21, rue d'Artois, F-75008 PARIS
http://www.cigre.org

CIGRE US National Committee
2025 Grid of the Future Symposium

Enhancing Transient Stability Analysis with Machine Learning: Fault Classification in the IEEE 9-Bus System

Polat Goktas [1] and Shaibu Ibrahim [2]

[1]Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Türkiye

[2]Acciona Energia USA Global, Chicago, USA

cigre United States

2025 GRID OF THE FUTURE™
Denver, CO | November 10-13

POWER ENGINEERS MEMBER OF WSP

Presentation Outline

- Background
- Methodology: power model and machine learning
- System configurations and fault scenarios
- ML methods for Fault Classification
- Results and Discussion
- Conclusion and future direction

Why Transient Stability Matters ?

Northeast Blackout of 2003

(August 14, 2003)

Roughly 63 GW of load was interrupted

Over 400 transmission lines,

531 generating units, and

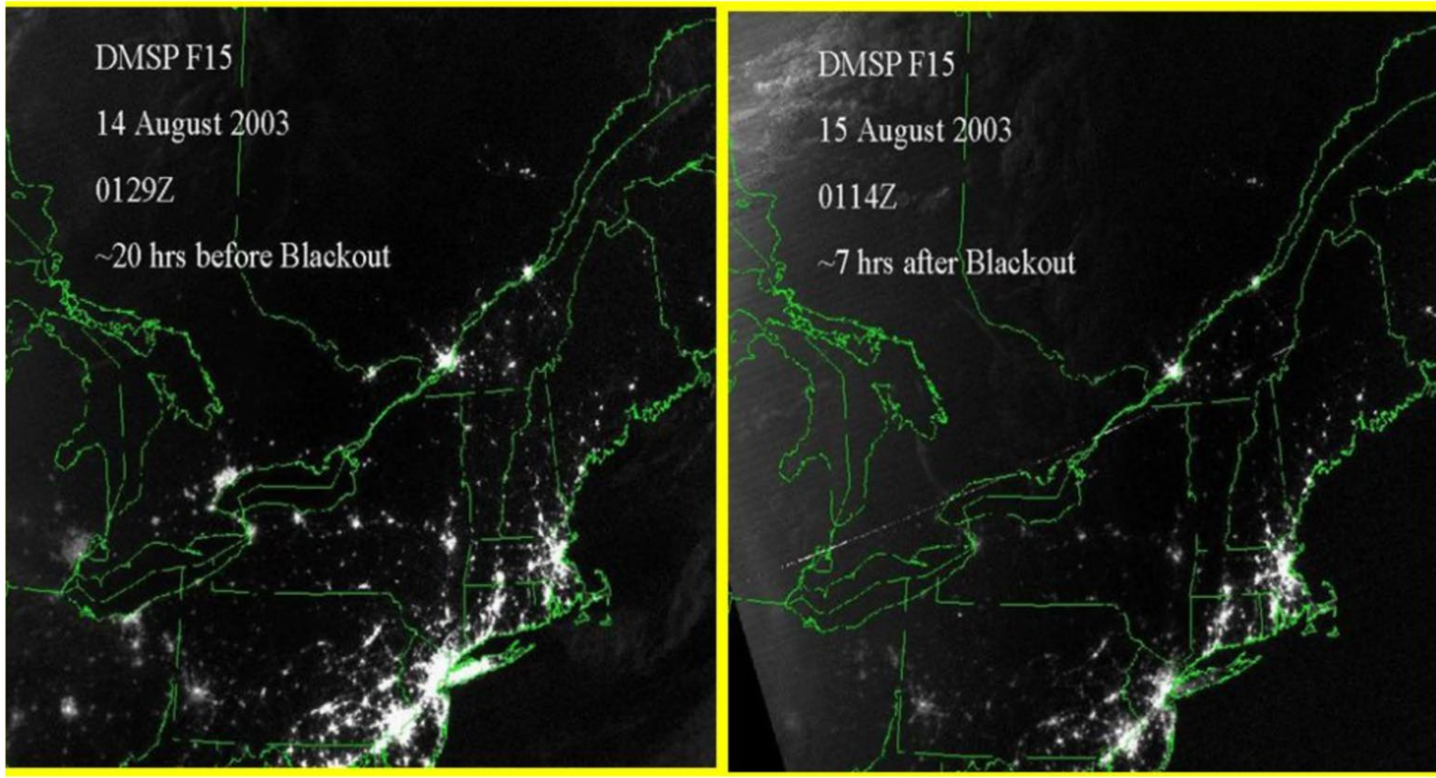
261 power plants tripped

~ 50 million people affected in 8 U.S states and

2 Canadian Provinces

Source: Andersson et al., “Causes of the 2003 Major Grid Blackouts,” IEEE Trans. Power Syst., 2005

Ability to return to the state of operating equilibrium after a severe disturbance



Source: Spectrum News Rochester, Aug 2022

Background: Evolving Grid Complexities

Situational awareness
Human errors

With the high penetration of renewables into the grid, coupled with their intermittency and lack of inertia.
This may further complicate the grid's future operation.

This therefore calls for

intelligent, data-driven fault detection and classification

Methodology and Model Set-Up

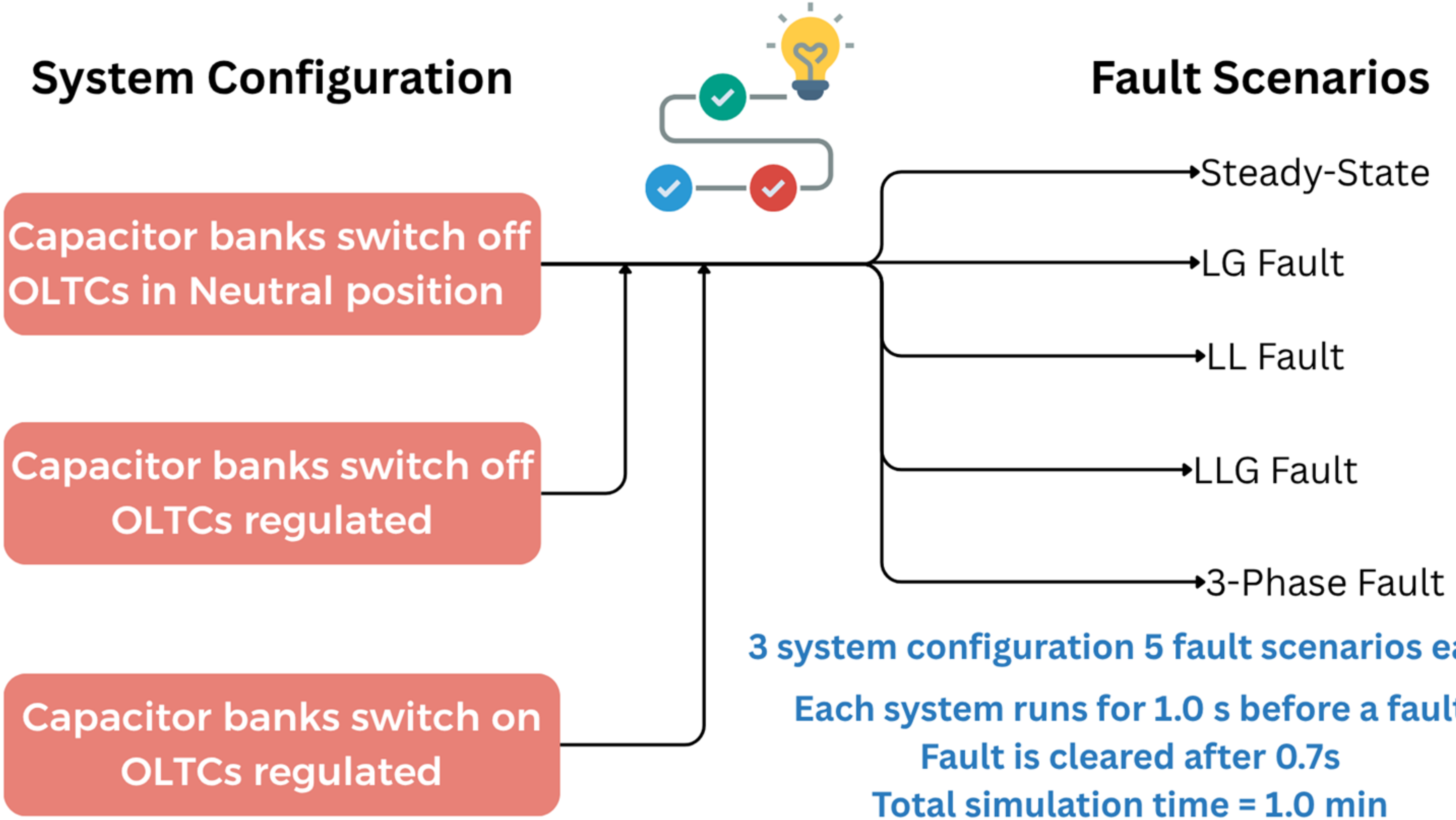
Modified IEEE-9 Bus

Software - ETAP

Parameters retrieved:

- Bus voltages
- Bus angles
- Bus Frequency
- Active power
- Reactive power
- voltage/frequency

etc



3 system configuration 5 fault scenarios each

Each system runs for 1.0 s before a fault

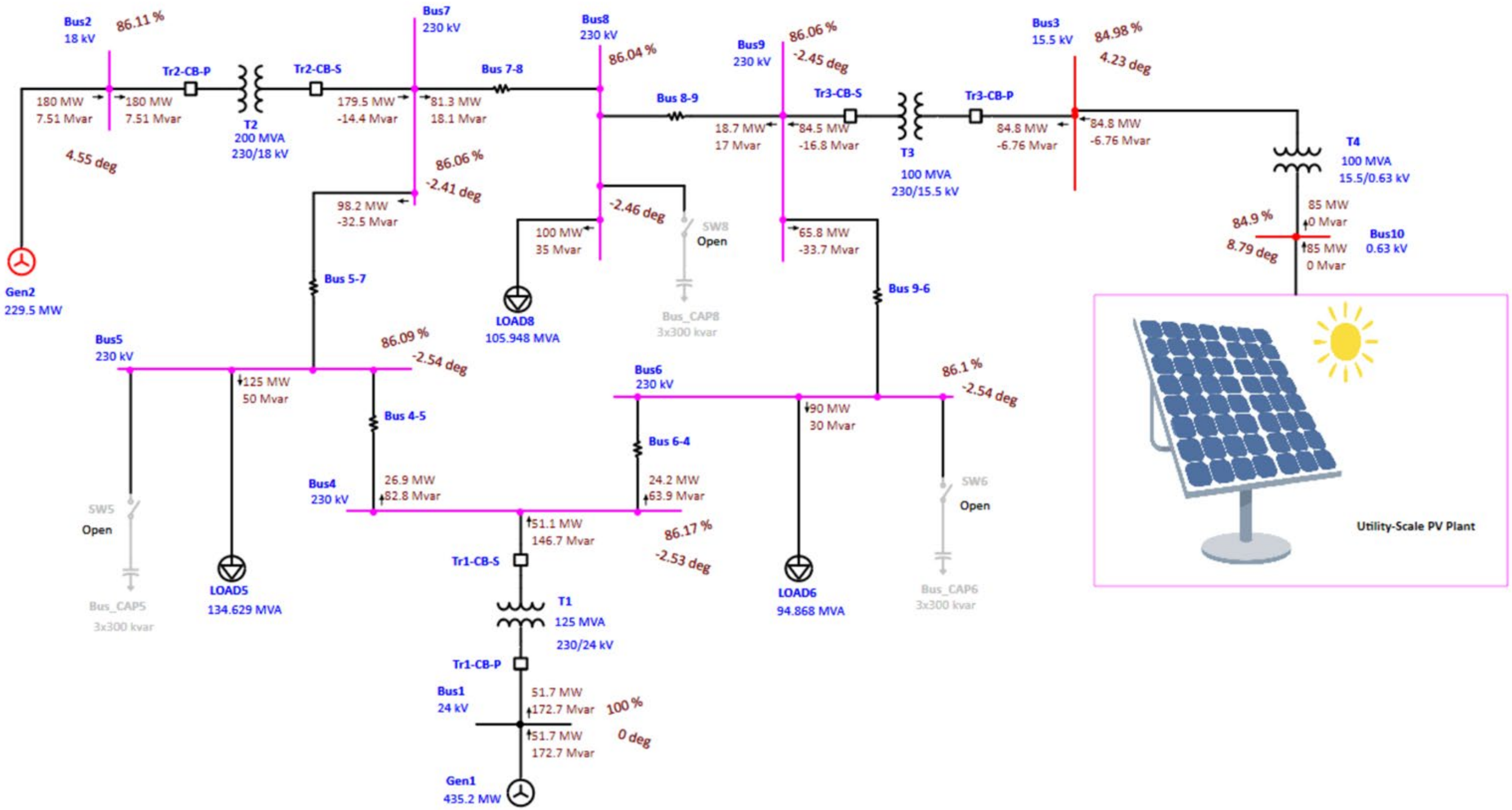
Fault is cleared after 0.7s

Total simulation time = 1.0 min

Initial load flow Scenario

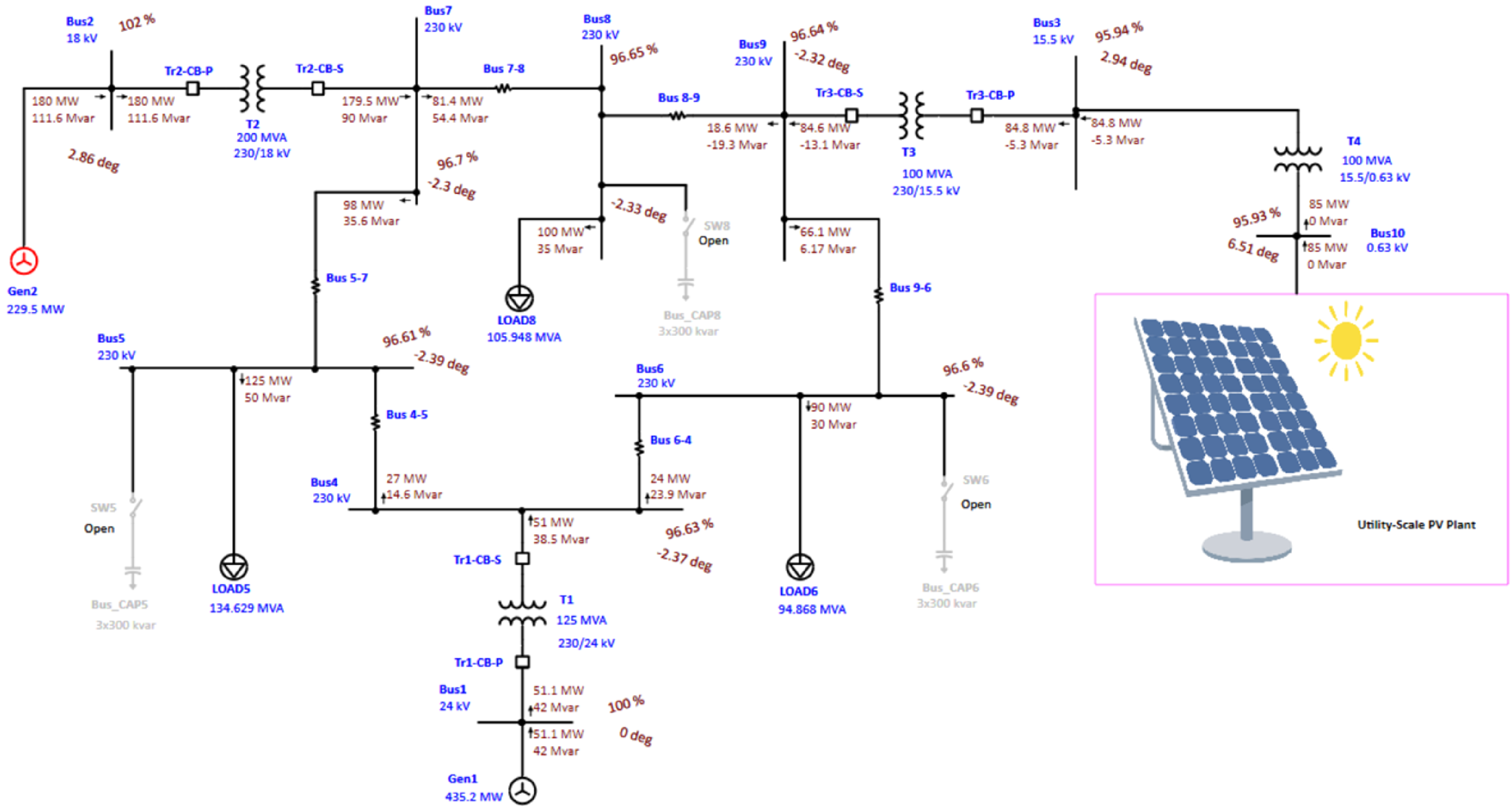
Red – Bus voltage condition is **critical**,
<0.85 pu V
>1.10 pu V

Magenta - Bus voltage condition is **marginal**,
< 0.95 pu V
> 1.05 pu V

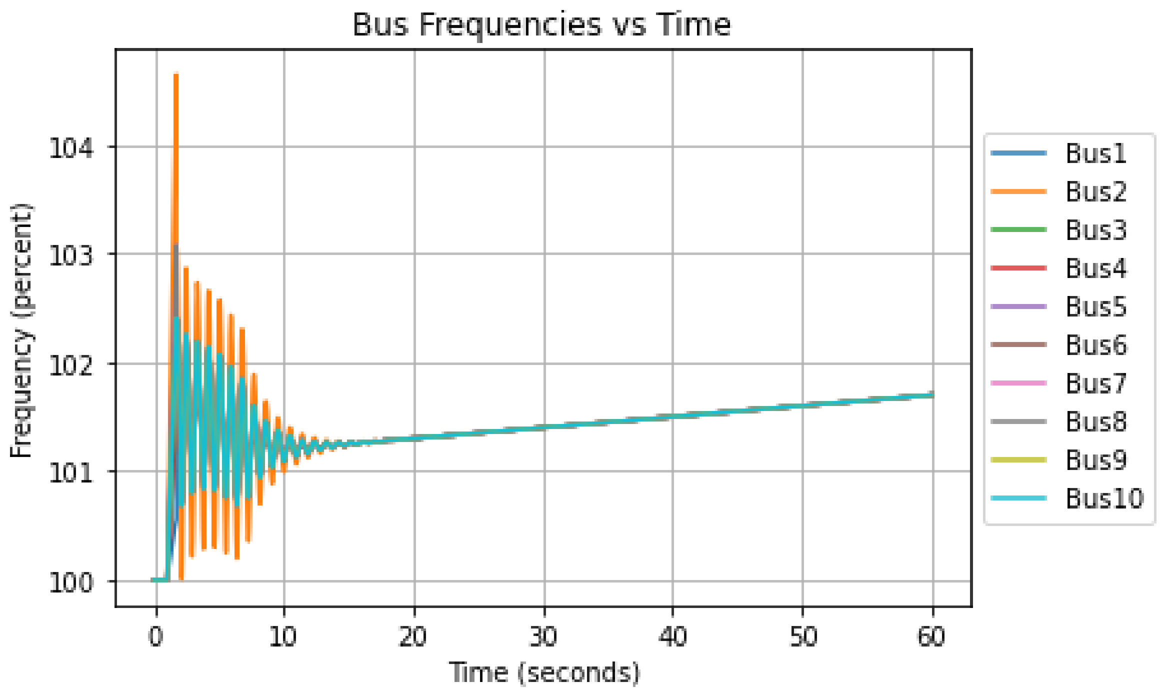


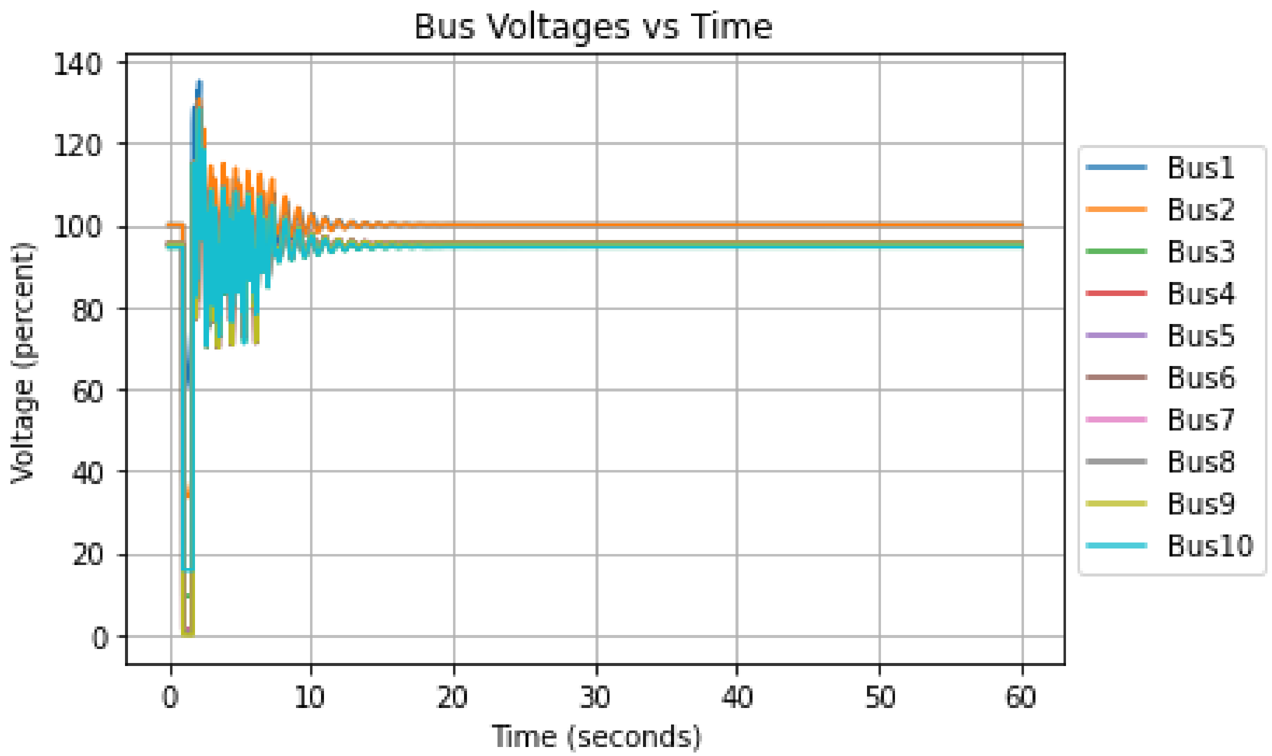
Based Case Condition

Improved Load Flow scenario, Gen 2, PF control at 0.85 and PV at 1.0 PF

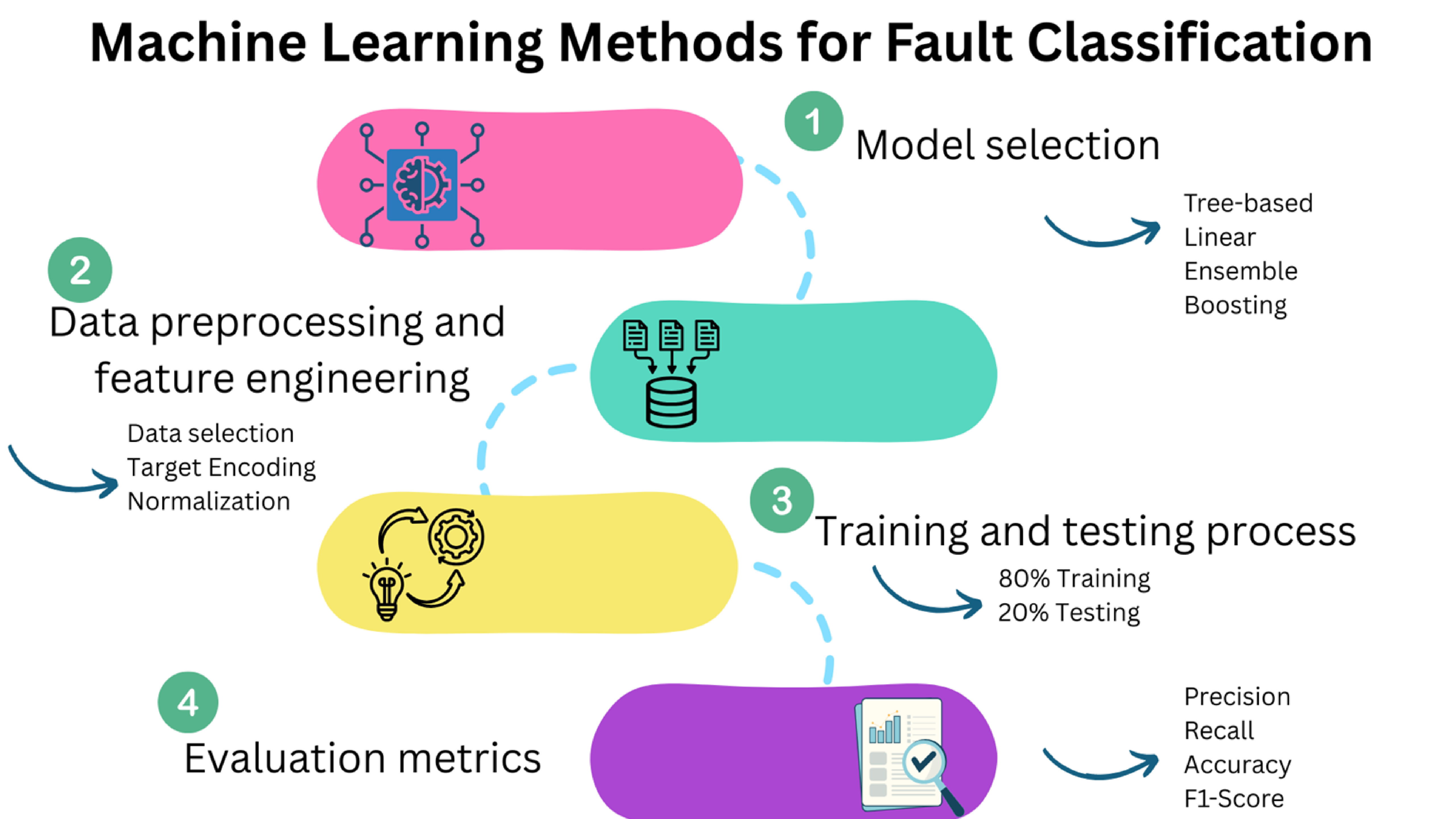


Graphs post fault condition





15 plots across different scenarios.


Machine Learning Methods for Fault Classification
1
Model selection
Tree-based
Linear
Ensemble
Boosting
2
Data preprocessing and feature engineering
Data selection
Target Encoding
Normalization
3
Training and testing process
80% Training
20% Testing
4
Evaluation metrics
Precision
Recall
Accuracy
F1-Score

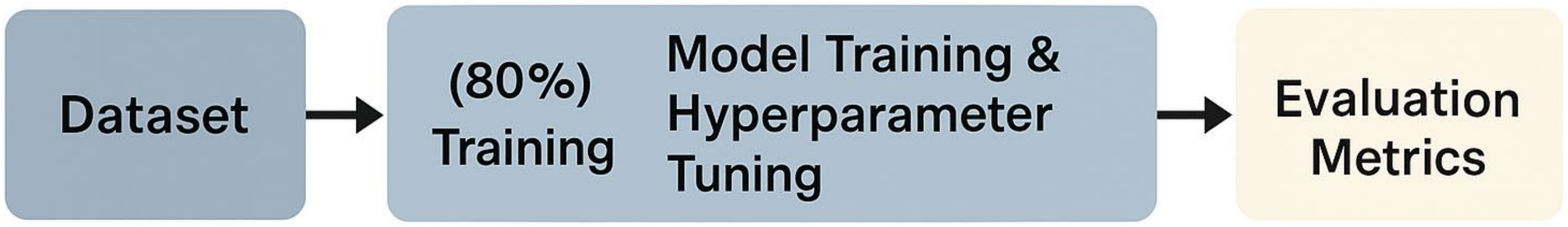
Training and Testing Process

- The dataset was divided into **80% training** and **20% testing** to ensure a balanced validation approach.

- Each ML model was trained on the training subset and **optimized using hyperparameter tuning** to enhance classification performance.

- **Evaluation metrics:** Accuracy, Precision, Recall, and F1-score guided the tuning process.

- The objective was to **achieve consistent and reliable fault classification** across diverse fault scenarios.



Evaluation Metrics

Precision (P): Accuracy of true positives among predicted positives.

Recall (R): Sensitivity in identifying all actual fault instances.

F1-Score: Harmonic mean of Precision and Recall, balancing both aspects.

Accuracy: Overall proportion of correctly classified fault types.

Log-Loss (Cross-Entropy): Evaluates **reliability of probabilistic predictions**, with lower values indicating higher confidence.

$$Precision(P) = \frac{TP}{TP + FP} \tag{1}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{2}$$

$$F1\,Score = \frac{2 * P * R}{P + R} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

Additionally, we utilized log-loss (cross-entropy loss) as a metric to measure the reliability of probabilistic predictions. Log-loss is calculated as follows:

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N}[y_i.\log(p_i) + (1 - y_i).\log(1 - p_i)] \tag{5}$$

where N is the number of samples, y_i is the true label of the i-th instance, and p_i is the predicted probability for that instance.

Results and Discussion

In this study, we evaluated state-of-the-art ML classifiers to assess their effectiveness in detecting fault types, specifically categorizing **1LG, 2LLG, and 3LLL faults, across different scenarios within the modified IEEE 9-bus system.**

Performance was measured based on accuracy, precision, recall, F1-score, and confusion matrices, summarizing each classifier's predictive capability. The study considered three configurations:

1. **NO CAP NO TAP**: Without capacitor banks and transformer tap settings.
2. **TAP, NO CAP**: Only transformer tap settings were used, excluding capacitor banks.
3. **WITH CAP TAP**: Both capacitor banks and transformer tap settings were utilized.

Results and Discussion

TABLE 1: VOLTAGE ANGLE - CLASSIFICATION PERFORMANCE METRICS

Model	No CAP, No TAP				No CAP, TAP Used				With CAP, TAP Used			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Decision Tree	84.11%	84.39%	84.11%	83.71%	85.76%	86.25%	85.76%	85.49%	82.51%	82.72%	82.51%	82.16%
Logistic Regression	52.98%	55.50%	52.98%	47.67%	56.29%	59.75%	56.29%	52.73%	45.87%	42.62%	45.87%	38.71%
Ridge Classifier	52.65%	55.38%	52.65%	47.37%	44.70%	39.96%	44.70%	37.30%	46.20%	43.13%	46.20%	38.88%
AdaBoost	84.11%	84.97%	84.11%	83.63%	64.57%	62.48%	64.57%	63.13%	73.60%	73.54%	73.60%	71.49%
Random Forest	83.44%	83.65%	83.44%	83.02%	**86.09%**	**86.62%**	**86.09%**	**85.81%**	84.49%	85.35%	84.49%	84.19%
Extra Trees	**85.10%**	**85.38%**	**85.10%**	**84.78%**	**86.09%**	**86.46%**	**86.09%**	**85.83%**	**84.82%**	**85.50%**	**84.82%**	**84.53%**
Gradient Boosting	**84.44%**	**84.94%**	**84.44%**	**84.04%**	**87.75%**	**88.67%**	**87.75%**	**87.53%**	**85.48%**	**86.94%**	**85.48%**	**85.24%**
KNeighbors	83.77%	84.51%	83.77%	83.36%	**86.09%**	**86.62%**	**86.09%**	**85.81%**	78.55%	84.44%	78.55%	79.32%
SVC	64.24%	70.88%	64.24%	60.92%	64.90%	72.66%	64.90%	61.61%	65.68%	71.39%	65.68%	63.16%

Results and Discussion

TABLE II: BUS FREQUENCY - CLASSIFICATION PERFORMANCE METRICS

Model	No CAP, No TAP				No CAP, TAP Used				With CAP, TAP Used			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Decision Tree	**94.63%**	**95.33%**	**94.63%**	**94.56%**	89.71%	91.63%	89.71%	89.85%	90.98%	92.25%	90.98%	91.02%
Logistic Regression	94.21%	95.02%	94.21%	94.14%	71.60%	77.70%	71.60%	68.54%	90.57%	91.80%	90.57%	90.57%
Ridge Classifier	64.05%	47.29%	64.05%	52.62%	59.67%	39.91%	59.67%	47.60%	63.93%	77.02%	63.93%	52.16%
AdaBoost	**94.63%**	**95.33%**	**94.63%**	**94.56%**	63.37%	70.75%	63.37%	55.48%	61.89%	43.77%	61.89%	50.93%
Random Forest	**94.63%**	**95.33%**	**94.63%**	**94.56%**	**90.53%**	**92.57%**	**90.53%**	**90.73%**	**92.21%**	**93.66%**	**92.21%**	**92.31%**
Extra Trees	**94.63%**	**95.33%**	**94.63%**	**94.56%**	90.12%	92.09%	90.12%	90.29%	**92.21%**	**93.66%**	**92.21%**	**92.31%**
Gradient Boosting	**94.63%**	**95.33%**	**94.63%**	**94.56%**	**90.53%**	**92.57%**	**90.53%**	**90.73%**	90.98%	92.25%	90.98%	91.02%
KNeighbors	**94.63%**	**95.33%**	**94.63%**	**94.56%**	89.71%	91.63%	89.71%	89.85%	91.80%	93.18%	91.80%	91.89%
SVC	64.05%	47.29%	64.05%	52.62%	87.24%	88.79%	87.24%	87.18%	89.75%	90.87%	89.75%	89.75%

Results and Discussion

Overall Performance: Ensemble and boosting models (ExtraTrees, RandomForest, GradientBoosting, AdaBoost) consistently outperformed linear models (Logistic Regression, RidgeClassifier), achieving the highest accuracy, precision, and recall across all configurations.

NO CAP NO TAP: Tree-based models, especially ExtraTrees and DecisionTree, achieved ~84–95% accuracy, while linear models underperformed (<55%), indicating their limitations in capturing nonlinear fault patterns.

NO CAP TAP: Ensemble methods like ExtraTrees and RandomForest maintained strong performance (~86–90% accuracy), while SVM and Logistic Regression lagged, confirming the need for nonlinear, ensemble-based approaches.

WITH CAP TAP: GradientBoosting and RandomForest yielded the best results (~85–90% accuracy), particularly excelling under complex conditions with capacitor banks and tap changers, where fine-tuning further enhanced performance.

Conclusion

Summary of Findings

- Ensemble and boosting models (e.g., **ExtraTrees, RandomForest, GradientBoosting**) consistently outperformed linear classifiers in **fault** **classification**.
- Accuracy ranged from **85–95%** for nonlinear models across all configurations.
- Linear models (e.g., Logistic Regression, Ridge Classifier) underperformed, particularly in **complex transient scenarios**.

Key Insights

- **Tree-based and ensemble approaches** effectively capture nonlinear fault behavior in power systems.
- Configurations involving **capacitor banks and tap changers** increase system complexity, best managed by adaptive, high-capacity models.
- **Model selection** must align with **system configuration** to ensure optimal classification and stability assessment.

Ensemble and boosting ML models provide robust, high-precision solutions for transient stability fault classification, enabling reliable and efficient power system operation.

Future Direction

Data-Driven to predict events

adaptability

accuracy > 95%.

ML-based fault classification has the potential to enhance transient stability significantly.

Future direction: Integrate with real-time PMU streams and an AI fault location model.

This research bridges data science and power system protection for a smarter, more resilient grid.


THANK YOU
Q&A