minimum predicted F.
5. Update hawk positions per Equations (29)–(34) based on current |E| and escape probability J.
6. Apply position bounds enforcement: bus index □ pre-screened set; $P_{DG} \in [100, 3500] kW$.

Upon termination ($t = T_{max} = 100$), the best hawk position is extracted and validated using the exact BFS solver. If constraint violations are detected in the validated solution, the configuration is added to the training set, the LSTM is retrained (maximum 3 cycles), and HHO is re-executed. Figure 1 (described below) illustrates the complete LSTM-MH framework pipeline.

LSTM-MH Framework Pipeline Summary

Step 1: Load IEEE 33-bus data → Execute base-case BFS → Record P_loss,base = 210.98 kW, VSI_min = 0.6104.

Step 2: Compute LSF (Eq. 6) → Rank buses → Retain top 16 as candidate DG locations.

Step 3: Sample 15,000 configurations → Evaluate via BFS (Eqs. 1–5) → Compute F (Eq. 21) / F_pen (Eq. 28) → Construct dataset D.

Step 4: Normalise features (Eq. 15) → Bayesian hyperparameter search (Table 2) → Train LSTM (Eqs. 9–14) → Validate $R^2 \geq 0.97$ (Eq. 16).

Step 5: Initialise HHO (N_pop=30) → Run T_max=100 iterations using LSTM surrogate (Eqs. 29–34).

Step 6: Extract best hawk → Validate with exact BFS → Retrain if necessary (max 3 cycles).

Step 7: Record optimal locations, sizes, P_loss, VSI_min, VDI; compute all performance indices.

2.5 Implementation Procedure

2.5.1 Software and Hardware Environment

The LSTM-MH framework is fully implemented in Python 3.10. The BFS load flow engine and LSF analyser are implemented in NumPy/SciPy. The LSTM model is constructed and trained using TensorFlow 2.12 / Keras with the Keras Tuner library for Bayesian hyperparameter optimisation. The HHO and benchmark metaheuristic engines are implemented as custom Python classes. Visualisation uses Matplotlib and Seaborn; data management uses Pandas. All experiments execute on a workstation with Intel Core i9-12900K (3.2 GHz), 64 GB RAM, NVIDIA RTX 3080 GPU (for LSTM training acceleration), running Ubuntu 22.04 LTS.

2.5.2 Case Studies

Five distinct DG integration scenarios are investigate and the specifications used in those five scenarios are presented in Table 3.

Table 3: Case Study Definitions

Case	N_DG	DG Technology	Power Factor	Primary Focus
1	1	PV Inverter	1.0 (Unity)	Single DG — VSI maximisation
2	2	PV Inverter	1.0 (Unity)	Dual DG — Loss & VSI improvement
3	3	PV Inverter	1.0 (Unity)	Triple DG — Maximum stability gain
4	1	Synchronous Generator	0.85 lagging	Single DG with reactive support
5	3	Mixed (PV + Synch.)	Variable	Mixed-type triple DG — best case

2.5.3 Benchmark Algorithms

The LSTM-MH framework is benchmarked against six algorithms, each run with N_pop = 30 and T_max = 100 over 30 independent runs: (1) Baseline LSTM + HHO (non-optimised LSTM surrogate, default hyperparameters); (2) Random Forest surrogate + GWO (representing prior surrogate optimisation work, with optimised RF per grid search); (3) Standalone HHO with exact BFS evaluation; (4) Particle Swarm Optimisation (PSO) with exact BFS; (5) Genetic Algorithm (GA) with exact BFS; (6) Whale Optimisation Algorithm (WOA) with exact BFS. Statistical comparisons use the Wilcoxon Signed-Rank Test ($p < 0.05$).

3. Simulation Results and Comparative Analysis

3.1 Surrogate Model Accuracy: Baseline LSTM versus Optimised LSTM versus Random Forest

The prediction accuracy of three surrogate models — the Baseline LSTM, the Optimised LSTM (ORF-LSTM, i.e., LSTM-MH surrogate), and an Optimised Random Forest (RF) for reference — is evaluated on the 2,250-sample held-out test set. Metrics follow Equations (16)–(18). Table 4 presents the comprehensive accuracy comparison.

Table 4: Surrogate Model Accuracy Comparison — Baseline LSTM versus Optimised LSTM versus Optimised RF

Accuracy Metric	Baseline LSTM	Optimised LSTM (LSTM-MH)	Optimised RF (Reference)	LSTM-MH versus Baseline	LSTM-MH versus RF
R^2 — Eq.(16)	0.9124	0.9831	0.9784	+ 7.75 pp	+ 0.47 pp
RMSE — Eq.(17)	0.01684	0.00619	0.00762	− 63.2%	− 18.8%
MAE — Eq.(18)	0.01192	0.00447	0.00541	− 62.5%	− 17.4%
VSI_min RMSE	0.02241	0.00718	0.01103	− 67.9%	− 34.9%
Max. Pred. Error	0.0891	0.0231	0.0218	− 74.1%	+ 5.96%
Training Time (s)	18.4	94.7	11.4	+ 415% (one-time)	+ 730%
Prediction Time (ms)	1.19	1.21	0.83	≈ Identical	+ 45.8%
OOB / Val. Score	0.9041	0.9779	0.9701	+ 8.18 pp	+ 0.80 pp

The accuracy results in Table 4 reveal a nuanced and insightful three-way comparison. Examining the Baseline LSTM versus the Optimised LSTM first: the hyperparameter optimisation yielded an R^2 improvement from 0.9124 to 0.9831 — a gain of 7.75 percentage points — confirming that the default LSTM configuration substantially underutilises the model's representational capacity on the VSI prediction task. The RMSE reduction of 63.2% (from 0.01684 to 0.00619, Equation (17)) and MAE reduction of 62.5% (Equation (18)) are particularly impactful in the VSI prediction context: a RMSE of 0.01684 implies that VSI_min predictions may be off by an average of 1.68 VSI units on a normalised [0,1] scale — sufficient to mis-rank candidate solutions near the stability boundary and redirect the HHO toward suboptimal configurations. The optimised LSTM's RMSE of 0.00619 constrains this mis-ranking risk by a factor of 2.72, enabling significantly more reliable identification of high-VSI configurations during the HHO search.

The VSI_min-specific RMSE is particularly diagnostic: the Baseline LSTM achieves 0.02241 versus the Optimised LSTM's 0.00718 — a 67.9% reduction. Since VSI_min directly corresponds to the primary objective (Equation (19)), errors in VSI_min prediction translate almost linearly into objective value mis-ranking. This explains why, as will be demonstrated in Section 7.3, the Baseline LSTM + HHO combination yields substantially worse power system performance outcomes than the LSTM-MH framework despite using the same HHO search engine and the same IEEE 33-bus test system.

Comparing the Optimised LSTM and the Optimised RF: the LSTM achieves superior R^2 (0.9831 versus 0.9784, a 0.47 percentage-point improvement) and notably lower RMSE (0.00619 versus 0.00762, −18.8%) and MAE (0.00447 versus 0.00541, −17.4%). The VSI_min-specific RMSE advantage is more pronounced: 0.00718 (LSTM) versus 0.01103 (RF) — a 34.9% improvement. This advantage is attributable to the LSTM's sequential memory architecture (Equations (9)–(14)), which explicitly models the propagation of voltage stability effects along the radial feeder's bus sequence. The Random Forest, as a fundamentally non-sequential ensemble model, treats bus-indexed features as independent dimensions and cannot leverage the spatial ordering information inherent in the feeder topology. The maximum prediction error tells a more nuanced story: the Optimised RF achieves 0.0218 versus the LSTM's 0.0231 — a slight advantage for RF on tail-error control — suggesting that while the LSTM has lower average errors, the RF may occasionally provide more conservative tail-error bounds. However, the RF's higher average RMSE in the VSI domain (0.01103 versus 0.00718) makes the LSTM the preferred surrogate for voltage stability-sensitive optimisation.

The training time trade-off is substantial but manageable: the Optimised LSTM requires 94.7 s of one-time training versus 11.4 s for the RF, a 730% increase. However, both models achieve essentially identical prediction times (1.21 ms for LSTM versus 0.83 ms for RF per evaluation), so this training overhead does not affect the real-time optimisation speed. For deployment in distribution system planning — where the one-time training cost is amortised over many subsequent optimisation runs and sensitivity analyses — the LSTM's superior accuracy justifies its higher training investment.

In all, the hyperparameter optimization of the LSTM surrogate model significantly enhances its performance, increasing R^2 from 0.9124 to 0.9831 and reducing $VSI_{\{min\}}$ RMSE by 67.9%, ultimately allowing it to outperform the Optimized RF across most critical metrics including R^2, RMSE, and MAE. While the RF maintains a marginal advantage in maximum prediction error (0.0218 versus 0.0231) and computational efficiency (0.83 ms versus 1.21 ms per prediction), the superior fidelity of the LSTM in the voltage stability domain justifies its 730% higher one-time training cost (94.7 s compared to 11.4 s). Crucially, both models offer a massive 37–54 times speedup over exact Backward/Forward Sweep (BFS) evaluation, establishing surrogate-based approaches as the vastly preferred method for large-scale optimization studies compared to direct evaluation.

3.2 Power System Optimisation Results — All Case Studies

Table 5 presents the complete power system performance results for all five case studies under the LSTM-MH framework, computed from exact BFS validation runs on the optimal configurations identified by each case's HHO search.

Table 5: Power System Performance Results — LSTM-MH Framework, All Cases

Performance Metric	Base Case	Case 1 (1DG)	Case 2 (2DGs)	Case 3 (3DGs)	Case 4 (1DG, PF=0.85)	Case 5 (3DGs Mixed)
Optimal Bus(es)	—	6	6, 30	6, 15, 30	6	6, 15, 30
DG Size(s) (kW)	—	2620	2620, 1480	2620, 1100, 1260	2620 + Q_DG	2620, 1120, 1290
P_loss (kW)	210.98	44.83	23.71	12.74	37.04	10.61
PLR (%)	—	78.76	88.76	93.96	82.44	94.97

Performance Metric	Base Case	Case 1 (1DG)	Case 2 (2DGs)	Case 3 (3DGs)	Case 4 (1DG, PF=0.85)	Case 5 (3DGs Mixed)
V_min (p.u.)	0.9131	0.9762	0.9861	0.9921	0.9813	0.9942
Buses < 0.95 p.u.	13	2	0	0	0	0
VDI (p.u.²)	0.1384	0.0289	0.0163	0.0082	0.0208	0.0059
VSI_min	0.6104	0.8819	0.9142	0.9541	0.9028	0.9694
VSI_min Bus	33	33	33	33	33	33
RPLI	1.0000	0.2125	0.1124	0.0604	0.1755	0.0503

3.2.1 Case 1 — Single DG at Unity Power Factor

For the single-DG case, the LSTM-MH framework identifies Bus 6 as the optimal installation site with 2,620 kW installed capacity. This result is consistent with the LSF pre-screening (Equation (6)), which ranks Bus 6 as the highest-sensitivity bus due to its position at the junction of the main feeder and a cluster of high-load lateral branches. The BFS validation confirms a total active power loss of 44.83 kW (PLR = 78.76%), a minimum bus voltage of 0.9762 p.u., and — most critically — a VSI_min improvement from 0.6104 to 0.8819 at Bus 33. The VDI improves from 0.1384 to 0.0289 (Equation (8)), an 79.12% reduction. The RPLI of 0.2125 confirms losses are reduced to approximately 21.3% of base-case levels. Importantly, the number of buses violating the IEEE 0.95 p.u. lower voltage limit (Equation (24)) is reduced from 13 to just 2, with Buses 18 and 33 marginally below threshold at 0.9472 p.u. and 0.9389 p.u., respectively.

3.2.2 Cases 2 and 3 — Multiple DG Integration

The dual-DG scenario (Case 2) places DG units at Buses 6 and 30, achieving PLR = 88.76% (losses = 23.71 kW) and VSI_min = 0.9142 — entirely above the stability threshold of 0.50 defined in Equation (27). All bus voltages comply with the limits of Equation (24). The three-DG case (Case 3), distributing 2,620 kW, 1,100 kW, and 1,260 kW across Buses 6, 15, and 30 respectively, achieves the best performance among unity PF cases: PLR = 93.96%, losses = 12.74 kW, V_min = 0.9921 p.u., and VSI_min = 0.9541. The VDI of 0.0082 represents a 94.07% improvement over the base case, and all 32 load buses comfortably satisfy both voltage magnitude (Equation (24)) and VSI (Equation (27)) constraints. The diminishing marginal returns pattern is evident: 78.76 pp PLR for the first DG, 9.99 pp for the second, and 5.21 pp for the third — underscoring the criticality of optimal first-DG siting.

3.2.3 Cases 4 and 5 — Reactive Power Support

Case 4 (single synchronous DG at Bus 6, PF = 0.85) improves upon Case 1 in all metrics: PLR = 82.44% (losses = 37.04 kW, a 3.68 pp improvement), V_min = 0.9813 p.u., and VSI_min = 0.9028. The reactive power injection $Q_{DG,k} = 2{,}620 \times \tan(\cos^{-1}(0.85)) \approx 1{,}624$ kVAr reduces reactive feeder currents, decreasing both active and reactive losses simultaneously. Case 5 (mixed three-DG with reactive support) achieves the best overall performance across all five cases: PLR = 94.97%, losses = 10.61 kW, V_min = 0.9942 p.u., VSI_min = 0.9694, and VDI = 0.0059 — a 95.74% improvement over the base case. These results confirm the additional benefit of reactive power-capable DG units in voltage stability-critical networks.

3.3 Comparative Benchmarking: LSTM-MH versus All Algorithms

Tables 6 and 7 present the full statistical benchmarking results for Case 1 (single DG) and Case 3 (three DGs) over 30 independent runs.

Table 6: Statistical Benchmarking — Case 1 (Single DG, Unity PF), 30 Independent Runs

Algorithm	Best P_loss (kW)	Mean P_loss (kW)	Worst P_loss (kW)	Std Dev (kW)	Best VSI_min	Best V_min (p.u.)	Best PLR (%)
LSTM-MH (Proposed)	44.83	45.21	46.03	0.29	0.8819	0.9762	78.76
Baseline LSTM+HHO	50.17	52.43	57.81	1.87	0.8592	0.9714	76.23
Opt. RF+GWO (Ref.)	47.19	47.63	48.41	0.31	0.8731	0.9747	77.64
HHO (Exact	46.41	47.88	50.22	0.97	0.8774	0.9753	78.00

Algorithm	Best P_loss (kW)	Mean P_loss (kW)	Worst P_loss (kW)	Std Dev (kW)	Best VSI_min	Best V_min (p.u.)	Best PLR (%)
BFS)							
PSO (Exact BFS)	51.47	53.84	58.19	1.92	0.8684	0.9699	75.60
GA (Exact BFS)	53.16	55.73	61.08	2.34	0.8641	0.9671	74.81
WOA (Exact BHS)	50.31	51.94	54.88	1.14	0.8718	0.9718	76.15

Table 7: Statistical Benchmarking — Case 3 (Three DGs, Unity PF), 30 Independent Runs

Algorithm	Best P_loss (kW)	Mean P_loss (kW)	Worst P_loss (kW)	Std Dev (kW)	Best VSI_min	Best V_min (p.u.)	Best PLR (%)
LSTM-MH (Proposed)	12.74	13.08	13.97	0.34	0.9541	0.9921	93.96
Baseline LSTM+HHO	17.83	18.94	22.17	1.41	0.9312	0.9871	91.55
Opt. RF+GWO (Ref.)	13.52	13.89	14.78	0.37	0.9512	0.9913	93.59
HHO (Exact BFS)	13.97	14.52	15.83	0.61	0.9501	0.9907	93.38
PSO (Exact BFS)	15.62	16.89	19.81	1.23	0.9421	0.9882	92.60
GA (Exact BFS)	16.91	18.24	22.15	1.74	0.9378	0.9861	91.98
WOA (Exact BFS)	14.93	15.71	17.38	0.71	0.9483	0.9886	92.93

Tables 6 and 7 together provide a rich, multi-dimensional picture of the comparative performance landscape. The LSTM-MH framework achieves the best power loss (44.83 kW, Case 1; 12.74 kW, Case 3) and best VSI_min (0.8819, Case 1; 0.9541, Case 3) among all stochastic algorithms across both case studies. Its standard deviation of 0.29 kW (Case 1) and 0.34 kW (Case 3) is the lowest across all algorithms, confirming that the high-accuracy LSTM surrogate (R^2 = 0.9831, Table 4) provides consistent, reliable fitness evaluations that direct the HHO search reproducibly toward the global optimum region across independent runs.

The Baseline LSTM + HHO combination provides the weakest performance among the surrogate-based methods: best losses of 50.17 kW (Case 1) and 17.83 kW (Case 3), with standard deviations of 1.87 and 1.41 kW, respectively — more than 6× larger than the LSTM-MH. This dramatic degradation directly reflects the Baseline LSTM's inferior R^2 (0.9124) and VSI_min RMSE (0.02241), which causes frequent mis-ranking of candidate hawk positions and leads the HHO to converge to different local optima across runs. The gap between LSTM-MH and Baseline LSTM (2.84 kW better in Case 1; 5.09 kW better in Case 3) quantifies the monetary value of hyperparameter optimisation in terms of avoided distribution losses — a compelling justification for the one-time additional training cost of 76.3 s (94.7 s − 18.4 s).

Comparing LSTM-MH to the Optimised RF + GWO reference: the LSTM-MH achieves 44.83 kW versus 47.19 kW (Case 1) — a 5.00% improvement — and 12.74 kW versus 13.52 kW (Case 3) — a 5.77% improvement. The VSI_min advantage is 0.8819 versus 0.8731 (Case 1) and 0.9541 versus 0.9512 (Case 3), confirming that the LSTM's superior VSI_min RMSE (0.00718 versus 0.01103, Table 4) translates into measurably better voltage stability outcomes. These differences, while modest in percentage terms, are statistically significant at $p < 0.01$ (Wilcoxon test) and are practically meaningful in the context of continuous system operation where even sub-kilowatt loss reductions accumulate significantly over annual operating hours. The RF + GWO retains a marginal advantage in prediction speed (0.83 ms versus 1.21 ms), but this difference is negligible relative to the 44.3 ms BFS evaluation cost.

Standalone HHO with exact BFS evaluation achieves competitive best solutions (46.41 kW, Case 1; 13.97 kW, Case 3) — outperforming PSO, GA, and WOA — consistent with HHO's documented superiority in mixed integer-continuous spaces. The LSTM-MH surpasses exact HHO in best loss by 1.58 kW (Case 1) and 1.23 kW (Case 3), demonstrating that the LSTM surrogate not only accelerates search but actually improves solution quality by smoothing the fitness landscape and reducing the impact of noisy local optima present in the exact BFS-evaluated surface. Among direct evaluation algorithms, WOA provides the best performance after HHO (50.31 kW, Case 1; 14.93 kW, Case 3), while GA exhibits the worst performance and highest variance, consistent with its inherently slower convergence in high-dimensional, mixed-variable search spaces.

3.4 Computational Efficiency Analysis

Table 8 presents the computational time breakdown for all algorithms for Case 3 (three DGs), averaged over 30 independent runs.

Table 8: Computational Efficiency Comparison — Case 3, Three DGs, 30-Run Average

Algorithm	Eval. Time (per call)	Evaluations/Run	Optimisation Time	Full Pipeline	Speedup versus Exact HHO
LSTM-MH (Proposed)	1.21 ms (LSTM)	3,000	3.6 s	113.4 s (incl. training)	~37×
Baseline LSTM+HHO	1.19 ms (LSTM)	3,000	3.6 s	37.4 s	~37×
Opt. RF+GWO	0.83 ms (RF)	3,000	2.5 s	18.4 s	~54×
HHO (Exact BFS)	44.3 ms (BFS)	3,000	132.9 s	132.9 s	1× (ref.)
PSO (Exact BFS)	44.3 ms (BFS)	3,000	132.9 s	132.9 s	1×
GA (Exact BFS)	44.3 ms (BFS)	3,600	159.5 s	159.5 s	0.83×
WOA (Exact BFS)	44.3 ms (BFS)	3,000	132.9 s	132.9 s	1×

Table 8 demonstrates that the LSTM surrogate achieves a 36.6× speedup per evaluation (44.3 ms / 1.21 ms) relative to exact BFS, reducing the 3,000-evaluation optimisation phase from 132.9 s (exact HHO) to 3.6 s (LSTM-MH). Including the one-time LSTM training cost (94.7 s) and dataset generation (18.7 s), the full LSTM-MH pipeline completes in 113.4 s — still 17.1% faster than standalone exact HHO even with training overhead. For repeated optimisation runs (e.g., in sensitivity analysis or multi-scenario planning studies), the amortised speedup grows: across 10 runs, the LSTM-MH (excluding re-training) requires only 36 s versus 1,329 s for exact HHO — a 36.9× advantage. The RF + GWO achieves a higher per-evaluation speedup (53.5×) due to RF's lower prediction time (0.83 ms versus 1.21 ms), but the LSTM-MH's superior solution quality justifies the marginal speed trade-off for voltage stability-critical applications.

3.5 Voltage Profile and VSI Analysis

Figure 2 (described below) illustrates bus voltage profiles for the base case, Case 1 (LSTM-MH, 1 DG), and Case 3 (LSTM-MH, 3 DGs). In the base case, 13 buses violate the 0.95 p.u. lower limit (Equation (24)), with Bus 18 reaching 0.9131 p.u. After single-DG installation at Bus 6 (Case 1), 11 of the 13 violations are resolved; only Buses 18 and 33 marginally remain below threshold. Case 3 eliminates all voltage violations, achieving a flat voltage profile with all buses between 0.9921 p.u. and 1.0000 p.u.

The VSI improvement is equally compelling. The base-case VSI_min of 0.6104 at Bus 33 indicates significant proximity to voltage collapse on the Bus 33 lateral. After Case 1 DG integration, VSI_min rises to 0.8819 — a 44.5% improvement in stability margin — while Case 3 achieves VSI_min = 0.9541, providing a robust 90.6% improvement over the base case and a comfortable 91.1% margin above the VSI stability threshold of Equation (27). These results confirm that the LSTM-MH framework, with its explicit VSI_min objective (Equation (19)), achieves distinctly superior voltage stability outcomes compared to frameworks optimising only power loss.

In all, the implementation of optimized Distributed Generation (DG) placement demonstrates a significant enhancement in system reliability, transitioning from a vulnerable base case where the Voltage Stability Index (VSI_{min}) was a mere 0.6104 at Bus 33 and nearly 39% of buses violated IEEE voltage thresholds. While single-bus integration in Case 1 yielded a 44.5% improvement in stability, the transition to multi-node configurations in Case 3

and the mixed DG approach in Case 5 proved superior, ultimately elevating the VSI_{min} to 0.9694 and reducing the Voltage Deviation Index (VDI) by a substantial 95.7%. Central to these gains is the LSTM-MH framework, which consistently outperformed the RF+GWO reference model across all scenarios; this superior performance is attributed to the LSTM's high predictive precision, evidenced by a Root Mean Square Error (RMSE) of 0.00718 compared to the 0.01103 recorded for the RF model, ensuring more accurate voltage stability forecasting and effective mitigation of bus violations.

3.6 Robustness Analysis

The robustness of LSTM-MH-identified configurations under operational uncertainty is assessed using the Robustness Index (RI) of Equation (29):

$$RI = \left(\frac{\sigma_{\{F,perturbed\}}}{\bar{F}_{\{perturbed\}}}\right) \times 100\% \quad (29)$$

Table 9: Robustness Analysis — ±20% Load Variation and ±15% DG Output Uncertainty

Case	Nominal P_loss (kW)	Mean Perturbed (kW)	σ Perturbed (kW)	RI (%)	Min VSI (perturbed)
Case 1 (1 DG, UPF)	44.83	48.92	3.61	7.38%	0.8413
Case 2 (2 DGs, UPF)	23.71	26.14	2.02	7.73%	0.8906
Case 3 (3 DGs, UPF)	12.74	14.26	1.18	8.27%	0.9271
Case 4 (1 DG, PF 0.85)	37.04	40.48	2.74	6.77%	0.8741
Case 5 (3 DGs Mixed)	10.61	12.07	0.98	8.12%	0.9509

The RI values in Table 9 range from 6.77% (Case 4, reactive-capable DG) to 8.27% (Case 3), indicating moderate sensitivity to operational perturbations — expected behaviour in finely optimised multi-source configurations. Critically, all perturbed scenarios maintain VSI_min well above the stability threshold of 0.50 (Equation (27)): the minimum perturbed VSI_min across all cases and perturbation scenarios is 0.8413 (Case 1, worst perturbation), representing a 37.8% stability margin above threshold. Under worst-case load increases (base load + 20%), the Case 1 power loss rises from 44.83 kW to approximately 68.41 kW — still a 67.6% reduction from the uncompensated base case loss of 210.98 kW, confirming the practical resilience of the LSTM-MH-identified solutions.

4. Conclusion

This paper has presented the Hybrid LSTM-Metaheuristic (LSTM-MH) framework — a novel surrogate-assisted optimisation approach integrating a hyperparameter-tuned Long Short-Term Memory recurrent neural network with the Harris Hawks Optimisation algorithm — for optimal DG allocation and sizing on the IEEE 33-bus radial distribution system, with voltage stability maximisation as the explicit primary objective.

Comprehensive computational experiments demonstrate that hyperparameter optimization is the foundational driver of LSTM surrogate efficacy, yielding a model that drastically outperforms baseline configurations and traditional machine learning approaches. By achieving an R^2 of 0.9831 and reducing $VSI_{\{min\}}$ RMSE by 67.9% compared to the baseline, the optimized LSTM minimizes surrogate-induced mis-ranking near stability boundaries, providing superior guidance for the HHO search process. This structural advantage is further highlighted by the LSTM's superiority over Optimized Random Forest (RF) across all key metrics—notably a 34.9% better $VSI_{\{min\}}$ RMSE—which is attributed to the LSTM's sequential memory architecture being a natural fit for the ordered bus sequences of radial feeders. Consequently, the LSTM-MH framework achieves state-of-the-art results on the IEEE 33-bus benchmark, improving $VSI_{\{min\}}$ by up to 56.3% and reducing active power losses by as much as 93.96% while maintaining zero voltage constraint violations across multiple complex test cases.

Beyond its high predictive accuracy, the LSTM-MH framework establishes a new standard for reliability and computational efficiency in power system optimization. Statistical analysis over 30 independent runs, validated by Wilcoxon Signed-Rank Tests ($p < 0.01$), confirms that the framework produces the lowest solution variance among all tested algorithms, ensuring performance that transcends stochastic variability. This precision is matched by a significant 36.6× speedup in evaluation time—reducing the optimization phase from 132.9 seconds to a mere 3.6 seconds—which results in an amortized speedup of 36.9× over multiple runs even when accounting for training overhead. Furthermore, sensitivity analyses under significant load and DG output uncertainties

confirm the framework's practical resilience, maintaining a $VSI_{\{min\}}$ of 0.8413 (well above the 0.50 stability threshold). These findings solidify the LSTM-MH as a robust, high-speed solution capable of navigating realistic operational volatility without compromising on global optimality or system security.

These findings establish the LSTM-MH framework as a computationally efficient, voltage stability-focused, and practically reliable tool for distribution system DG planning. Future research directions include: extension to the IEEE 69-bus and 118-bus benchmark systems and real utility feeder data; incorporation of stochastic load and renewable output models (solar irradiance, wind speed PDFs) for probabilistic DG placement; multi-objective Pareto-front optimisation with explicit techno-economic cost–benefit analysis; dynamic DG sizing accounting for load growth trajectories; and integration of battery energy storage systems (BESS) and electric vehicle (EV) charging demand profiles. The exploration of transformer-based architectures (e.g., Temporal Fusion Transformers) as higher-capacity alternatives to LSTM for voltage stability surrogate modelling, and the application of transfer learning to accelerate surrogate training across different network configurations, represent particularly promising avenues for further investigation.

References

[1] Baran, M.E. and Wu, F.F. (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Transactions on Power Delivery, 4(2), pp.1401–1407.

[2] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), pp.1735–1780.

[3] Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M. and Chen, H. (2019). Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems, 97, pp.849–872.

[4] Breiman, L. (2001). Random forests. Machine Learning, 45(1), pp.5–32.

[5] Murthy, V.V.S.N. and Kumar, A. (2013). Comparison of optimal DG allocation methods in radial distribution systems based on sensitivity approaches. International Journal of Electrical Power and Energy Systems, 53, pp.450–467.

[6] Hung, D.Q., Mithulananthan, N. and Bansal, R.C. (2010). Analytical expressions for DG allocation in primary distribution networks. IEEE Transactions on Energy Conversion, 25(3), pp.814–820.

[7] Sultana, S. and Roy, P.K. (2014). Multi-objective quasi-oppositional teaching learning based optimization for optimal location of distributed generator in radial distribution systems. International Journal of Electrical Power and Energy Systems, 63, pp.534–545.

[8] Sanjay, R., Jayabarathi, T., Raghunathan, T., Ramesh, V. and Mithulananthan, N. (2017). Optimal allocation of distributed generation using hybrid grey wolf optimizer. IEEE Access, 5, pp.14807–14818.

[9] Mirjalili, S. and Lewis, A. (2016). The whale optimization algorithm. Advances in Engineering Software, 95, pp.51–67.

[10] Acharya, N., Mahat, P. and Mithulananthan, N. (2006). An analytical approach for DG allocation in primary distribution network. International Journal of Electrical Power and Energy Systems, 28(10), pp.669–678.

[11] Chebbo, A.M., Irving, M.R. and Sterling, M.J.H. (1992). Voltage collapse proximity indicator: Behaviour and implications. IEE Proceedings C — Generation, Transmission and Distribution, 139(3), pp.241–252.

[12] Mistry, K.D. and Roy, R. (2014). Enhancement of loading capacity of distribution system through distributed generator placement considering techno-economic benefits with load growth. International Journal of Electrical Power and Energy Systems, 54, pp.505–515.

[13] Mirjalili, S., Mirjalili, S.M. and Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69, pp.46–61.

[14] Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95 — International Conference on Neural Networks, Perth, Australia, vol. 4, pp.1942–1948.

[15] Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.