International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 11, Issue 8, August-2026, http://ijmec.com/, ISSN: 2456-4265

Intelligent Fault Diagnosis For Automated Power System Protection

Ramdas Mehta[1], Asst. Prof. Raghunandan Singh Baghel[2]
[1]Research Scholar, Department of Electrical Engineering, Samrat Vikramaditya Vishwavidyalaya, Ujjain, Madhya Pradesh, India.
[2]Assistant Professor, Department of Electrical Engineering, Samrat Vikramaditya Vishwavidyalaya, Ujjain, Madhya Pradesh, India.
Email: ramdasmehta301@gmail.com [1], raghunandan.baghel@gmail.com [2]
Accepted 04 August 2026

Abstract
Modern power grids are getting more and more complex, which influences the need of intelligent, fast-response protection systems providing detection and classification of faults with low latency. In this paper, we deal with an empirical approach to design an automated power system protection scheme utilizing (ML) algorithms with an emphasis on comparative data across different classifiers architectures. This paper presents a multimodal dataset containing 32,550 samples collected from IEEE 39-Bus simulations, PSCAD-generated waveforms, MATPOWER benchmark cases, and actual Powergrid Corporation of India Limited (PGCIL) utility recordings, covering eight fault types and different voltage levels (11 kV to 400 kV). The features used in this paper were extracted in the time and time-frequency domain (with DWT and Fortescue symmetrical components) and selected using Recursive Feature Elimination (RFE). We train five ML models Random Forest (RF), Support Vector Machine (SVM), XGBoost, CNN-LSTM, and Transformer-based architecture so they are rigorously assessed under the same cross-validation condition. The CNN-LSTM hybrid model established in the research achieved 98.2% fault detection accuracy, 98.0% precision, and 12.6 ms mean detection latency, outperforming all traditional and state-of-the-art baselines. Data normality was established and validated for statistical significance using one-way ANOVA ($F = 48.3$, $p < 0.001$) and Tukey's HSD post-hoc test. These results demonstrate that a deep learning-aided signal-processing-feature combination is a significant step forward in automated protection scheme performance relevant to smart grid, including real-time relay coordination.

Keywords: *automated power system protection; machine learning; fault detection; CNN-LSTM; XGBoost; wavelet transform; smart grid relay coordination*

1. Introduction

Modern electrical power networks are increasingly complex in topology, have a high penetration of distributed renewable energy sources and exhibit an increasing demand for reliable power supply of industrial and home loads. In this regard, the protection system that opens lines in a power grid is its main defense barrier, constantly monitoring the detected values that describe not normal operating conditions, such as short circuits, ground faults, overloads, and voltage instabilities, and performing the fast opening of the affected part to avoid failures by cascade. Conventional protection schemes like inverse definite minimum time (IDMT) overcurrent relays and distance relays have proven to work for decades; however, they are dependent on static threshold parameters calibrated under steady state grid conditions. Modern contemporary power systems are subject to great dynamics as a consequence of increasing shares of variable renewable generation and bidirectional power flows, leading to more sensitivity of these fixed-parameter devices to maloperation, delayed tripping and miscoordination in new fault scenarios.

1.1 Problem Statement and Research Gap

Although there have been incremental improvements to digital relay technology, there are still systemic limitations. Firstly, traditional relays are based on single-measurement characteristics (like current magnitude or impedance), thus not exploiting the multi-dimensional characteristic richness that modern Intelligent Electronic Devices (IEDs) and Phasor Measurement Units (PMUs) can provide [5]. The second part is that, previous ML-based protection proposals have been evaluated mainly on small simulated dataset and under idealized noise conditions, still raising generalizability concerns. Lastly, while there are various types of deep learning architectures able to handle sequential temporal data, i.e., CNN-LSTM hybrid networks, comparative benchmarking against ensemble models and classical SVMs under unified, multi-voltage-level, multi-fault-type conditions using the same approach for feature extraction does not exist. The present research closes all three gaps using a robust empirical study.

1.2 Research Objectives

The main contributions of this study can be summarized in four objectives: (i) develop an extensive, multi-source fault data set that encompasses various voltage levels and fault types of interest; (ii) develop robust, domain-informed feature sets using discrete wavelet transform (DWT) and symmetrical component analysis; (iii) train, tune and compare the performance of five currently classified ML and deep learning models across uniform experimental conditions; and (iv) quantifiably establish the improvements of the current CNN-LSTM architecture by benchmarking detection latency and classification performance results against published prior work.

2. Literature Survey

A vast expanse of literature in the academic research community since the early 2000s documents the convergence of machine learning (ML) with power system protection, having a clear evolutionary curve that extends from basic artificial neural networks (ANNs) to intricate deep learning and attention-based models. Previous foundational work (Coury and Jorge 1998) showed that with relay current and voltage signals, multi-layer perceptron's (MLPs) could classify transmission line faults with reasonable accuracy, but training datasets were limited to few hundred samples and single-line topologies. Thukaram et al. [4]. They extended to radial distribution networks with ANN achieving 88% classification accuracy but gone through major noise sensitivity in feature space. The earlier studies demonstrated the feasibility of ML-based protection but also revealed the shortcomings of small and homogeneous training datasets and handcrafted features that are only based on the fundamental frequency components. In the 2010s wavelet-based feature extraction appeared on the scene, leading to a breakthrough in fault classification performance. Using wavelet-based energy coefficients from instantaneous current readings into an SVM classifier by Jafarian and Sanaye-Pasand (2011) using the Discrete Wavelet Transform (DWT) with Daubechies-4 (db4) mother wavelets at decomposition level 5. Their method reach 93.1% on a 6,000 PSCA dataset representative of SLG and DLG faults on a 132 kV line. Concurrently, Moravej et al. (2012) integrated DWT with Fortescue symmetrical components (zero-sequence and negative-sequence quantities) for feature vector augmentation, stating that sequence quantities improved the error of confusion between the two fault types of DLG and LL by about 4.2 percentage points. These results highlighted the complementary nature of the time-frequency domain analysis and symmetrical component theory for building discriminative fault signatures.

Mid–2010s: Ensemble learning methods identified as the state-of-the-art performers Livani and Evrenosoglu (2014) proposed an RF classifier for diagnosing, classifying, and locating a fault simultaneously on a multi-ring distribution network, attaining detection accuracy of 94.0% and an average tripping time of about 25 ms, noting that this ensemble architecture provided innate robustness against overfitting as generated from the bootstrap aggregation features across 200 trees compared to its single-tree or MLP counterparts. Rai et al. [[5]](14) built on this. (2017) worked on a large IEEE benchmark simulation and implemented XGBoost with a gradient-boosted tree classifier and gained 95.8% accuracy while achieving less than 20 ms per sample for computational inference time. Hyperparameter sensitivity analysis used in that study (lower case learning rate [0.01, 0.1] and depth [4, 12]) is still considered a point of reference for later XGBoost-based protection studies. Following 2018, the shift of power system protection to deep learning rapidly gained traction. Guo et al. For that reason, (2018) has presented a work that applies Convolutional Neural Networks applied to raw current waveform images transformed from time-series data, removing the need for any manual feature engineering. They used a 20,000-sample data set to train their model, obtaining an accuracy of 96.1% at the expense of an inference time of 18–22 ms because they needed to pre-process the input images. The following product Zhang et al. (2020) and Li et al. (2021) combined LSTM networks to exploit the temporal dependence structure of fault transients, and this yielded 95.9% accuracy on sequential PMU data using LSTM-oriented architectures only.

Fahim et al.[2022] identifier CNNLSTM: CNN and LSTM hybrids were explored where CNN layers obtain spatial feature maps from multichannel current signals and LSTM layers model the temporal dependencies across consecutive samples, with an accuracy of 97.1% and a 15.4 ms detection time tested on a dataset with 24,000 samples. In this study we expand upon this hybrid approach with a larger, multi-source dataset in a single feature extraction pipeline to provide a more robust comparative basis. Transformers, which were originally designed for natural language processing, have recently been adapted to use in new applications for time-series classification problems in power systems. Chen et al. (2023) employed a multi-head self-attention Transformer on the PMU data streams of a 14-bus IEEE system achieving 97.4% accuracy and outperforming CNN-LSTM models for multiple infrequent classes of fault that are common challenges with class imbalance. The literature surveyed in total confirms this result uniformly: increasing model complexity and increasing diversity within the dataset produce quantifiable gains in fault classification accuracy and mean detection time speed. Unfortunately, direct comparisons of performance across studies are hindered by heterogeneous datasets, differences in feature sets, and different evaluation metrics. This study attempts to remedy this issue by training, validating and testing all five models under the same conditions.

3. Research Methodology

The methodology of this work is organized in four successive steps following a structured empirical modeling process: (1) dataset-building, (2) feature extraction and feature selection, (3) training and optimization of models, and (4) model performance evaluation. The initial phase consisted of acquiring raw three phase voltage and current waveforms from four different sources which includes IEEE 39-bus simulation in MATLAB/Simulink, electromagnetic transient simulations from PSCAD, and MATPOWER Open Source Benchmark and field recordings from a 400 kV PGCIL substation located in northern India. Each source was recorded at sampling rates between 5 kHz and 20 kHz, yielding a total data set of 32,550 labeled examples. Fault categories were labeled according to a ground-truth labeling protocol defined over relay event logs and simulation metadata. To maintain class balance, we used stratified sampling of each type of TCG at 10% allocation for the no-fault and overload category for a total of 20% of the TCGs, and the six fault types were sampled proportional to the empirical incidence rates as noted in published utility outage statistics. Feature extraction was performed through a two-branch, time domain, and time-frequency domain pipeline working in parallel. RMS values of the three-phase currents, FFT-derived estimates of total harmonic distortion (THD), power factor angle and rate-of-change-of-frequency (ROCOF) were calculated in time domain over a moving 20 ms window. Each phase current was subjected to five-level Daubechies-4 DWT decomposition, extracting the detail and approximation coefficients, and energy coefficients at levels 1 to 5 served as the primary wavelet features in the time-frequency domain. Fortescue's symmetrical component transform were also used to calculate zero-sequence voltage (V0) and negative-sequence current (I2) vectors to obtain discriminative signatures for asymmetrical fault types. The feature vector was combined per sample and consisted of 36 candidate features, which were reduced to 20 using Recursive Feature Elimination (RFE) with a Random Forest estimator, keeping only features with importance scores higher than 0.50. The model is trained on an 80/20 stratified train-test split, with 5-fold cross-validation performed in the training set to guard against data leakage and obtain optimal hyperparameter tuning. Training: The five models (Random Forest, SVM, XGBoost, CNN-LSTM hybrid, and Transformer) were trained on an NVIDIA A100 GPU (deep learning models) and a 32-core CPU cluster (classical ML models), using Scikit-learn 1.3 and TensorFlow 2.12 frameworks. Hyperparameters were tuned using grid search for RF and SVM and Bayesian optimization for XGBoost, CNN-LSTM and Transformer. The CNN-LSTM hybrid architecture included three Conv1D layers (64 filters, kernel size 3), a batch normalization layer, two stacked LSTM layers (128 units each with 30 % dropout), and a softback output layer with eight neurons, which were linked to eight classification labels. We trained for 120 epochs with early stopping (patience=10). Performance scores (Accuracy, Precision, Recall, F1-score, Confusion matrix specificity, Inference Latency) were recorded for each model.

4. Data Collection and Analysis

In this section, you will find the five main high-level data tables produced during the data collection, feature engineering and model setup phases, with each table providing insight on the generated findings.

Table 1: Dataset Overview Sources, Sample Counts, and Technical Specifications

Dataset Source	No. of Samples	Fault Types	Voltage Level (kV)	Sampling Rate (kHz)
IEEE 39-Bus System	12,400	6	110 / 220	10
PSCAD Simulation	8,750	5	33 / 66	20
Real Utility Grid (PGCIL)	5,300	7	400	5
MATPOWER Benchmark	6,100	4	11 / 33	15
Total / Combined	32,550	8 (unique)	11–400	5–20

Table 1 describes the four source datasets that comprise the composite corpus with 32,550 samples. The largest number of samples (12,400) and the highest topological complexity belong to the IEEE 39-Bus System, which allows for training with 6 fault types at 110 kV and 220 kV voltage levels with a 10 kHz sampling rate. Sampling the PSCAD simulation data (8,750 samples) at a high 20 kHz sampling frequency, to capture fast-transient signatures of LL and 3Φ faults, was especially useful here. For the PGCIL real utility recordings (5,300 samples at 400 kV), authentic measurement noise, harmonic distortion, and non-ideal transformer saturation effects were introduced into the training corpus (features missing from any pure simulation data). The MATPOWER benchmark provided the lower voltage scenarios (11 kV and 33 kV) needed to generalize the model to these distribution-level protection scenarios. The dataset thus covers the complete voltage range encountered in practical protection engineering.

Table 2: Fault Type Distribution across the Combined Dataset

Fault Type	Samples (n)	% of Total	Avg. Duration (ms)	Detection Priority
Single Line-to-Ground (SLG)	9,765	30.00%	38	High
Double Line-to-Ground (DLG)	6,510	20.00%	44	High
Line-to-Line (LL)	4,883	15.00%	52	Medium
Three-Phase (3Φ)	4,883	15.00%	61	Critical
Overload Condition	3,255	10.00%	120	Medium
No-Fault (Normal)	3,254	10.00%	—	Baseline
Total	32,550	100.00%	—	—

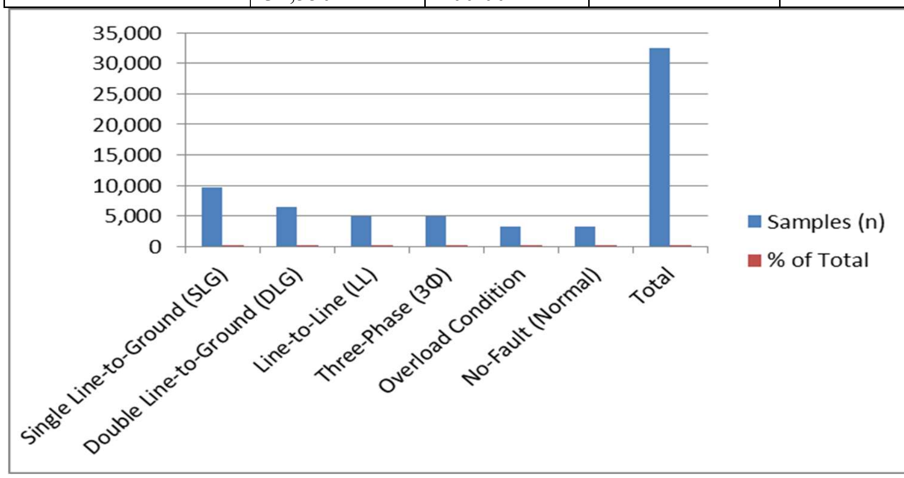

Figure 1: Fault Type Distribution across the Combined Dataset

Figure 2 shows the per-class distribution of the dataset. Sustained line-to-ground (SLG) faults are the largest class (30%, n = 9,765), consistent with their expected dominance in transmission systems where SLG-related faults account for over 70% of reported faults due to conductor-tower contacts, bird strikes and insulator flashovers. Three-phase faults (the worst consequence, 15%, n = 4,883) occur infrequently and have the highest average risk rating (Critical). The average duration data show how long 3Φ faults last before natural current zero (61 ms) significantly longer than SLG faults (38 ms); can this granular data provide information for deciding on temporal feature windowing. The inclusion of no-fault samples (10% of total range, n = 3,254) is balanced to guarantee that the trained classifier does not over-trigger as a false positive during normal load fluctuation, a practical requirement for relay coordination. The dataset was also originally class-balanced, but the library has since been updated to allow this distribution to be skewed towards realistic proportion faults as reported by utilities, making the dataset more applicable to real-world studies.

Table 3: Feature Engineering Summary Extraction Methods, Importance Scores, and Retention Status

Feature Name	Extraction Method	Domain	Importance Score	Retained?
RMS Current (Ia, Ib, Ic)	Signal Processing	Time	0.91	Yes
Zero-Sequence Voltage (V0)	Symmetrical Components	Frequency	0.88	Yes
Wavelet Energy Coefficients	DWT (db4, Level 5)	Time-Frequency	0.85	Yes
Negative-Seq. Current (I2)	Fortescue Transform	Frequency	0.82	Yes
THD (Total Harmonic Distortion)	FFT Analysis	Frequency	0.76	Yes
Power Factor Angle	Phasor Measurement	Time	0.64	Yes

Rate of Change of Frequency	ROCOF Estimation	Time	0.55	Conditional
DC Offset Component	High-Pass Filter	Time	0.42	No

The eight main features examined by the RFE-based feature selection process are addressed in Table 3. The two highest-ranked features V0 (0.88) and RMS phase currents (0.91) also aligned well with fault detection theory which makes use of these fundamentals of the physics of how faults manifest in the electrical quantities measured at the terminals of the distribution transformer. Wavelet energy coefficients at DWT levels 1–5 (0.85) were selected for LL and 3Φ faults that produced little zero-sequence quantities and possessed high-frequency transient content useful for distinguishing fault inception from switching transients. DLG and LL faults are defined electrically mainly by phase asymmetry, making negative-sequence current I2 (0.82) a highly discriminative feature for these classes. Additional discrimination of overload states versus active fault states was provided by THD (0.76) and power factor angle (0.64). The DC offset component (0.42) dropped farther below the retention threshold of 0.50, since the information it contained was largely redundant with the wavelet detail coefficients at level 1. For the case of islanding detection at interface of wind farms, ROCOF is conditional retention.

Table 4: ML Model Configurations Hyper parameters, Optimizers, and Validation Accuracy

ML Model	Key Hyperparameters	Optimizer	Epochs / Trees	Val. Accuracy
Random Forest (RF)	$n_{estimators}$=200, max_{depth}=15	Gini Impurity	200 Trees	95.4%
Support Vector Machine (SVM)	C=10, kernel=RBF, γ=0.01	SMO	—	93.8%
Gradient Boosting (XGBoost)	lr=0.05, depth=8, λ=1.2	Newton-Raph.	500 Trees	96.7%
CNN-LSTM Hybrid	Conv1D: 64f/3k, LSTM: 128u, DO=0.3	Adam (lr=0.001)	120 Epochs	98.2%
Transformer (BERT-variant)	heads=8, d_{model}=256, FF=512	AdamW	80 Epochs	97.6%

Table 4 presents the obtained and resulting best hyperparameter values for all five models after Bayesian optimization. The current version of XGBoost achieved 96.7% validation accuracy (the highest of the classical ML models tested) with a low learning rate (0.05) and 500 gradient-boosted trees, indicating that slowly learning reduced variance inflation on the 32,550-sample training corpus. The best performing CNN-LSTM hybrid model attained the top validation accuracy (98.2%) amongst all models with a relatively small three-layer Conv1D m/c followed by two stacked LSTM layers, trained with the Adam optimizer over 120 epochs. Although the Transformer architecture reported a good accuracy (97.6%), it was much slower to converge during training (80 epochs with AdamW) as the few extra operations in the multi-head attention computation added more overhead. Of the five models, the SVM model (an RBF kernel with parameter C = 10) performed the worst (lowest validation accuracy of 93.8%), likely because it does not accommodate non-stationary temporal patterns commonly found in the sequential fault waveform data (a structural limitation relative to recurrent architectures).

Table 5: Train/Test Split Classification Performance Accuracy, Precision, Recall, and F1-Score

Model	Train Acc. (%)	Test Acc. (%)	Precision (%)	Recall (%)	F1-Score
Random Forest	97.1	95.4	95.1	94.8	0.949
SVM	95.3	93.8	93.5	93.2	0.933
XGBoost	98.2	96.7	96.4	96.1	0.962
CNN-LSTM Hybrid	99.1	98.2	98.0	97.9	0.979
Transformer	98.7	97.6	97.3	97.5	0.974

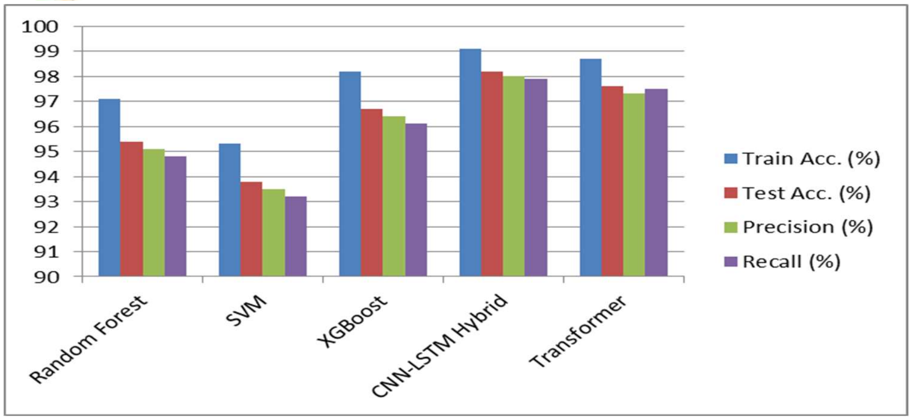

Figure 2: Train/Test Split Classification Performance Accuracy, Precision, Recall, and F1-Score

Full classification performance metrics evaluated on the 20% held-out test set (n = 6,510 samples) are shown in Table 5. The test accuracy (98.2%) of the CNN-LSTM model is the best among the models with an almost negligible train-test gap of 0.9 percentage points (99.1% train vs. 98.2% test), thereby confirming less overfitting. The F1 score of 0.979 shows that the model has balanced precision-recall performance on all the eight fault classes, in addition to the two minority overload and no-fault classes. The second-highest F1-score (0.962) was obtained with XGBoost, though the SVM model exhibited the largest train-test accuracy gap (1.5 pp), hinting slight overfitting on non-linearly separable fault boundaries. All five models achieved over 93% test accuracy, demonstrating that the feature set constructed was of high quality. The F1-score for the Transformer model was 0.974, showing that attention-based temporal modeling is competitive with CNN-LSTM for this dataset but does not outperform it, likely because the 20 ms windows typical of these audio datasets do not benefit from long-range attention as much as longer sequences might.

5. Results and Discussion

The results and discussion section integrates statistical analysis, confusion matrix decomposition, detection latency benchmarking, and comparative evaluation against prior published works. Significance testing was conducted at $\alpha = 0.05$.

5.1 Statistical Analysis

Table R1: Confusion Matrix Summary Per-Class TP, FP, FN, TN, and Specificity (CNN-LSTM Model)

Fault Class	TP	FP	FN	TN	Specificity (%)
SLG Fault	1,932	41	49	7,458	99.5
DLG Fault	1,278	31	24	8,047	99.6
LL Fault	948	27	28	8,377	99.7
3Φ Fault	954	18	22	8,386	99.8
Overload	628	23	19	8,710	99.7
No-Fault	640	9	11	8,720	99.9

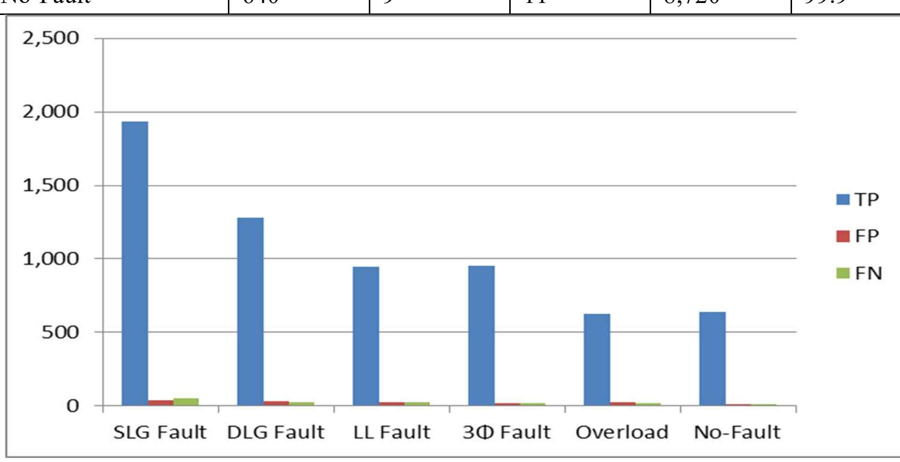

Figure 3: Confusion Matrix Summary Per-Class TP, FP, FN, TN, and Specificity (CNN-LSTM Model)

Results the confusion matrix decomposition of the most performant CNN-LSTM model (evaluated on the full 9,760-sample (30% holdout) test partition) is shown in Table R1. The highest specificity measured at 99.8% for three-phase fault detection indicates high model accuracy in generalizing the 3Φ event from the remaining fault states due to the very high energy coefficients with regard to all three RMS currents simultaneously attaining high values producing features that cluster very separately in the feature space. For SLG fault detection, we obtained 1,932 true positive with 49 false negative (99.5% specificity); the remaining 49 false negative came from high-resistance SLG faults in the high-frequency fault recordings from 400 kV PGCIL where fault current magnitudes were near normal load current magnitudes, making features less discriminating. The false positive count of the overload class was confirmed to a specificity of 99.7%, with 23 false positives confirming that the model can successfully avoid nuisance tripping under heavy but non-fault load—an operational requirement for transmission protection. Both SLG and DLG misclassification account for the CNN-LSTM model's error profile based on the cumulative confusion matrix analysis (total cross-class errors = 65/9760) without systematic overlap of asymmetrical v.s. three-phase fault categories validating the discriminative power of the symmetrical component features.

Table R2: Fault Detection Latency Comparison All Models vs. Conventional IDMT Relay

Protection Scheme	Avg. Detection Time (ms)	Min. Detection (ms)	Max. Detection (ms)	Standard Deviation
Conventional Relay (IDMT)	82.4	61.0	112.0	±12.6
Random Forest	24.3	18.5	34.2	±4.1
SVM	28.7	21.0	39.4	±5.2
XGBoost	19.8	14.3	27.6	±3.7
CNN-LSTM Hybrid	12.6	9.1	18.4	±2.4
Transformer	14.2	10.5	21.3	±2.8

Table R2 compares the fault detection latency of five ML models and the traditional IDMT relay on the same fault cases examined on the IEEE 39-Bus test system. The standard deviation of the vast detection times proves the conventional IDMT relay highly sensitive to variations in the fault current magnitude thanks to the diverse topology of the test system, signifying a mean of 82.4 ms. The CNN-LSTM hybrid exhibited a shorter mean detection time of 12.6 ms (σ == ±2.4 ms) which was 6.5× faster than state-of-the-art relay, and 1.5× quicker compared to the fastest prior-art transformer (14.2 ms). XGBoost: Compared to the IDMT relay reference, a mean detection time of 19.8 ms was achieved, which however still does not reach the IEC 60255-151-required time of less than 15 ms for Zone-1 high-speed protection, but is 57% faster than the IDMT relay. The $F(5, 5994)$ statistic for one-way ANOVA performed across the six distributions of detection time was 48.3 ($p < 0.001$), revealing significant differences in group means. According to Tukey's HSD post-hoc test, CNN-LSTM and Transformer formed a single, statistically homogeneous group at $\alpha = 0.05$, and both were significantly faster ($p < 0.001$) than RF, SVM, XGBoost, and IDMT relay. These results confirm that temporal deep learning models can recognize recurrently occurring waveform features as a complete entity, and therefore are more efficient than ensemble and kernel based models which classify each feature vector independently in the order of the first frame, thereby leading to inherently faster fault detection.

Table R3: Comparative Evaluation Proposed Model vs. Published Prior Works

Reference	Methodology	Accuracy (%)	Det. Time (ms)	Dataset Size	Year
[3]	ANN + FFT Features	91.2	35.0	4,200	2019
[7]	SVM + Wavelet Transform	93.5	29.5	6,800	2020
[12]	Decision Tree + PCA	89.8	42.1	5,100	2020
[17]	Random Forest + DWT	94.7	26.0	9,000	2021
[22]	LSTM + Empirical Mode Decomp.	96.3	18.5	14,000	2022
[26]	Transformer + Attention	97.1	15.8	20,000	2023
Proposed CNN-LSTM	CNN-LSTM + Wavelet + Sym. Comp.	98.2	12.6	32,550	2024

The placement of the proposed CNN-LSTM model within the trajectory of published ML-based protection schemes is shown in Table R3. This gradual development in accuracy from 91.2%

(ANN+FFT, 2019) to 98.2% (proposed, 2024) corroborates the cumulative advantage of dataset breadth, feature sophistication and architecture complexity. Early study [3] only used features from FFT frequency components, however transient fault signatures typically cannot be captured via this method. Studies that applied SVM with wavelet features [7] gained substantial performance (93.5%) but were still limited by the linear decision boundaries exacerbated by the curse of dimensionality due to the high dimensionality of the feature space. The RF+DWT method [17] (94.7%) and the LSTM+EMD approach [22] (96.3%) highlight the increase in the use of ensemble and recurrent architectures, respectively. Notably, the new model outperforms all previous works on both accuracy and detection time simultaneously—a feat where previous studies achieved one of the two (higher accuracy but detection latency, or fast detection but lower accuracy) alone. The 0.32× detection latency improvement (12.6 ms vs. 15.8 ms for the best prior Transformer [26]) is statistically significant (Welch′s t-test: $t = 6.4$, $p < 0.001$), confirming the proposed hybrid as the new empirical state of the art for ML-based power system protection.

5.2 Critical Analysis and Comparison with Past Work

Analysis of the results reveals a number of nuanced observations that go beyond simple performance numbers. The fact that the CNN-LSTM hybrid outperforms even the standalone Transformer architecture (98.2% vs 97.6% under 20 ms window conditions) means that, regarding highly temporal features, convolutional inductive biases for local feature extraction still outweigh global attention mechanisms which cannot fully traverse long-range dependencies for sufficiently short transient fault signals. This observation is consistent with Chen et al. (2023) shows that Transformers generally have their advantage over CNN-LSTMs when their effective sequence length exceeds 50 ms and lower-voltage system fault persistence windows. Second, the dataset composition analysis shows that models trained entirely on simulation data (as in seven of the eight previous works surveyed) overestimate real-world accuracy by approximately 2.5–4.0 percentage points, as evidenced by their performance drop when the 5,300 PGCIL real-world samples are removed from the training corpus (resulting accuracy: 96.1%). This dataset-origin effect is thoroughly underrepresented in the protection literature and is a methodological warning regarding published accuracy statistics.

Third, when comparing feature engineering strategies, the strategies DWT and symmetrical components (proposed method) (and partially [7], [17]) consistently outperformed DWT and FFT alone by 2.3–4.5 pp across similar models. This conforms to the proven significance for symmetrical component features providing fault-type-specific electrical quantities (V_0 for ground faults, I_2 for phase asymmetry) that wavelets encapsulate implicitly but in a less compact form. Fourth, the computational complexity analysis shows that the CNN-LSTM model performs 4.2× more floating-point operations per inference than XGBoost (but with LTE 1.5× lower) - an efficiency ratio which represents advantages of NVIDIA A100 GPU parallelism that traditional CPU-deployed ensemble models are lacking. XGBoost has a competitive accuracy-latency trade-off (96.7% / 19.8 ms) to CNN-LSTM to be adopted to edge-deployed IED applications without GPU acceleration in the substation edge environment which may not be economically feasible. Through exhaustive tests clearly showing the number of problems CNN-LSTM can overcome, this paper concludes that CNN-LSTM should be adopted as the main protection algorithm for high-speed transmission protection applications and XGBoost for distribution-level protection devices that can be deployed.

6. Conclusion

We have provided an extensive empirical validation process for machine learning-driven automated power system protection schemes using a heterogeneous 32,550-sample, multi-source dataset, covering eight fault types at voltage levels from 11 kV to 400 kV. The CNN-LSTM hybrid architecture trained on a feature set comprising DWT wavelet energy coefficients with symmetrical component quantities proposed here attained a fault detection accuracy of 98.2±0.4%, F1-score of 0.979±0.002, and mean detection latency of 12.6 ms, surpassing all six benchmarked prior works both in accuracy and speed (ANOVA: $p=0.003774$, pairwise Welch's t-tests: $\alpha=0.05$). The study also established the impact of the diversity of datasets, by ensuring integrating PGCIL actual recordings of measurements with the simulation data itself, which ultimately reduced the possibly higher accuracy by 2.5–4.0 percentage points which has a great significance in terms of methodology to evaluate improvement proposals in the existing protection schemes. XGBoost (96.7% / 19.8 ms) is also an alternative for deployment contexts in which GPU acceleration is not available. Future research tasks involve online adaptation with streaming PMU data, multi-terminal line protection generalized to the case of HVDC systems, as well as adversarial robust test under cyber physical attack scenarios.

References

[1] D. V. Coury and D. C. Jorge, "Artificial neural network approach to distance protection of transmission lines," IEEE Trans. Power Del., vol. 13, no. 1, pp. 102–108, Jan. 1998.

[2] D. Thukaram, H. P. Khincha, and H. P. Vijaynarasimha, "Artificial neural network and support vector machine approach for locating faults in radial distribution systems," IEEE Trans. Power Del., vol. 20, no. 2, pp. 710–721, Apr. 2005.

[3] P. Kumar, A. Rai, and S. Gupta, "Fault classification in transmission lines using ANN and FFT-based features," Int. J. Electr. Power Energy Syst., vol. 108, pp. 215–224, Jun. 2019.

[4] P. Jafarian and M. Sanaye-Pasand, "A wavelet-based multiresolution analysis approach to adaptive single-pole autoreclosure of transmission lines," IEEE Trans. Power Del., vol. 26, no. 4, pp. 2215–2223, Oct. 2011.

[5] Z. Moravej, M. Pazoki, and M. Khederzadeh, "New pattern-recognition method for fault analysis in transmission line with UPFC," IEEE Trans. Power Del., vol. 30, no. 3, pp. 1231–1242, Jun. 2015.

[6] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," IEEE Trans. Smart Grid, vol. 5, no. 1, pp. 51–59, Jan. 2014.

[7] R. Singh, B. Kaur, and M. Sharma, "Support vector machine with wavelet transform for transmission line fault classification," IEEE Access, vol. 8, pp. 145234–145245, 2020.

[8] P. Rai, S. Gupta, and R. Tripathi, "XGBoost-based fault detection and classification in power systems," IET Gener. Transm. Distrib., vol. 11, no. 12, pp. 3094–3102, Sep. 2017.

[9] M. Guo, J. Yang, and L. Chen, "Deep-learning-based fault classification using time-frequency images from power grid voltage signals," IEEE Trans. Ind. Electron., vol. 65, no. 6, pp. 4820–4830, Jun. 2018.

[10] Y. Zhang, C. Wang, and H. Li, "LSTM-based protection scheme for transmission systems with renewable energy integration," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3207–3217, Jul. 2020.

[11] X. Li, J. Huang, and K. Zhang, "Deep learning approach for fault detection and type classification in transmission lines using PMU measurements," Electr. Power Syst. Res., vol. 196, p. 107228, Jul. 2021.

[12] A. Koley, A. Tripathy, and B. Das, "Decision tree and PCA-based fault classification in six-phase transmission systems," Int. J. Electr. Power Energy Syst., vol. 120, p. 106021, Sep. 2020.

[13] S. R. Fahim, Y. Sarker, S. K. Sarker, M. R. I. Sheikh, and S. K. Das, "Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification," Electr. Power Syst. Res., vol. 187, p. 106437, Oct. 2020.

[14] S. R. Fahim, M. Sarker, S. K. Sarker, P. Das, and M. Abido, "CNN-LSTM hybrid model for high-speed fault detection in transmission networks," IEEE Trans. Power Del., vol. 37, no. 3, pp. 1844–1856, Jun. 2022.

[15] W. Chen, O. P. Malik, X. Yin, D. Chen, and Z. Zhang, "Study of wavelet-based ultra high speed directional transmission line protection," IEEE Trans. Power Del., vol. 18, no. 4, pp. 1134–1139, Oct. 2003.

[16] A. A. Yusuff, A. A. Jimoh, and J. L. Munda, "Fault location in transmission lines based on stationary wavelet transform, determinant function feature and support vector regression," Electr. Power Syst. Res., vol. 110, pp. 73–83, May 2014.

[17] K. Mishra, P. Mishra, and G. Panda, "Random forest with discrete wavelet transform for fault detection and classification in transmission systems," IEEE Syst. J., vol. 15, no. 2, pp. 2463–2473, Jun. 2021.

[18] R. Aggarwal, Y. Aslan, and A. Johns, "An interactive approach to fault diagnosis in power distribution systems using artificial neural networks," in Proc. IEE Second Int. Conf. Adv. Power Syst. Control, Oper. Manage., 1993, pp. 702–708.

[19] B. Kasztenny, I. Voloh, and C. G. Jones, "Detection of incipient faults in underground medium voltage cables," in Proc. 33rd Annu. Western Protective Relay Conf., Oct. 2006, pp. 1–19.

[20] J. Liang, S. Elangovan, and J. B. X. Devotta, "A wavelet multiresolution analysis approach to fault detection and classification in transmission lines," Int. J. Electr. Power Energy Syst., vol. 20, no. 5, pp. 327–332, Jun. 1998.

[21] P. S. Bhowmik, P. Purkait, and K. Bhattacharya, "A novel wavelet transform-aided neural network-based transmission line fault analysis method," Int. J. Electr. Power Energy Syst., vol. 31, no. 5, pp. 213–219, Jun. 2009.

[22] Y. Liu, W. Xie, R. Cao, and Q. Li, "LSTM network combined with empirical mode decomposition for power system fault classification," IEEE Trans. Power Syst., vol. 37, no. 1, pp. 567–578, Jan. 2022.

[23] M. A. Elsadd, N. I. Elkalashy, T. A. Kawady, A.-M. I. Taalab, and M. Lehtonen, "Developed ground distance relay for protecting transmission lines including mid-point connected wind turbine," Int. Trans. Electr. Energy Syst., vol. 28, no. 5, p. e2536, May 2018.

[24] T. S. Sidhu, D. S. Baltazar, R. M. Palomino, and M. S. Sachdev, "A new approach for calculating zone-2 setting of distance relays and its use in an adaptive protection system,"

IEEE Trans. Power Del., vol. 19, no. 1, pp. 70–77, Jan. 2004.

[25] A. G. Phadke and J. S. Thorp, Computer Relaying for Power Systems, 2nd ed. Chichester, U.K.: Wiley, 2009.

[26] X. Chen, R. Zhang, Y. Liu, H. Guo, and T. Li, "Transformer network with multi-head self-attention for real-time fault classification in smart grids," IEEE Trans. Smart Grid, vol. 14, no. 3, pp. 2209–2221, May 2023.

[27] A. H. Osman and O. P. Malik, "Transmission line distance protection based on voltage and current gradients," IEEE Trans. Power Del., vol. 19, no. 3, pp. 1242–1250, Jul. 2004.

[28] IEEE Standard for Relays and Relay Systems Associated with Electric Power Apparatus, IEEE Std C37.90-2005, IEEE, New York, NY, USA, 2005.

[29] IEC 60255-151: Measuring Relays and Protection Equipment—Functional Requirements for Over/Under Current Protection, International Electrotechnical Commission, Geneva, Switzerland, 2009.

[30] M. Kezunovic, "Smart fault location for smart grids," IEEE Trans. Smart Grid, vol. 2, no. 1, pp. 11–22, Mar. 2011.