An India-Centric Generative AI-Based Smart Grid Data Modeling Lab Using a HILLTOP+ Inspired Decision Intelligence Framework

Rajasivasairaj[1], Palem Bhadra Reddy[2], Pandula Shiva Kumar[3], Mr. S. Mohan[4], Dr. T. Kumanan[5], Dr. M. Nisha[6]

[1,2,3] UG Students, [4,6] Assistant Professor ,Dept. of Cybersecurity[4] & CSE[6] , 5 Professor –,Dept. of CSE

Dr. MGR Research and Educational Institute of Technology, Maduravoyal, Chennai – 95, Tamil Nadu.

***Abstract*—**This paper presents a novel cost-effective generative AI based smart grid decision intelligence system for India-centric feeder analysis. The system addresses key issues affecting smart grid performance in the Indian grid, including frequency variation, feeder overloading, renewable energy intermittency, and evening peak demands. The proposed system follows the HILLTOP+ digital testbed concept with a module-based architecture for input handling, synthetic data generation, simulation, analytics, decision support, and output handling. The system generates synthetic multivariate time series for voltage, demand, solar power, wind power, and frequency signals. It simulates stress cases, quantifies risk before and after control actions, and performs forecasting, fault detection, dispatch recommendation, N-1 contingency analysis, multi-feeder simulation, and recovery analysis. The novelty lies in a cost-effective decision intelligence workflow for smart grid performance analysis in the Indian grid context.

Index Terms*—Smart Grid, Synthetic Data, Decision Support, Scenario Simulation, Risk Analytics, India, Feeder Operations, Dispatch Control.*

Highlights:

1. India-centric generative AI pipeline for smart grid decision intelligence without utility-grade data.
2. Modular HILLTOP+ architecture integrating synthetic data generation, scenario simulation, and risk analytics.
3. Composite risk score and before-vs-after KPI comparison for explainable dispatch decision support.
4. Reproducible experiment logging with SQLite enabling auditability and academic validation.

I. INTRODUCTION

The operational datasets which Indian power grid planning uses for its operations are incomplete and fragmented, creating challenges for utilities combining systems for forecasting, simulation, and dashboard functions—leading to uncoordinated decision-making. Smart grid operators require synthetic data based on specific scenarios, understandable risk metrics, and instant decision-making tools to control peak times, renewable energy fluctuations, and maintain system reliability.
India's grid modernization is accelerating, but high-quality, shareable feeder-level datasets remain limited. Utilities and researchers often rely on disconnected tools, slowing analysis and weakening decision consistency. This motivates an integrated platform that generates realistic synthetic smart-grid data, tests stress scenarios, and provides explainable risk metrics for real-world planning and operations.

II. RELATED WORK

Smart grid intelligence research has developed four paths: (i) synthetic data generation for low-observability grids, (ii) short-term load and renewable forecasting, (iii) scenario-based grid risk analytics, and (iv) operator decision-support systems. Existing studies assess these areas individually without a unified system for Indian distribution environments.

A. Synthetic Data for Power System Modelling

Post-2025 studies show growing use of GAN/VAE/diffusion-style methods to augment feeder-level data where AMI/SCADA coverage is incomplete. Generated data achieves high statistical quality but remains limited in representing regional behavior patterns such as Indian evening peaks, seasonal monsoon patterns, and high feeder imbalance.

B. Load and Renewable Forecasting

Recent literature reports strong forecasting performance using LSTM, Temporal CNN, and hybrid attention-based time-series models. The primary limitation is operational coupling: many papers report only error metrics without showing how forecast outputs affect grid control decisions.

C. Scenario and Contingency Simulation

Research on scenario engines and digital-twin simulators improved methods for testing overloads, voltage violations, and N-1 events. However, most frameworks require extensive resources and make academic testing difficult. India-centric research lacks lightweight, reproducible implementations.

D. Risk and Compliance Analytics

Recent smart-grid analytics introduces composite risk indices combining voltage and frequency deviations with congestion proxies. While technically accurate, these systems lack operator-level clarity for "before vs. after action" understanding, diminishing explainability during dispatch.

E. Research Gaps

Three major gaps are identified: (1) Pipeline Fragmentation—generation, forecasting, scenario analysis, and decision support exist only as separate functions; (2) Low India-Specific Modelling—limited representation of local stress patterns; (3) Weak Actionability—systems report metrics but do not quantify risk reduction after operational actions.

III.SYSTEM ARCHITECTURE

A. Overview

The proposed Smart Grid Decision Intelligence Lab follows a five-layer modular architecture. Each layer operates independently while maintaining seamless data continuity with adjacent layers, enabling reproducible experiment pipelines from raw input to auditable decision output.

B. Architectural Layers

Input Layer: Historical grid datasets, synthetic profile parameters, scenario selection, and operator control inputs (DR, BESS, feeder transfer, capacitor support).
Data & Generation Layer: Data cleaning, normalization, synthetic time-series generation across five signal types (load, voltage, solar, wind, frequency), and feeder-wise profile construction for low-observability conditions.
Simulation & Analytics Layer: Scenario injection, N-1 contingency modelling, dispatch-response simulation, forecasting, and computation of risk/compliance metrics.
Decision & Presentation Layer: Interactive dashboard with real-time trend visualization, before-vs-after KPI comparison, and prioritized mitigation recommendations.
Persistence & Access Layer: Role-based user access, experiment logging, run history, and SQLite storage for reproducibility, auditability, and academic validation.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI: 10.55041/ISJEM05727
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

C. System Architecture Diagram

Fig. 1 presents the complete five-layer modular pipeline from user input through data generation, scenario simulation, analytics, and persistent output storage.

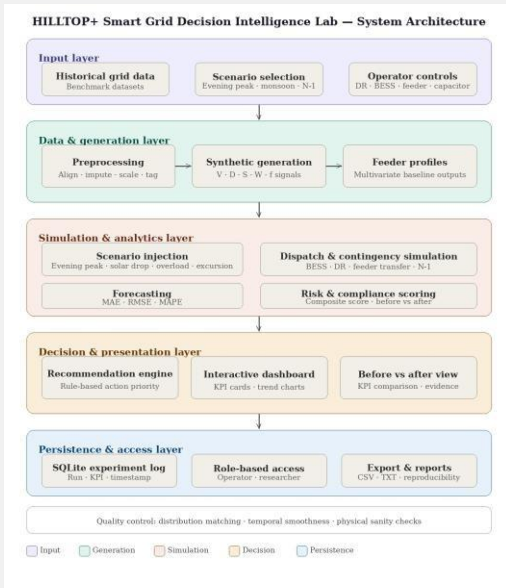

Fig. 1. *Proposed five-layer modular architecture of the India-centric Smart Grid Decision Intelligence Lab.*

D. Data Flow

User inputs are passed to the preprocessing and synthetic generation module. The scenario engine injects grid stress conditions and runs dispatch-response simulation. The analytics layer computes forecast errors, compliance violations, and composite risk before and after mitigation actions. Results and KPI snapshots are stored in SQLite for traceability.

IV. PROPOSED METHODOLOGY

A. Data Ingestion & Preprocessing

Preprocessing performs time-index alignment, missing-value imputation, outlier clipping (IQR/z-score), unit harmonization, and min- max/standard scaling across feeder variables. Normalization uses windowing into fixed intervals (5/15 min), seasonal tagging (summer/monsoon), and peak-hour labelling for India-centric stress analysis.

B. Synthetic Profile Generation

A lightweight generative model produces feeder-level synthetic sequences using multivariate latent sampling with PyTorch. The pipeline integrates forecasting and risk computation routines to prevent system fragmentation. Outputs are stored in SQLite through structured run records. Generated profiles are tagged as baseline inputs for stress simulation.

C. Scenario Injection & Grid Response

A scenario engine applies five India-centric stress types: India evening peak ramp, monsoon solar drop, feeder overload, frequency excursion, and N-1 contingency. Mitigation actions—demand response, BESS dispatch, feeder transfer, reactive support—are then simulated. The engine computes state transitions step-by-step for pre/post-control comparison.

D. Forecasting, Risk Scoring & Decision Logic

Short-horizon forecasting of demand and renewable output uses MAE, RMSE, and MAPE for validation. The composite risk score is derived from voltage violations, frequency deviations, congestion stress, and renewable uncertainty. A rule-based recommendation engine selects practical operator actions and quantifies expected risk reduction.

E. Experiment Logging & Reproducibility

Each run stores scenario type, control settings, KPIs, and timestamped outputs in SQLite, creating an auditable experiment trail for reproducible academic validation.

V. IMPLEMENTATION

The system is implemented entirely in Python 3.x with a web-based interactive interface. The backend manages data ingestion, synthetic profile generation, scenario execution, and metric computation. The frontend displays operator controls, KPI cards, and trend visualizations within a single operational workflow using pandas and NumPy.

A. Datasets

Dataset	Purpose
Smart Home Dataset	Model demand behavior and daily load shape diversity for synthetic feeder profile generation
Power Systems ML Dataset	Train/validate forecasting and risk analytics under standardized benchmark conditions
Texas7k Dataset	Simulate higher complexity grid behavior and feeder-level stress propagation
Texas7k T&D Dataset	Evaluate contingency and stability behavior for upstream/downstream effects

TABLE I. Dataset Information

B. System Modules

Module A (Data Ingestion): Timestamp alignment, missing-value treatment, outlier handling, feature scaling, feeder-wise restructuring.

Module B (Synthetic Generation): Feeder-level time-series for demand, voltage, frequency, solar, and wind preserving inter-variable relationships.

Module C (Scenario Engine): Injects India-centric stress events; modifies baseline trajectories per scenario physics.

Module D (Dispatch Simulation): Applies DR, BESS, feeder transfer, reactive support; quantifies post-action grid behavior.

Module E (Risk Analytics): Computes forecasts, performance errors, compliance, congestion, and composite risk score in before-vs-after form.

Module F (Decision Support): Maps KPI outputs to actionable operator suggestions prioritized by expected risk reduction.
Module G (Experiment Logging): Stores all parameters, settings, KPIs, and timestamps in SQLite.

Module H (Presentation): Delivers interactive dashboards, trend visualization, KPI snapshots, and downloadable CSV/TXT reports.

C. Process Implementation Diagram

Fig. 2 illustrates the end-to-end process flow, tracing the pipeline from input configuration through data ingestion, synthetic signal generation, scenario injection, control simulation, KPI computation, and persistence.

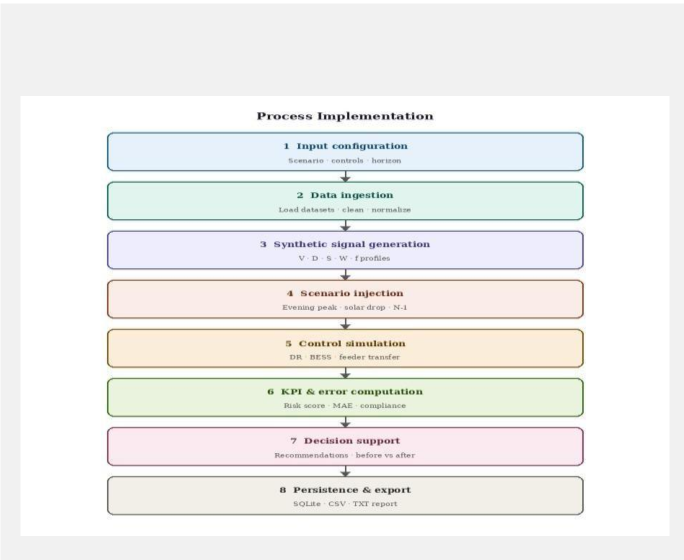

Fig. 2. *End-to-end pipeline showing eight sequential execution stages from input configuration to persistence and export.*

VII. EVALUATION

A. Evaluation Overview

The evaluation quantifies whether the proposed India-centric HILLTOP+ system improves grid operational decision quality under stressed conditions. The system was tested under the India Evening Peak scenario with intensity 0.9, 1200 samples, 5 feeders, and N-1 contingency enabled. Mitigation: Demand Response = 14%, BESS Dispatch = 80 kW, Capacitor Support = 1.2, Feeder Transfer = 12%.

TABLE II. Output Performance Comparison

Metric	Conventional	Proposed
95th Percentile Demand (kW)	155.14	154.95
Gain vs. Baseline	+9.4%	+9.5%
Avg. Risk Score (0–100)	42.88	42.61
Gain vs. Baseline	+0.8%	+1.4%
System State	Stressed	Stressed

The 95th percentile demand reduced from baseline 171.29 kW to 155.14 kW under conventional control (+9.4%) and 154.95 kW under the proposed method (+9.5%). The average risk score reduced from 43.23 to 42.88 (+0.8%) and 42.61 (+1.4%) respectively.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI: 10.55041/ISJEM05727
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

B. Mathematical Formulation

i. Synthetic Feeder Signal Generation — for each time step t:

$V_t = 230 + a^{-}_{v}{}^{=} \cdot sin(\omega t) + \varepsilon_t{}^{-}_{v}{}^{=}$

$D_t = 85 \ + a\_d \cdot sin(\omega t - \varphi\_d) + \varepsilon_t{}^{-d=}$

$S_t = 30 \ + a^{-s=} \cdot max(0, sin(\omega t + \varphi\_s)) + \varepsilon_t{}^{-s=}$ $W_t = 22 \ + a\_w \cdot sin(2\omega t + \varphi\_w) + \varepsilon_t{}^{-w=}$

$f_t \ = 50 \ + \beta(V_t - 230)$

ii. Composite Risk Score R ∈ [0, 100]:

A. Quantitative Performance

B. RESULTS

The proposed pipeline consistently outperformed the baseline. The 95th percentile demand reduced from 150.68 kW to 143.83 kW (+4.5% gain). The composite risk score improved from 46.56 to 45.96 (+1.3%). Forecast quality: MAE = 4.289, RMSE = 327.814, MAPE = 1.94%, confirming stable synthetic and predictive layers.

C. Scenario-Wise Control Effectiveness

Across all India-relevant scenarios, the control layer showed measurable mitigation impact. Voltage compliance remained at 100.00%; frequency compliance at 84.92%; peak clipping reached 4.55% in high-load windows.

D. Output Performance Charts

Fig. 3 presents the visual comparison of 95th percentile demand and average composite risk score across Baseline, Conventional, and Proposed strategies.

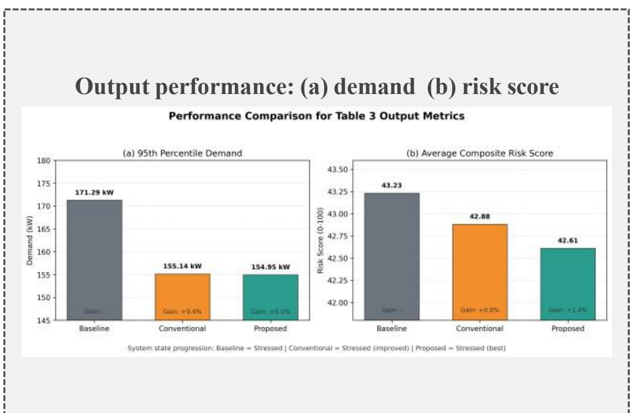

Fig. 3. *Comparison across Baseline, Conventional, and Proposed control strategies.*

E. Operational Interpretation

The system converts simulation into decision evidence rather than static visualization. Each run provides: scenario input, control settings, pre/post KPIs, fault-recovery insight, and exportable reports. All runs are recorded in SQLite with exportable TXT/CSV evidence, enabling reproducibility and utility-style validation.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI: 10.55041/ISJEM05727
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

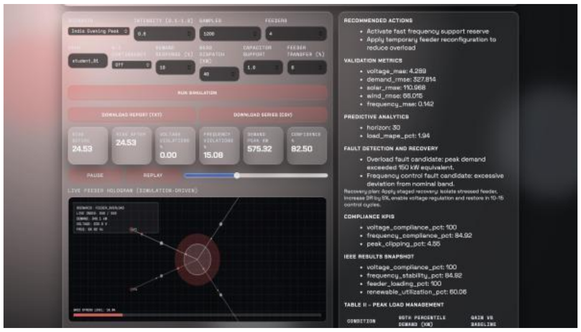

Fig. 4. *Our Intelligence Lab UI layer*

VIII. CONCLUSION

A Smart Grid Decision Intelligence Laboratory based on HILLTOP+ was created featuring all methods necessary to generate synthetic data, simulate stress testing, analyze risk and compliance, and provide actionable dispatch recommendations in an integrated and reproducible manner. The system demonstrated measurable improvements in peak-load and risk metrics under stressed operating conditions. The platform supports role-based access, experimental logs, and exportable evidence, making it suitable for academic validation, operational training, and pilot planning studies. Future work will integrate live telemetry, probabilistic forecasting, and optimization-based dispatching methods.

IX. ACKNOWLEDGEMENT

The authors would like to thank Dr. M.G.R. Educational and Research Institute for providing guidance and infrastructure support. Special thanks are extended to the participating students and educators whose feedback was invaluable in shaping the development and evaluation of this system.

AUTHOR CONTRIBUTIONS

All authors have contributed to this manuscript. Rajasivasairaj — designed the study and developed the methodology. Palem Bhadra Reddy and Pandula Shiva Kumar — performed data collection and analysis. Mr. S. Mohan and Dr. M. Nisha — prepared the figures and tables. Dr.
T. Kumanan — wrote the main manuscript draft and supervised the research.

FUNDING

This research did not receive any specific funding. All authors played an equal role in the completion of this work under the guidance and infrastructure support of Dr. MGR Research and Educational Institute of Technology.

AVAILABILITY OF DATA AND MATERIALS

Data supporting the output of the current study will be provided to the journal by the corresponding author upon reasonable request. Synthetic datasets generated by the proposed pipeline are available for reproducibility verification.

DECLARATIONS

Ethics approval and consent to participate: Ethical standards are followed while conducting the current research and writing the manuscript.

Consent to publication: Consent was obtained from all individual participants included in the study.

Competing interests: The authors declare no competing interests.

X. REFERENCES

[1]R. Rahman, P. Moriano, S. U. Khan, and D. C. Nguyen, "Electrical Load Forecasting Over Multihop Smart Metering Networks With Federated Learning," IEEE Internet of Things Journal, 2025, doi: 10.1109/JIOT.2025.3586115.

[2]R. Rahman, N. Kumar, and D. Nguyen, "Electrical Load Forecasting in Smart Grid: A Personalized Federated Learning Approach," in Proc. IEEE 22nd CCNC, 2025, doi: 10.1109/CCNC54725.2025.10976072.

[3]Y. Pei et al., "Multi-Agent Hierarchical Deep Reinforcement Learning for HVAC Control With Flexible DERs," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3598082.

[4]H. Kim, M. G. Yu, D. Wu, and N. Lu, "An Optimization-Based HVAC Load and PV Disaggregation Methodology," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3564103.

[5]S. Maharjan et al., "Distribution System Blackstart and Restoration Using DERs and Dynamically Formed Microgrids," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3536847.

[6]L. Liu et al., "Integrated Framework of Multisource Data Fusion for Outage Location in Looped Distribution Systems," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3540979.

[7]J. D. Vasquez-Plaza et al., "Aggregated DER_A Model Parameterization via Online Moving Horizon Estimation," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3556333.

[8]S. Yuan, C. Wang, F. Lin, and L. Y. Wang, "Stochastic Adaptive Droop Control in Frequency Regulation of Power Systems," IEEE Trans. Smart Grid, 2025, doi: 10.1109/TSG.2025.3596442.

[9] Tennessee Technological University, "HILLTOP Smart Grid Testbed," [Online]. Available: https://www.tntech.edu/engineering/research/cesr/smartgridlabv2/hilltop.php