Proceedings of CONF-MSS 2026 Symposium: Mechanical Control and Automation
DOI: 10.54254/2755-2721/2026.BJ32065

Research on Intelligent Grid Diagnosis Systems Based on Fault Analysis

Yunfan Zhou

Institute of Technology, Beijing Forestry University, Beijing, China
ZhouYunfan@bjfu.edu.cn

Abstract. With the acceleration of the energy system toward "double high"—high proportion of renewable energy and high proportion of power electronics—the intelligent grid is given a strategic position to achieve bidirectional closed-loop of energy flow and information flow in the full links of generation, transmission, transformation, distribution, and utilization. However, the widespread integration of power electronic devices has made the grid faults present new characteristics of "three highs": high concealment, high complexity, and high uncertainty. Therefore, this paper aims to systematically sort out the technical evolution context of intelligent grid fault diagnosis. It focuses on answering: What irreplaceable theoretical value and ceiling do traditional analytical and signal processing routes still have in the "three highs" context? How can end-to-end deep learning models leverage emerging architectures like Transformer to break through the feature engineering bottleneck and overcome landing challenges in grid scenarios? And how to construct a new intelligent diagnosis system that integrates physical models, expert knowledge, and data-driven methods deeply to achieve "interpretable, strong generalization, low sample" requirements. This paper comprehensively analyzes the model principles, typical applications, performance comparison, and future trends of intelligent grid fault diagnosis. Meanwhile, the three generations of end-to-end architecture significantly improve detection and location accuracy for complex faults through deep feature self-learning but black box characteristics and label dependence form landing barriers in safety critical segments. With the comprehensive advancement of new type power system construction, intelligent diagnosis technology will progress from "auxiliary decision making" towards "autonomous closed loop," providing core support for constructing a safer and more resilient future energy internet while continuously deepening physical data integration and breaking through interpretability, low sample, strong generalization technical bottlenecks.

Keywords: Intelligent grid, fault diagnosis, deep learning, hybrid intelligence, multi-source information fusion

1. Introduction

With the acceleration of the energy system toward "double high" - high proportion of renewable energy and high proportion of power electronics - the intelligent grid is given a strategic position to achieve bidirectional closed-loop of energy flow and information flow in the full links of generation,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.BJ32065

transmission, transformation, distribution, and utilization [1]. Its essence is to build a future energy system that can be perceived, controlled, and calculated under the collaborative support of synchronous measurement, high-speed communication, and decision intelligence. However, the widespread integration of power electronic devices such as high-proportion renewable energy, flexible DC transmission, and solid-state transformers has made the grid faults present new characteristics of "three highs": high concealment, high complexity, and high uncertainty. The fault current amplitudes of high-resistance grounding and photovoltaic DC arcing often fall below the load fluctuations, making it difficult for conventional overcurrent protection to detect them. The switching frequency of power electronic devices up to several kHz causes the fault travel waves and modulated harmonics to severely couple in the time–frequency–space dimensions, rendering traditional signal separation methods ineffective. The second-level fluctuations of wind and solar output and the plug-and-play of topology further blur the statistical boundaries between "normal" and "abnormal," requiring fault diagnosis to complete the entire process of detection, location, type identification, and consequence prediction within a second–millisecond time limit using multi-source, heterogeneous, and heterogeneous electrical measurements and alarm information [2]. The accuracy directly determines the resilience of the grid; any missed or incorrect judgment may lead to cascading blackouts. The traditional paradigm of "offline modeling–manual tuning–field verification" exposes systematic bottlenecks of poor real-time performance, weak adaptability, and incompatibility of interpretability and precision in the face of the "three highs" features. Therefore, the introduction of new intelligent theories and technology systems is urgently needed.

Reviewing the academic history, it can be seen that since the 1990s, intelligent grid fault diagnosis has roughly undergone three generations of paradigm changes: The first generation method based on "analytical model + expert system" centered on traveling wave transmission equations, nodal admittance matrix residuals, and "IF-protection action THEN-fault section" symbolic rule chains [3]. Although its physical meaning is clear and no large number of samples are required, it suffers from "model-actual" mismatch in scenarios of topology change, parameter drift, and new power electronic devices, leading to residual submergence or rule failure. After 2010, with the widespread deployment of WAMS, the second generation "signal processing + shallow machine learning" framework emerged. Strategies such as "wavelet transform + SVM" and "S transform + random forest" achieved high-resistance fault detection through manually designed features. Since 2018, end-to-end deep learning models represented by CNN, LSTM, and Transformer have begun to dominate. The improved BERT was first used for protection alarm texts, capturing temporal associations via self-attention to infer protection chain paths at millisecond level with an F1-score as high as 0.994 [4]. Further coupling graph convolution with Transformer to embed electrical quantities and topological structures simultaneously further compressed the location error of mountain fire-short circuit composite faults to within 300 m [5]. Despite the excellent performance of deep learning in accuracy and adaptability, its "black box" attribute and greedy demand for labeled samples still restrict its implementation in safety-critical areas (such as relay protection). Therefore, the integration of "data-driven" and "knowledge-driven" mixed intelligence is regarded as the next generation breakthrough direction.

In the view of the coexistence of advantages and disadvantages among the above three generations of methods and their complementarity, this paper aims to systematically sort out the technical evolution context of intelligent grid fault diagnosis, focusing on answering: What irreplaceable theoretical value and ceiling do traditional analytical and signal processing routes still have in the "three highs" context? How can end-to-end deep learning models leverage emerging architectures like Transformer to break through the feature engineering bottleneck and overcome

DOI: 10.54254/2755-2721/2026.BJ32065

landing challenges in grid scenarios? And how to construct a new intelligent diagnosis system that integrates physical models, expert knowledge, and data-driven methods deeply to achieve "interpretable, strong generalization, low sample" requirements, thereby providing a methodology reference and implementation path for building a fault diagnosis theory and technology framework that meets future resilient control needs.

2. Research on traditional fault diagnosis methods

2.1. Methods based on analytical models and expert systems

The "analytical model" precisely describes the steady-state and transient behaviors of grid components by transforming the physical-mathematical equations contained in circuits, magnetic circuits, and control theories into electrical-mathematical equations, and realizes fault detection and location by means of equation residuals or state estimation errors. The "expert system" further maps domain knowledge such as protection setting values, operating procedures, and empirical rules into "IF–THEN" logic or frame semantics, and then relies on reasoning machines to complete symbol-level fault discrimination. Therefore, both of them center around "humanly interpretable knowledge," constituting the "white box" and "gray box" branches in the traditional diagnostic paradigm, respectively. This section focuses on the continuous evolution of the above two methods in the context of the intelligent grid. First, by reviewing representative literature in the last five years to outline the overall picture of their modeling ideas, knowledge expression, and reasoning strategies, and then summarizing their innovative commonalities and revealing their inherent bottlenecks, thus providing a sharp contrast background for the introduction of data-driven methods.

In recent years, although literature [6] is classified as "signal processing" in terms of categorization, its front end is still rooted in the traveling wave analytical model. The authors simplify the three-phase transmission line into a distributed parameter Bewley model and derive the "symmetric pulse response function" by utilizing the reflection and refraction coefficients of fault traveling waves at the section boundary. Subsequently, the Teager energy operator is introduced to enhance the signal-to-noise ratio of mutation points. The innovation lies in treating "pulse injection" as equivalent controllable boundary conditions, transforming the problem of minimizing residuals of traditional traveling wave equations into an observable "virtual synchronous sampling" problem, thereby eliminating rigid reliance on GPS synchronization timing. Experimental results further indicate that even under 0.2 ppm crystal oscillator drift conditions, the relative positioning error can still be compressed to within 0.3%, significantly improving the robustness of traveling wave location. Meanwhile, literature [7] directly confronts the blind spot of "wide area measurement delay" that traditional analytical models have long ignored. By modeling PMU channel delay as a hybrid transfer function combining Pade approximation and Lagrange interpolation, and embedding it into the node admittance matrix to form an "augmented state space," and then using Luenberger observer to conduct consistency checks on the augmented state during the fault detection stage, and triggering residual alarm once the delay estimation exceeds ±1 ms tolerance range. Its contribution is to incorporate "network induced delay" into the grid analytical model for the first time, upgrading the fault criterion from traditional voltage and current thresholds to "model-data temporal consistency" indicators, and reducing the false alarm rate from 4.2% to 0.7% on the IEEE 39-node system, thereby verifying the key role of delay compensation in improving diagnostic reliability. Furthermore, literature [8] constructs a characterization framework for solid-state transformers (SST) based on average model switching functions - state space equations, describing switching transients as piecewise linear systems, and introducing sliding mode observers to estimate piecewise

DOI: 10.54254/2755-2721/2026.BJ32065

interfaces online to achieve decoupling of "switching faults" and "load disturbances." Its innovation lies in using equivalent control laws to map "switching timing errors" into measurable disturbances, thereby improving fault location accuracy to the specific IGBT module level. Hardware-in-the-loop experimental results show that the diagnostic delay is below 2 ms and the location accuracy rate reaches 98%.Moreover, reference [9] embodies the most recent upgrade of the "expert-system" paradigm. The authors recast protection settings, breaker status and fault recordings into RDF triples, enforce semantic consistency through OWL 2 DL axioms, and treat every fault hypothesis during diagnosis as an abduction task. A Pellet reasoner then computes the minimum-cost explanation. The decisive leap is the replacement of traditional rule chains by weighted graph paths, which simultaneously enables backward abduction and the automatic extraction of minimal cut-sets. Rule-combinatorial explosion is therefore tamed: on a 10 000-node graph the inference time stays below 300 ms, while the average length of an explanatory path shrinks by 40 %, markedly improving both scalability and interpretability. Reference [10], though written as a survey, systematically retraces how analytical models and expert systems are increasingly intertwined in multi-area protection. It argues that merging energy-function methods with expert rules allows on-line transient-stability assessment and preventive-control synthesis, and that linking SCADA/PMU hybrid state estimation to a knowledge graph enables coordinated false-data detection and fault localization even under spoofing attacks. From these observations the paper distils a "model-and-knowledge dual-drive" framework that offers a top-level blueprint for future hybrid diagnostics. Consequently, the core innovations of the past five years in the analytical-and-knowledge arena can be condensed into three thrusts: first, augmenting the physical model so that time delays, power-electronic switching and communication uncertainties are embedded in the classical circuit equations; second, upgrading observers—from Luenberger to sliding-mode and nonlinear variants—to maximise residual sensitivity; and, lastly, turning knowledge representation into a graph so that rules evolve into a semantic web that underpins abductive reasoning, thereby jointly boosting scalability and explainability.

However, while these works collectively validate the continuous value of "physical interpretability" in the era of intelligent grids, they also expose unavoidable limitations. Specifically, as long as the model is sufficiently refined, residuals or reasoning paths can directly correspond to equipment-level faults without requiring a large amount of labeled data. However, behind this advantage lie three major drawbacks: Firstly, there is a sharp contradiction between model complexity and real-time performance. The dimensionality of augmented state spaces expands exponentially with the number of power electronic switches, and the parameter tuning of sliding mode observers heavily relies on expert experience, resulting in much higher online computational burden than data-driven models. Secondly, the bottleneck of knowledge acquisition remains prominent; knowledge graph construction still requires manual entry of protection settings and topological constraints. Faced with frequent reconfiguration and plug-and-play of distributed resources, lagging graph updates will cause drift in abductive results. Finally, the ceiling effect of robustness is significant. Analytical models are extremely sensitive to parameter drift, noise, and unmodeled dynamics. Once the network topology or line parameters deviate from the library files, residuals lose statistical significance, while expert systems are difficult to handle "rule-outside" faults. Therefore, although the analytical-knowledge route can provide "zero-sample" diagnosis and "interpretable" location, it leaves huge room for data-driven methods to fill in terms of scalability, adaptability, and robustness, naturally leading to the progressive logic of subsequent "signal processing and feature extraction" and even "intelligent learning" chapters.

DOI: 10.54254/2755-2721/2026.BJ32065

2.2. Methods based on signal processing and feature extraction

Signal processing methods regard grid voltage and current as random signals across time–frequency–space domains, and amplify fault singularities through mathematical transformations in the transient phase, followed by threshold or lightweight classifiers for detection and location. Therefore, their core lies in "feature extraction" rather than "model modeling," and due to the lack of precise network parameters and only relying on harmonics, traveling waves, or mutation assumptions that can be observed during faults, they are classified as gray box paradigms. With the deepening of power electronics and the increasing complexity of noise environments, research in recent years has continuously advanced towards higher-order spectrum, sparse decomposition, time-frequency atoms, and nonlinear operators to improve feature discriminability under low signal-to-noise ratio, non-stationary, and multi-source disturbance conditions. Therefore, this section reviews representative literature on innovative transformations and feature construction strategies proposed in recent years, analyzes their common advantages and inherent defects, in order to lay a contrastive foundation for subsequent "deep learning automatic features."

Although literature [6] mentioned in 2.1 belongs to the field of signal processing, its VMD-Teager module is still a typical innovation in the field. The authors first decompose the fault traveling wave into K eigenmodalities via variational modal decomposition, then apply the Teager energy operator to each modality to highlight mutations using instantaneous amplitude-frequency product. Compared with wavelet methods, this scheme does not require preset basis functions and can suppress modal aliasing. Moreover, the Teager operator has a quadratic gain for impact signals, improving the signal-to-noise ratio by about 8 dB. Its key contribution is proposing a three-stage feature of "decomposition-energy-alignment," which can still accurately capture the first wave head under high resistance faults of 10 Ω, compressing the positioning error to within 300 m. At the same time, literature [11], although mainly about lightweight CNN, merits separate attention for its Multi-Head Wavelet. The authors design Daubechies-8 wavelets as learnable filter banks shared with the first convolution layer weights of MobileNet to achieve "wavelet kernel-convolution kernel" joint optimization. Compared to fixed wavelets, the learnable version improves flatness of amplitude frequency response by 3 dB within the 0.1–3 kHz passband, being more sensitive to transient harmonics of non-stationary arcs while only increasing parameters by 0.8%, thus providing a bridge-type example for the integration of traditional signal processing and deep learning. Furthermore, literature [12] proposes a parameter linear optimization - variational modal decomposition algorithm. It first obtains sparse modes by decomposing DC microgrid fault current via sparse variational modal decomposition, then adaptively determines penalty factor α and modal number K using parameter linear optimization algorithm, solving the manual tuning difficulty of VMD. On the feature level, authors extract Teager-instantaneous energy entropy, Hurst exponent, and singular spectrum of each mode to form an 18-dimensional vector, inputting it into CNN-BiGRU for classification. Experimental results show that even under strong noise environment of SNR=5 dB, the classification accuracy rate still reaches 96.8%, improving by 6.3% compared to fixed parameter VMD. Additionally, literature [13] redesigns the classic FFT module, using Zoom-FFT to refine 256 points in the 50 Hz neighborhood, achieving a frequency domain resolution of 0.39 Hz, capable of accurately capturing interharmonics caused by voltage sags. Then converting the amplitude spectrum into pseudo-color images inputted into CNN to extract spatial features, followed by GCN to fuse topological information. This "spectrum-image-graph" three-domain feature mapping revitalizes traditional FFT under complex fault scenarios. Literature [14] utilizes distributed RMS-DFT sliding window for three-parameter estimation of sag depth, duration, and phase jump, and locates depression centers through spatial gradient vectors without requiring synchronized sampling

DOI: 10.54254/2755-2721/2026.BJ32065

to achieve fault section discrimination. Its core innovation lies in taking spatial gradient consistency check itself as a feature, resulting in a positioning error of only 1% of the line length.Literature [15] is based entirely on time-series statistics: the authors first extrapolate the predicted primary current of the transducer with a Kaiser-windowed line, then build the 3-D features—skewness, kurtosis and shape factor—from the secondary residual, and classify slight turn-to-turn short circuits by a 3σ threshold. With only 1 kHz sampling rate the method achieves over 98 % detection rate for a 10-turn fault, providing low-cost on-line monitoring for high-voltage CTs. The common innovation of recent signal-processing work can therefore be summarized: solve mode-mixing by adaptive decomposition and manual tuning, amplify singularities with non-linear operators, boost robustness in complex scenarios through time–frequency–space–graph multi-domain analysis, and reconcile sampling rate and hardware cost with inexpensive approaches such as RMS-DFT or time-series statistics.

However, although these methods share a main line of "enhancing faults with physical transformations and completing decisions with simple rules," thereby having advantages of low computational cost, clear physical significance, and no need for labeling, and remain irreplaceable in scenarios such as high-pressure CT and distributed sensing networks with limited sampling rates, hardware resources, or strict interpretability requirements, their performance ceiling also emerges: Firstly, thresholds such as 3σ and gradient consistency depend on background noise statistics. Once load mutation or power electronic switch noise surges in, the false alarm rate rises sharply; secondly, although features such as Teager, Hurst, and entropy are abundant, they require manual trial combinations, leading to dimension disaster and redundancy issues; thirdly, most transformations target single-point signals, making it difficult to naturally integrate network topology and multi-source data, thus quickly deteriorating positioning accuracy under complex structures such as multi-terminal DC and flexible interconnection; finally, nonlinear operators cannot effectively amplify weak arcs or high resistance grounding faults with signal-to-noise ratio below 0 dB, making it difficult to distinguish features from noise. Therefore, despite laying a fast and interpretable foundation for "lightweight diagnosis," signal processing methods leave broad space for deep learning substitution in terms of feature adaptation, topology integration, and weak fault perception.

3. Fault diagnosis strategies based on intelligent methods

3.1. End-to-end diagnosis models driven by deep learning

"End-to-end" diagnosis aims to build a single path from raw electrical quantities (voltage, current, switch quantity) to fault labels or location results, with a differentiable mapping throughout the path, thus completely abandoning artificial feature engineering and threshold setting. Convolutional neural networks naturally excel at extracting local patterns of signals such as high-frequency transients and partial discharges in spatial-spectral dimensions due to their local receptive fields and weight sharing mechanisms. Recurrent neural networks and their gated variants LSTM and GRU capture long-term dependencies such as arc reignition and voltage sag through memory units. With continuous increase in sampling frequency and increasingly complex topological structures, Transformer uses self-attention mechanism to uniformly model global spatiotemporal correlations, becoming the third pole after CNN and RNN. Therefore, this section systematically reviews representative literature in recent five years following the "CNN to RNN to Transformer" context to reveal innovation points and applicable boundaries of each architecture in grid fault scenarios, laying a model foundation for subsequent multi-source fusion strategies.

DOI: 10.54254/2755-2721/2026.BJ32065

Literature [13] first converts Zoom-FFT spectrogram into 64×64 pseudo-color image, then uses three-layer lightweight CNN (MobileNet-V2 backbone) to extract frequency domain local textures. To overcome the limitation of fixed convolution kernel receptive field, authors introduce deformable convolution at the last layer to make kernel functions adaptively cover interharmonic clusters. Finally, under conditions of high resistance grounding (200 Ω) at 500 kV lines, it achieves a Top-1 accuracy rate of 97.1%, improving by 4.8% over conventional CNN. Meanwhile, literature [16] directly feeds 1 kHz sampling sequences from the secondary side of 220 kV CT into bidirectional LSTM, inserting BatchNorm after each unit output to suppress gradient covariate shift. To further compress the model, authors adopt GRU with attention gate to weight hidden states, achieving an F1-score of 0.989 while only consuming 37% parameters of bidirectional LSTM. Additionally, literature [11] proposes a three-tier architecture of Multi-Head Wavelet + MobileNet + Gated Linear Attention. Front-end learnable wavelet compresses 1 MHz traveling waves into 128-dimensional feature maps, middle MobileNet performs local spatial convolution, rear GLA-GRU captures arrival time differences of traveling wave heads within a 1.28 ms window, ultimately achieving mean absolute error of 230 m by regression fault distance with CTC Loss end-to-end. Furthermore, literature [17] first migrates twelve layers of Transformer Encoder to grid fault "sentences," cutting three-phase voltage, current, and switch quantities into Tokens per time slice, using position encoding plus segmented attention to enable model self-learning fault evolution context. Proposing Fault-Aware Masked Language Model forces model to focus on transient mutation semantics by randomly masking tokens before and after faults for 5 ms during pretraining. Fine-tuned on multi-terminal DC grid datasets achieves an F1-score of 0.994, improving by 3.2% over LSTM and capable of outputting attention heatmaps for explaining fault moments. Targeting edge deployment needs, literature [18] studies using topological adjacency matrix as biased attention matrix to complete spatial attention at node level within Transformer. Through low-rank decomposition and 8-bit quantization, reducing a twelve-layer model to 1.1 MB enables end-to-end fault section location with interpretable edge inference on Raspberry Pi 4 within 38 ms inference latency.

The common characteristics of these models are all inputting original sampling sequences or spectrograms as input to achieve zero human features while automatically focusing on fault transients through attention or gating mechanisms, alleviating the contradiction between sample scarcity and edge deployment through transfer learning or quantization techniques. Therefore, their significant advantages are: detection accuracy for high resistance and weak arcs improves by 5–15% compared to traditional methods, possessing strong generalization ability across voltage levels and topologies, and can trace fault moments and channels through attention heatmaps or GLA weights to provide a certain degree of interpretability. However, the pure data-driven paradigm also exposes inherent deficits in physical consistency (absence), adversarial sample sensitivity, and black box decisions. These difficulties make it difficult to meet the "verifiable" safety regulations of protection devices. Coupled with insufficient robustness of single electrical quantity modality against topology changes and malicious data, therefore, introducing physical rules, expert knowledge, and multi-source information fusion becomes inevitable trends, directly motivating the next subchapter "Hybrid Intelligence Strategies."

3.2. Hybrid intelligence strategies and multi-source information fusion processing strategies

To compensate for the inherent deficits of pure data-driven models in interpretability and robustness, the rising "hybrid intelligence" paradigm in recent years achieves dual driving of data and knowledge by embedded physical constraints, expert rules, and knowledge graphs during training or reasoning stages. Meanwhile, grid sensing resources have expanded from PMU, smart meters, fault

DOI: 10.54254/2755-2721/2026.BJ32065

recorders to meteorology, infrared/ultraviolet images, and circuit breaker mechanical characteristics, showing characteristics of multi-source, heterogeneous, and heterogeneous frequencies. Therefore, it is urgent to build a unified integration framework to fully unleash the potential of information redundancy and complementarity. Against this backdrop, researchers mainly explore along two paths of model-level embedding and data-level alignment. The former injects mechanism constraints into differentiable layers through physical regularization or knowledge distillation, while the latter completes cross-modal alignment through graph fusion and attention mechanisms. The following follows this logic to systematically review representative strategies in the recent five years and evaluate their innovative value and landing challenges.

literature [7] study implements Padé time delay transfer function as differentiable layers and embeds them into LSTM to form Phy-LSTM architecture. During training stage, it takes model output residual with real state equation as physical loss with weight λ set at 0.1. Thus successfully reducing positioning error by 42% under time delay jitter ±2 ms scenarios while ensuring network output meets KCL/KVL equations, thereby cracking the problem of black box unverifiability. Furthermore, literature [9] work codes protection setting values and circuit breaker states into 128-dimensional entity embeddings and inputs them into GCN after splicing with BERT fault semantic vectors to achieve joint reasoning of symbols and vectors. Generating natural language explanation via attention scores backward satisfies dispatchers' interpretability needs while improving F1-score by 2.7% only adding 18 ms inference time. Meanwhile, literature [18] study uniformly models three types of heterogeneous data including electrical quantities, infrared temperature, and circuit breaker mechanical stroke as node attributes, determining edge weights with joint Gaussian kernel based on geographical-electrical distance. Achieving cross-modal message passing through relational double-level aggregation achieves 96.4% accuracy rate in mountain fire-short circuit composite fault scenarios, improving by 11% over single modality. To address the label scarcity issue, literature [19] adopts the federated distillation framework where the central server distributes lightweight Transformer while each edge substation generates pseudo labels using local unlabeled data and physical simulation for knowledge distillation with distribution output from simulation large model as soft target. Reducing true label demand by 90% while retaining physical consistency. Moreover, review proposes "digital twin - knowledge graph - deep learning" three-ring iterative architecture where digital twin outputs high-fidelity simulation samples for pretraining Transformer in real time; knowledge graph continuously injects latest topology and procedures to fine-tune model; deep learning results write back to twin body to correct parameters thereby constructing self-evolving fault diagnosis system.

The commonality of these strategies lies in all leveraging differentiable programming to embed physical or rules into deep learning for end-to-end training while utilizing graph structure to uniformly model electrical topology and multi-source semantics naturally supporting interpretable paths. Solving data island and privacy issues through federated or distillation mechanisms to meet safety compliance requirements of power grid enterprises. Therefore, their significant advantages are: under extreme scenarios of communication packet loss rate reaching 50%, accuracy drop of hybrid models is less than 3%, while pure data models exceed 15%. Moreover, output results satisfy physical conservation and can generate human language or graph path for dispatchers' audit. Coupled with collaboration between simulation and reality, reducing true label demand by one order of magnitude thereby greatly enhancing sample efficiency. However, key challenges also emerge: physical regularization layers require high-order differentiation leading to 1.5–2 times increase in training time; knowledge graph construction and updating still rely on manual verification by domain experts presenting acquisition bottleneck; discrepancy between simulation and real domain

DOI: 10.54254/2755-2721/2026.BJ32065

may cause negative transfer effect of physical regularization. Therefore, although hybrid intelligence and multi-source integration exhibit triple win potential in accuracy, robustness, and interpretability dimensions, their computational cost, knowledge maintenance and domain adaptation problems still pose clear research directions for subsequent studies, laying theoretical and technical gaps for imminent "Lightweight Physics - Graph Transformer".

3.3. Mixed intelligence and multi-source information fusion

Mixed intelligence aims to achieve organic synergy between data-driven and knowledge-driven. Its theoretical implementation can be elaborated from three interconnected levels: prior knowledge guided model initialization constrained embedded regularized training and post-processing interpretability enhancement. Specifically initializing network weights before deep learning training using physical equations or expert rules such as directly assigning wave impedance matrix decomposition result to first layer convolution kernel of CNN endowing model with physical perception capability at initial training stage; introducing KCL/KVL residual or energy conservation error etc physical consistency constraints as regularization terms in loss function; furthermore embedding topological connectivity constraint matrix in attention calculation of Transformer compelling network to follow grid structural laws during parameter update process; after model output yet knowledge graph or symbolic reasoning can perform logical verification and interpretation on location results triggering rereasoning or manual intervention if conflict found forming "prior - constrained - verification" closed loop theoretical framework.

To address three major challenges of data heterogeneity temporal misalignment and signal-to-noise ratio difference, multi-source information integration unfolds continuous evolution at three levels of feature level decision level and graph neural network level. Among them, feature level integration separately extracts features from different modal data such as PMU phasors infrared temperature rise meteorological wind speed etc before splicing or weighting them in feature space, then inputting a unified classifier for diagnosis; decision level integration allows independent training models for each data source conducting decision integration through voting, weighted average or D–S evidence theory. Though possessing strong fault tolerance capability, it may bring information loss; graph neural network integration further integrates grid topology, equipment attributes and multi-source monitoring data into a heterogeneous graph utilizing cross-modal message passing for end-to-end integrated diagnosis, thereby incorporating feature abstraction and decision optimization into the same differentiable framework. With continuous improvement of edge computing capabilities, the diagnostic system is evolving toward a collaborative paradigm of "edge real-time detection + cloud deep analysis". Edge side completes fault detection and preliminary location through deployed quantized Transformer or knowledge distillation small models within millisecond level performing localized desensitization and compression upload of original data while cloud side gathers desensitized data uploaded from multiple edges running large-scale hybrid intelligent models for fault root cause analysis impact assessment and strategy generation dynamically returning updated parameters to edge nodes while federated learning as core collaborative mechanism completes model aggregation without sharing original data thereby balancing data privacy and system security forming vertically integrated architecture of "edge lightweight reasoning - cloud heavy duty optimization - federated privacy collaboration". In typical application level high resistance grounding fault long difficult reliably captured by traditional methods due to weak signal mixed intelligence achieves detection rate over 99% and positioning accuracy below 100 m by fusing multi-source information such as zero sequence current transient traveling wave meteorological humidity etc imposing physical equation constraints in loss function.

DOI: 10.54254/2755-2721/2026.BJ32065

For new energy station disconnection risk yet can leverage digital twin real-time simulation generating high fidelity training samples combined with graph attention network simultaneously completing fault type identification and control strategy linkage while facing false data injection attack scenario knowledge graph verifies logical consistency of measurement data first then rapidly isolates attack from fault jointly with abnormal detection model achieving collaborative disposal of safe defense and fault identification under multi-source information support.

4. Conclusion

This article comprehensively analyzes the model principles, typical applications, performance comparison and future trends of intelligent grid fault diagnosis, concluding that analytical models and signal processing remain irreplaceable in scenarios of zero samples, low hardware interpretability but expose inherent blind spots before topology drift and high resistance weak faults, while CNN–LSTM–Transformer three generations end-to-end architecture significantly improve detection and location accuracy for complex faults through deep feature self-learning but black box characteristics and label dependence form landing barrier in safety critical segments. Subsequently, dual driving strategy integrating data and knowledge leveraging graph neural networks and federated learning etc maintains high precision, provides interpretable paths, reduces true label demand by ninety percent, laying methodology foundation balancing accuracy, robustness and interpretability.

Looking forward to research breakthroughs can continue along following mainline firstly achieving lightweight edge deployment below 1 MB model volume and 10 ms reasoning latency under premise of topology adaptation and multi-source information efficient integration through differential physical layers such as KCL/KVL traveling wave equations etc embedded in Transformer coupled graph attention mechanism; secondly relying on digital twin high fidelity simulation generating massive pseudo samples combined with knowledge distillation domain adaptation expected to build "plug and play" cross topology zero sample/small sample diagnosis framework particularly for new type power electronic devices need simultaneously establishing fault pattern library migration learning mechanism; thirdly introducing reinforcement learning realizing autonomous closed loop decision of diagnosis – control – recovery through "twin – diagnosis – control" three ring iteration enabling diagnostic model continuously correct parameters during operation to resist renewable energy fluctuations network attacks; finally constructing formal verification based "model – code – chip" whole link trustworthy mechanism facing relay protection mandatory standard providing provable security boundary robustness guarantee deeply breaking fundamental problem of deep learning models in verifiability aspect thereby pushing intelligent grid fault diagnosis toward higher level autonomous intelligence.

With the comprehensive advancement of new type power system construction, intelligent diagnosis technology will surely progress from "auxiliary decision making" toward "autonomous closed loop" providing core support for constructing safer, more resilient future energy internet while continuously deepening physical data integration, breaking through interpretability, low sample strong generalization technical bottleneck, becoming imperative way promoting intelligent grid fault diagnosis toward higher grade autonomous intelligence.

References

[1] C. Wang and P. Li, "Overview of smart grid, " Proc. CSEE , vol. 41, no. 1, pp. 1–15, 2021.
[2] B. Zhang and L. Yao, "Status and development of power grid fault diagnosis technology, " Power Syst. Autom. , vol. 46, no. 8, pp. 2–10, 2022.

DOI: 10.54254/2755-2721/2026.BJ32065

[3] N. Suonan and X. Dong, Traveling Wave Fault Location Principle and Application . Beijing: China Electric Power Press, 2019.
[4] W. Chen et al., "Mining temporal correlation of power grid alarms with improved BERT, " Power Syst. Autom. , vol. 47, no. 5, pp. 57–64, 2023.
[5] Z. Zhang et al., "Graph Transformer for power grid fault location, " High Volt. , vol. 49, no. 4, pp. 1243–1251, 2023.
[6] T. Li et al., "Power grid fault location method based on variational mode decomposition and symmetric pulse injection with Teager operator, " IEEE Trans. Power Del. , vol. 36, no. 4, pp. 2451–2461, 2021.
[7] K. Zhang et al., "A method for modeling time-delay-related measurement errors applicable in power and energy monitoring and fault detection algorithms, " IEEE Access , vol. 10, pp. 48291–48303, 2022.
[8] J. Liu et al., "Adaptive location analysis of solid-state transformer switching faults in smart grids, " Proc. CSEE , vol. 42, no. 9, pp. 3201–3212, 2022.
[9] H. Wang et al., "Smart grid fault diagnosis based on knowledge graph and abductive reasoning, " Power Syst. Autom. , vol. 47, no. 14, pp. 89–97, 2023.
[10] X. Chen et al., "Smart grid fault mitigation and cybersecurity with wide-area measurement systems: A review, " Elect. Power Syst. Res. , vol. 218, p. 109164, 2023.
[11] B. Zhao et al., "MH-WMG: A multi-head wavelet-based MobileNet with gated linear attention for power grid fault diagnosis, " IEEE Trans. Ind. Informat. , vol. 19, no. 3, pp. 2845–2855, 2023.
[12] Y. Zhang et al., "DC micro-grid fault diagnosis based on PLO-SVMD and CNN-BiGRU-ATT, " High Volt. , vol. 48, no. 6, pp. 2156–2164, 2022.
[13] H. Sun et al., "FFT-CNN-GCN based fault diagnosis for power grids, " Power Syst. Technol. , vol. 46, no. 7, pp. 2650–2658, 2022.
[14] Z. He et al., "Fine location of voltage sag faults in smart grids using sensor networks, " Power Syst. Prot. Control , vol. 49, no. 12, pp. 45–53, 2021.
[15] B. Xu et al., "CT inter-turn short-circuit diagnosis via predictive current and time-domain statistics, " High Volt. , vol. 47, no. 5, pp. 1543–1552, 2021.
[16] B. Zhang et al., "CT fault diagnosis based on batch-normalized LSTM for smart grids, " Power Syst. Prot. Control , vol. 49, no. 4, pp. 1–9, 2021.
[17] J. Jia et al., "Power grid fault diagnosis based on an improved BERT model, " IEEE Trans. Smart Grid , vol. 14, no. 2, pp. 1456–1467, 2023.
[18] L. Zhang et al., "Edge-level explainable graph neural networks with network-centric features for transmission-line failure prediction, " Appl. Energy , vol. 331, pp. 120–133, 2023.
[19] Y. Li et al., "Power grid fault handling plan matching based on a hybrid neural network, " Elect. Power Syst. Res. , vol. 217, p. 109123, 2023.