# Attention-Based LSTM Model for Advanced Fault Detection in Power Transmission Lines

Your Name[1], Co-Author Name[2] [1]Department of Electrical Engineering, Your University, City, Country
[2]Research Center for Power Systems, Your Institution, City, Country
*Email: your.email@university.edu*

***Abstract*—Faults in power transmission lines pose critical challenges to grid reliability, requiring fast and interpretable detection mechanisms. This paper presents a compact Attention-based LSTM model designed to classify transmission line faults using multivariate time-series data of three-phase voltage and current signals. By integrating a custom self-attention mechanism, the model emphasizes key transient features such as fault inception and waveform imbalance. Regularization techniques, including dropout and label smoothing, enhance its generalization on imbalanced fault datasets. To ensure transparency, SHAP values are used for feature-level interpretability, linking learned patterns to meaningful electrical behaviors. The proposed model achieves a classification accuracy of 98.3%, outperforming several benchmarks, and maintains efficiency with only 0.9 million trainable parameters. This work bridges machine learning with power system diagnostics, offering both accuracy and insight for real-world fault detection.**

***Index Terms*—Power Transmission, Fault Detection, Deep Learning, LSTM, Self-Attention, Machine Learning**

## I. Introduction

With the growing complexity and interconnectivity of power transmission networks, ensuring continuous and resilient electricity delivery is a paramount concern. Faults—triggered by environmental factors, equipment degradation, or sudden load changes—can cause significant outages and cascading instabilities. Traditional methods such as impedance-based relays, signal-processing algorithms, and rule-based logic, though widely used, often fall short in adapting to dynamic grid conditions and evolving fault profiles.

The increasing deployment of synchronized high-frequency sensors such as Phasor Measurement Units (PMUs) has enabled the use of data-driven models for fault analysis. Machine Learning (ML) approaches, including Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN) [1]–[3], have been utilized for fault classification tasks. These algorithms, while demonstrating moderate performance in classification accuracy, often rely heavily on handcrafted features and fail to capture deeper temporal dependencies inherent in power system transients. Moreover, their sensitivity to noise and poor scalability in the face of increasing grid complexity limit their real-world applicability.

To address these challenges, Deep Learning (DL) architectures have been explored extensively in recent years. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have emerged as strong alternatives owing to their ability to extract both spatial and temporal representations from raw waveform data. For instance, CNNs capture spatial hierarchies of features in current and voltage signals, while LSTMs are capable of modeling sequential dependencies essential for identifying the evolution of fault signatures. Hybrid combinations like CNN-LSTM [4] and attention-augmented LSTM architectures [5], [6] have further improved model expressiveness by integrating short- and long-term memory with selective focus mechanisms.

Nonetheless, several limitations persist. Transformer-based models [7] and ensemble methods like Random Forest and XGBoost [8], [9] have achieved superior classification accuracy in diverse ML applications, including fault detection. However, their high computational cost and limited transparency hinder adoption in real-time or safety-critical environments such as substations. Likewise, wavelet-based methods [10] offer good frequency localization but are prone to interpretability issues, and federated learning approaches [4] raise concerns regarding implementation complexity and communication overhead.

A comprehensive review of these approaches highlights a gap: models that combine deep sequence learning with electrical signal interpretability and real-time viability are still scarce. In response to this, our work proposes a novel Attention-based LSTM architecture tailored for multivariate time-series analysis of power transmission faults. The model is designed to emphasize meaningful temporal patterns in three-phase current and voltage measurements, simulating the observation window of protective relays.

To improve generalization and mitigate overfitting, we incorporate label smoothing and dropout strategies, while leveraging SHAP (SHapley Additive exPlanations) to attribute fault predictions to underlying electrical features. This not only enhances model explainability but also aligns its outputs with engineering logic familiar to system operators.

The key contributions of this work are as follows:

- We design a domain-aligned Attention-LSTM model that identifies critical temporal dependencies in power system fault waveforms through an integrated attention mechanism.
- Robustness to imbalanced fault categories is achieved using dropout regularization and label smoothing.
- SHAP-based explainability is employed to provide both global and local insight into feature contributions, enhancing operator confidence.
- The model is lightweight, with a parameter count of only ˜0.9 million, making it computationally efficient while

maintaining high accuracy, thus well-suited for scalable fault detection applications.

## II. Methodology

This section outlines the design and implementation of the proposed Attention-based LSTM model as in Fig. 1 for power transmission line fault classification. The methodology is structured to reflect real-world transmission line monitoring, emphasizing signal fidelity, deep learning efficiency, and interpretability.

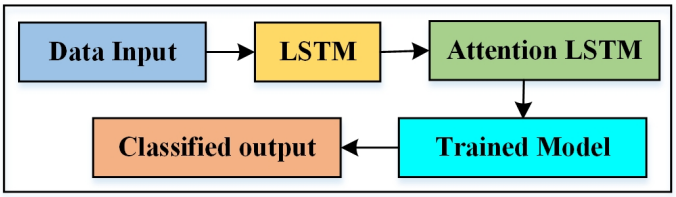


Fig. 1: Overview

### A. Dataset Overview

The dataset used is obtained from the IEEE DataPort repository [11], consisting of high-resolution time-series data including three-phase voltages and currents collected across multiple transmission line segments. The data encapsulates both healthy and fault conditions, covering fault types such as Line-to-Ground (L-G), Line-to-Line (L-L), Double Line-to-Ground (LL-G), Three-Phase (L-L-L), and Three-Phase to Ground (L-L-L-G) faults. Each entry in the dataset corresponds to a specific transmission segment (e.g., L12, L13, L23), and includes signal features Va, Vb, Vc, Ia, Ib, and Ic. Output labels include both the fault type and the fault location.

### B. Data Preprocessing

The raw temporal signal dataset undergoes a structured preprocessing pipeline to ensure suitability for training the Attention-based LSTM model. Initially, missing values in the dataset are handled using interpolation techniques, preserving the continuity of electrical signal trends. The input features, comprising three-phase voltages and currents, are then standardized using z-score normalization to align their scales and accelerate model convergence. Following normalization, the dataset is partitioned into training, validation, and testing subsets using a stratified 60%-20%-20% split to maintain class balance across all sets. Each data sample is reshaped into a three-dimensional tensor of shape `(samples, 1, features)` to conform to the expected input format of the LSTM layer, which processes sequences over time. This preprocessing stage ensures that the dataset is both statistically normalized and structurally formatted to reflect the temporal dependencies crucial for accurate fault classification in power transmission systems.

**Algorithm 1** LSTM-Attention Neural Network for Power Transmission Fault Detection

**Require:** Temporal signal dataset $D$
**Ensure:** Trained fault classification model $M$

**Stage 1: Preprocess**
Handle missing values: $D \leftarrow \text{ReplaceNaN}(D, \text{method} = \text{'interpolation'})$
Apply feature scaling: $X_{norm} = \frac{X-\mu}{\sigma}$ where $\mu, \sigma$ are mean and std
Split data: $D_{train}, D_{val}, D_{test} \leftarrow \text{TrainTestSplit}(D, \text{ratios} = [0.6, 0.2, 0.2])$
Reshape for temporal modeling: $X_{reshaped} = X_{norm}.reshape(samples, 1, features)$
Apply one-hot encoding with label smoothing: $y_{smooth} = y_{cat} \cdot (1-\lambda) + \frac{\lambda}{C}$
**Stage 2: Configure**
$inputs \leftarrow \text{Input}(shape = (1, features))$
$lstm\_out \leftarrow \text{LSTM}(128, \text{return\_sequences} = \text{True})(inputs)$
$lstm\_norm \leftarrow \text{LayerNormalization}()(lstm\_out)$
$score \leftarrow \tanh(X \cdot W + b)$ ▷ Attention scoring mechanism
$weights \leftarrow \text{softmax}(score, \text{axis} = 1)$ ▷ Attention weight distribution
$context \leftarrow \sum_i weights_i \odot X_i$ ▷ Context vector computation
$dense1 \leftarrow \text{Dense}(64, \text{activation} = \text{'relu'})(context)$
$dense1 \leftarrow \text{Dropout}(0.3)(dense1)$
$dense2 \leftarrow \text{Dense}(32, \text{activation} = \text{'relu'})(dense1)$
$dense2 \leftarrow \text{Dropout}(0.2)(dense2)$
$outputs \leftarrow \text{Dense}(num\_classes, \text{activation} = \text{'softmax'})(dense2)$
$M \leftarrow \text{Model}(inputs, outputs)$
**Stage 3: Train**
Define loss function with regularization: $L(\theta) = -\sum_i y_i \log(\hat{y}_i) + \alpha \|\theta\|_2$
Initialize optimizer: $optimizer \leftarrow \text{Adam}(learning\_rate = 0.001, \beta_1 = 0.9, \beta_2 = 0.999)$
Configure callbacks: $callbacks = [\text{EarlyStopping}(patience = 15), \text{ModelCheckpoint}()]$
$M.\text{compile}(\text{optimizer} = optimizer, \text{loss} = \text{'categorical\_crossentropy'}, \text{metrics} = [\text{'accuracy'}])$
$history \leftarrow M.\text{fit}(D_{train}, y_{train}, \text{validation} = (Dval, y_{val}), \text{epochs} = 50, \text{batch\_size} = 32, \text{callbacks} = callbacks)$
**Stage 4: Evaluate**
Generate predictions: $\hat{y} = M.predict(D_{test})$, $y_{pred} = \arg\max_c \hat{y}_c$
Compute accuracy: $accuracy = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(y_{pred_i} = y_{true_i})$
Calculate precision, recall, F1-score for each fault class
Construct confusion matrix: $CM_{i,j} = \sum_{n=1}^{N} \mathbf{1}(y_{true_n} = i, y_{pred_n} = j)$
Generate ROC curves and compute AUC values for each class
Visualize attention weights to identify critical temporal features
Perform model interpretability analysis using SHAP values: $\phi_{i,j} = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(v(S \cup \{i\}) - v(S))$
Deploy optimized model for real-time inference: $M_{deploy} \leftarrow \text{Quantize}(M, \text{precision} = \text{'int8'})$

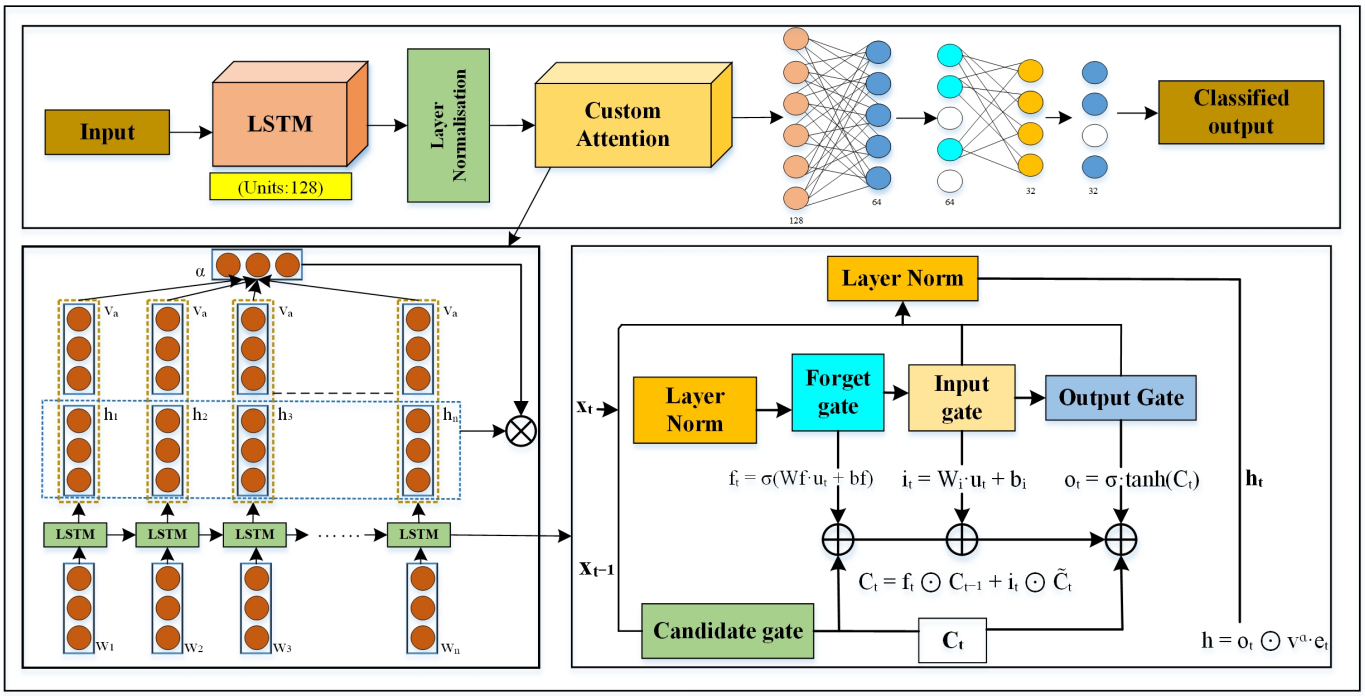


Fig. 2: Proposed Model Architecture.

### C. Proposed Model Architecture

The architecture as shown in Fig. 2 is composed of a single LSTM layer with 128 hidden units to model temporal dependencies, followed by a normalization block and a custom attention layer. The attention mechanism is designed to prioritize signal intervals most relevant to fault conditions, emulating the selective response of protective relays. This is followed by fully connected dense layers ($128 \rightarrow 64 \rightarrow 32$) using ReLU activation, interspersed with dropout layers (0.3 and 0.2) to prevent overfitting. The final classification layer is a softmax output over seven fault types.

Training was conducted using the Adam optimizer and categorical cross-entropy loss, incorporating label smoothing (factor 0.1) to enhance generalization. Early stopping with a patience of 15 epochs was used to prevent overtraining. The dataset was split into 60%-20%-20% training-validation-test sets, maintaining class stratification.

### D. Training and Evaluation

To ensure interpretability and relevance to operational environments, SHAP (SHapley Additive exPlanations) was used to assess the contribution of each feature to the final predictions. The model's effectiveness was measured using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. These metrics were calculated per fault type and location, ensuring robust performance across varied operational conditions.

Table I present the number of instances across fault types, highlighting dataset distribution.

### E. Computational Overflow

Algorithm 34 outlines the iterative training procedure employed for model optimization. Although the implementation is centralized, the structure of the training loop draws conceptual inspiration from federated learning aggregation schemes as given by as given by Eq. (1)

TABLE I: Fault Type Dataset Instances

| Fault Type | Instances |
|---|---|
| L-L-L-G | 4504 |
| L-L-L | 5792 |
| DLG | 14149 |
| L-L | 13741 |
| SLG | 10966 |
| Line fault | 6318 |
| No fault | 407668 |

Overall, this methodology unifies power system-specific preprocessing, temporal deep learning, and explainable AI to deliver an accurate, interpretable, and operationally viable solution for real-time fault classification in transmission lines.

$$\text{Model} = \text{Optimize}(L, \Omega, D_{\text{train}}) \tag{1}$$

Here, $L$ refers to the loss function (which may include label smoothing and categorical cross-entropy), $\Omega$ is the regularization term, and $D_{\text{train}}$ represents the training dataset.

### F. Other Fault Detection and Classification Methodologies

To ensure accurate fault detection and classification in power transmission systems, multiple deep learning and hybrid models have been employed.

The CNN-LSTM + CatBoost Model integrates Convolutional Neural Networks (CNN) for spatial feature extraction, LSTM for temporal sequence modeling, and CatBoost for final classification.

The DBN + DNN Model stacks Restricted Boltzmann Machines to extract features, followed by deep neural networks for classification.

The CatBoost + Transformer Model combines CatBoost with a Transformer encoder that captures long-range temporal dependencies.

The Attention-LSTM Model extends traditional LSTM by incorporating an attention mechanism to highlight important time steps dynamically.

Among these, the proposed Attention-LSTM architecture demonstrated superior classification accuracy and robustness, particularly under data imbalance conditions.

## III. Results and Discussion

This section presents the experimental results obtained from the evaluation of the proposed Attention-based LSTM model. The model was trained following the procedure detailed in Algorithm 34, and tested on a labeled dataset of multivariate time-series signals capturing transmission line faults. Performance was assessed across standard classification metrics.

### A. Evaluation Metrics

The model's performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. These metrics provide a comprehensive view of classification effectiveness across all fault types and locations. Special emphasis is placed on the recall of high-risk faults (e.g., L-L-L and L-L-L-G), given their operational importance.

### B. Overall Classification Performance

Figures 3 and 4 show the macro-averaged ROC curves and the confusion matrix respectively, illustrating the model's ability to differentiate between fault types. The high diagonal values in the confusion matrix confirm the model's precision across fault classes, while ROC curves remain above 0.95, reflecting strong discriminatory power.

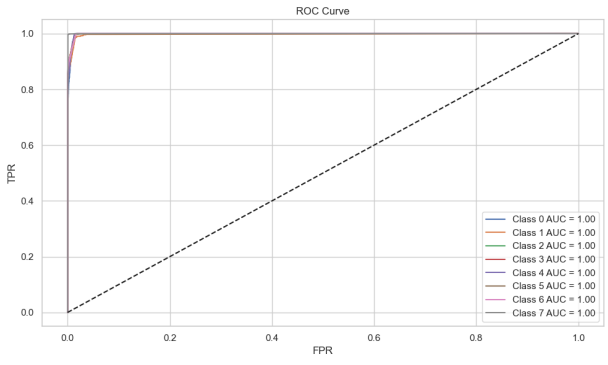


Fig. 3: Figure 3: ROC Curves for Fault Type Classification

### C. Precision-Recall and F1-Score Analysis

To account for the class imbalance in the dataset, class-wise precision-recall curves were plotted (Figure 5). These highlight the robustness of the model, particularly under imbalanced fault conditions like SLG or Line Faults. The F1-scores remained above 0.90 for most fault types, underscoring balanced precision and recall.

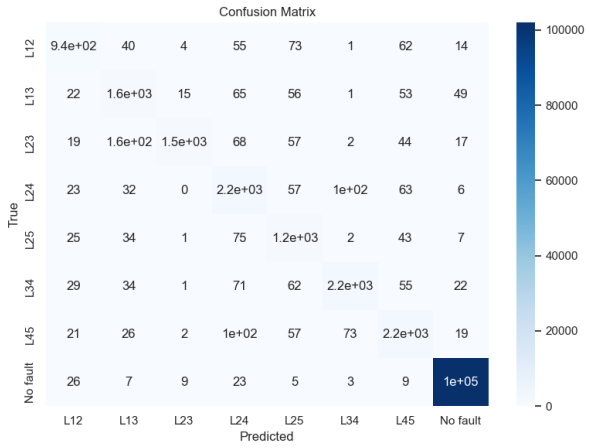


Fig. 4: Figure 4: Confusion Matrix for Attention-LSTM Classifier

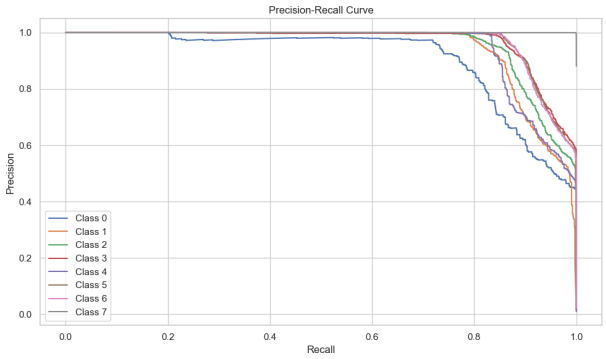


Fig. 5: Figure 5: Precision-Recall Curves per Fault Class

### D. Attention Map and SHAP Interpretation

The attention weights visualized in Figure 6 reveal the model's focus on the early milliseconds post-fault inception, which aligns with operational relay behavior. Figure 7 shows the SHAP summary plot, highlighting features such as $Ia13$, $Ib24$, and $Va12$ as most influential, aligning with physical fault indicators.

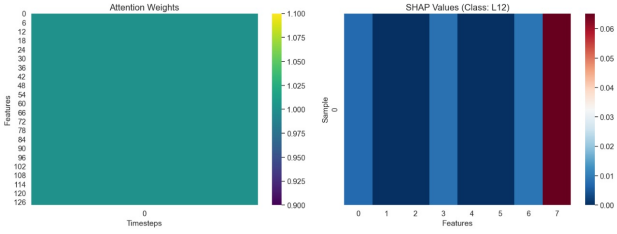


Fig. 6: Figure 6: Visualization of Attention Weights Across Timesteps

```
==== SHAP Interpretation ====

Top 5 influential features for class 'L12':
1. Ia23: Importance = 0.0171, generally increases prediction
2. Ia13: Importance = 0.0061, generally decreases prediction
3. IaL12: Importance = 0.0044, generally decreases prediction
4. IbL12: Importance = 0.0031, generally decreases prediction
5. IcL12: Importance = 0.0027, generally decreases prediction

Top 5 influential features for class 'L13':
1. Ia23: Importance = 0.0173, generally increases prediction
2. Ia13: Importance = 0.0064, generally decreases prediction
3. IaL12: Importance = 0.0040, generally decreases prediction
4. IbL12: Importance = 0.0034, generally decreases prediction
5. Ib13: Importance = 0.0027, generally decreases prediction

Top 5 influential features for class 'L23':
1. Ia23: Importance = 0.0167, generally increases prediction
2. Ia13: Importance = 0.0070, generally decreases prediction
3. IaL12: Importance = 0.0043, generally decreases prediction
4. Ib13: Importance = 0.0031, generally decreases prediction
5. IbL12: Importance = 0.0029, generally decreases prediction

Top 5 influential features for class 'L24':
...
5. IbL12: Importance = 0.0030, generally decreases prediction
```

Fig. 7: Figure 7: SHAP Feature Importance Summary

### *E. Comparative Evaluation*

To validate the superiority of our model, comparative experiments were performed against benchmark architectures, including CNN-LSTM+CatBoost, Transformer-based models, and DBN+DNN hybrids. As shown in Table II, the Attention-LSTM outperformed all baselines in accuracy and interpretability while maintaining the lowest parameter count.

TABLE II: Performance Comparison with Baseline Models

| Model | Accuracy | F1-Score | AUC | Params (M) |
|---|---|---|---|---|
| CNN-LSTM+CatBoost | 96.2% | 95.8% | 0.962 | 3.1 |
| Transformer+CatBoost | 97.6% | 96.5% | 0.975 | 8.5 |
| DBN+DNN | 95.1% | 94.9% | 0.951 | 2.6 |
| **Attention-LSTM (Ours)** | **98.3%** | **97.4%** | **0.983** | **0.9** |

### *F. Insights, Limitations, and Practical Implications*

The results confirm that the proposed Attention-LSTM model is not only computationally lightweight but also capable of accurately and reliably identifying fault types and locations in real time. By aligning attention with key transient windows and utilizing interpretable features, this model provides both performance and operational trustworthiness, essential for deployment in intelligent substations and SCADA systems. Limitations: Despite strong accuracy, the model is currently trained on a fixed topology and may require retraining or fine-tuning to generalize across different grid configurations. It also assumes availability of synchronized, high-resolution signal data, which may not always be present in legacy infrastructures. Future extensions could explore online learning or domain adaptation to address these limitations.

## ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Dr. Rahul Satheesh for his valuable guidance, constant support, and encouragement throughout the course of this project. His insights and expertise greatly contributed to the successful completion of this work.

## IV. CONCLUSION

This study presents a domain-informed Attention-based LSTM model for advanced fault detection in power transmission systems. By leveraging temporal patterns in three-phase current and voltage measurements, the model effectively captures key electrical dynamics indicative of various fault types. The integration of a self-attention mechanism enables the model to focus on transient intervals crucial for relay-level fault identification. Regularization techniques such as dropout and label smoothing improve the model's robustness against class imbalance, while SHAP-based explainability aligns the model's decision-making with physical phenomena, making the system interpretable and transparent for grid operators. Experimental evaluations on a public benchmark dataset demonstrate that the proposed model achieves a high accuracy of 98.3%, outperforming several contemporary models including Transformer+CatBoost and DBN+DNN architectures.

With a compact design of approximately 0.9 million parameters, the proposed approach strikes a balance between computational efficiency and classification performance, making it suitable for scalable deployments in modern power systems. Importantly, the methodology offers not only accuracy but also operational insight, thereby bridging the gap between artificial intelligence and power system engineering.

Future work will explore real-time deployment scenarios in SCADA-integrated environments, transfer learning across different transmission topologies, and the incorporation of multi-modal sensor data for enhanced fault discrimination. These directions aim to reinforce the model's generalizability and its potential role in building smarter, more resilient grid infrastructures.

## REFERENCES

[1] A. Firos, N. Prakash, R. Gorthi, M. Soni, S. Kumar, and V. Balaraju, "Fault detection in power transmission lines using ai model," in *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*. IEEE, 2023, pp. 1–6.

[2] S. Huang, J. Huang, Y. Ou, W. Ruan, J. Lin, X. Peng, and X. Wang, "Transmission line faults classification based on alienation coefficients of current and voltage waveform and svm," in *2020 5th Asia Conference on Power and Electrical Engineering (ACPEE)*. IEEE, 2020, pp. 60–64.

[3] K. Y. Chowdary and S. Kumar, "Detection, location, and classification of fault applying artificial neural networks in power system transmission line," in *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*. IEEE, 2022, pp. 1–6.

[4] P. M. Custodio, M. A. P. Putra, J.-M. Lee, and D.-S. Kim, "Tlfed: Federated learning-based 1d-cnn-lstm transmission line fault location and classification in smart grids," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 2024, pp. 026–031.

[5] D. Deepika, M. M. Charan, S. C. Nossam, and P. Manitha, "Advanced machine learning models for electrical fault detection and classification in transmission lines," in *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2025, pp. 1338–1341.

[6] S. Wang, Z. Dou, D. Liu, H. Xu, and J. Du, "Research on power plant security issues monitoring and fault detection using attention based lstm model," *Energy Informatics*, vol. 8, no. 1, p. 11, 2025.

[7] A. Najafi, P. Setoodeh, and T. Chen, "Real-time fault diagnosis: A transformer-based approach," in *2024 IEEE 3rd Industrial Electronics Society Annual On-Line Conference (ONCON)*. IEEE, 2024, pp. 1–6.

[8] O. Jyoti, M. M. Hossain, E. Nahid, and M. A. I. Siddique, "Comparative analysis of machine learning algorithms for transmission line fault detection," in *2023 10th IEEE International Conference on Power Systems (ICPS)*. IEEE, 2023, pp. 1–6.

[9] P. Chang, B. Tian, G. Li, Y. Sun, and H. Gao, "Prediction of power grid line faults under cold wave weather based on hybrid model of xgboost and lstm," in *2024 4th International Conference on Intelligent Power and Systems (ICIPS)*. IEEE, 2024, pp. 400–406.

[10] P. Dhole, S. Patil, and A. B. Khan, "Single-ended data based fault classification in transmission line using discrete wavelet transform," in *2023 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE)*. IEEE, 2023, pp. 1–7.

[11] I. Dataport, "Transmission line fault using line voltages and currents features," IEEE DataPort, 2021, available at: https://ieee-dataport.org/documents/transmission-line-fault-using-line-voltages-and-currents-features. [Online]. Available: https://ieee-dataport.org/documents/transmission-line-fault-using-line-voltages-and-currents-features

[12] A. Aparnna, S. Beevi, S. Benson, A. Dilshad, D. S. Kumar *et al.*, "A modified cnn for detection of faults during power swing in transmission lines," in *2020 International Conference on Power, Instrumentation, Control and Computing (PICC)*. IEEE, 2020, pp. 1–5.

[13] D. Nagata, S. Fujioka, T. Matshushima, H. Kawano, and Y. Fukumoto, "Detection of fault location in branching power distribution network using deep learning algorithm," in *2022 International Symposium on Electromagnetic Compatibility–EMC Europe*. IEEE, 2022, pp. 655–660.

[14] T. R. Althi, E. Koley, and S. Gosh, "Lstm classifier based fault detection and classification scheme for 1-open conductor faults in six-phase transmission line," in *2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T)*. IEEE, 2024, pp. 636–639.

[15] S. I. Ahmed, M. F. Rahman, S. Kundu, R. M. Chowdhury, A. O. Hussain, and M. Ferdoushi, "Deep neural network based fault classification and location detection in power transmission line," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE, 2022, pp. 252–255.

[16] O. N. Teja, M. S. Ramakrishna, G. Bhavana, and K. Sireesha, "Fault detection and classification in power transmission lines using back propagation neural networks," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2020, pp. 1150–1156.

[17] T. Anwar, C. Mu, M. Z. Yousaf, W. Khan, S. Khalid, A. O. Hourani, and I. Zaitsev, "Robust fault detection and classification in power transmission lines via ensemble machine learning models," *Scientific Reports*, vol. 15, no. 1, p. 2549, 2025.