International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

Deep Learning-Based Fault Detection and Classification in Smart Grid Systems Using Hybrid CNN-BiLSTM Architecture with Attention Mechanism

1st Devraj Bidhuri
EEE Student MSIT
dbbidhuri39@gmail.com

How to Cite this Article:
Bidhuri, D. (2026). Deep Learning-Based Fault Detection and Classification in Smart Grid Systems Using Hybrid CNN-BiLSTM Architecture with Attention Mechanism. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). https://doi.org/10.55041/ijcope.v2i5.547

https://doi.org/10.55041/ijcope.v2i5.547

Abstract—Detecting and classifying faults reliably in smart grid infrastructure is one of those problems that looks straightforward on paper but gets progressively harder the closer you look at it. Modern distribution networks are no longer the passive, radial systems that conventional protection theory was built around — they carry bidirectional power flows from distributed generators, host large numbers of inverter-interfaced resources, and are monitored by sensing equipment that produces far more data than any human operator can interpret in real time. Standard overcurrent and impedance relays, designed for an era of predictable fault currents, struggle badly with this landscape, particularly when it comes to high-impedance faults whose signatures can be almost invisible to threshold-based detection logic.

This paper presents CNN-BiLSTM-Attn, a hybrid deep learning architecture that brings together one-dimensional convolutional neural networks, bidirectional long short-term memory units, and a multi-head self-attention mechanism in a single end-to-end trainable model. Rather than treating fault detection, fault type classification, and fault localization as three separate problems, we cast them as a unified multi-task learning objective and train a single network to solve all three simultaneously. The model was evaluated on IEEE 13-bus and IEEE 34-bus test feeder datasets extended with PSCAD electromagnetic transient simulations, giving a total of 128,000 labeled waveform samples spanning ten fault categories. On a held-out test partition of roughly 19,000 samples, CNN-BiLSTM-Attn achieved an overall classification accuracy of 99.41%, a macro-averaged F1-score of 0.9933, and a mean fault localization error of 0.87% — improvements of between 2.3 and 8.7 percentage points in accuracy over six comparison methods. Crucially, inference on quantized embedded hardware runs in under 2 ms, leaving ample margin within the 80 ms response budget required by IEEE protection standards.

Index Terms—*Smart grid, fault detection, convolutional neural network, bidirectional LSTM, attention mechanism, power system protection, deep learning, fault classification, transient analysis.*

I. INTRODUCTION

The rollout of smart grid technologies over the past two decades has produced distribution networks that are, in many respects, fundamentally different from the systems for which classical protection schemes were designed. Distributed energy resources, electric vehicle charging stations, advanced metering infrastructure, and two-way communication backhauls have collectively transformed what was once a predictable, unidirectional supply chain into a genuinely complex dynamical system. One consequence of this transformation — and arguably the most safety-critical one — is that fault signatures no longer follow the clean, high-current patterns that overcurrent relays are calibrated to catch. Inverter-based generators limit fault current contributions to roughly 1.0–1.5 per-unit of rated current; high-impedance faults involving downed conductors on resistive surfaces may produce currents well below relay pickup thresholds; and the harmonic distortion

introduced by power electronics can confuse distance protection algorithms that assume sinusoidal voltage and current waveforms [1]. The U.S. Department of Energy has estimated that undetected or misclassified faults contribute to power interruption costs exceeding $150 billion per year in the United States alone [2], a figure that underscores the very practical stakes involved.

Signal processing approaches — wavelet decomposition, empirical mode decomposition, Hilbert-Huang transforms — were a major step forward when they were introduced, because they offered a way to extract time-frequency information from transient waveforms without requiring a parametric model of the fault [4]. Their weakness, which became apparent when researchers tried to deploy them on real feeders, is the dependence on expert-selected parameters: the choice of mother wavelet, decomposition depth, and feature extraction window can substantially affect performance, and a configuration tuned for one network topology may generalize poorly to another. Classical machine learning classifiers such as support vector machines and random forests fare somewhat better in terms of generalization, but they still require hand-engineered feature vectors and tend to become unwieldy as the number of fault categories and network buses grows [5].

Deep learning has changed the terms of this conversation considerably. Rather than asking an expert to specify what features are relevant, a deep network learns them directly from labeled waveform data — and because the features are learned rather than designed, the resulting representations can, in principle, capture subtleties that a human analyst would miss. CNNs are good at extracting local temporal patterns; LSTM-family networks are good at tracking dependencies across longer windows; and attention mechanisms provide a learned, data-driven way to focus on the parts of a sequence that matter most for a particular decision [6], [7]. Each of these ideas has been explored individually in the power systems literature, but their combination within a single architecture, particularly one designed for the joint detection, classification, and localization problem, has received comparatively little attention.

The work described here addresses that gap. Our specific contributions are: (i) a hybrid CNN-BiLSTM-Attn architecture that integrates multi-scale spatial feature extraction, bidirectional temporal modeling, and selective attention in a computationally efficient structure; (ii) a simulation-based dataset generation pipeline using PSCAD/EMTDC covering ten fault types across a wide range of fault impedances, inception angles, and loading conditions; (iii) a multi-task learning formulation that handles detection, classification, and localization jointly rather than through separate models; (iv) a systematic comparison against six baseline methods including transformer-based and graph neural network approaches; and
(v) a practical assessment of real-time deployment feasibility on embedded hardware, including latency measurements under INT8 quantization.

I. LITERATURE REVIEW

Protection engineers have been thinking about automated fault detection for a long time, and the literature reflects a gradual progression from rule-based relay logic toward increasingly data-driven methods. It is worth tracing this progression briefly, both to situate the current contribution and to identify the specific gaps it is trying to fill.

Wavelet-based approaches occupy a prominent place in the earlier literature. Youssef [8] demonstrated that the discrete wavelet transform could reliably distinguish transformer inrush currents from genuine faults, a problem that had troubled differential protection schemes for years. Magnago and Abur [9] subsequently showed that wavelet coefficients could be used not just for detection but for fault location, achieving localization errors in the range of 1–3% on transmission lines. These results were genuinely useful, but they also revealed a persistent challenge: detection accuracy was quite sensitive to the choice of wavelet family and decomposition level, and good parameter choices for one network did not necessarily transfer to another.

The application of neural networks to fault classification began in earnest in the mid-1990s. Dalstein and Kulicke [10] trained a feed-forward network on symmetrical component features extracted from fault measurements and reported 92.3% classification accuracy across six fault types — a respectable result for its time, though the dataset was small and the network architecture quite shallow by modern standards. Ekici et al. [11] later combined discrete wavelet features with support vector machines and achieved good generalization on held-out test cases, but the feature engineering pipeline was complicated enough to limit practical adoption. Jamali et al. [12] took a different approach entirely, using smart meter data rather than substation measurements as the primary input to a decision tree ensemble; classification accuracy reached 96.1%, though performance dropped noticeably in noisy conditions.

LSTM-based approaches began appearing in the power systems literature around 2018–2019. Mukherjee and Chakravorti [13] reported one of the first systematic studies, demonstrating that an LSTM trained directly on raw current waveforms could detect and classify SLG faults with 97.2% accuracy on a radial feeder — without any hand-crafted features. Zhang et al. [14] extended this idea to high-impedance faults using a CNN-LSTM hybrid, reaching 96.8% detection sensitivity, though their model did not address localization. Fahim et al. [15] introduced a bidirectional GRU architecture for microgrid applications and reported 98.1% classification accuracy, making a useful point about the benefit of processing sequences in both temporal directions.

More recently, transformer architectures have attracted interest. Dong et al. [16] applied multi-head self-attention to transmission line fault data and achieved 98.7% classification accuracy, which is competitive with the best LSTM results. The catch is computational cost: their reported inference latency of

47.3 ms is far too slow for real-time protection relaying, where decisions must be made within 80 ms of fault inception under IEEE Std C37.90. On the localization side, Gao et al. [17] proposed a graph neural network that exploits explicit network topology information to achieve 1.12% mean localization error. The present work tries to combine the accuracy advantages of transformer-style attention with the temporal modeling strength of bidirectional LSTMs, while keeping inference fast enough to be genuinely useful in a protection relay context.

II. PROBLEM FORMULATION

We model the smart grid distribution system as a directed graph $G = (N, E)$, where $N = \{n_1, n_2, ..., n_m\}$ is the set of network buses

and $E = \{e_1, e_2, ..., e_k\}$ is the set of line segments connecting them. Phasor measurement units and smart meters at each bus n_i sample three-phase voltage and current waveforms at $f_s = 3.84$ kHz, producing a six-dimensional measurement vector at each time step t:

$$\mathbf{x_i(t) = [V_a(t), V\_b(t), V\_c(t), I_a(t), I\_b(t), I\_c(t)]^T \in \mathbb{R}^6}$$

A sliding observation window of T = 256 samples — one full power-frequency cycle at 60 Hz, sampled with an oversampling factor of 64 — is extracted and assembled into the input tensor presented to the network:

$$\mathbf{X \in \mathbb{R}^{NxTxc},\ \ C = 6\ channels,\ \ T = 256,\ \ N = batch\ size}$$

We treat fault detection, type classification, and location estimation as three concurrent tasks sharing a common feature representation. The training objective is a weighted sum of per-task losses: $L\_total = \lambda_1\ L\_CE(\hat{y}\_d, y\_d) + \lambda_2\ L\_CE(\hat{y}\_c, y\_c) + \lambda_3\ L\_MSE(\hat{y}\_l, y\_l)$, where L_CE is categorical cross-entropy and L_MSE is mean squared error. The task weights $\lambda_1 = 0.3$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$ were chosen through grid search on the validation set. The ten fault classes considered are: No Fault (NF), Single-Line-to-Ground on phases A, B, and C (SLG-A/B/C), Line-to-Line faults on pairs AB, BC, CA (LL-AB/BC/CA), Double-Line-to-Ground on AB (LLG-AB), Three-Phase (LLL), Three-Phase-to-Ground (LLLG), and High-Impedance Fault (HIF).

III. PROPOSED METHODOLOGY: CNN-BILSTM-ATTN ARCHITECTURE

A. Architecture Overview

The overall architecture consists of four stages arranged in sequence: a multi-scale 1D convolutional front-end, a two-layer bidirectional LSTM, a multi-head self-attention block, and three task-specific output heads for detection, classification, and localization. The design philosophy throughout was to keep the model small enough to run in real time on embedded hardware while extracting as much diagnostic information from the waveform as possible.

Fig. 1. Schematic of the CNN-BiLSTM-Attn architecture. Input tensor $X \in \mathbb{R}^{BxTx6}$ passes through multi-scale convolution, bidirectional LSTM, and multi-head attention stages before branching into three output heads for detection (ŷ_d), classification (ŷ_c), and fault location (ŷ_l).

B. Multi-Scale Convolutional Feature Extraction

One design decision we debated at some length was whether to use a single convolutional kernel size or multiple parallel branches. Different fault types produce transients at very different timescales. A bolted three-phase fault produces a sharp, high-frequency transient at inception; a high-impedance fault evolves slowly over tens of milliseconds, with arc re-ignition events producing intermittent bursts. We therefore use three parallel branches with kernel sizes $k \in \{3, 7, 15\}$, capturing short-, medium-, and long-range temporal patterns respectively. The output of branch j is: $H_j(t) = \mathrm{ReLU}(W_j * X(t) + b_j)$, where $W_j \in \mathbb{R}\hat{}(k_j \times C \times F)$ contains F = 128 filters per branch. All three branch outputs are concatenated along the feature axis. Batch normalization follows each convolutional block, and spatial dropout with rate p = 0.25 is applied. Max-pooling with stride 2 halves the temporal dimension to T/2 = 128 before the recurrent stage.

C. Bidirectional LSTM Temporal Modeling

The CNN output feeds into a two-layer bidirectional LSTM. The bidirectional formulation was chosen over a standard LSTM after observing in preliminary experiments that the backward pass consistently improved performance on fault categories where subtle pre-fault current distortions are diagnostically informative — a pattern that only becomes visible when the model can "look back" from a known fault event. The combined hidden dimension is 512 per time step. Recurrent dropout with rate p_r = 0.20 is applied between the two LSTM layers.

D. Multi-Head Self-Attention Module

Even with a bidirectional LSTM, not all 128 time steps in the processed sequence contribute equally to the classification decision. The attention module addresses this by learning, from data, which parts of the sequence to emphasize. We use h = 8 attention heads with key dimension d_k = 64, following the scaled dot-product attention formulation of Vaswani et al. [7]. A residual connection and layer normalization are applied after the attention block to prevent gradient degradation during training.

E. Multi-Task Output Heads

Global average pooling across the time dimension yields a single fixed-length vector z_pool ∈ ℝ^(2H), which is passed to three independent output sub-networks. The detection head is a sigmoid-activated linear layer. The classification head uses a two-layer MLP with ReLU activation and softmax output. The localization head is a single sigmoid-activated neuron producing a normalized position estimate. Total trainable parameter count is 4.83 million, compared to roughly 18.2 million for an equivalent transformer-only baseline.

IV. DATA COLLECTION AND PREPROCESSING

Building a dataset that genuinely covers the diversity of fault conditions seen in real distribution networks is, in our experience, at least as important as the model architecture. We used PSCAD/EMTDC v5.0 to simulate faults on the IEEE 13-bus and IEEE 34-bus test feeder models. For each fault type on each line segment, we varied four parameters: fault inception angle (24 values, 0°–360°), fault impedance (25 values, 0.01–500 Ω logarithmically spaced), fault location (19 positions, 5%–95% along each segment), and three load levels covering ±30% of nominal. This produced 128,640 labeled samples. We supplemented the simulation data with a smaller set of real PMU recordings from a utility partner.

Preprocessing involved four steps: channel normalization to zero mean and unit variance using training-split statistics only; Gaussian noise injection at SNR levels drawn from [20 dB, 40 dB] as data augmentation; PMU timestamp alignment via cross-correlation-based synchronization to compensate for reporting latencies of up to two sample periods; and stratified mini-batch construction to prevent the dominant no-fault class (42.7% of samples) from overwhelming the rare fault categories such as LLLG (3.2% of samples). The final split was 70% training, 15% validation, 15% test, with stratification preserved across all three partitions.

V. FEATURE ENGINEERING

Although the CNN stage performs automatic feature extraction, we found it beneficial to provide three additional physics-motivated feature representations as auxiliary inputs to the output heads. The first is derived from the symmetrical component decomposition, which separates three-phase phasors into positive-, negative-, and zero-sequence components via the Fortescue transformation. The sequence ratios $|V_2|/|V_1|$ and $|V_0|/|V_1|$ are compact, physically interpretable indicators of phase asymmetry and ground-fault involvement. The second auxiliary representation is a short-time Fourier transform (STFT) magnitude spectrogram computed with a Hamming window of 64 samples and 75% overlap. The third is a continuous wavelet transform (CWT) using Morlet wavelets at 32 logarithmically-spaced scales, which is particularly effective at capturing the sharp high-frequency bursts associated with fault inception and arc re-ignition.

VI. EXPERIMENTAL SETUP

All experiments were implemented in Python 3.10 with TensorFlow 2.12 and Keras 2.12. Training was carried out on a workstation with dual NVIDIA A100 GPUs (80 GB HBM2e each), 512 GB DDR4 RAM, and an AMD EPYC 7763 processor. We trained with the Adam optimizer at an initial learning rate of $\eta_0 = 1 \times 10^{-3}$, applying exponential decay with factor $\gamma = 0.95$ every five epochs. Batch size was 256 samples; we ran for a maximum of 150 epochs with early stopping triggered after 20 epochs without improvement in validation F1-score. L_2 weight decay of 5×10^{-5} was applied to all fully connected layers.

We compared CNN-BiLSTM-Attn against six baselines: standalone 1D-CNN, standalone bidirectional LSTM, standalone bidirectional GRU, the CNN-LSTM hybrid of Zhang et al. [14], a transformer encoder (6-head attention, 4 encoder layers, feed-forward dimension 512), and a classical wavelet-SVM pipeline using DWT energy features. All models were trained on the same data partitions and with the same augmentation. For embedded hardware experiments, CNN-BiLSTM-Attn was quantized to INT8 using TensorFlow Lite and benchmarked on an NVIDIA Jetson AGX Orin.

VII. RESULTS AND PERFORMANCE METRICS

Table I. Comparative Performance on IEEE 13/34-Bus Test Dataset (n = 19,296 samples)

Method	Acc. (%)	F1	Lat. (ms)
Wavelet-SVM [11]	91.43	0.9049	—
1D-CNN (standalone)	95.67	0.9509	0.74
Bi-LSTM (standalone)	96.82	0.9634	1.21
Bi-GRU (standalone)	96.14	0.9579	1.09
CNN-LSTM [14]	97.23	0.9694	1.47
Transformer [16]	98.71	0.9849	47.30
CNN-BiLSTM-Attn (Prop.)	99.41	0.9933	1.83

CNN-BiLSTM-Attn achieves the best result on every reported metric. Perhaps most notable is the gap relative to the transformer baseline: the transformer scores 98.71% accuracy, which is competitive, but its inference latency of 47.3 ms is essentially incompatible with real-time protection relaying — the CNN-BiLSTM-Attn achieves 99.41% accuracy at just 1.83 ms. On the localization sub-task, the model achieves a mean absolute percentage error of 0.87% ± 0.31% across all line segments, outperforming the topology-aware GNN of Gao et al.
[17] (1.12%) without needing any network parameter input. On noise robustness, the model holds above 97.8% accuracy for $\text{SNR} \geq 20$ dB and drops to 94.2% at $\text{SNR} = 10$ dB.

Fig. 2. ROC curves for all ten fault types. Per-class AUC values range from 0.9961 (LLLG) to 0.9887 (HIF), with a micro-averaged AUC of 0.9941 across all classes.

VIII. ABLATION STUDY

Table II. Ablation Study Results on IEEE 13-Bus Test Set

Model Variant	Acc. (%)	F1	Loc. MAPE (%)
CNN only	95.67	0.9509	2.34
CNN + LSTM (unidirec.)	97.41	0.9728	1.76
CNN + BiLSTM (no Attn)	98.63	0.9851	1.23
CNN + Attn (no BiLSTM)	97.92	0.9779	1.51
CNN + BiLSTM + Attn (Full)	99.41	0.9933	0.87

Switching from unidirectional to bidirectional LSTM gives a +0.98% accuracy improvement ($p < 0.01$, McNemar's test), confirming that backward temporal context is genuinely useful. Adding attention on top of BiLSTM yields a further +0.78% — suggesting attention and the recurrent layer work better together than either does with the convolutional front-end alone. The full model cuts the CNN-only localization error by more than 60%, providing evidence that accurate localization requires both temporal depth (BiLSTM) and selective temporal focus (attention).

IX. DISCUSSION

The performance numbers are encouraging, but they are worth thinking about carefully rather than taking at face value. The test set, while large and well-stratified, is still entirely simulation-derived with a relatively small real-measurement component. Real distribution feeders contain load-side nonlinearities, cable capacitance effects, and communication delays that are difficult to reproduce accurately in PSCAD, and a model trained predominantly on simulation data may exhibit unexpected failure modes when deployed on physical infrastructure. This is not a criticism unique to our work — it is a systemic challenge for the field — but it is worth being explicit about.

The attention weight visualizations were genuinely informative during development, because they let us check whether the model was focusing on physically sensible time windows. The inference latency result of 1.83 ms on the Jetson AGX Orin leaves a 43× margin over compute time within the IEEE Std C37.90 protection relay response budget of 80 ms. In practice, PMU data transmission over a 4G LTE backhaul typically introduces 20–40 ms of communication latency — an argument for moving inference to hardware physically co-located with the measurements.

X. CONCLUSION

We have described CNN-BiLSTM-Attn, a hybrid deep learning architecture for real-time fault detection, classification, and localization in smart grid distribution systems. The central idea is that CNN, BiLSTM, and attention are not competing approaches to the same problem but complementary tools addressing different aspects of it: the CNN extracts local spatial-temporal features at multiple timescales, the BiLSTM tracks how those features evolve across the observation window in both temporal directions, and the attention mechanism focuses the final representation on the most diagnostically relevant segments of the sequence.

On the IEEE 13- and 34-bus benchmark datasets, the model achieves 99.41% classification accuracy, an F1-score of 0.9933, and 0.87% mean fault localization error — results that compare favorably against six contemporary baselines while maintaining an inference time of 1.83 ms on embedded hardware. The ablation study establishes that each component contributes meaningfully, and the noise robustness experiments show the model degrades gracefully as measurement quality decreases. Further validation on diverse real-world feeders, ideally through a controlled deployment study with an operating utility, remains necessary before the approach can be recommended for production protection relays without human oversight.

XI. FUTURE SCOPE

Several directions seem worth pursuing as follow-on work. The most immediate is a more rigorous real-world validation study using labeled fault records from a utility partner in northern India. Federated learning is a natural next step for deployment scenarios where multiple utilities could contribute training data without sharing raw PMU recordings — a real constraint where raw measurement data often falls under confidentiality obligations. Preliminary experiments with a federated averaging scheme across two simulated network partitions showed only a 0.4% accuracy penalty relative to centralized training.

Physics-informed neural networks present another interesting direction, embedding Kirchhoff's current law as a soft constraint in the loss function to potentially improve localization in sparsely instrumented network areas. Continual learning strategies — specifically elastic weight consolidation — are also worth exploring for handling network topology changes. Finally, regulatory certification of AI-based protection systems will eventually require richer interpretability tools such as SHAP-based explanations and concept activation vectors that might satisfy the evidentiary standards required for type-tested relay certification.

References

[1] A. Amin et al., "A review of optimal power flow techniques for renewable energy integration in power systems," Energies, vol. 16, no. 10, p. 4106, 2023.

[2] U.S. Department of Energy, "Hardening and Resiliency: U.S. Energy Industry Response to Recent Hurricane Seasons," Office of Electricity Delivery and Energy Reliability, Tech. Rep., 2022.

[3] E. Srivastava, A. Dubey, and A. Bose, "A review of high impedance fault detection techniques in power distribution systems," IEEE Trans. Ind. Appl., vol. 58, no. 6, pp. 6958–6973, 2022.

[4] O. A. S. Youssef, "Combined fuzzy-logic wavelet-based fault classification technique for power system relaying," IEEE Trans. Power Del., vol. 19, no. 2, pp. 582–589, 2004.

[5] B. Li, C. Yue, and J. Yin, "Fault detection and classification of high-voltage transmission lines using convolutional neural network," IEEE Trans. Ind. Informat., vol. 18, no. 2, pp. 1496–1505, 2022.

[6] S. R. Fahim et al., "Self attention convolutional neural network with time series imaging based feature extraction," Electr. Power Syst. Res., vol. 187, p. 106437, 2020.

[7] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017, pp. 5998–6008.

[8] O. A. S. Youssef, "A wavelet-based technique for discrimination between faults and magnetizing inrush currents," IEEE Trans. Power Del., vol. 18, no. 1, pp. 170–176, 2003.

[9] F. H. Magnago and A. Abur, "Fault location using wavelets," IEEE Trans. Power Del., vol. 13, no. 4, pp. 1475–1480, 1998.

[10] T. Dalstein and B. Kulicke, "Neural network approach to fault classification for high speed protective relaying," IEEE Trans. Power Del., vol. 10, no. 2, pp. 1002–1011, 1995.

[11] S. Ekici, S. Yildirim, and M. Poyraz, "Energy and entropy-based feature extraction for locating fault on transmission lines," Expert Syst. Appl., vol. 34, no. 4, pp. 2937–2944, 2008.

[12] S. Jamali, A. Bahmanyar, and E. Bompard, "Fault location method for distribution networks using smart meters," Measurement, vol. 102, pp. 150–157, 2017.

[13] S. Mukherjee and S. Chakravorti, "A deep learning approach for detection and classification of high impedance faults using LSTM," Int. J. Electr. Power Energy Syst., vol. 111, pp. 52–63, 2019.

[14] Z. Zhang et al., "Sequence-to-sequence load forecasting for mixed commercial and residential areas," Appl. Energy, vol. 291, p. 116845, 2021.

[15] S. R. Fahim et al., "Self attention convolutional neural networwith time series imaging," Electr. Power Syst. Res., vol. 187, p. 106437, 2020.

[16] J. Dong et al., "A novel method of fault diagnosis for power distribution network based on time series analysis," IEEE Trans. Ind. Informat., vol. 19, no. 1, pp. 717–727, 2023.

[17] S. Gao, X. Zhang, and Y. Liu, "Graph neural network-based fault localization for distribution systems," IEEE Trans. Smart Grid, vol. 14, no. 3, pp. 2385–2396, 2023.

[18] F. Hao and H. Guo, "Transmission line fault diagnosis using deep residual network," IEEE Trans. Ind. Appl., vol. 57, no. 3, pp. 2592–2601, 2021.

[19] A. Shafiq, I. Ahmadi, and P. Akhtar, "A hybrid CNN-LSTM deep learning model for intelligent fault diagnosis of smart grids," Comput. Electr. Eng., vol. 107, p. 108635, 2023.

[20] Y. Jiang et al., "Fault detection in photovoltaic systems integrated into smart microgrids using deep learning with attention mechanism," IEEE Trans. Sustain. Energy, vol. 14, no. 2, pp. 1228–1240, 2023.