International Journal of Safety and Security Engineering
Vol. 16, No. 10, October, 2026, pp. 2189-2203
Journal homepage: http://iieta.org/journals/ijsse

Dual-Domain Signal Representation and Recurrent Deep Learning for Electrical Tracking Fault Detection in Medium-Voltage Switchgear

Mustafa Mahmood Yahya[1*], Yaseen Ahmed Mohammed Alsumaidaee[2], Maha Safar Abdulmajed[3]

[1] Department of Technical Mechatronics Engineering, Technical College of Engineering, Al-Kitab University, Kirkuk 36015, Iraq
[2] Department of Computer Science, Al Imam Al Adham University College, Baghdad 36001, Iraq
[3] Department of Laboratory Sciences, College of Pharmacy, University of Tikrit, Tikrit 34001, Iraq

Corresponding Author Email: eng.yassin.ahmed@imamaladham.edu.iq

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijsse.161001

Received: 28 July 2026
Revised: 10 September 2026
Accepted: 16 September 2026
Available online: 31 October 2026

Keywords:
medium-voltage switchgear, Electrical Tracking, gated recurrent unit, surface discharge, ultrasonic condition monitoring, time-domain representation, frequency-domain representation, fault diagnosis

ABSTRACT

Electrical Tracking is a progressive surface-insulation fault that can form a carbonized conductive path, reduce creepage strength, and ultimately promote flashover in medium-voltage (MV) switchgear. This paper presents two gated recurrent unit (GRU)-based computational pipelines for distinguishing Tracking from a heterogeneous Non-Tracking class comprising Normal, Corona, Arcing, and Mechanical records. Field-derived airborne-ultrasonic data previously reported for MV switchgear are processed as separate time-origin and frequency-origin branches. The time-origin branch contains 438 computational records represented by 40 × 128 tensors, whereas the frequency-origin branch contains 160 records represented by 20 × 40 tensors. Under the original computational 70/15/15 split, the program produced test-partition accuracies of 95.45% and 95.83%, with macro F1-scores of 0.930 and 0.947, respectively. Because the Tracking records were bootstrap-resampled with replacement before the random split, identical computational records may occur in more than one subset, creating potential cross-partition leakage. The percentages are therefore reported only as split-specific outputs of the original program and are not interpreted as reliable held-out generalization performance. The time-origin test partition detects all 51 Tracking records but produces three false tracking alarms, while the frequency-origin test partition correctly recognizes all 17 Non-Tracking records and misses one of seven Tracking records. The two branches differ in sample population, preprocessing, tensor dimensions, and dropout; their numerical differences are interpreted descriptively rather than as a controlled causal comparison of signal domains. Leakage-controlled generalization and uncertainty estimation require exact duplicate removal followed by grouped and repeated validation.

1. INTRODUCTION

Switchgear integrates switching, protection, isolation, and control functions throughout electric-power networks. Low-voltage assemblies serve utilization and secondary-distribution circuits, medium-voltage (MV) assemblies occupy the distribution and industrial range, and high-voltage equipment supports transmission and bulk supply; the exact voltage boundaries vary among standards and utilities [1-3]. This work focuses on metal-enclosed MV switchgear, where busbars, circuit breakers, cable terminations, current transformers, and solid-insulation supports operate within compact clearances and under simultaneous electrical, thermal, mechanical, and environmental stress [4-6].

A defect in MV switchgear can interrupt several downstream loads and can expose personnel and adjacent equipment to dielectric failure, fire, or arc-related hazards [7, 8]. Modern condition-based maintenance therefore combines periodic inspection with partial-discharge (PD) sensing, transient-earth-voltage (TEV) measurements, acoustic or ultrasonic examination, thermal observation, and health-index assessment [4, 8, 9]. The sensitivity of each route, however, depends on enclosure geometry, sensor position, electromagnetic interference, background noise, and the operating state of the equipment [10, 11].

Electrical Tracking is a progressive surface-discharge mechanism rather than an instantaneous puncture of the insulation [12, 13]. Moisture, conductive contamination, ageing, local electric-field enhancement, and surface damage can increase leakage current along an insulating path; the resulting Joule heating creates partially dry regions across which repetitive dry-band arcs may develop [14-16]. Continued thermal and chemical attack erodes the material and forms a carbon-rich path, progressively reducing creepage strength and increasing the likelihood of sustained arcing or flashover [2, 17].

Tracking is difficult to identify in service because corona, arcing, mechanical vibration, and normal switching activity can contain overlapping transient and spectral components [18-20]. Airborne-ultrasonic monitoring is attractive because

it is non-contact and can be applied without interrupting service, although the measured waveform remains sensitive to sensor model, distance, gain, cabinet construction, and ambient noise [6, 21, 22]. A practical diagnostic system must therefore separate tracking from healthy operation and from competing abnormal conditions rather than from a single idealized background class [23, 24].

Deep learning reduces dependence on manually selected thresholds by learning discriminative structure directly from waveforms, spectrograms, and phase-resolved patterns [2, 6, 20]. Recent switchgear research includes pre-trained acoustic models, waveform-spectrogram fusion, denoising with optimized neural classifiers, cloud-edge recognition, continuous-wavelet image learning, semi-supervised diagnosis, transformer-based feature extraction, and state-space models [25-28]. These developments improve recognition capability, but they also make it difficult to determine whether a reported gain comes from the signal representation, the classifier capacity, or the validation protocol [29-31].

Recurrent learning is appropriate when diagnostic evidence is represented as an ordered sequence [32-34]. A gated recurrent unit (GRU) controls the contribution of previous states through reset and update gates, while avoiding the separate memory-cell structure of a long short-term memory (LSTM) network [35, 36]. This relative compactness is useful for limited switchgear datasets and for experiments intended to separate the effect of signal representation from the effect of an increasingly elaborate classifier [36-39].

Accordingly, this study evaluates two domain-specific sequence pipelines using a common GRU decision principle [2, 37-39]. Its contributions are: (1) a documented binary reformulation of a field-derived five-condition switchgear database; (2) separate time-origin and frequency-origin sequence pipelines; (3) explicit training, validation, and test-partition outputs under the original computational split, together with disclosure of potential cross-partition duplication from pre-split bootstrap resampling; (4) corrected class ordering in all confusion matrices; (5) class-wise interpretation of false alarms and missed tracking cases; and (6) a transparent branch-wise comparison that reports diagnostic performance together with model size and recorded training cost without attributing the observed differences solely to signal domain [30, 31, 34, 35].

The remainder of the paper is organized as follows. Section 2 critically discusses related studies and establishes the research motivation. Section 3 presents the dataset provenance, signal representations, GRU formulation, and evaluation procedure. Section 4 reports and interprets the experimental results. Section 5 compares the proposed approach with representative switchgear and PD studies, and Section 6 concludes the paper and identifies the next validation steps.

2. RELATED WORK AND RESEARCH MOTIVATION

2.1 Evolution from engineered features to deep sequence learning

Ultrasonic switchgear diagnosis has progressed from engineered descriptors with shallow classifiers to deep models that learn directly from signal-derived representations. Early time- and frequency-domain extreme-learning-machine studies established the feasibility of automated fault recognition but remained dependent on fixed descriptors and comparatively simple class formulations [19, 40]. The practical difficulty is greater for tracking detection because the negative class must reject not only normal operation but also corona, arcing, and mechanical signatures that can share transient or spectral characteristics [15, 20, 29].

Convolutional and recurrent models subsequently improved feature learning on the same general switchgear data lineage. Hybrid 1D-CNN-LSTM studies reported high split-specific performance for multi-fault, tracking, arcing, and corona tasks [18, 20, 41]. These results demonstrate that the airborne-ultrasonic corpus contains highly discriminative structure, but architecture depth, preprocessing, class formulation, sampling, and signal representation often changed simultaneously. Consequently, a high percentage should be interpreted together with the experimental protocol rather than assigned to a single architectural or domain choice [2, 35, 37].

For the present tracking task, the most relevant earlier result is the 1D-CNN-LSTM surface-discharge study [18], which reported 100% accuracy in both signal domains under a 70/15/15 split. The broader multi-fault study [41] and a later mechanical-fault study [21] also reported high split-specific performance. Related work addressed switch-cabinet temperature prediction [42], switchgear digitalization [43], hybrid discharge sensing [44], and cable-PD classification [45]. This cross-study consistency shows that high split-specific percentages have been reported on this corpus, but it does not validate independence of the current partitions. The present study therefore reports the exact GRU program outputs as split-specific results and explicitly notes that pre-split bootstrap resampling may place identical computational records in different subsets; the reported percentages are not treated as reliable held-out generalization estimates.

2.2 Recent deep learning for acoustic and partial-discharge diagnosis

Recent switchgear and PD research can be grouped around several methodological directions rather than individual algorithms. One direction targets noise robustness and operating-state variability through denoising, pre-training, acoustic feature interaction, recurrent encoders, and multi-scale adaptive processing [25-28, 46]. These methods can improve robustness under difficult sensing conditions, although their multi-stage designs make it harder to isolate the contribution of the original representation.

A second direction transforms discharge measurements into richer spatial representations, including continuous-wavelet images, PRPD/PSA patterns, waveform-spectrogram fusion, transformer-enhanced feature maps, and state-space encodings [28, 29, 32, 47, 48]. Such representations expose local and global structure effectively, but image construction, larger backbones, or ensemble inference increase computational complexity relative to a compact recurrent baseline.

A third direction addresses limited labels, deployment constraints, privacy, and domain shift through semi-supervised learning, architecture search, cloud-edge diagnosis, federated learning, digital twins, and transfer learning [30, 33, 34, 49-52]. These studies broaden the deployment envelope of intelligent condition monitoring, but they generally solve sensing or validation problems that differ from binary airborne-ultrasonic tracking detection.

Taken together, the recent literature shows that high

recognition accuracy can arise from multiple interacting sources: representation design, model capacity, denoising, transfer, additional unlabeled data, or optimization strategy [25, 29, 46, 51]. This motivates a deliberately compact recurrent baseline in which branch construction, model size, class-wise errors, and evaluation limitations are reported explicitly rather than summarized by a headline accuracy alone.

From an evaluation perspective, the literature also shows why split-specific results should be separated from claims of field generalization. Small or correlated datasets can produce stable-looking point estimates even when uncertainty remains at the event level [36-39]. In the present implementation, bootstrap resampling with replacement precedes the random split, so identical computational records may occur in more than one subset. The test percentages are therefore reported only as original split-specific program outputs. Because record independence is not established, confidence intervals based on independent binomial observations are not reported. Event-grouped validation additionally requires identifiers that are not available in the current computational dataset.

This positioning does not diminish the practical value of the observed results. Instead, it distinguishes two questions that are often conflated: whether a compact GRU can discriminate the available tracking and Non-Tracking computational records, and whether the same percentage will generalize across physically independent events, sites, sensors, and operating conditions. The first question is answered by the current experiments; the second is identified as the next validation stage [44, 51].

2.3 Rationale for the proposed gated recurrent unit-based dual-domain study

Time-frequency representations preserve the local evolution of short-lived impulses, whereas compact spectral representations emphasize persistent band-energy structure [53, 54]. Across PD research, adaptive wavelets, recurrent embeddings, few-shot CNN-LSTM models, and image-based classifiers all show that useful diagnostic information can be reorganized before classification [55, 56]. The present study adopts this general principle but applies it through two separately implemented recurrent inputs so that their branch-specific behavior can be described without claiming a controlled paired-domain experiment.

The GRU provides a compact recurrent mechanism in which reset and update gates regulate the contribution of historical information [35]. Reviews of recurrent learning, time-series classification, and industrial fault diagnosis emphasize both the effectiveness of sequence models and the need for transparent validation when datasets are limited or correlated [36-39]. Accordingly, the present analysis reports the implemented split procedure, class-wise errors and precision-recall measures [57], parameter counts, and the limitations created by possible duplicate records and unavailable event metadata.

The proposed approach is therefore not presented as universally superior to recent deep architectures. Its value lies in an implementation-grounded comparison of two GRU pipelines: both use the same recurrent layer topology, while their input populations, preprocessing, tensor sizes, and regularization settings are reported explicitly. This framing allows predictive behavior and computational cost to be examined transparently without treating a branch-wise difference as proof of intrinsic time-origin or frequency-origin superiority [18, 21, 41]. The conceptual framework in Figure 1 synthesizes the physical tracking mechanism, ultrasonic sensing context, representation choices, and learning/validation considerations described in prior work [9, 13, 18, 37].

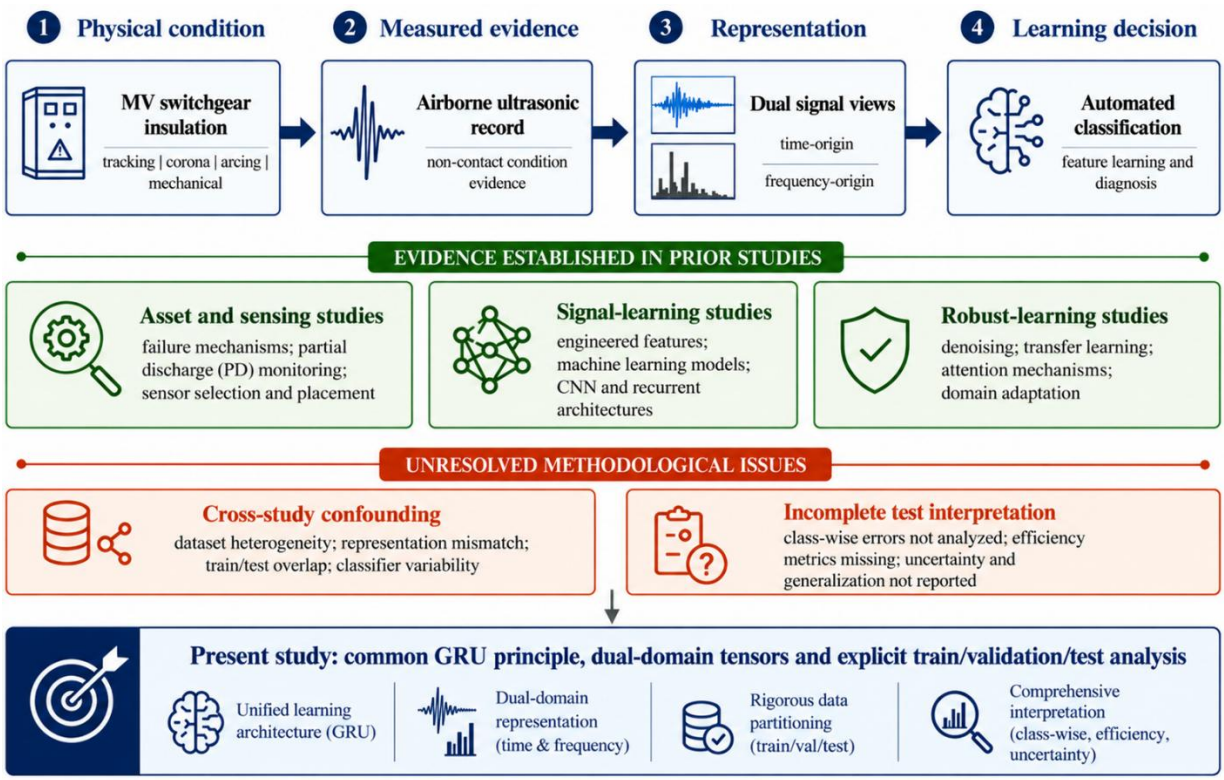

Figure 1. Conceptual framework of the study linking switchgear insulation faults, ultrasonic evidence, dual-domain signal representation, learning-based diagnosis, and the identified methodological gap

2.4 Research gap

Despite rapid progress in deep-learning-based switchgear diagnosis, four connected methodological gaps remain. First, many studies optimize a representation and a classifier together, so reported gains cannot be assigned cleanly to temporal evolution, spectral concentration, or model capacity [26, 31, 33, 51]. Second, image encodings, transfer learning, architecture search, and model fusion can improve recognition but also increase computational burden and complicate interpretation of compact recurrent baselines [28, 29, 34, 51].

Third, earlier electrical-tracking studies from the same data lineage established the feasibility of high-accuracy deep classification [18, 41], but they did not provide a paired GRU experiment in which identical physical events, identical preprocessing, and identical model settings were compared across domains. Fourth, class-specific tracking sensitivity, false-alarm behavior, uncertainty, and event-level split provenance are not always reported together. These gaps motivate a transparent branch-wise analysis that preserves the original program outputs while clearly distinguishing split-specific computational results under possible duplication from leakage-controlled and event-independent generalization.

To address the above gaps, the principal contributions of this study are summarized as follows:

1. Data provenance and task reformulation: A field-derived five-condition ultrasonic switchgear corpus is transparently reformulated as a binary Tracking-versus-Non-Tracking task while preserving source provenance, class definitions, and the practical fault context [19, 40, 41].

2. Branch-specific sequence construction: Two recurrent inputs are implemented: a 40×128 time-origin tensor and a 20×40 frequency-origin tensor. The populations, preprocessing operations, Mel parameters, tensor construction, and class composition are reported explicitly and are not treated as interchangeable or paired observations.

3. Common recurrent topology with branch-specific settings: Both pipelines use two 32-unit GRU stages followed by a 32-unit dense ReLU layer and a two-class softmax output. Because the input dimensions and dropout rates differ, the comparison is presented as an implementation-level comparison rather than a controlled causal test of signal domain.

4. Stage-wise and class-wise evaluation: Training, validation, and test-partition outputs are reported separately, together with corrected confusion matrices, precision, recall, and F1-score, plus explicit interpretation of false alarms, missed Tracking cases, potential cross-partition duplication, and the limits on statistical uncertainty claims.

5. Performance-complexity reporting: The recorded runs yield 95.45% test accuracy for the time-origin branch and 95.83% for the frequency-origin branch. The frequency-origin implementation also contains fewer trainable parameters and has a shorter recorded training time, while the time-origin branch achieves complete Tracking recall in its test subset. These are reported as branch-specific observations, not as evidence that one signal domain is intrinsically superior.

Table 1. Critical comparison of representative deep-learning studies

Study	Diagnostic Target	Method / Representation	Reported Result	Principal Limitation
Ishak et al. [19]	Corona vs. non-corona	TD/FD engineered inputs + ELM	87.5%	Shallow model; one discharge class
Wang et al. [27]	GIS mechanical faults	Pre-trained acoustic transformer system	Competitive	Different fault physics; multi-stage system
Zhang et al. [28]	GIS mechanical faults	Waveform-spectrogram feature fusion	99.65% reported	Mechanical target; larger fusion pipeline
Ishaq et al. [29]	MV-switchgear PD defects	CWT + optimized CNN	Superior reported	Image conversion and higher complexity
Jia et al. [30]	Switchgear PD defects	Cloud-edge DBN	88.03%	Laboratory PD; different sensing objective
Fei et al. [31]	Three switchgear PD types	Improved-MSA optimized BPNN	High reported	Engineered features and optimizer coupled
Chang and Lin [32]	Six GIS/cable PD defects	PRPD/PSA CNN ensemble	97.5%	Different apparatus and image patterns
Yang et al. [33]	GIS PD with scarce labels	Semi-supervised UHF diagnosis	94.59%	Requires unlabeled data; GIS target
Han et al. [51]	Field GIS PD	ResNet50 transfer learning	93.7%	Laboratory-to-field transfer; image backbone
Present study	Tracking vs. heterogeneous Non-Tracking	Dual-domain GRU sequence learning	95.45% / 95.83%	Pre-split bootstrap may duplicate records across partitions; paired/grouped validation needed

Note: Medium-voltage (MV), partial-discharge (PD), gated recurrent unit (GRU).

Table 1 and the integrated literature synthesis show that recent accuracy gains are frequently coupled with image encodings, pre-training, transfer learning, architecture search, or model fusion [27, 29, 34, 51]. The present work therefore emphasizes traceable tensor construction, compact recurrent topology, stage-wise metrics, class-wise errors, and explicit limits on the interpretation of its original computational split, including the possibility of duplicate records across partitions [18, 21, 41]. A broader cross-study comparison is retained in Section 5, where methodological qualifications are reported alongside representative results rather than used for absolute ranking.

3. MATERIALS AND METHODS

3.1 Dataset provenance, construction, and class definition

The present work is a secondary computational analysis of the field-derived airborne-ultrasonic switchgear database reported by Ishak et al. [40]. The source measurements were collected at utility installations in Kedah, Kuala Lumpur,

Melaka, Selangor, Perak, Negeri Sembilan, and Johor [19, 40] and stored as audio records before conversion to numerical matrices in MATLAB. The time-origin matrices contain ordered signal samples, whereas the frequency-origin records in the source workflow were derived from the time records using a one-dimensional fast Fourier transform (FFT).

The source database contains five working labels: Normal, Corona, Arcing, Mechanical, and Electrical Tracking [40]. For the present analysis, Electrical Tracking is the positive class (Tracking), while Normal, Corona, Arcing, and Mechanical records are combined into a heterogeneous negative class (Non-Tracking). This formulation is more demanding than a healthy-versus-fault task because the negative class includes competing abnormal signatures rather than normal operation alone [18-21, 40, 41].

Figure 2 summarizes the source provenance, binary reformulation, and branch-specific computational datasets. The time-origin branch contains 438 records (125 Non-Tracking and 313 Tracking), and the frequency-origin branch contains 160 records (120 Non-Tracking and 40 Tracking) [40]. In the implemented preprocessing, the Tracking matrix in each branch is bootstrap-resampled with replacement before feature extraction and before the subsequent random train/validation/test split. Because n_samples is not specified, the class count is unchanged (313 remains 313 in the time-origin branch and 40 remains 40 in the frequency-origin branch). However, replacement sampling can select the same source row multiple times. When the resampled matrix is then randomly partitioned, identical computational records can occur in more than one subset, creating a potential cross-partition leakage pathway. The current computational dataset does not retain a physical-event or group identifier. Consequently, the 95.45% and 95.83% values are reported only as original split-specific program outputs and are not interpreted as reliable held-out or event-independent generalization estimates.

As summarized in Table 2, the two branches are not synchronized representations of an identical sample population. They differ in record count, class balance, record length, feature construction, tensor dimensions, and dropout. Their results therefore characterize two implemented diagnostic pipelines rather than a controlled paired-domain experiment. This distinction is maintained throughout the paper, so the observed differences in accuracy, recall, parameter count, and training time are not attributed solely to time-origin versus frequency-origin representation.

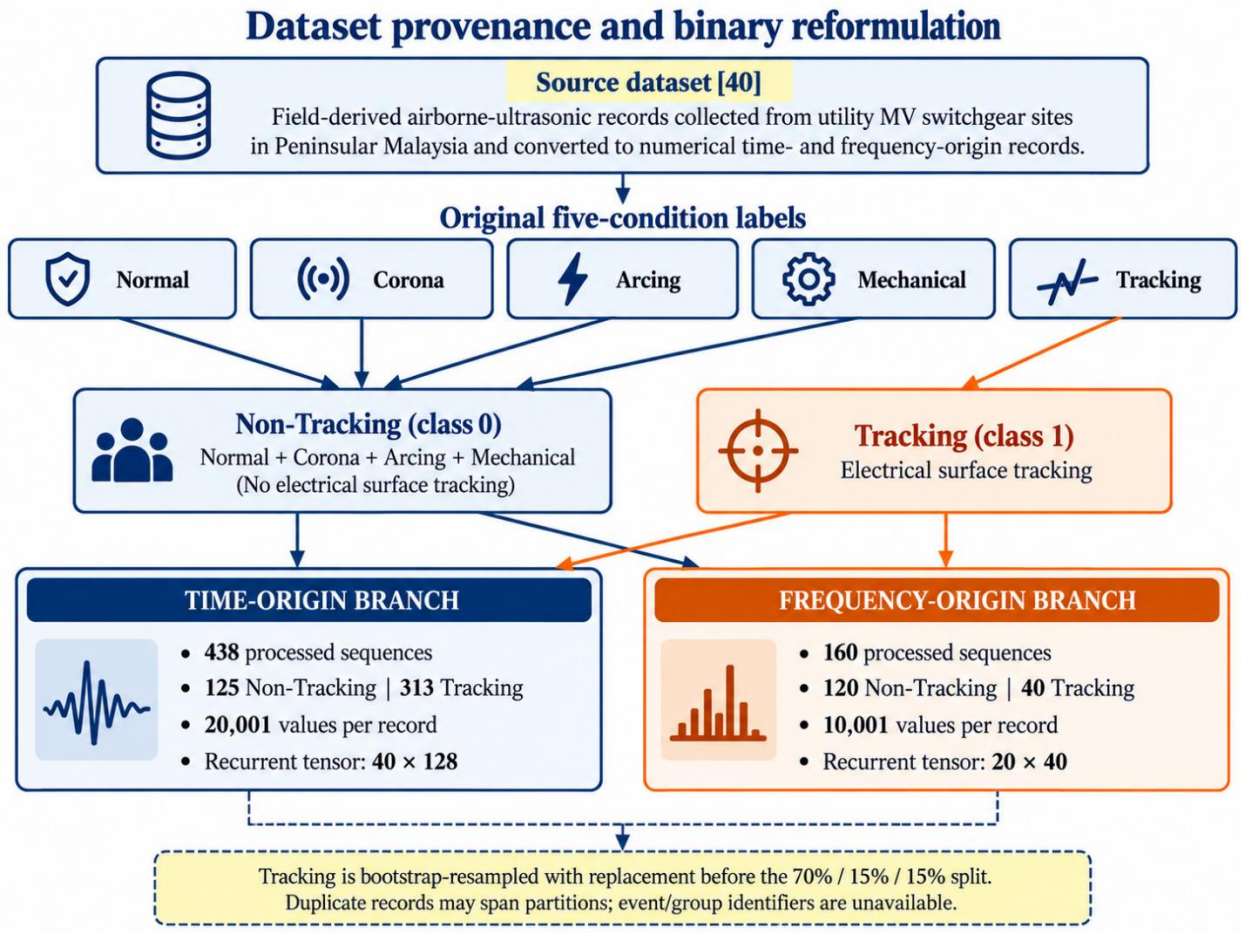

Figure 2. Provenance of the field-derived ultrasonic database, binary class reformulation, and branch-specific computational datasets used in the present study

Table 2. Derived dataset composition

Domain	Non-Tracking Composition	n	Positive Class	n	Total	Values / Record
Time-origin	Normal + corona + arcing + mechanical	125	Tracking	313	438	20,001
Frequency-origin	Normal + corona + arcing + mechanical	120	Tracking	40	160	10,001

3.2 Time-origin signal representation

Each time-origin record contains 20,001 ordered values. No median/IQR RobustScaler or other train-fitted scaling transform is applied in the implemented time-origin pipeline. The Non-Tracking records are processed from the loaded source matrices, whereas the Tracking matrix is divided by 255 and bootstrap-resampled with replacement at unchanged class size before feature extraction and before subsequent dataset partitioning. Because replacement sampling can repeat rows, identical Tracking records may later be assigned to different subsets. The recurrent input is generated with librosa.feature.melspectrogram(y = x, sr = 20001). The effective parameters are n_fft = 2048, win_length = 2048, Hann windowing, hop_length = 512, center = True, power = 2.0, n_mels = 128, fmin = 0 Hz, and fmax = sr/2 (approximately 10,000.5 Hz) [58].

$$X(m,k) = \Sigma_n x[n]w[n - mH]exp(-j2\pi kn/N) \quad (1)$$

The Mel transform internally applies windowed short-time Fourier analysis and projects the resulting power spectrum onto the Mel filter bank. The output for a 20,001-value record is a 128 × 40 Mel-power matrix; it is transposed so that the recurrent sequence dimension is 40 steps and each step contains 128 features. The power_to_db operation present elsewhere in the notebook is used only for visualization and is not the model input. Thus, the implemented feature should be described as a Mel power-spectrogram rather than a log-Mel tensor.

$$M(m,b) = \Sigma_k |X(m,k)|^2 H_{b(k)} H \quad (2)$$

The final time-origin tensor therefore has shape 40 × 128, where 40 is the recurrent sequence length and 128 is the feature vector at each step. Figure 3 summarizes the implemented workflow from the 20,001-value time record to Mel-power extraction, tensor transposition, GRU training, and test-partition evaluation under the original computational split.

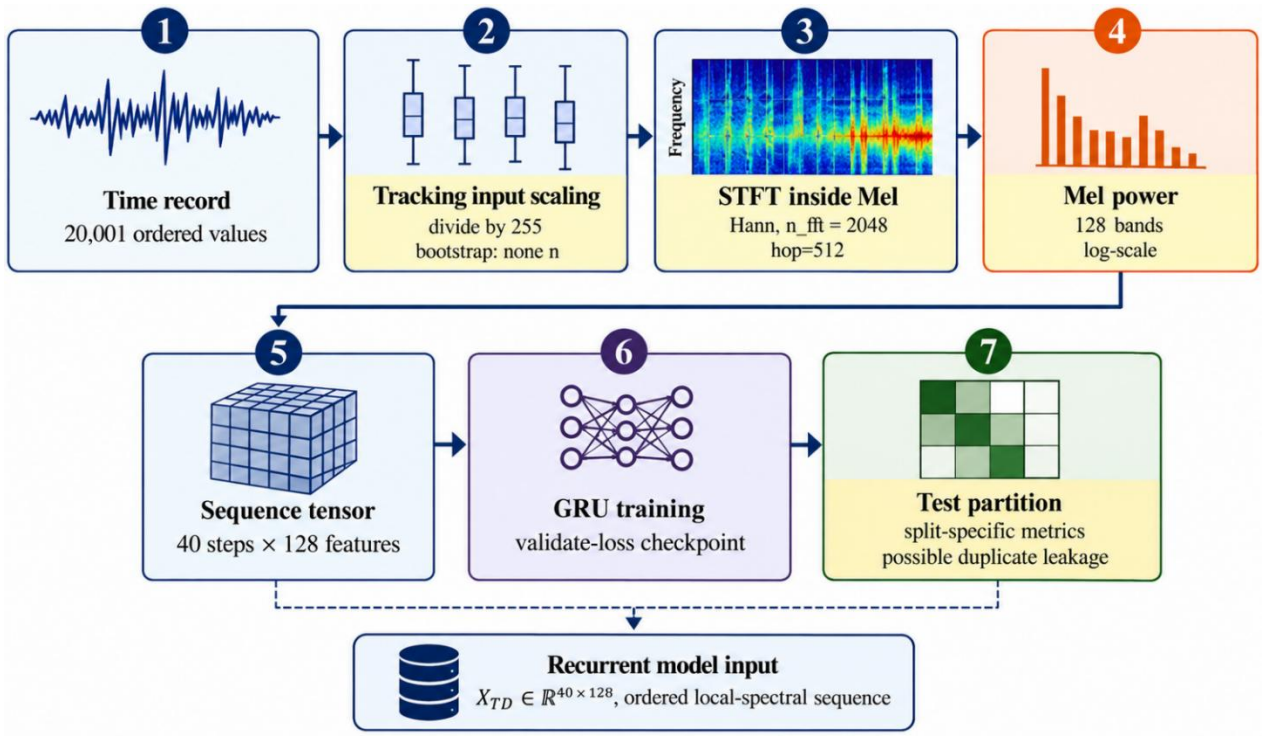

Figure 3. Implemented time-origin workflow from the 20,001-value record to the 40 × 128 Mel-power recurrent tensor and test-partition evaluation under the original computational split

3.3 Frequency-origin signal representation

The source workflow provides frequency-origin records derived by a one-dimensional FFT of the time-origin measurements, with 10,001 values per record [40]. No median/IQR RobustScaler or other train-fitted scaling transform is applied in the implemented frequency-origin pipeline. The Non-Tracking records are processed from the loaded source matrix, whereas the Tracking matrix is divided by 255 and bootstrap-resampled with replacement at unchanged class size before feature extraction and before subsequent dataset partitioning. Because replacement sampling can repeat rows, identical Tracking records may later be assigned to different subsets. The implementation then applies librosa.feature.melspectrogram(y = x, sr = 10001, n_mels = 40, n_fft = 512). With the remaining defaults, the effective settings are win_length = 512, Hann windowing, hop_length = 512, center = True, power = 2.0, fmin = 0 Hz, and fmax = sr/2 (approximately 5,000.5 Hz) [58]. Because the input is already spectral, this is a secondary Mel-power embedding rather than a spectrogram of the original waveform; Mel enhancement approaches developed for other acoustic tasks [59] are not applied.

For a 10,001-value frequency-origin record, the implementation produces a 40 × 20 Mel-power matrix, which is transposed to 20 recurrent steps with 40 features per step. This smaller tensor reduces recurrent input dimensionality, but the branch also contains fewer records and uses a different dropout setting; therefore, lower parameter count or shorter recorded training time cannot be assigned to spectral

representation alone. Figure 4 summarizes the implemented frequency-origin pipeline, and Figure 5 illustrates the dimensional and informational differences between the two recurrent inputs.

Table 3 summarizes the branch-specific representation characteristics. The comparison is intentionally descriptive: the branches differ in source population, record length, Mel configuration, tensor size, and diagnostic emphasis. Using the same GRU layer topology provides a common recurrent baseline, but it does not convert the experiment into a paired domain-ablation study because the data populations and regularization settings are not identical.

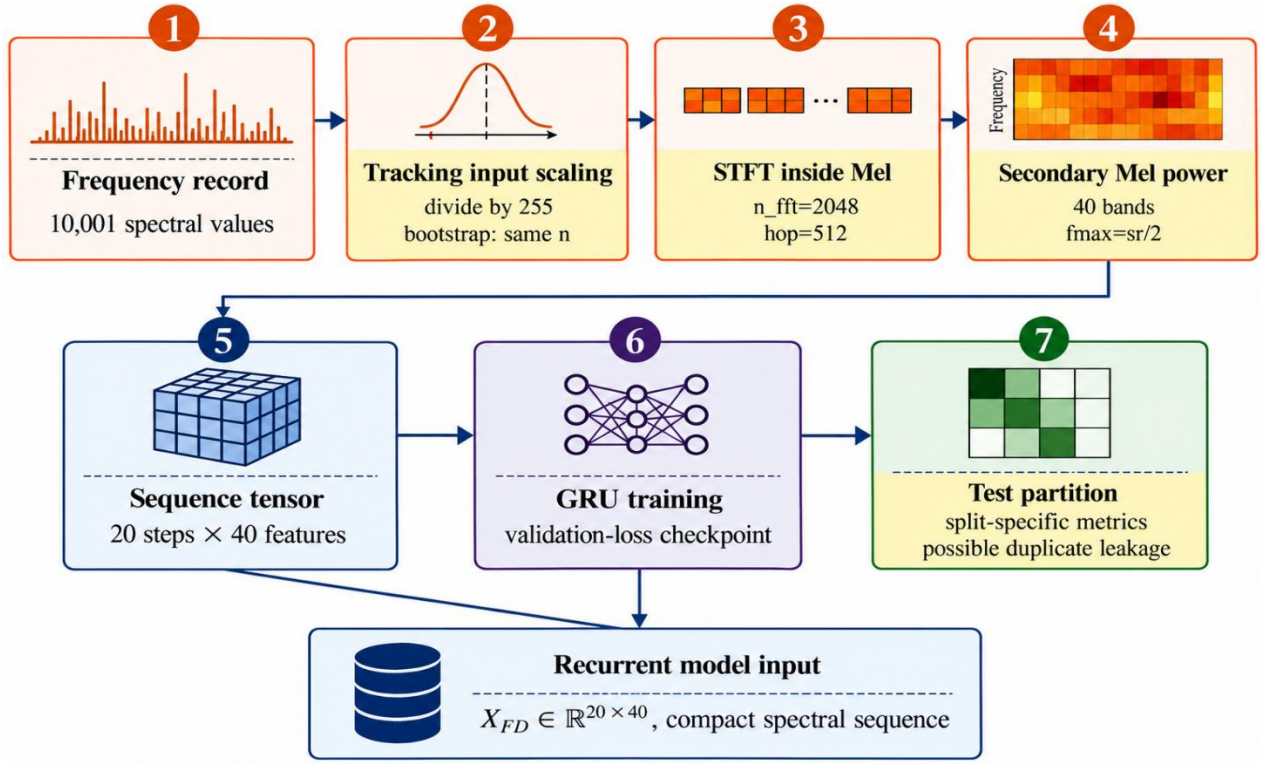

Figure 4. Implemented frequency-origin workflow from the 10,001-value spectral record to the 20 × 40 secondary Mel-power recurrent tensor and test-partition evaluation under the original computational split

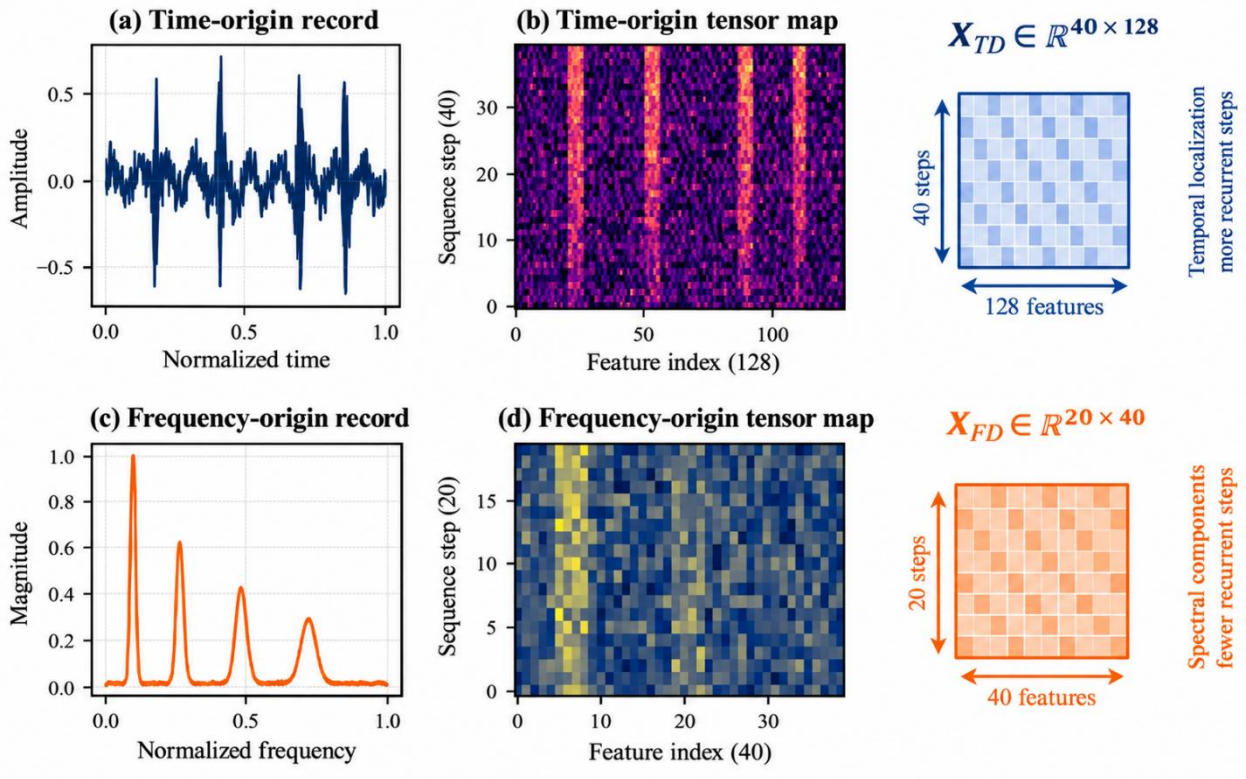

Figure 5. Schematic comparison of the time-origin and frequency-origin sequence representations, with explicit recurrent-step and feature dimensions

3.4 Gated recurrent unit-based recurrent classifier

Let x_t denote the feature vector at sequence step t and h_{t-1} the preceding hidden state. The GRU update gate z_t and reset gate r_t are calculated using Eqs. (3) and (4), respectively [35, 36].

$$z_t = \sigma\left(W_{zx_t} + U_{zh_{t-1}} + b_z\right)$$
$$r_t = \sigma(W_{rx_t} + U_{rh_{t-1}} + b_r) \quad (3)$$

The candidate hidden state and the updated state are then obtained from Eq. (5). These gates allow the model to retain diagnostically useful sequence information while suppressing irrelevant history [37].

$$\tilde{h}_t = tanh(W_{hx_t} + U_h(r_t \odot h_{t-1}) + b_h) \quad (4)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (5)$$

The implemented classifier contains two recurrent GRU stages followed by a dense rectified linear unit (ReLU) mapping and a two-class softmax output. Both branches use 32 units in each GRU stage and 32 units in the dense ReLU layer. The time-origin branch uses dropout = 0.2 and contains 23,010 trainable parameters, whereas the frequency-origin branch uses dropout = 0.4 and contains 14,562 trainable parameters because its input feature dimension is smaller. Figure 6 therefore represents a shared layer topology with independently trained, branch-specific input dimensions and regularization settings [35-37].

Table 3. Domain-specific representation characteristics

Property	Time-Origin Branch	Frequency-Origin Branch
Starting record	Ordered time-origin sequence	FFT-derived frequency-origin sequence
Record length	20,001 values	10,001 values
Embedding	Mel power spectrogram: 128 bands, n_fft = 2048, hop = 512	Secondary Mel power embedding: 40 bands, n_fft = 512, hop = 512
Final tensor	40 × 128	20 × 40
Diagnostic emphasis	Transient order and local spectral change.	Spectral concentration and band-energy structure.
Main risk	False tracking alarms from competing transients.	Missed weak Tracking evidence after compact embedding.

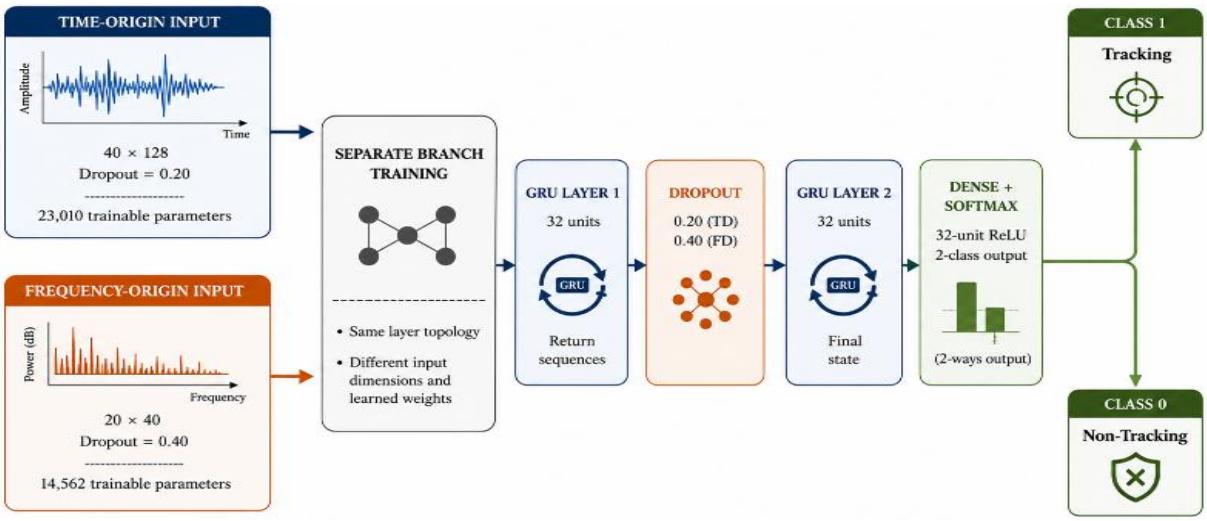

Figure 6. Shared gated recurrent unit (GRU)-based decision topology with separately trained time-origin and frequency-origin branches, domain-specific tensor dimensions, dropout rates, output classes, and trainable parameter counts

3.5 Training protocol and evaluation metrics

Each branch is partitioned at the computational-record level into 70% training, 15% validation, and 15% testing subsets using scikit-learn train_test_split. In the time-origin implementation, the first split uses random_state = 0 and the validation/test split uses random_state = 21; in the frequency-origin implementation, both split calls use random_state = 21. However, the Tracking matrix is bootstrap-resampled with replacement before these split calls. Consequently, repeated copies of the same computational record may be assigned to different partitions. Although the test partition is not used directly in model fitting after the split, it cannot be regarded as an independent held-out set in the usual generalization sense. No stratify or groups argument is used, and no physical-event identifier is available in the current computational dataset. Accordingly, the reported accuracies are retained only as the original split-specific program outputs and are not used as estimates of leakage-free or event-independent generalization. Both models are trained for 60 epochs with batch size 16, Adam at a learning rate of 10^{-4}, categorical cross-entropy loss, and a validation-loss checkpoint [60].

Table 4 reports the complete implementation settings. Recorded training time is interpreted only as a property of the recorded runs because the branches differ in sample count, input size, and dropout. Because record independence cannot be established and exact bootstrap duplicates may span partitions, Wilson 95% confidence intervals are not reported in the present analysis. Treating all 66 and 24 test-partition records as independent observations could overstate the effective sample size and produce intervals that are too narrow. Any future uncertainty analysis should be performed only after exact duplicate removal and, when event identifiers are available, repeated event-grouped partitioning.

Table 4. Training and implementation settings

Setting	Time-Origin	Frequency-Origin
Split ratio	70 / 15 / 15	70 / 15 / 15
Epochs	60	60
Batch size	16	16
Optimizer	Adam, learning rate 10^{-4}	Adam, learning rate 10^{-4}
Loss	Categorical cross-entropy	Categorical cross-entropy
Dropout	0.2	0.4
Trainable parameters	23,010	14,562
Recorded training time	89.56 s	27.80 s
Split method	Seeds: 0, 21	Seeds: 21, 21
Event groups	Unavailable	Unavailable

Because class distributions are unequal and a missed Tracking case is operationally different from a false tracking alarm, evaluation includes precision, recall, F1-score, and macro F1 in addition to accuracy. For the positive Tracking class, the standard definitions are given in Eq. (6) [61].

$$\begin{aligned} Precision &= TP/(TP + FP) \\ Recall &= TP/(TP + FN) \\ F1 &= 2(Precision * Recall)/(Precision + Recall) \end{aligned} \quad (6)$$

Macro F1 is the unweighted mean of the class-specific F1-scores and therefore prevents the larger class from dominating the summary [62]. Because pre-split replacement sampling can duplicate records across partitions, the effective number of independent test observations is unknown; therefore no Wilson or other independent-binomial confidence interval is attached to the current test proportions.

4. RESULTS AND DISCUSSION

4.1 Time-origin results

The time-origin model converges with a small separation between training and validation loss in Figure 7. Using the checkpoint retained by minimum validation loss, the training, validation, and test-partition accuracies are 98.69%, 100.00%, and 95.45%, respectively. The test partition contains 63 correct predictions out of 66 records. These figures are the original split-specific program outputs under the potential-duplication protocol described in Section 3.5; no independent-observation confidence interval is assigned. Table 5 reports all three partitions so that model fitting, validation, and test-partition performance remain clearly separated.

Table 5. Time-origin performance across data partitions

Stage	Samples	Loss	Accuracy
Training	306	0.0247	98.69%
Validation	66	0.0059	100.00%
Testing	66	0.0871	95.45%

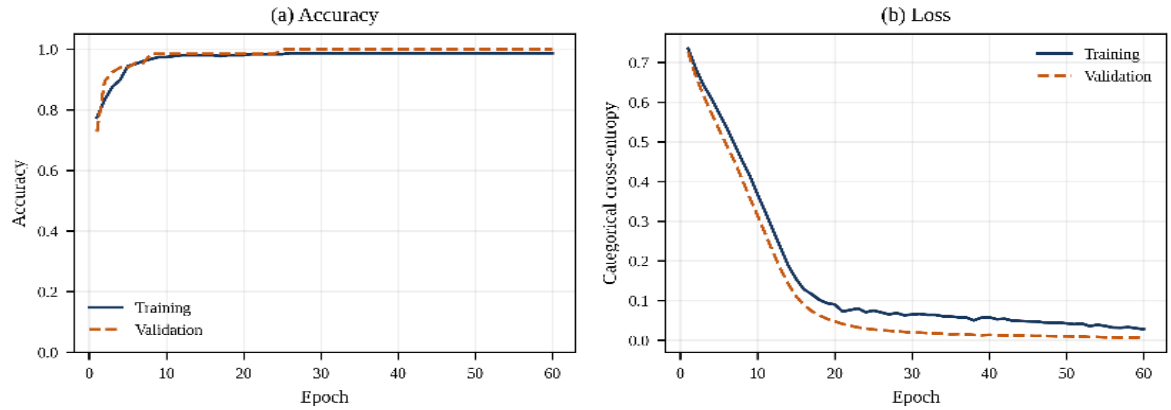

Figure 7. Training and validation accuracy and loss for the time-origin gated recurrent unit (GRU) model

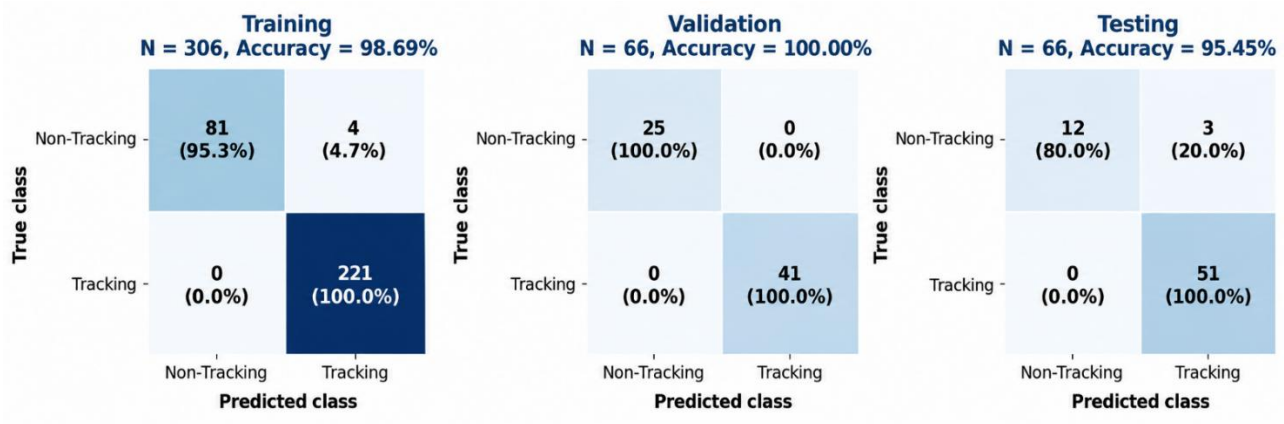

Figure 8. Corrected time-origin confusion matrices for training, validation, and testing; class order is Non-Tracking followed by Tracking

The corrected confusion matrices in Figure 8 follow the encoder order: Non-Tracking is class 0 and Tracking is class 1. In the test partition, all 51 Tracking records are detected and 12 of 15 Non-Tracking records are correctly rejected, producing three false tracking alarms. Tracking recall is therefore 100.0%, while Non-Tracking recall is 80.0%. These proportions describe this specific partition only and are not accompanied by independent-binomial confidence intervals. The error pattern favors fault sensitivity at the cost of additional inspection-triggering false alarms.

4.2 Frequency-origin results

The frequency-origin learning curves in Figure 9 show greater early-epoch variability, consistent with the smaller and differently balanced branch. Using the saved validation-loss checkpoint, the training and validation accuracies are both 100.00%, while the test-partition accuracy is 95.83% (23/24). As with the time-origin branch, this is a split-specific program output under potential duplication and no independent-observation confidence interval is reported. The higher test loss relative to training loss indicates that confidence calibration remains worth examining even when the predicted class is usually correct as shown in Table 6 [36-38].

As shown in Figure 10, all 17 Non-Tracking test records are classified correctly, while one of seven Tracking records is missed. The corresponding Tracking recall is 85.71%, and Non-Tracking recall is 100.0%. These proportions are descriptive for this specific partition; no independent-observation confidence interval is claimed. Thus, this split contains no false tracking alarms but does contain one missed Tracking case, an error that is operationally more consequential and should be examined under larger, repeated, event-grouped tests.

Table 6. Frequency-origin performance across data partitions

Stage	Samples	Loss	Accuracy
Training	112	0.0902	100.00%
Validation	24	0.1247	100.00%
Testing	24	0.1105	95.83%

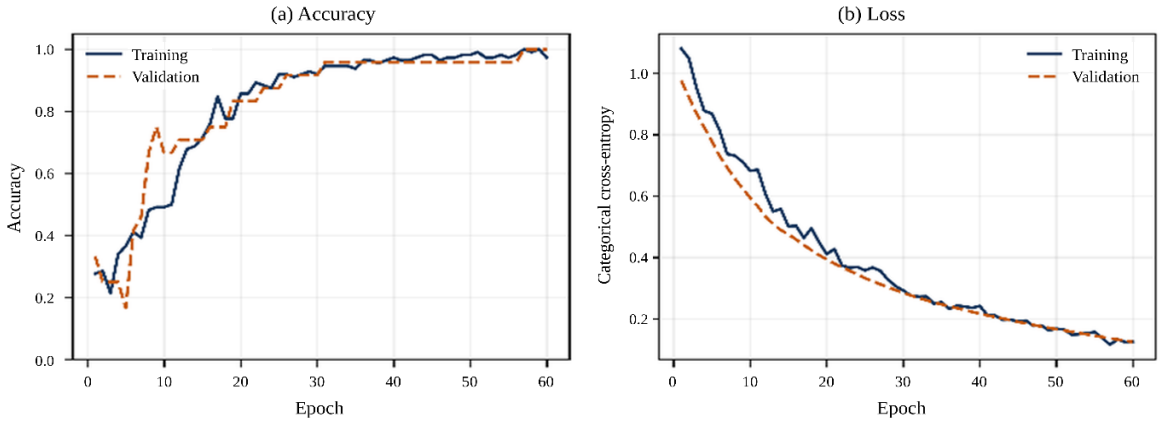

Figure 9. Training and validation accuracy and loss for the frequency-origin gated recurrent unit (GRU) model

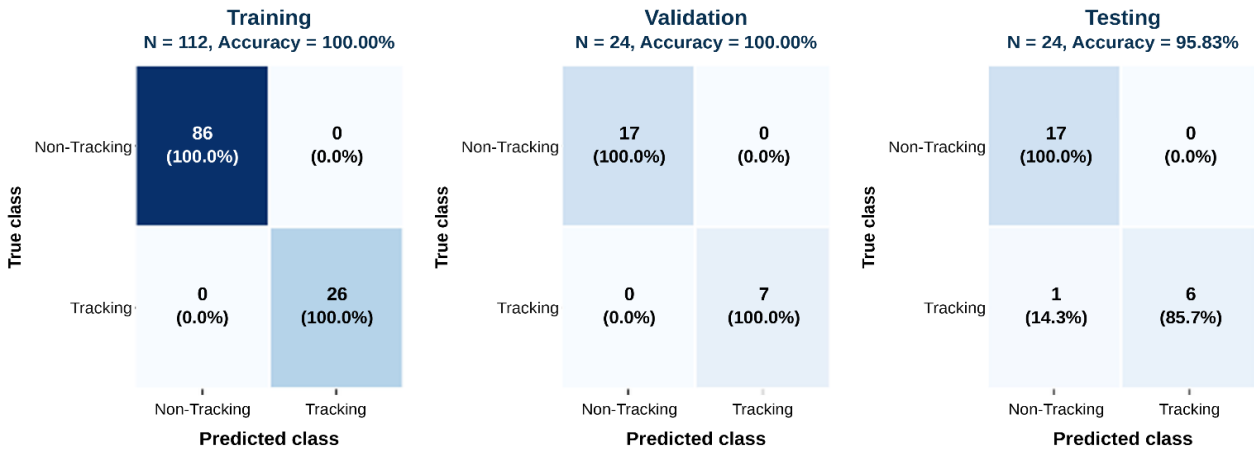

Figure 10. Corrected frequency-origin confusion matrices for training, validation, and testing; class order is Non-Tracking followed by Tracking

4.3 Class-wise and computational comparison

Table 7 makes the complementary error patterns explicit. The time-origin branch has Non-Tracking recall of 0.800 and Tracking recall of 1.000, whereas the frequency-origin branch has Non-Tracking recall of 1.000 and Tracking recall of 0.857. Their macro F1-scores are 0.930 and 0.947, respectively. Because the test partitions contain only 66 and 24 records, arise from non-equivalent branch populations, and may include duplicate records introduced before splitting, the 0.38-percentage-point difference in accuracy is treated as descriptive only rather than evidence of statistical or causal superiority of one signal domain.

Table 7. Test-stage class-wise performance

Branch	Class	Precision	Recall	F1-Score
Time-origin	Non-Tracking	1.000	0.800	0.889
Time-origin	Tracking	0.944	1.000	0.971
Frequency-origin	Non-Tracking	0.944	1.000	0.971
Frequency-origin	Tracking	1.000	0.857	0.923

Table 8. Direct comparison of the implemented branches

Criterion	Time-Origin	Frequency-Origin	Interpretation
Test-partition accuracy	95.45%	95.83%	0.38 percentage-point split-specific difference; no domain-superiority or generalization inference
Macro F1	0.930	0.947	0.017 descriptive difference
Tracking recall	1.000	0.857	Time-origin detects all tracking cases
False tracking alarms	3	0	Frequency-origin is more specific
Trainable parameters	23,010	14,562	Frequency-origin uses 36.7% fewer
Recorded training time	89.56 s	27.80 s	69.0% lower in the recorded run; not hardware-normalized or domain-isolated

The frequency-origin recorded run uses fewer trainable parameters and less recorded training time than the time-origin run (Table 8). These differences are useful implementation descriptors, but they are jointly affected by the smaller input tensor, smaller branch dataset, different dropout rate, and feature construction. Accordingly, the computational comparison is reported as branch-specific cost rather than as evidence that frequency-origin information alone is responsible for the observed reduction.

5. COMPARISON WITH EXISTING STUDIES

Cross-study accuracy comparisons should be interpreted cautiously because equipment type, sensing route, fault taxonomy, sample independence, class balance, model capacity, and validation protocol differ substantially across publications [36-39]. Table 9 therefore provides contextual evidence rather than an absolute numerical ranking. This distinction is particularly important for the present data lineage, where several studies report very high split-specific accuracy under 70/15/15 partitions, but such percentages do not resolve sample dependence, possible pre-split duplication, or the need for event-grouped external validation.

Within the same switchgear data lineage, earlier work provides context for the numerical scale of the present program outputs.

The tracking 1D-CNN-LSTM study reported 100% test accuracy in both domains [18]; the corona study used 438 time-domain and 160 frequency-domain samples with a 70/15/15 split and reported 98.4% and 100% hybrid-model test accuracy [63]; the broader multi-fault study likewise reported very high performance [41]; and the mechanical-fault study used a 70/15/15 protocol with oversampling and reported 100% classification in its evaluated domains [21]. These earlier figures provide context only; they do not validate the present split because similar preprocessing and partitioning choices may carry related dependence or leakage risks. Accordingly, the present 95.45% and 95.83% values are cited only as outputs of the original program, not as corroborated estimates of held-out generalization.

Table 9. Comparison with representative switchgear and partial-discharge (PD) studies

Study	Target	Method / Input	Test Result	Qualification
Alsumaidaee et al. [18]	Tracking / non-tracking	TD/FD + 1D-CNN-LSTM	100% reported	Single split; hybrid capacity and domain coupled
Ishak et al. [19]	Corona / non-corona	Ultrasound TD/FD + ELM	87.5%	Simple baseline; one target fault
Wang et al. [27]	GIS mechanical faults	Pre-trained acoustic transformer model	Competitive	Different fault physics and model scale
Zhang et al. [28]	GIS mechanical faults	Waveform-spectrogram fusion	99.65% reported	Mechanical target; multi-module architecture
Ishaq et al. [29]	MV-switchgear PD defects	CWT + optimized CNN	Superior reported	2-D image conversion and higher complexity
Jia et al. [30]	Switchgear PD defects	Cloud-edge DBN	88.03%	Laboratory PD and different sensor route
Fei et al. [31]	Three switchgear PD types	Improved-MSA BPNN	High reported	Engineered descriptors + optimizer
Chang and Lin [32]	Six GIS/cable defects	PRPD/PSA CNN ensemble	97.5%	Different apparatus and defect taxonomy
Yang et al. [33]	GIS PD, limited labels	Semi-supervised UHF model	94.59%	Needs unlabeled data; GIS not tracking
Han et al. [51]	Field GIS PD	ResNet50 transfer learning	93.7%	Transfer setting; image backbone
Present-TD	Tracking / heterogeneous background	40 × 128 + GRU	95.45%	Original split-specific output; potential duplicate leakage; 51/51 Tracking detected; 3 false alarms
Present - FD	Tracking / heterogeneous background	20 × 40 + GRU	95.83%	Original split-specific output; potential duplicate leakage; 1 of 7 Tracking records missed

Note: partial-discharge (PD), gated recurrent unit (GRU), long short-term memory (LSTM).

Beyond this data lineage, recent PD studies obtain high recognition performance using denoising, wavelet images, ensemble CNNs, semi-supervised learning, transfer learning, transformers, and state-space models [25-34, 46-52]. These approaches are valuable when noise suppression, image structure, scarce labels, cross-site transfer, or high-capacity feature learning is central to the application.

The present GRU pipelines address a narrower objective: to provide compact recurrent baselines for airborne-ultrasonic Tracking-versus-Non-Tracking classification with explicit stage-wise metrics, class-wise errors, tensor dimensions, parameter counts, and split limitations. This narrower scope allows the observed results to be reproduced and interpreted without implying that a compact GRU should replace architectures designed for different sensing or deployment conditions [35-41].

The two implemented pipelines therefore exhibit different numerical profiles rather than a proven ranking: the frequency-origin run has a slightly higher point accuracy and lower recorded model cost, whereas the time-origin run has complete Tracking recall in its test partition. Because the branches are not paired and the current test outputs arise from a protocol with possible cross-partition duplication, no claim of intrinsic domain superiority or reliable between-branch generalization difference is made. A controlled comparison would require exact duplicate removal, synchronized representations of the same physical events, matched preprocessing and regularization, repeated event-grouped partitions, and a common external benchmark.

6. CONCLUSION AND FUTURE WORK

This study evaluated two compact GRU-based computational pipelines for detecting Electrical Tracking against a heterogeneous Non-Tracking class containing Normal, Corona, Arcing, and Mechanical records. Under the original computational split, the time-origin program output was 95.45% test-partition accuracy (63/66) with a macro F1-score of 0.930; it detected all 51 Tracking records and produced three false tracking alarms. The frequency-origin program output was 95.83% test-partition accuracy (23/24) with a macro F1-score of 0.947; it correctly rejected all 17 Non-Tracking records and missed one of seven Tracking records.

These values should be read only as the original program outputs. Because Tracking records were bootstrap-resampled with replacement before the random split, identical computational records may occur across training, validation, and test partitions. This potential cross-partition leakage prevents the 95.45% and 95.83% values from being interpreted as reliable held-out generalization performance. In addition, the absence of physical-event group identifiers prevents event-independent validation. The two branches also differ in sample population, class balance, record length, feature construction, tensor dimensions, and dropout, so their numerical comparison remains descriptive.

Within the original split, the branch-specific error patterns suggest a hypothesis for future system design rather than a validated deployment conclusion. The time-origin run shows higher Tracking sensitivity in its current test partition, whereas the frequency-origin run shows no false tracking alarms in its current test partition while using 36.7% fewer trainable parameters and less recorded training time. A future monitoring system could examine confidence-aware fusion or a hierarchical alarm rule, provided that the decision logic is first validated on duplicate-controlled, synchronized measurements of the same physical events.

The next validation stage should first remove exact duplicate computational records, then preserve physical-event groups during partitioning and repeat the evaluation across multiple grouped splits. Uncertainty intervals and calibration should be recomputed only on these leakage-controlled partitions, followed by external testing with data from different substations, sensors, distances, operating voltages, enclosure geometries, and noise conditions. Comparisons with LSTM, 1D-CNN, hybrid CNN-recurrent, transformer, and transfer-learning baselines should use the same event-level benchmark. Such testing would determine whether the original split-specific program outputs extend to a reproducible field-ready diagnostic system.

REFERENCES

[1] Subramaniam, A., Sahoo, A., Manohar, S.S., Raman, S.J., Panda, S.K. (2021). Switchgear condition assessment and lifecycle management: Standards, failure statistics, condition assessment, partial discharge analysis, maintenance approaches, and future trends. IEEE Electrical Insulation Magazine, 37(3): 27-41. https://doi.org/10.1109/MEI.2021.9399911

[2] Alsumaidaee, Y.A.M., Yaw, C.T., Koh, S.P., Tiong, S.K., Chen, C.P., Ali, K. (2022). Review of medium-voltage switchgear fault detection in a condition-based monitoring system by using deep learning. Energies, 15(18): 6762. https://doi.org/10.3390/en15186762

[3] Hussain, A.S.T., Ahmed, S.A., Taha, T.A. (2022). Early fault identification for operating circuit breaker based on classifier model system. Indonesian Journal of Electrical Engineering and Computer Science, 26(2): 699-706. https://doi.org/10.11591/ijeecs.v26.i2.pp699-706

[4] Mukhlif, S.K., Nassar, A.K., Al-Dulaimi, O., et al. (2026). Deciphering domestic energy demand through household electricity and gas consumption. Energy and Built Environment, 7(2): 502-517. https://doi.org/10.1016/j.enbenv.2025.06.004

[5] Hassan, Q., Algburi, S., Sameen, A.Z., Salman, H.M., Jaszczur, M. (2023). A review of hybrid renewable energy systems: Solar and wind-powered solutions: Challenges, opportunities, and policy implications. Results in Engineering, 20: 101621. https://doi.org/10.1016/j.rineng.2023.101621

[6] Ramadan, G.M., Abosinnee, A.S., Al Dulaimi, M.K.H., Jaffer, Z., Alkhayyat, A., Naser, Z.L. (2023). Hybrid deep learning enabled load prediction for energy storage systems. In 2023 6th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, pp. 466-472. https://doi.org/10.1109/IICETA57613.2023.10351206

[7] Zhang, C.X., Dong, M., Ren, M., et al. (2018). Partial discharge monitoring on metal-enclosed switchgear with distributed non-contact sensors. Sensors, 18(2): 551. https://doi.org/10.3390/s18020551

[8] O, S., Bimaridi, A., Park, J., et al. (2022). A study on the monitoring parameters for power asset management of switchgear. In 2022 9th International Conference on Condition Monitoring and Diagnosis (CMD),

Kitakyushu, Japan, pp. 305-308. https://doi.org/10.23919/CMD54214.2022.9991476
[9] Álvarez Gómez, F., Albarracín-Sánchez, R., Garnacho Vecino, F., Granizo Arrabé, R. (2018). Diagnosis of insulation condition of MV switchgears by application of different partial discharge measuring methods and sensors. Sensors, 18(3): 720. https://doi.org/10.3390/s18030720
[10] Iadarola, G., Mingotti, A., Negri, V., Spinsante, S. (2025). Fault detection in MV switchgears through unsupervised learning of temperature conditions. Sensors, 25(15): 4818. https://doi.org/10.3390/s25154818
[11] Chen, L., Li, Q., Long, G.H., Zou, Y., Deng, J. (2025). Cable partial discharge identification network based on adaptive residual diffusion denoising and morphological attention. Scientific Reports, 15(1): 42848. https://doi.org/10.1038/s41598-025-25197-9
[12] Riba, J.R., Moreno-Eguilaz, M., Bogarra, S. (2023). Tracking resistance in polymeric insulation materials for high-voltage electrical mobility applications evaluated by existing test methods: Identified research needs. Polymers, 15(18): 3717. https://doi.org/10.3390/polym15183717
[13] Nazir, M.T., Khalid, A., Akram, S., et al. (2023). Electrical tracking, erosion and flammability resistance of high voltage outdoor composite insulation: Research, innovation and future outlook. Materials Science and Engineering: R: Reports, 156: 100757. https://doi.org/10.1016/j.mser.2023.100757
[14] Alqudsi, A.Y., Ghunem, R.A., David, É. (2021). Analyzing the role of filler interface on the erosion performance of filled RTV silicone rubber under DC dry-band arcing. IEEE Transactions on Dielectrics and Electrical Insulation, 28(3): 788-796. https://doi.org/10.1109/TDEI.2021.009337
[15] Ayubi, B.I., Zhang, L., Xu, H.K., Yan, Q.Y. (2022). Surface discharge characteristics and influencing factors of polyimide insulations under high-frequency sinusoidal voltages. IET Generation, Transmission & Distribution, 16(5): 1000-1012. https://doi.org/10.1049/gtd2.12344
[16] Nazir, M.T., Butt, F.T., Phung, B.T., et al. (2020). Simulation and experimental investigation on carbonized tracking failure of EPDM/BN-based electrical insulation. Polymers, 12(3): 582. https://doi.org/10.3390/polym12030582
[17] Ramesh, M., Cui, L., Shtessel, Y., Gorur, R. (2021). Failure studies of polymeric insulating materials using sliding mode observer. International Journal of Electrical Power & Energy Systems, 126: 106539. https://doi.org/10.1016/j.ijepes.2020.106539
[18] Alsumaidaee, Y.A.M., Koh, S.P., Yaw, C.T., Tiong, S.K., Chen, C.P. (2023). Detecting surface discharge faults in switchgear by using hybrid model. Indonesian Journal of Electrical Engineering and Computer Science, 32(1): 413-422. https://doi.org/10.11591/ijeecs.v32.i1.pp413-422
[19] Ishak, S., Koh, S.P., Tan, J.D., Tiong, S.K., Chen, C.P. (2020). Corona fault detection in switchgear with extreme learning machine. Bulletin of Electrical Engineering and Informatics, 9(2): 558-564. https://doi.org/10.11591/eei.v9i2.2058
[20] Mohammed Alsumaidaee, Y.A., Yaw, C.T., Koh, S.P., et al. (2023). Detecting arcing faults in Switchgear by using deep learning techniques. Applied Sciences, 13(7): 4617. https://doi.org/10.3390/app13074617
[21] Alsumaidaee, Y.A.M., Yahya, M.M., Yaseen, A.H. (2024). Reliable mechanical fault diagnosis in medium voltage electrical switchgear using a 1D-convolutional neural network. Journal Européen des Systèmes Automatisés, 57(5): 1461-1470. https://doi.org/10.18280/jesa.570521
[22] Jia, Z.J., Fan, S.H., He, D.M., Wang, J.Y., Dong, H.B., Tang, Z.G. (2021). Multi-physical tracking and sensing technology of switchgear partial discharge based on edge computing and smart gateway. Journal of Physics: Conference Series, 1952(4): 042101. https://doi.org/10.1088/1742-6596/1952/4/042101
[23] Xing, Y.Q., Wang, Z.W., Liu, L., et al. (2022). Defects and failure types of solid insulation in gas-insulated switchgear: In situ study and case analysis. High Voltage, 7(1): 158-164. https://doi.org/10.1049/hve2.12127
[24] Zhao, L.H., Wu, Y.Z., Huang, X.L., et al. (2022). Research on the temperature rise characteristics of medium-voltage switchgear under different operation conditions. IEEJ Transactions on Electrical and Electronic Engineering, 17(5): 654-664. https://doi.org/10.1002/tee.23553
[25] Chen, S.Y., Illias, H.A., Wong, J.K.R., Mansor, N.N. (2024). Denoising of partial discharges in switchgear insulation material using hybrid wavelet denoising-optimization-machine learning. Ain Shams Engineering Journal, 15(11): 103032. https://doi.org/10.1016/j.asej.2024.103032
[26] Dong, H.K., Zhang, Y.X., Zhang, B., et al. (2026). A noise-robust xLSTM-based method for fault diagnosis in gas-insulated switchgear using acoustic signals under strong noise interference. Expert Systems with Applications, 304: 130827. https://doi.org/10.1016/j.eswa.2025.130827
[27] Wang, Z.H., Zhang, Z.P., Shao, Y.Y., et al. (2024). Gas-insulated switch-gear mechanical fault detection based on acoustic feature analysis using a multi-state pre-trained neural network. Measurement Science and Technology, 35(7): 076121. https://doi.org/10.1088/1361-6501/ad3d78
[28] Zhang, Z.P., Liu, H.G., Yuan, G.G., et al. (2024). Gas-insulated switch-gear mechanical fault detection based on acoustic using feature fused neural network. Electric Power Systems Research, 230: 110226. https://doi.org/10.1016/j.epsr.2024.110226
[29] Ishaq, A., Junaid, M., Hussain, G.A., Khan, S.U., Chen, Y., Yu, D.S. (2025). Partial discharge defect classification in MV switchgear by using CWT and deep learning approach. IEEE Transactions on Instrumentation and Measurement, 74: 1-12. https://doi.org/10.1109/TIM.2025.3562981
[30] Jia, Z.J., Fan, S.H., Wang, Z.C., Shao, S., He, D.M. (2025). Partial discharge defect recognition method of switchgear based on cloud-edge collaborative deep learning. Scientific Reports, 15(1): 10956. https://doi.org/10.1038/s41598-024-81478-9
[31] Fei, Z.J., Li, Y.Y., Yang, S.Y. (2024). Pattern recognition of partial discharge faults in switchgear using a back propagation neural network optimized by an improved mantis search algorithm. Sensors, 24(10): 3174. https://doi.org/10.3390/s24103174
[32] Chang, C.K., Lin, Y.H. (2025). Defect recognition for

partial discharge patterns of gas insulated switchgear and cable joint based on deep learning methods. IEEE Transactions on Dielectrics and Electrical Insulation, 32(2): 1147-1154. https://doi.org/10.1109/TDEI.2024.3431431

[33] Yang, J.R., Hu, K.L., Zhang, J., Bao, J.S. (2024). Semi-supervised learning for gas insulated switchgear partial discharge pattern recognition in the case of limited labeled data. Engineering Applications of Artificial Intelligence, 137: 109193. https://doi.org/10.1016/j.engappai.2024.109193

[34] Jing, Q.Z., Yan, J., Wang, Y.X., He, R.X., Lu, L. (2022). A novel differentiable neural network architecture automatic search method for GIS partial discharge pattern recognition. Measurement, 195: 111154. https://doi.org/10.1016/j.measurement.2022.111154

[35] Cho, K., Van Merriënboer, B., Gulçehre, Ç., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1724-1734. https://doi.org/10.3115/v1/D14-1179

[36] Hewamalage, H., Bergmeir, C., Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1): 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[37] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A. (2019). Deep learning for time series classification: A review. Data Mining and Knowledge Discovery, 33(4): 917-963. https://doi.org/10.1007/s10618-019-00619-1

[38] Iqbal, R., Maniak, T., Doctor, F., Karyotis, C. (2019). Fault detection and isolation in industrial processes using deep learning approaches. IEEE Transactions on Industrial Informatics, 15(5): 3077-3084. https://doi.org/10.1109/TII.2019.2902274

[39] Zhang, S., Zhang, S.B., Wang, B.N., Habetler, T.G. (2020). Deep learning algorithms for bearing fault diagnostics—A comprehensive review. IEEE Access, 8: 29857-29881. https://doi.org/10.1109/ACCESS.2020.2972859

[40] Ishak, S., Yaw, C.T., Koh, S.P., Tiong, S.K., Chen, C.P., Yusaf, T. (2021). Fault classification system for switchgear CBM from an ultrasound analysis technique using extreme learning machine. Energies, 14(19): 6279. https://doi.org/10.3390/en14196279

[41] Alsumaidaee, Y.A.M., Paw, J.K.S., Yaw, C.T., et al. (2023). Fault detection for medium voltage switchgear using a deep learning hybrid 1D-CNN-LSTM model. IEEE Access, 11: 97574-97589. https://doi.org/10.1109/ACCESS.2023.3294093

[42] Sun, X., Chen, Y., Wei, J., Liu, Q., Guo, H., Cheng, R.J. (2025). Switch cabinet temperature prediction using a fusion of CNN and LSTM neural networks. Applied System Innovation, 8(5): 157. https://doi.org/10.3390/asi8050157

[43] Kaštelan, N., Vujović, I., Krčum, M., Assani, N. (2022). Switchgear digitalization—Research path, status, and future work. Sensors, 22(20): 7922. https://doi.org/10.3390/s22207922

[44] Hussain, G.A., Zaher, A.A., Hummes, D., Safdar, M., Lehtonen, M. (2020). Hybrid sensing of internal and surface partial discharges in air-insulated medium voltage switchgear. Energies, 13(7): 1738. https://doi.org/10.3390/en13071738

[45] Kumar, H., Shafiq, M., Kauhaniemi, K., Elmusrati, M. (2024). A review on the classification of partial discharges in medium-voltage cables: Detection, feature extraction, artificial intelligence-based classification, and optimization techniques. Energies, 17(5): 1142. https://doi.org/10.3390/en17051142

[46] Zheng, S.P., Liu, J.F., Zeng, J. (2025). A partial discharge pattern recognition method based on multi-scale adaptive denoising network and stacking ensemble learning. Electric Power Systems Research, 241: 111392. https://doi.org/10.1016/j.epsr.2024.111392

[47] Liu, Y.Q., Li, W.D., Hua, X.C., Liu, H.Y., Li, H., Zhang, G.J. (2025). A Swin transformer enhanced depthwise convolution with dynamic channel reweighting for partial discharge diagnosis in gas insulated switchgear. Electric Power Systems Research, 248: 111993. https://doi.org/10.1016/j.epsr.2025.111993

[48] Deng, Y., Xie, Q., Chen, J.W., Tan, D.P., Liu, H. (2025). KSMNet: Learn the core discriminant features and semantic relationships for partial discharge pattern recognition via Mamba. Measurement, 246: 116683. https://doi.org/10.1016/j.measurement.2025.116683

[49] Wang, Y.X., Yan, J., Jing, Q.Z., Wang, J.H., Geng, Y.S. (2022). A novel federated deep learning framework for diagnosis of partial discharge in gas-insulated switchgear. Measurement Science and Technology, 33(9): 095112. https://doi.org/10.1088/1361-6501/ac7a09

[50] Yan, J., Wang, Y.X., Zhou, Y., Wang, J.H., Geng, Y.S. (2024). A novel meta-learning network for partial discharge source localization in gas-insulated switchgear via digital twin. IET Generation, Transmission & Distribution, 18(9): 1785-1794. https://doi.org/10.1049/gtd2.13156

[51] Han, X.T., Wang, H.T., Cui, J., et al. (2025). Gas-insulated switchgear partial discharge classification method based on deep transfer learning using experimental and field data. High Voltage, 10(4): 845-855. https://doi.org/10.1049/hve2.70088

[52] Sun, H.B., Otake, Y., Matsuyama, K., Raghunathan, A. (2025). Switchgear partial discharge diagnosis using scarce fault records. In 2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Valletta, Malta, pp. 1-5. https://doi.org/10.1109/ISGTEurope64741.2025.11305301

[53] Almehdhar, A., Prochazka, R. (2025). Classification of multiple partial discharge sources using time-frequency analysis and deep learning. Applied Sciences, 15(10): 5455. https://doi.org/10.3390/app15105455

[54] Li, Q.Y., Luo, H., Cheng, H., et al. (2023). Incipient fault detection in power distribution system: A time–frequency embedded deep-learning-based approach. IEEE Transactions on Instrumentation and Measurement, 72: 1-14. https://doi.org/10.1109/TIM.2023.3250220

[55] Florkowski, M. (2020). Classification of partial discharge images using deep convolutional neural networks. Energies, 13(20): 5496. https://doi.org/10.3390/en13205496

[56] Zheng, Q.H., Wang, R.Y., Tian, X.Y., et al. (2023). A

real-time transformer discharge pattern recognition method based on CNN-LSTM driven by few-shot learning. Electric Power Systems Research, 219: 109241. https://doi.org/10.1016/j.epsr.2023.109241

[57] Miao, J.J., Zhu, W. (2022). Precision–recall curve (PRC) classification trees. Evolutionary Intelligence, 15(3): 1545-1569. https://doi.org/10.1007/s12065-021-00565-2

[58] McFee, B., Raffel, C., Liang, D., et al. (2015). Librosa: Audio and music signal analysis in Python. In Proceedings of the 14th Python in Science Conference, Austin, Texas, USA, pp. 18-24. https://doi.org/10.25080/Majora-7b98e3ed-003

[59] Shao, N., Zhou, R., Wang, P.Y., et al. (2025). CleanMel: Mel-spectrogram enhancement for improving both speech quality and ASR. IEEE Transactions on Audio, Speech and Language Processing, 33: 3202-3214. https://doi.org/10.1109/TASLPRO.2025.3592333

[60] Zhang, Z.J. (2018). Improved Adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, pp. 1-2. https://doi.org/10.1109/IWQoS.2018.8624183

[61] Goutte, C., Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Advances in Information Retrieval, Springer, Berlin, Heidelberg, pp. 345-359. https://doi.org/10.1007/978-3-540-31865-1_25

[62] Saito, T., Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One, 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432

[63] Alsumaidaee, Y.A.M., Yaw, C.T., Koh, S.P., et al. (2023). Detection of corona faults in switchgear by using 1D-CNN, LSTM, and 1D-CNN-LSTM methods. Sensors, 23(6): 3108. https://doi.org/10.3390/s23063108