

ORIGINAL RESEARCH

Open access

Efficient Fault Detection in Urban Power Grids Using Semi-Supervised Learning with Adaptive Thresholds

Amal Al-Balushi^{1*}, Saif Al-Hinai¹

Abstract

With the rapid expansion of urban power grids and the widespread adoption of renewable energy sources, conventional fault-diagnosis methods for power grids face significant challenges in handling the complexities of modern, highly interconnected systems. While artificial intelligence approaches offer promising alternatives for power grid fault diagnosis, the scarcity of labeled grid data severely hinders the application of AI techniques in this domain. To tackle these issues, this study introduces a semi-supervised learning framework featuring self-supervised learning and an adaptive threshold mechanism (SAT-SSL) designed for fault detection and classification in power grids. Compared with existing approaches, the proposed method substantially reduces reliance on labeled data while maintaining excellent recognition performance. First, a frequency-domain analysis is applied to power grid data to identify and filter anomalous events, which are then classified and labeled based on their visual characteristics, thereby constructing a dedicated power grid dataset. Next, the Yule–Walker algorithm is employed to extract features from the power grid data. A semi-supervised learning framework is subsequently developed, integrating a self-supervised loss and a dynamic threshold mechanism to enhance the model's information extraction capacity and adaptability across diverse operational scenarios. Finally, the constructed power grid dataset, together with two standard benchmark datasets, is utilized to evaluate the model's effectiveness. Experimental results demonstrate that the proposed model maintains a low error rate across multiple scenarios and varying label proportions. Specifically, on the power grid dataset, retaining only 5% of the labels yields an error rate of just 6.15%, confirming the method's ability to achieve precise fault detection and classification even with very limited labeled data.

Keywords Smart grid, Power grid fault detection, Semi-supervised learning, Data-driven, Deep learning

*Correspondence:

Amal Al-Balushi
amal.balushi@squ.edu.om

¹ Department of Smart Energy and Grid Systems, Faculty of Engineering, Sultan Qaboos University, Muscat, Oman

Introduction

Urban power grids serve as the essential backbone of city life and economic activity, bearing the critical responsibility of delivering reliable and continuous electricity to households and enterprises. As renewable energy technologies advance, urban energy structures are undergoing continuous optimization and transformation. As a result, the proportion of photovoltaic and wind power in urban power grids continues to rise, making substantial contributions to improving environmental quality. Nevertheless, this transition introduces significant new challenges. The intermittent and variable characteristics of renewable energy sources increase the complexity and unpredictability of urban grid dynamics, thereby raising the difficulty of system maintenance. To ensure high-quality power delivery to consumers and enhance overall supply reliability, rapid and accurate detection and classification of grid faults have become a key research priority [1].

In power grid systems, fault occurrences lead to detectable alterations in electrical parameters such as current, voltage, and grid frequency [2]. Consequently, prompt analysis of these electrical signals enables not only a thorough comprehension of fault origins and evolution but also early identification of underlying causes. Moreover, swift detection and analysis establish an effective early-warning system for the power network, allowing operators to promptly implement appropriate protective actions and remedial measures, thereby preventing more severe consequences [3].

Conventional fault diagnosis techniques for power grids are predominantly based on electrical signal interpretation and encompass well-known strategies such as the Instantaneous Value Method [4], Distance Protection [5], Waveform Comparison Method [6], Traveling Wave Method [7], Harmonic Analysis Method [8], and State Estimation [9], along with various related procedures. Moreover, numerous composite strategies that merge several of these techniques have been proposed over time [10].

While these classic fault-detection approaches delivered strong performance during the initial phases of power grid diagnostics, the continuous growth in scale and topological complexity of today's grids has exposed serious drawbacks. In particular, they encounter substantial obstacles when attempting to represent very large-scale systems or to efficiently process massive datasets, which severely restricts their suitability for reliable fault detection and categorization tasks within modern metropolitan power networks.

The escalating structural and operational complexity of city-level power grids has rendered legacy supervisory systems and conventional digital acquisition equipment inadequate for fulfilling current monitoring requirements. Fortunately, the emergence of Phasor Measurement Units (also referred to as synchronous phasor measurement units, PMUs) has provided a transformative sensing technology that markedly improves both precision and temporal resolution of grid observation [11]. These devices operate at sampling rates far exceeding those of earlier measurement instruments. With ongoing improvements in performance and reductions in deployment cost, PMU technology is receiving steadily increasing recognition within the power system monitoring community [12]. Researchers have even released an open-source counterpart to proprietary PMU hardware—termed OpenPMU—specifically tailored to academic and experimental applications [13].

In contrast to standard metering devices, PMUs facilitate broad-area, synchronized, high-accuracy observation across entire power networks, delivering precise measurements of phase angles, voltage magnitudes, current magnitudes, and frequency. This capability yields significantly finer-grained insight into grid behavior. Capitalizing on the rich, high-fidelity datasets produced by PMUs, the work presented in Ref. [14] effectively implemented islanding detection for distributed generation installations, achieving highly satisfactory performance. Likewise, Ref. [15] developed a real-time, multi-level scheme for detecting, classifying, and pinpointing faults throughout power systems by exploiting PMU measurements. In Ref. [16], low-frequency oscillation early-warning computations were performed by integrating PMU-derived data with the Fuzzy Kernel Hierarchical Analysis Evaluation approach. Collectively, the adoption of PMU technology significantly strengthens the foundation for reliable, granular fault condition assessment in power grids.

Fueled by substantial advances in artificial intelligence (AI), these systems now demonstrate remarkable capacity for autonomous parameter tuning, optimization, handling of enormous data volumes, and effective management of the strongly nonlinear phenomena typical of electrical power networks. As a direct consequence, AI methodologies have achieved broad penetration throughout the electric power domain [17]. This intrinsic strength positions AI as particularly well-suited to the demands of contemporary, highly intricate grid infrastructures [18].

Indeed, intelligent algorithms are now routinely deployed across diverse facets of grid analytics and control. For example, Ref. [19] adopted a Graph Neural Network architecture to forecast nodal power flows in transmission systems. Ref. [20] introduced a streamlined federated learning architecture to distinguish various categories of grid disturbances. A considerable body of work has likewise focused specifically on intelligent fault diagnosis. Ref. [21] combined an auto-encoder network with a multi-layer extreme learning machine to identify and categorize faults occurring on transmission lines. In Ref. [22], fault diagnosis in power systems was accomplished using Artificial Neural Networks that utilized nodal

voltage and current information extracted from network models, with particular attention to the effects of varying numbers of hidden layers. Ref. [23] applied Fast Fourier Transform preprocessing together with a Support Vector Machine whose parameters were tuned via Particle Swarm Optimization to diagnose load-side faults inside indoor distribution networks. The contribution in Ref. [24] fused Discrete Wavelet Transform features with an innovative self-attention-based Convolutional Neural Network to address transmission network fault identification. Ref. [25] similarly paired Discrete Wavelet Transform signal processing with Support Vector Machine classification for the purpose of detecting and classifying grid faults. Ref. [26] employed Deep Recursive Neural Networks in conjunction with LSTM units to perform fault zone determination, fault type recognition, and fault distance estimation. Finally, Ref. [27] proposed a weighted, corrective, fuzzy-reasoning spiking neural P system tailored specifically for power grid fault diagnosis.

All the methods discussed above leveraged deep learning models to directly extract features from raw data for fault diagnosis. In their respective investigations, these approaches delivered promising performance outcomes. Nevertheless, power grid monitoring infrastructure continuously generates enormous volumes of operational measurements. Manually annotating and categorizing such extensive datasets requires considerable human effort and financial investment [28]. As a result, these fully supervised deep learning techniques depend heavily on large volumes of labeled examples; their performance tends to degrade sharply when only a small number of labeled examples are available.

Semi-supervised learning (SSL) algorithms, by contrast, make effective use of a limited quantity of labeled instances combined with abundant unlabeled data during training, providing a powerful strategy for data-scarce conditions. Emerging in the 1970s [29], semi-supervised learning has since evolved into a diverse family of techniques. The rapid progress in deep learning has led to Deep Semi-Supervised Learning (DSSL), which has gained widespread adoption across many fields. As outlined in [29], contemporary DSSL approaches can be grouped into five main categories: generative-model-based methods, including GAN-derived frameworks such as Triple GAN [30], CT-GAN [31], and Variational Autoencoder-based models like SSSVAEs [32]; consistency-regularization-based techniques, for example the Ladder Network [33] and Π -model [34]; graph-based strategies such as DNCR [35] and Graph Convolutional Networks (GCN) [36]; pseudo-labeling-oriented methods, including Tri-Net [37] and Pseudo-Label [38]; and hybrid approaches like MixMatch [39] and FixMatch [40].

Semi-supervised learning has already been successfully applied to multiple fault diagnosis tasks. For example, Ref. [41] applied a semi-supervised hybrid classification autoencoder to diagnose faults in motor bearing vibration data. Ref. [42] developed a self-training semi-supervised deep learning procedure tailored for mechanical fault identification, while Ref. [43] introduced a deep neural network framework augmented with Adaptive Reinforcement (AR) semi-supervised transfer learning to address transformer fault diagnosis. Across these studies, semi-supervised strategies markedly decreased dependence on labeled samples while achieving higher diagnostic accuracy, thereby opening promising new avenues for fault diagnosis in electrical power grids.

To date, however, the adoption of semi-supervised learning specifically for power grid fault diagnosis remains relatively scarce. Although many existing SSL methods effectively combine consistency regularization with pseudo-labeling, they often fail to fully exploit the information in unlabeled samples. Moreover, the FixMatch algorithm adopts an overly stringent confidence threshold to improve pseudo-label reliability, thereby discarding a large proportion of candidate samples and substantially extending training time.

Drawing inspiration from the Double-Match framework [44], the present work builds on the FixMatch architecture to introduce a novel semi-supervised learning approach that combines self-supervision with a dynamic adaptive threshold (SAT-SSL). This proposed method simultaneously leverages information from both the feature representation and the pseudo-label spaces while adaptively modulating the confidence threshold for pseudo-label assignment. As a consequence, it improves overall model accuracy, increases the effective utilization of unlabeled data, and shortens training time. When applied to power grid fault diagnosis, experimental evaluations confirm that SAT-SSL attains high-precision fault identification and classification even under severely constrained labeled-data regimes.

The principal contributions of this study are summarized as follows:

(1) Given the intricate nature of power grid measurements and the challenges involved in directly interpreting grid events, this research combines Fast Fourier Transform (FFT) with the Yule–Walker algorithm to efficiently detect and characterize anomalous signals within the grid. Events are subsequently classified and labeled according to the visual patterns observed in the power grid data. By employing the Yule–Walker method for power spectral density estimation, the approach uncovers distinctive time-frequency signatures associated with different fault types, thereby markedly improving the precision of fault localization.

(2) In light of the massive scale of power grid data—which renders direct analysis difficult—and the prohibitive expense of manual annotation, this paper presents a semi-supervised learning framework featuring self-supervised pre-training and dynamic threshold tuning (SAT-SSL). The framework creatively fuses a small set of labeled examples with a much larger pool of unlabeled data, exhibiting outstanding performance in accurate grid fault diagnosis. Validation experiments conducted on both publicly available benchmark datasets and real-world power grid records highlight the clear superiority and practical applicability of the SAT-SSL method.

Materials and Methods

This section describes the proposed SAT-SSL approach, organized into three main components: data analysis, data preprocessing, and model construction and training.

The Data Analysis component processes operational measurements from phasor measurement units (PMUs) by applying the Fast Fourier Transform (FFT) and the Yule–Walker method to detect and remove irregularities in the grid data. After that step, the detected irregularities are grouped by visual pattern, and appropriate labels are assigned to the dataset accordingly.

In the data preprocessing component, raw measurements first pass through noise reduction and scaling procedures. Then the Yule–Walker technique is used to compute power spectral density estimates, from which time–frequency representations are created for successive time intervals.

During model construction and training, the SAT-SSL architecture combines consistency constraints, pseudo-label generation, a self-supervised objective, and adaptive threshold mechanisms to build an effective semi-supervised training pipeline. Training is performed on the labeled power grid dataset prepared earlier in the data analysis component. A schematic overview of how the SAT-SSL algorithm is applied to power grid fault identification appears in **Figure 1**.

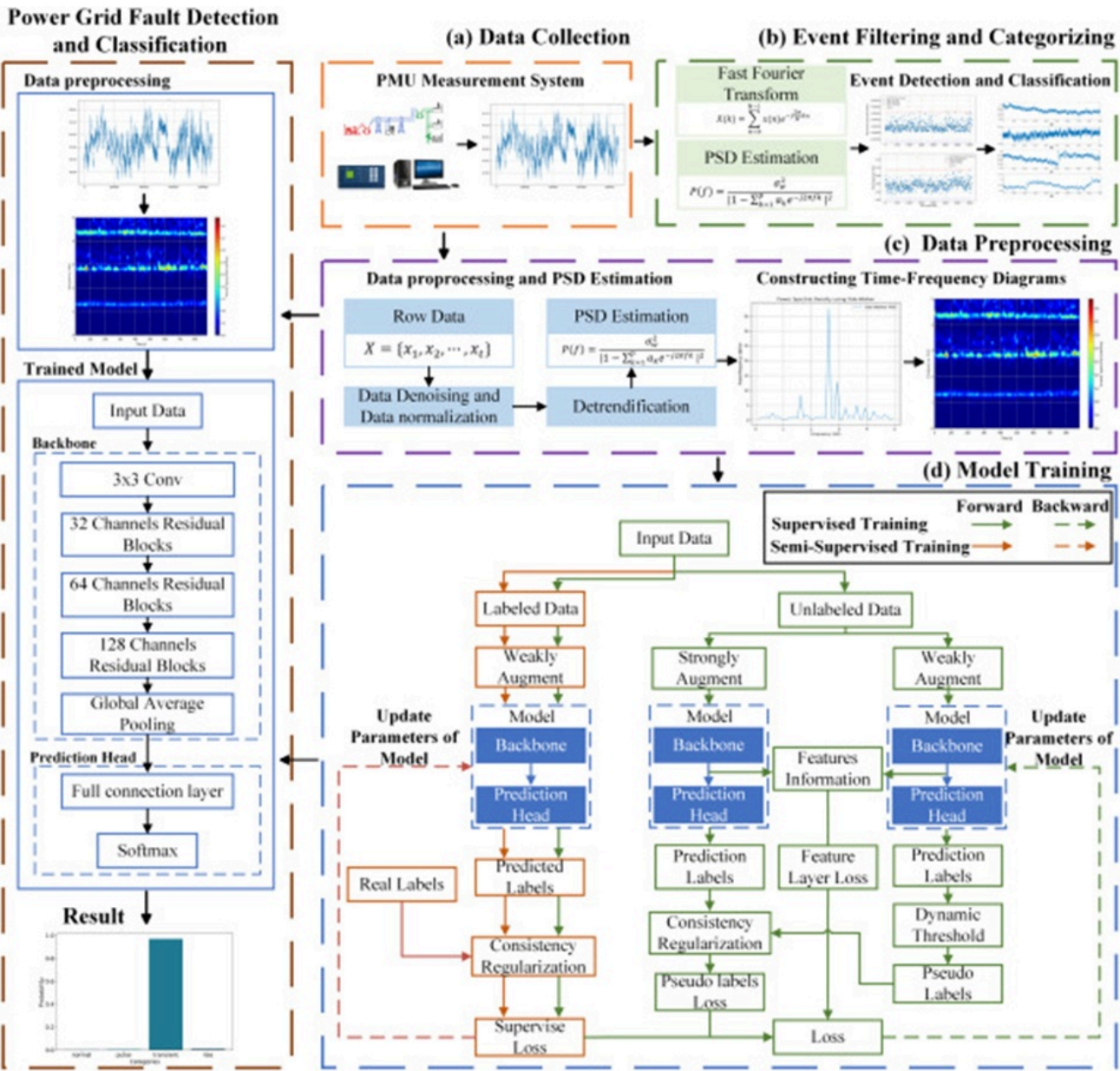


Figure 1. Overall structure of the SAT-SSL framework.

Data analysis

The measurements analyzed in this work were recorded from an operating electric power network. Recorded variables include PMU identifiers, time stamps, voltage magnitudes, frequency values, and voltage phase angles, all captured at a uniform sampling rate of 10 Hz.

Voltage signals exhibit high variability and are affected by many external factors, which makes them difficult to use for reliable event detection. Therefore, the present study focuses exclusively on frequency measurements to detect and categorize disturbances in the power grid.

With a sampling frequency of 10 Hz, the analysis is inherently limited to low-frequency oscillatory phenomena. Faster transients—such as those caused by capacitor bank switching or circuit breaker operations—cannot be reliably observed

from PMU recordings alone. Even so, collecting 10 samples per second still produces a large volume of data over a single day.

The core focus of the research lies in major system disturbances and sharp fluctuations originating from renewable generation. Relevant events encompass (among others) transmission line short circuits, relay tripping actions, and large step changes (upward or downward) in wind farm active power output.

Events of this magnitude can excite one or more electromechanical oscillation modes throughout the interconnected system. These modes manifest as sustained or decaying oscillations, which can be identified by examining the frequency components of PMU signals.

Disturbances affecting system frequency are broadly divided into three types: impulse-like events, transient oscillatory/damped events, and stepwise frequency excursions (rises or drops).

Impulse-like disturbances commonly result from brief local mismatches between generation and load, or from switching operations involving equipment. Possible causes include device activation, energization of large motors, line tripping, component failures, overload conditions, or short-circuit faults.

Transient oscillatory or damped events typically occur due to malfunctions in generating units or during electrical islanding within a region. Sharp frequency increases or decreases usually arise from sudden generation–load imbalances, unexpected variations in generator output (stemming from equipment faults, scheduled maintenance, or dispatch decisions), or faults on transmission corridors.

In most cases, these oscillations are dominated by a principal mode lying in the frequency band 0.1 Hz–2.0 Hz, and their amplitudes are noticeably larger than those produced by ordinary random load fluctuations. Provided the system remains stable, the oscillations gradually decay until the frequency returns to its nominal value.

To highlight low-frequency oscillatory activity, the frequency time series within each analysis window is differentiated to remove any constant (DC) bias. Although this operation increases the apparent noise level, genuine disturbance events produce frequency deviations large enough that background noise does not significantly interfere with detection.

The primary tools employed in this subsection for anomaly identification are spectral analysis and power spectral density estimation. Both methods operate using a sliding-window procedure. Each PMU data file is partitioned into consecutive segments of 10-second length, advanced in steps of 5 seconds.

The initial anomaly screening step relies on frequency-domain examination. Fast Fourier Transform (FFT) is applied to hourly blocks of PMU frequency data. Within every 10-second window, the FFT identifies the frequency bin between 0 Hz and 5 Hz with the highest magnitude and stores that peak amplitude.

After processing the complete dataset, all recorded peak amplitudes are collected and used to calculate their global mean and standard deviation. Any window whose peak amplitude exceeds the mean value by more than three standard deviations is flagged as a potential anomalous event.

Subsequently, power spectral density (PSD) estimation is applied to identify anomalies in power grid measurements. Given the large dataset and the need to minimize computational effort during anomaly filtering, the Yule–Walker algorithm is selected for PSD estimation in this work. The Yule–Walker approach relies on fitting an autoregressive (AR) model and is widely used to estimate its parameters. Here, it serves to estimate the power spectral density of the grid frequency time series.

An autoregressive (AR) model of order p is mathematically expressed as follows:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \quad (1)$$

Where X_t denotes the value of the modeled time series at time instant t , c represents the constant (intercept) term of the model, ϕ_i (for $i = 1, \dots, p$) are the autoregressive coefficients that quantify the dependence of the current value X_t on its previous p observations, and ϵ_t is a zero-mean white noise process.

To estimate the power spectral density (PSD) from observed data, the first step is to compute the sample autocovariance function. For a time series $X(t)$ of length N , the autocovariance at lag k is estimated by:

$$R(p) = \frac{1}{N} \sum_{t=1}^{N-p} x(t) x(t+p) \quad (2)$$

Where N is the total number of samples in the series.

Based on this autocovariance estimate, the Yule–Walker equations are formulated as:

$$R\phi = r \quad (3)$$

In the matrix notation above, R is a $p \times p$ symmetric Toeplitz matrix whose first column consists of the values $R(0), R(1), \dots, R(p-1)$; r is the column vector $\phi(0), \phi(1), \dots, \phi(p)$ containing the AR coefficients (note that $\phi_0 = 1$ by convention is often omitted); and r is the column vector $R(0), R(1), \dots, R(p-1)$.

Solving the linear system in [Eq. 3](#) yields the AR coefficient vector ϕ . For an AR model of order p , the corresponding power spectral density (P) is then given by:

$$P(f) = \frac{\sigma^2}{|1 - \sum_{i=1}^p \phi_i e^{-j2\pi f i}|^2} \quad (4)$$

Where σ^2 is the variance of the driving white noise, f represents frequency (ranging from 0 to half the sampling rate), and j is the imaginary unit.

The resulting PSD estimates are expressed in decibel (dB) scale, and the peak amplitude within each spectral estimate is identified and recorded across all data windows. The mean and standard deviation of these maximum peak amplitudes are subsequently computed. Any peak amplitude that exceeds the mean by more than three standard deviations is flagged as a candidate anomalous event.

After processing the PMU measurements with both spectral estimation techniques (FFT and Yule–Walker), all time instants associated with detected anomalous points are collected. A timestamp is considered an outlier only when both methods independently identify it as anomalous at the same time and location.

Model framework

This subsection presents the SAT-SSL algorithm, which stands for Semi-Supervised learning enhanced by Self-supervised learning and Adaptive Thresholding (SAT-SSL). The complete training procedure is illustrated in [Figure 2](#).

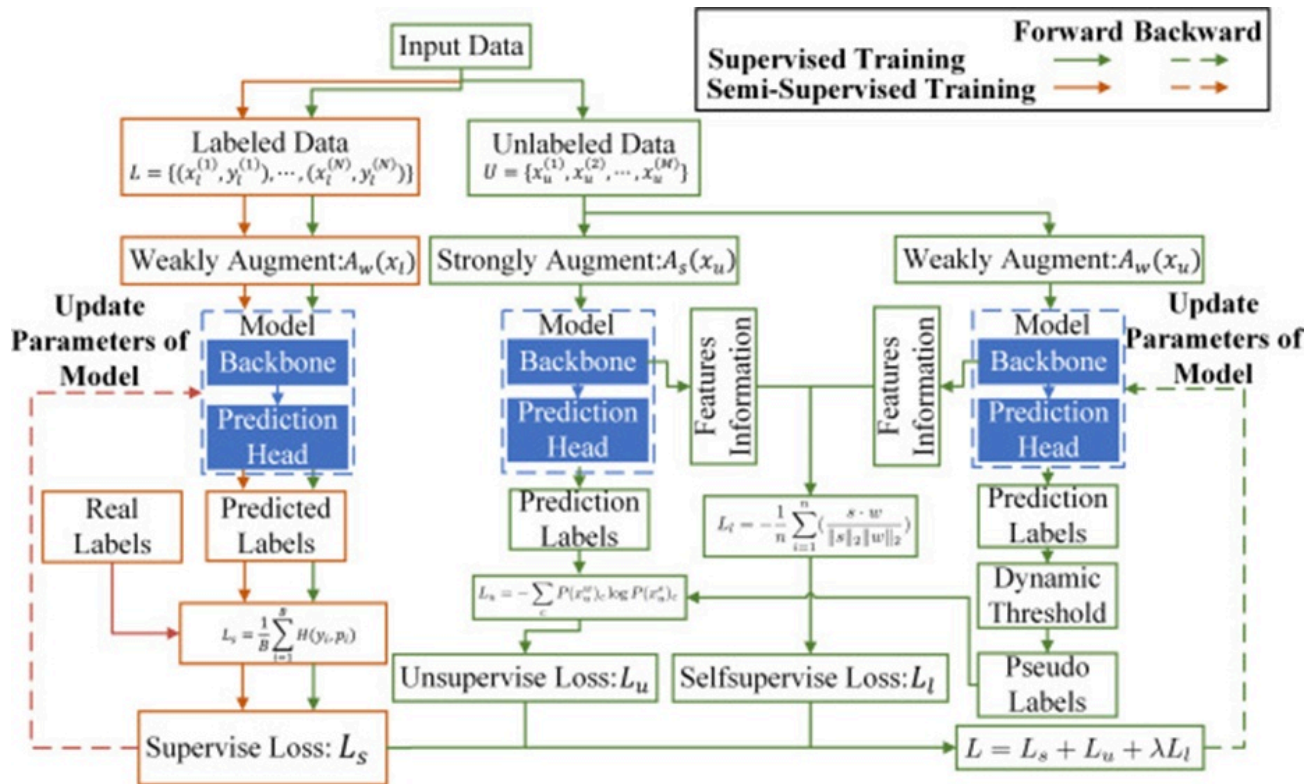


Figure 2. Training strategy of SAT-SSL.

Data augmentation

In semi-supervised learning, data augmentation is a key mechanism for improving model generalization by increasing data diversity. Labeled samples are typically limited and expensive to obtain, whereas unlabeled data is plentiful yet often under-exploited. Data augmentation helps close this gap by enabling effective use of unlabeled examples. In a semi-supervised setting, it supports the creation of pseudo-labels, which enlarge the effective training set and act as a form of regularization to prevent overfitting to the small amount of labeled data. Moreover, when combined with consistency regularization, data augmentation strengthens the model's resilience to input variations, ultimately enhancing performance on previously unseen samples.

The proposed approach adopts two categories of data augmentation: weak and strong. Weak augmentation involves lightweight transformations, such as horizontal flipping or small random rotations, which introduce modest changes while preserving the essential content of the data. These gentle modifications make weak augmentation appropriate for generating reliable pseudo-labels.

In contrast, strong augmentation applies more intensive and diverse transformations, including methods like RandAugment, Cutout, and color jitter. However, because Cutout carries a risk of removing informative regions of the data, this work primarily relies on RandAugment. RandAugment is a fully automated augmentation strategy that randomly selects and applies a sequence of standard operations—such as color adjustments, rotations, and other geometric or photometric changes—without requiring a separate optimization step to determine the best policy.

Pseudo-labeling and consistency regularization

This work builds on a semi-supervised learning framework inspired by FixMatch [40], which combines pseudo-labeling and consistency regularization to leverage unlabeled data. The two core components are described below.

Pseudo-labeling involves using the model's current predictions to assign labels to unlabeled samples. For an unlabeled input x_u , the model outputs a predicted probability distribution P . The pseudo-label $\tilde{y}$ is then defined as:

$$\tilde{y} = \operatorname{argmax}_c P(c|x_u) \quad (5)$$

Where c indexes the possible classes. This pseudo-label is only accepted if the maximum class probability exceeds a fixed confidence threshold τ :

$$\text{if } \max_c P(c|x_u) \geq \tau \text{ then use } \tilde{y} \text{ else discard} \quad (6)$$

Consistency regularization assumes that the model should produce similar predictions across different augmented versions of the same input. In other words, small perturbations to the data should not lead to inconsistent classification decisions. To enforce this property, both weak and strong augmentations are applied to the original unlabeled sample x_u , yielding (x_u) (weakly augmented) and (x_u) (strongly augmented). The model is expected to generate matching predictions p and q for these two versions. The inconsistency between them is penalized using cross-entropy loss, so the unsupervised loss term is formulated as:

$$\mathcal{L}_u = \frac{1}{N} \sum_{b=1}^N \mathbf{1}(\max(q) \geq \tau) \ell(q, \tilde{y}) \quad (7)$$

Where $\mathcal{L}_u$ denotes the cross-entropy loss between the prediction on the strongly augmented input (q) and the (sharpened) prediction or pseudo-label derived from the weakly augmented input (p), the indicator function $\mathbf{1}[\cdot]$ is 1 when the confidence condition is satisfied and 0 otherwise, and the overall unsupervised loss $\mathcal{L}_u$ encourages consistency across augmentations only for high-confidence pseudo-labels.

Adaptive threshold

In the standard FixMatch algorithm, a fixed confidence threshold is applied to filter pseudo-labels from unlabeled data. While this strategy helps maintain high-quality pseudo-labels, it often discards a large portion of potentially useful unlabeled samples, thereby limiting data utilization and slowing the model's overall convergence. To overcome this limitation and accelerate training progress, the present work introduces an adaptive (dynamic) thresholding mechanism into the framework.

Recognizing that different classes exhibit varying levels of classification difficulty and that model accuracy improves progressively during training, this study assigns class-specific, time-varying thresholds. This design better accommodates the distinct learning dynamics of each category and enables more effective exploitation of the available unlabeled data.

More precisely, for every class k and at each training step t , an adjustment factor (k) is computed to dynamically modulate the threshold:

$$\Theta_t(k) = \alpha_t(k) \cdot \theta \quad (8)$$

Where (k) is the adaptive threshold applied to class k at time t , (k) is the adjustment coefficient at that step, and θ denotes the fixed base threshold. A lower adjustment factor for a given class indicates poorer current learning performance for that class; in such cases, the threshold is lowered to admit more samples into the training process.

The key challenge is to define a reliable evaluation metric for computing this adjustment factor without introducing extra computational cost or requiring a separate validation set. At the same time, the mechanism should impose minimal overhead on training speed. Prior work has shown that excluding noisier pseudo-labels while retaining higher-confidence ones can substantially reduce bias in the model [45]. Building on this insight, the proposed approach uses the proportion of unlabeled samples within each class that receive high-confidence predictions (i.e., predictions exceeding the current threshold) as a proxy for the class's current learning effectiveness.

The evaluation metric is therefore calculated as:

$$\gamma_t(k) = \sum_{n=1}^N \mathbb{1} \left[p_{m,t}(y|\mathbf{u}_n) \geq \Theta_t(k) \wedge \arg \max_y p_{m,t}(y|\mathbf{u}_n) = k \right] \quad (9)$$

Where N is the total number of unlabeled samples, $p_{m,t}(y|\mathbf{u}_n)$ is the model's predicted probability distribution over classes for unlabeled sample $\mathbf{u}_n$ at training step t , τ is the current confidence threshold used for pseudo-label selection. The indicator function counts how many unlabeled samples are assigned high-confidence predictions to class k . This ratio serves as the basis for normalizing the learning effect and subsequently updating the adaptive threshold.

To prevent any single class from dominating the adjustment process too strongly during the early training phase, the adjustment factor is computed based on both the highest prediction confidence across all classes and the proportion of unlabeled samples that remain unassigned (i.e., below each class's threshold). The resulting normalized adjustment factor is given by:

$$\alpha_t(k) = \frac{\gamma_t(k)}{\max \left\{ \max_k \gamma_t(k), N - \sum_k \gamma_t(k) \right\}} \quad (10)$$

Where (k) is the number of unlabeled samples predicted as class k with confidence above the threshold at time t , N is the total number of unlabeled samples, and $\alpha_t(k)$ represents the normalized learning effectiveness score for class k . Following Eq. 8, this factor is then used to derive the class-specific dynamic threshold (k) .

During pseudo-label assignment, only predictions whose confidence exceeds the adaptive threshold (k) for the corresponding class k are accepted. By substituting (k) for the fixed threshold τ in Eq. 7, the unsupervised loss term becomes:

$$\mathcal{L}_u = \frac{1}{N} \sum_{b=1}^N \mathbb{1}(\max(q) \geq \Theta_t(k)) \ell(q, \tilde{y}) \quad (11)$$

Self-supervision

Within the proposed semi-supervised learning framework, incorporating consistency regularization directly at the feature level—through a dedicated self-supervised loss term—greatly improves the quality and robustness of the learned feature representations. By enforcing feature invariance to small input variations (typically induced by data augmentation), this mechanism markedly improves the model's ability to generalize to novel, unseen samples. It leverages the large pool of unlabeled data by encouraging the model to extract rich, meaningful patterns from inputs in a label-free manner, thereby substantially reducing reliance on large-scale annotated datasets.

To make full use of each unlabeled sample, a projection head is inserted before the network's final classification layer. This projection head extracts feature representations from both strongly augmented and weakly augmented versions of

each input, enabling direct computation of the self-supervised loss in the feature space while simultaneously leveraging information from both the feature representations and the pseudo-label predictions.

The overall architecture comprises three main components: the backbone network B , the prediction (classification) head G , and the projection head H . After the backbone extracts features (denoted after the p layer), the feature vector follows two separate paths:

- One path passes through the classification head G to produce class probability predictions, from which pseudo-labels are derived, and the corresponding classification loss is computed.
- The other path feeds into the projection head H , mapping the features into a lower-dimensional projection space where the self-supervised consistency loss is calculated.

The self-supervised loss is defined based on the cosine similarity between feature representations of differently augmented views of the same original sample. It is formulated as:

$$\mathcal{L}_l = -\frac{1}{N} \sum_{i=1}^N \left(\frac{h(s_i) \cdot w_i}{\|h(s_i)\|_2 \|w_i\|_2} \right) = -\frac{1}{N} \sum_{i=1}^N (\cos(h(s_i), w_i)) \quad (12)$$

Where s and w represent the projected features obtained from the strongly augmented and weakly augmented versions of the same input, respectively; N is the batch size (number of samples in the batch); h denotes a trainable linear projection head; and $\text{sim}(\cdot, \cdot)$ is the cosine similarity function (temperature scaling τ is typically applied but omitted here for brevity as in the original formulation).

For the supervised portion of the training (applicable to labeled data in a multi-class problem with L classes), the standard cross-entropy loss is used and can be written as:

$$\mathcal{L}_s = \frac{1}{M} \sum_{i=1}^M H(y_l^{(i)}, p_m(y|A_w(x_l^{(i)}))) \quad (13)$$

Where p_i is the ground-truth probability (usually 0 or 1 in one-hot encoding) for class i , and $\hat{p}_i$ is the model's predicted probability for class i .

By integrating all three loss components—supervised loss on labeled data, unsupervised consistency loss via pseudo-labeling, and self-supervised feature-level consistency—the total loss for each training batch is expressed as:

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \lambda \mathcal{L}_l \quad (14)$$

Where $\lambda \in \mathbb{R}$ is a hyperparameter that balances the contribution of the unsupervised (pseudo-label-based) loss term $\mathcal{L}_u$. The complete training procedure of the SAT-SSL algorithm is summarized in **Algorithm 1**.

```

1: Input: Strong agumentation  $A_s(x)$ , weak augmentation  $A_w(x)$ ,  $\mathcal{L} = \{(x_l^{(m)}, y_l^{(m)}) : m \in (1, \dots, M)\}$ ,  $\mathcal{U} = \{x_u^{(n)} : n \in (1, \dots, N)\}$ , self-supervise loss
   weight  $\lambda$ , base threshod  $\theta$ , backbone  $B$ , prediction head  $R$ , and projection layer  $W$ .
2: ▷ Supervise loss for labeled data
3: for  $i = 1, \dots, M$  do
4:    $p_i = B \circ R(A_w(x_l^{(i)}))$ 
5: end for
6:  $\mathcal{L}_s = \frac{1}{N} \sum_{i=1}^N H(y_l^{(i)}, p_m(y|A_w(x_l^{(i)})))$ 
7: ▷ Predictions on unlabeled data
8: for  $i = 1, \dots, N$  do
9:    $q_i = B \circ R A_s(x_u^{(i)})$ 
10:   $p_i = B \circ R A_w(x_u^{(i)})$ 
11:   $s_i = B A_s(x_u^{(i)})$ 
12:   $w_i = B A_w(x_u^{(i)})$ 
13:  if  $\max_c P(c|x_u) \geq \Theta_i(k)$  then
14:     $\bar{y} = \operatorname{argmax}_c P(c|x_u)$ 
15:  end if
16: end for
17: Update threshod by  $\Theta_i(k) = a_i(k) \cdot \theta$ 
18: ▷ Unsupervise loss for unlabeled data
19:  $\mathcal{L}_u = \frac{1}{N} \sum_{b=1}^N \mathbb{I}(\max(q) \geq \Theta_i(k)) \ell(q, \bar{y})$ 
20: ▷ Self-Supervise loss for unlabeled data
21:  $\mathcal{L}_l = -\frac{1}{N} \sum_{i=1}^N (\cos(h(s_i), w_i))$ 
22: return  $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \lambda \mathcal{L}_l$ 

```

Algorithm 1. SAT-SSL algorithm

Wide residual networks

Wide Residual Networks (Wide ResNets) constitute an alternative formulation of the classic Residual Network (ResNet) architecture. The primary innovation consists in substantially increasing the network's width—defined as the count of feature channels within each convolutional layer—rather than extending its depth. This widening strategy promotes greater training stability and frequently delivers competitive or superior accuracy. For these reasons, the current research selects Wide Residual Networks as the core architecture for the proposed classification system. A schematic overview of the network configuration implemented here is shown in [Figure 3](#). The overall structure comprises an input layer, a sequence of residual blocks, one or more fully connected layers, and a final output layer. For clarity, the architecture is conceptually split into two portions: (i) the backbone, which incorporates everything preceding the fully connected layer, and (ii) the prediction head, which consists of the last fully connected layer together with the associated activation operations.

Let the input image be I with dimensions $H \times W \times C$ (where H and W are the height and width of the image, respectively, and C is the number of channels, typically 3).

$$I \in \mathbb{R}^{H \times W \times C} \quad (15)$$

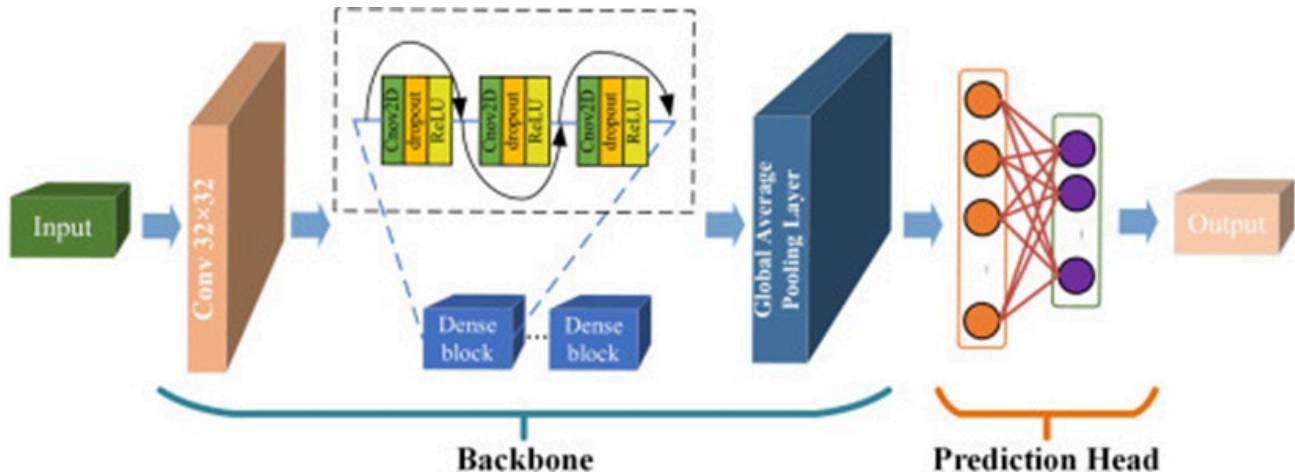


Figure 3. Architecture of wide residual networks.

The initial stage applies a convolutional filter K_1 , yielding the first feature representation F_1 according to:

$$F_1 = \text{ReLU}(I * K_1 + b_1) \quad (16)$$

Where $*$ stands for the 2D convolution operator, b_1 denotes the bias vector, and ReLU is the chosen point-wise non-linearity.

Inside each residual block that takes feature map F_{in} as input, the internal mapping is expressed as:

$$F_{\text{res}} = \text{ReLU}(F_{\text{in}} * K_{\text{res}} + b_{\text{res}}) \quad (17)$$

With K_{res} indicating the learnable convolutional kernel and b_{res} the corresponding bias.

Then, the output of the residual block is obtained by adding F_{res} to the input F_{in} :

$$F_{\text{out}} = F_{\text{res}} + F_{\text{in}} \quad (18)$$

Feature-map spatial resolution is subsequently reduced via either max-pooling or average-pooling operations:

$$F_{\text{pool}} = \text{MaxPool}(F_{\text{out}}) \quad (19)$$

Classification is finally performed by passing the processed features through one or several fully connected layers, after which the softmax non-linearity produces class probability estimates:

$$\mathbf{p} = \text{softmax}(W_{\text{fc}} F_{\text{pool}} + b_{\text{fc}}) \quad (20)$$

Experiments and Results Analysis

The experiments described in this section evaluate fault diagnosis performance on an electrical power network, employing phasor measurement unit (PMU) measurements collected from the United Kingdom electricity grid. These tests are specifically designed to assess the effectiveness of the newly introduced SAT-SSL approach. The following subsections detail the characteristics of the utilized dataset, the experimental protocol, and a comprehensive interpretation of the quantitative findings.

Data processing and experimental setup

The experimental datasets employed in this study originate from the phasor measurement unit (PMU) network managed by Queen's University of Belfast (QUB) across the island of Ireland and the British Isles. OpenPMU-designed measurement devices are installed at low-voltage (415 V) and medium-voltage (33 kV) connection points of distributed generators to capture variations in key electrical parameters of the power system. The recorded quantities include voltage magnitude, phase angle, and system frequency; these measurements are securely transmitted in real time to a central server at QUB via encrypted internet protocols.

Each measurement file covers a continuous 24-hour period from a single monitoring site. At a sampling rate of 10 Hz, each daily record comprises 864,000 data samples. Because real-world grid measurements are inevitably contaminated by noise and errors—stemming from sources such as electromagnetic interference, sensor inaccuracies, and transient disturbances—preprocessing is essential to minimize their impact on downstream state estimation and analysis tasks. As a preliminary cleaning step, anomalous zero values are detected, removed, and replaced by linear interpolation between the means of the immediately preceding and following valid samples.

Abnormal event screening and classification

This subsection outlines the practical implementation of event detection within the monitored power grid. While the underlying filtering techniques were presented in the preceding chapter, the focus here lies on their experimental application and outcomes.

The present work concentrates exclusively on frequency measurements for event identification. Observable disturbances in grid frequency can generally be grouped into four types: impulsive events, transient oscillations, and sustained increases or decreases in frequency. Manually scanning enormous volumes of continuous grid data to locate these targeted events is computationally expensive and impractical. To address this, the study adopts an automated event-detection pipeline that rapidly isolates time windows containing phenomena of interest.

Data are first organized into overlapping segments using a sliding-window technique. The chosen window length is 10 seconds (100 consecutive samples at 10 Hz), with an overlap of 5 seconds. Before spectral analysis, a first-order differencing operation is applied to the frequency time series. Each differentiated window is then processed independently using two complementary spectral estimation methods: the Fast Fourier Transform (FFT) and the Yule–Walker autoregressive approach, both applied to segments of equal length.

Data points falling within the expected normal range are plotted as blue circles, whereas those exceeding predefined thresholds are marked with orange crosses. The complete detection workflow is visualized in [Figure 4](#). In the left panel, the red and blue bounding boxes highlight the two distinct clusters of anomalous intervals identified by the respective methods during the illustrated time span.

To increase reliability, the abnormal periods flagged by FFT and Yule–Walker are merged, retaining only those time intervals that were simultaneously detected as anomalous by both algorithms. The consolidated abnormal segments are subsequently inspected and classified according to the morphological patterns observed in the frequency waveforms. Four principal categories emerge: normal steady-state behavior, impulse-like disturbances, transient oscillations, and monotonic frequency rise or fall, as exemplified in [Figure 5](#).

Following this labeling procedure, the curated dataset is assembled. In total, the final power grid event dataset contains 4000 labeled samples, evenly distributed as follows: 1000 instances of normal operation, 1000 instances of impulse events, 1000 instances of transient events, and 1000 instances of frequency rise or fall.

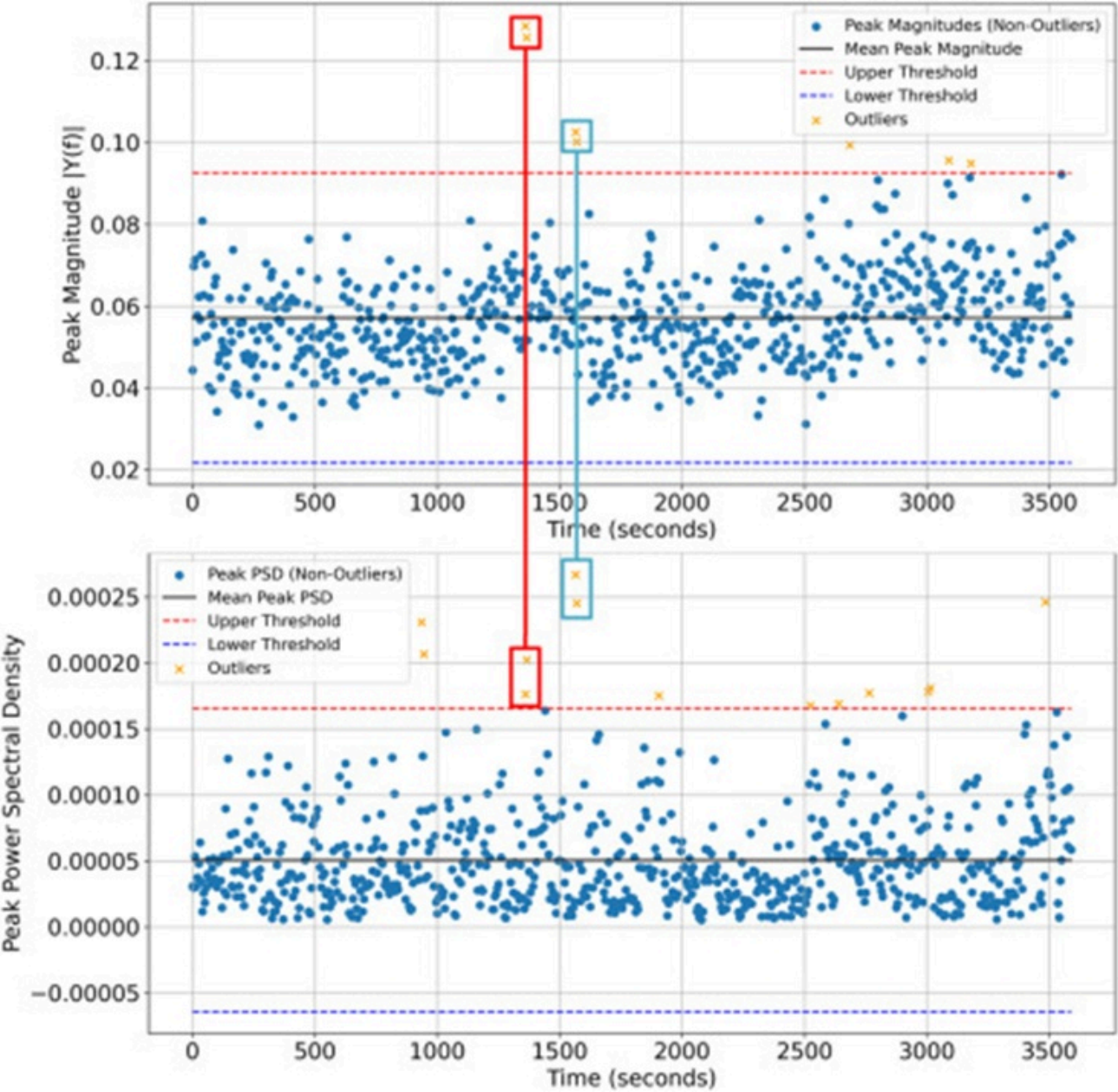


Figure 4. Anomalous event filtering process (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

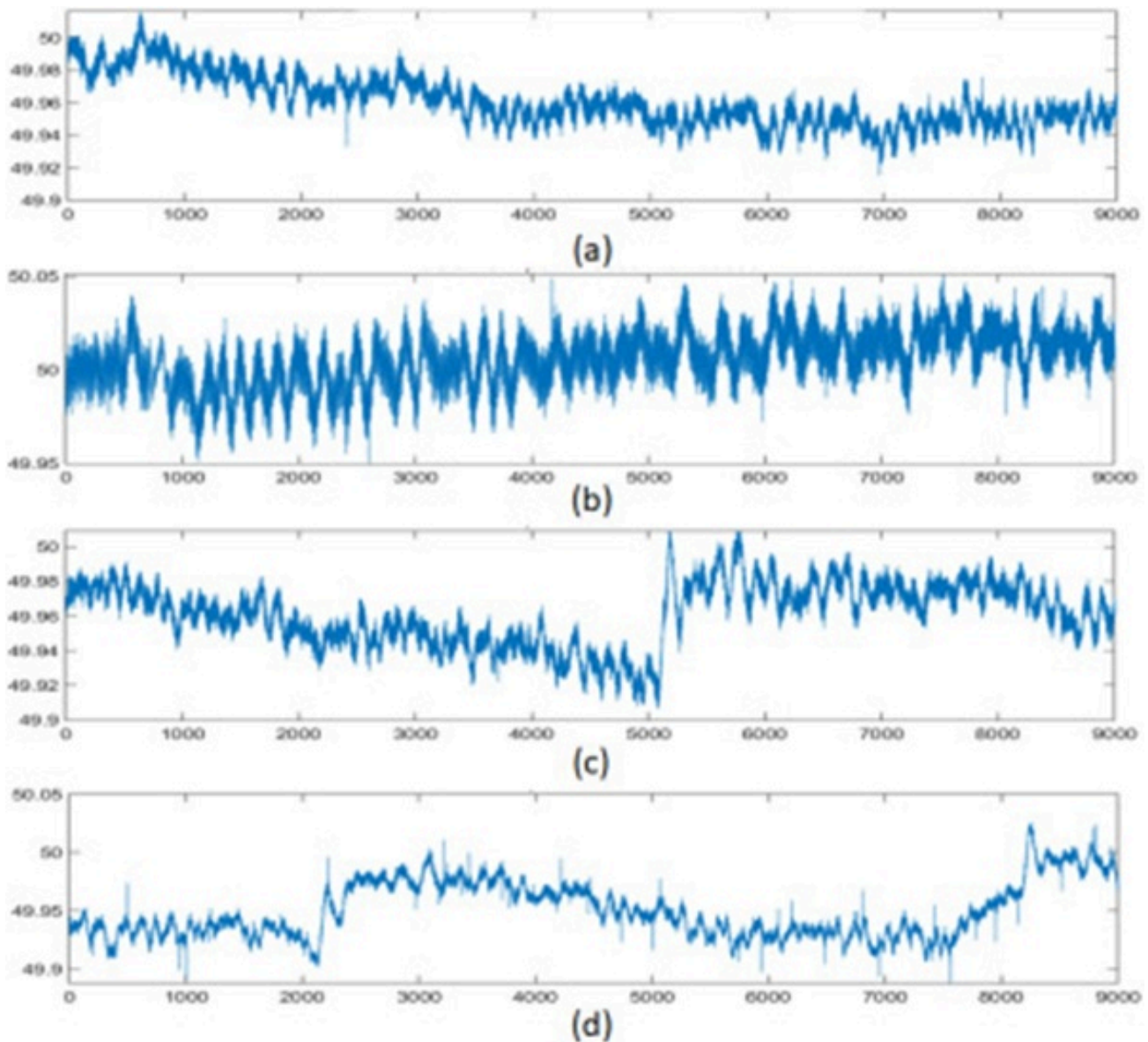


Figure 5. Grid frequency waveforms under different operating states. (a) is the normal operation state, (b) is the pulse state, (c) is the transient change, and (d) is the frequency rise state.

Time-frequency analysis

To reveal the joint time-frequency structure of the frequency measurements, the Yule–Walker parametric method was selected for power spectral density (PSD) estimation. Given the predominantly stationary nature of grid frequency signals, autoregressive (AR) modeling approaches, such as the Yule–Walker method, are well-suited for accurate PSD computation.

The long-term recordings were divided into non-overlapping 900-second blocks. Within each block, PSD estimates were computed using a 10-second (100-sample) sliding window, advancing in steps of 5 seconds. Based on preliminary experimentation and the data's statistical properties, the AR model order was set to 10. For every sliding window, a PSD spectrum was obtained, and the resulting time-frequency distributions were visualized as contour maps, as shown in [Figure 6](#). This representation effectively captures the evolution of spectral content over time. Notably, episodes of frequency decline are typically accompanied by a concentration of energy near 1 Hz in the higher-frequency bands.

To enable consistent downstream processing, a batch procedure was applied to generate time-frequency images for the entire dataset. All images were uniformly resized to 32×32 pixels. Finally, each time-frequency image was assigned a class label according to the event categorization scheme described in the previous subsection, thereby creating the labeled image dataset required for model training and evaluation.

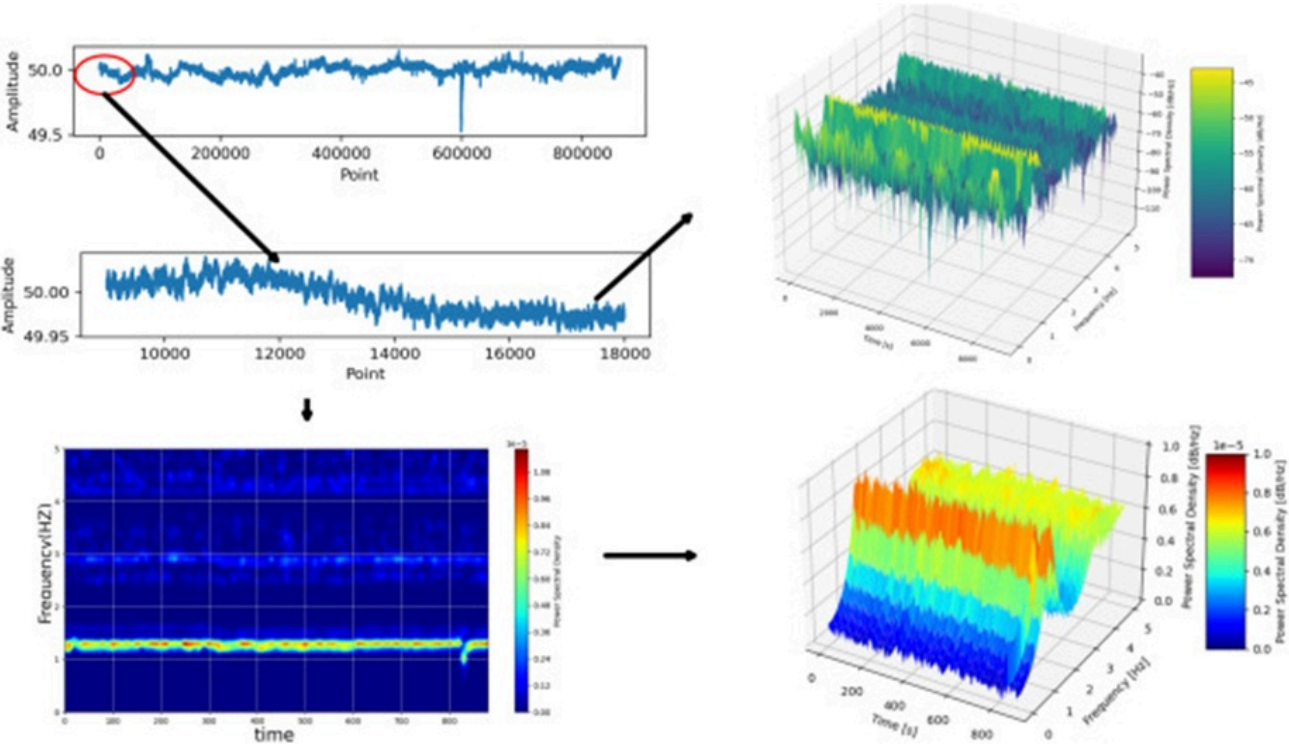


Figure 6. Time-frequency plot obtained by power spectral density analysis.

Model parameter setting

For fair benchmarking against existing semi-supervised learning methods, we conducted two experimental tracks using the widely adopted CIFAR-10 and SVHN datasets, as well as our custom power grid dataset. Performance was compared across four representative algorithms: SAT-SSL, FixMatch, ReMixMatch, and MixMatch. The complete hyperparameter configuration adopted for SAT-SSL is summarized in [Table 1](#). For FixMatch, the unsupervised (pseudo-label) loss weight w_u was set to 1, and the confidence threshold was set to 0.95. All remaining hyperparameters for ReMixMatch and MixMatch were kept identical to those listed for SAT-SSL in [Table 1](#).

Every dataset was partitioned into training, validation, and test subsets following a 7:1:2 ratio. All training and evaluation procedures were implemented using the TensorFlow framework and executed on a workstation equipped with an Intel Core i9-13900K processor and an NVIDIA GeForce RTX 4090 graphics card.

Table 1. Model parameter setting.

Parameter	Value
Number of residual layers	3
Starting confidence threshold	0.9

Batch size	64
Learning rate	0.03
Weight of pseudo-label loss	1
Weight decay	0.001
Number of filters in convolutional layers	16
Weight of self-supervised loss	1

Results analysis and discussion

To assess the effectiveness of the proposed SAT-SSL method, particularly on the real-world power grid dataset, the experimental evaluation was organized into two main phases. Phase one focused on standard computer-vision benchmarks (CIFAR-10 and SVHN) to enable direct comparison with published results. Phase two applied the same protocols to the power grid frequency-event dataset. In both phases, SAT-SSL was benchmarked against three established semi-supervised algorithms: MixMatch, ReMixMatch, and FixMatch.

Benchmark datasets

CIFAR-10 consists of 60,000 RGB images sized 32×32 pixels evenly distributed across 10 object classes. SVHN is a large-scale, real-world digit recognition dataset derived from Google Street View imagery, encompassing more than 600,000 labeled samples, also spanning 10 digit classes (0–9). Representative samples from both datasets are shown in [Figure 7](#).

To investigate label-efficiency, we simulated low- and medium-label regimes by retaining only a small fraction of ground-truth labels per class. For CIFAR-10, three label budgets were tested: 40, 250, and 4000 labels in total. For SVHN, the corresponding budgets were 40, 250, and 1000 labels. Because label selection introduces randomness, we fixed a random seed and repeated the subsampling procedure five times for each budget, reporting the average performance across these five independent runs.

Quantitative results on CIFAR-10 are reported in [Table 2](#) in terms of test error rates, while SVHN results appear in [Table 3](#). A side-by-side visual comparison of error rates across all four algorithms and label regimes is provided in the bar charts of [Figure 8](#).

Further insight into training dynamics is offered by the loss curves plotted in [Figure 9](#), which track the evolution of training and validation losses over optimization steps for each method. Model discrimination quality is visualized through confusion matrices in [Figure 10](#) (CIFAR-10) and [Figure 11](#) (SVHN).

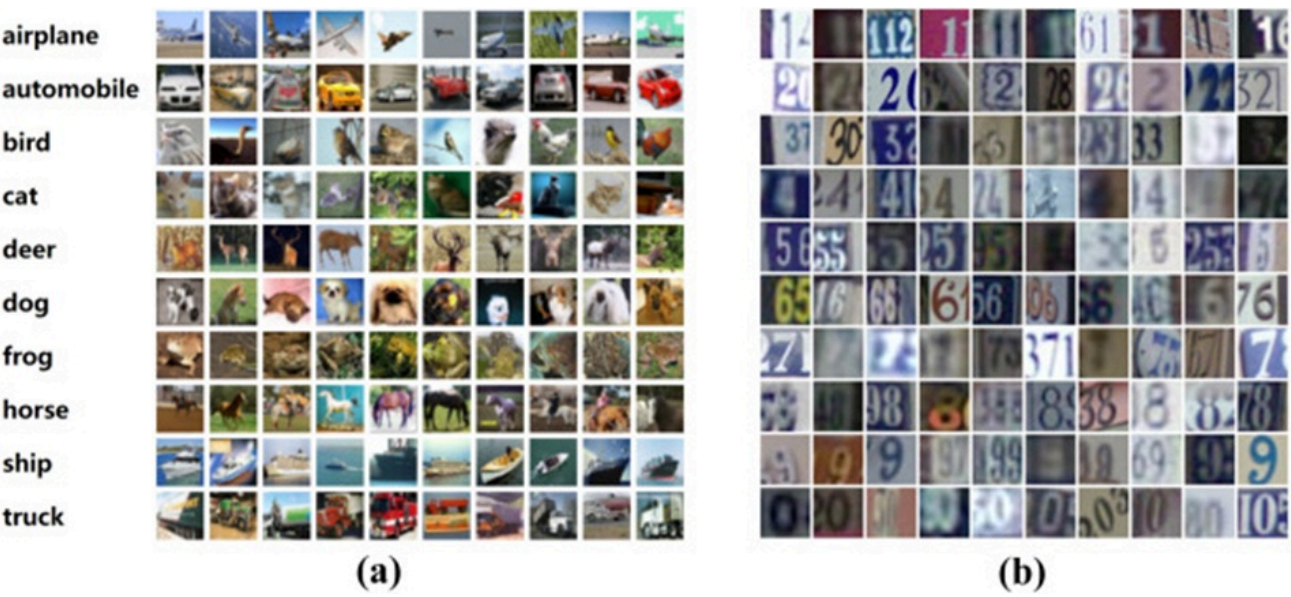


Figure 7. (a) CIFAR-10 dataset, and (b) SVHN dataset (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 2. Error rates are reported on the CIFAR-10 dataset.

Method	CIFAR-10 Accuracy (Error Rate ± Std)		
	40 labels	250 labels	4000 labels
MixMatch	43.60 ± 2.03	13.39 ± 0.96	7.36 ± 0.65
ReMixMatch	20.06 ± 5.39	6.03 ± 0.15	5.23 ± 0.14
FixMatch	13.62 ± 3.66	6.47 ± 0.31	4.83 ± 0.11
SAT-SSL	8.96 ± 2.11	4.69 ± 0.16	4.32 ± 0.11

Table 3. Error rates result from the SVHN dataset.

Method	SVHN Accuracy (Error Rate ± Std)		
	40 labels	250 labels	1000 labels
MixMatch	40.55 ± 1.32	6.05 ± 0.37	4.05 ± 0.33
ReMixMatch	4.56 ± 2.06	3.56 ± 0.31	3.03 ± 0.56
FixMatch	8.64 ± 3.01	3.20 ± 0.24	2.56 ± 0.11
SAT-SSL	4.23 ± 0.80	3.08 ± 0.26	2.33 ± 0.13

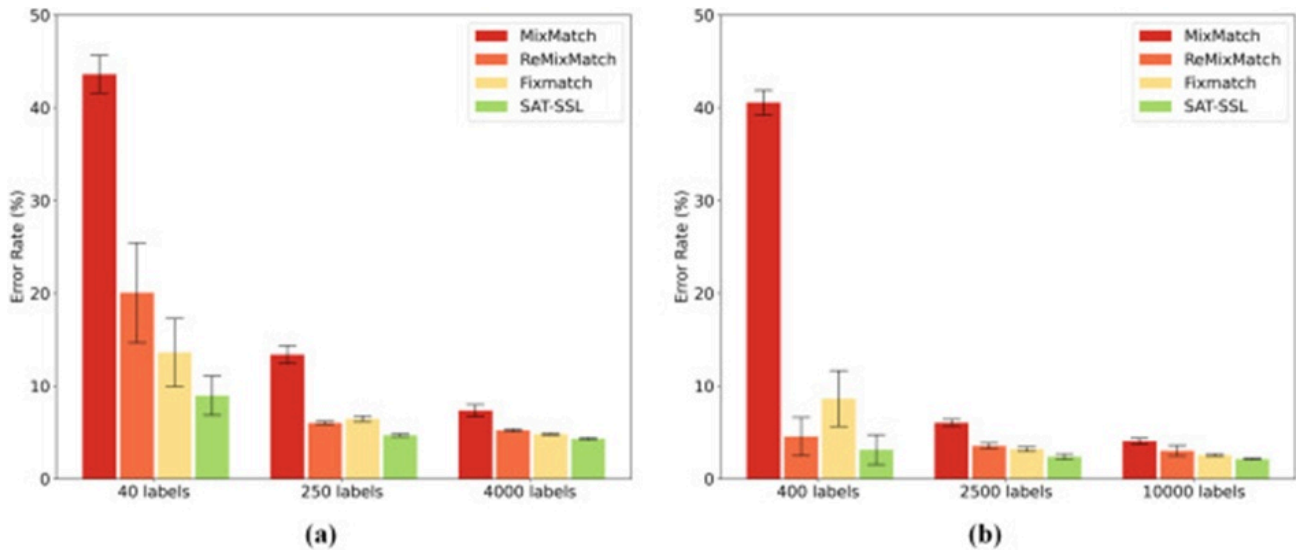


Figure 8. Comparative analysis of algorithm performance on standard datasets (a) Error rates on the CIFAR-10 dataset, and (b) Error rates on the SVHN dataset.

Multi-metric evaluation across accuracy, loss behavior, and confusion patterns consistently confirms the superiority of SAT-SSL. Across all tested label regimes on both CIFAR-10 and SVHN, SAT-SSL records the lowest error rates, as clearly evidenced by the bar plots in [Figure 8](#). The loss trajectories in [Figure 9](#) reveal that SAT-SSL converges faster and exhibits lower and more stable final loss values compared with the baselines, indicating improved optimization efficiency and reduced risk of overfitting even when labeled data are extremely scarce.

The confusion matrices in [Figures 10 and 11](#) further illustrate the enhanced classification reliability of SAT-SSL. Diagonal dominance is markedly stronger, while off-diagonal entries (misclassifications) are substantially reduced, reflecting superior class separation and discriminative capability under semi-supervised conditions.

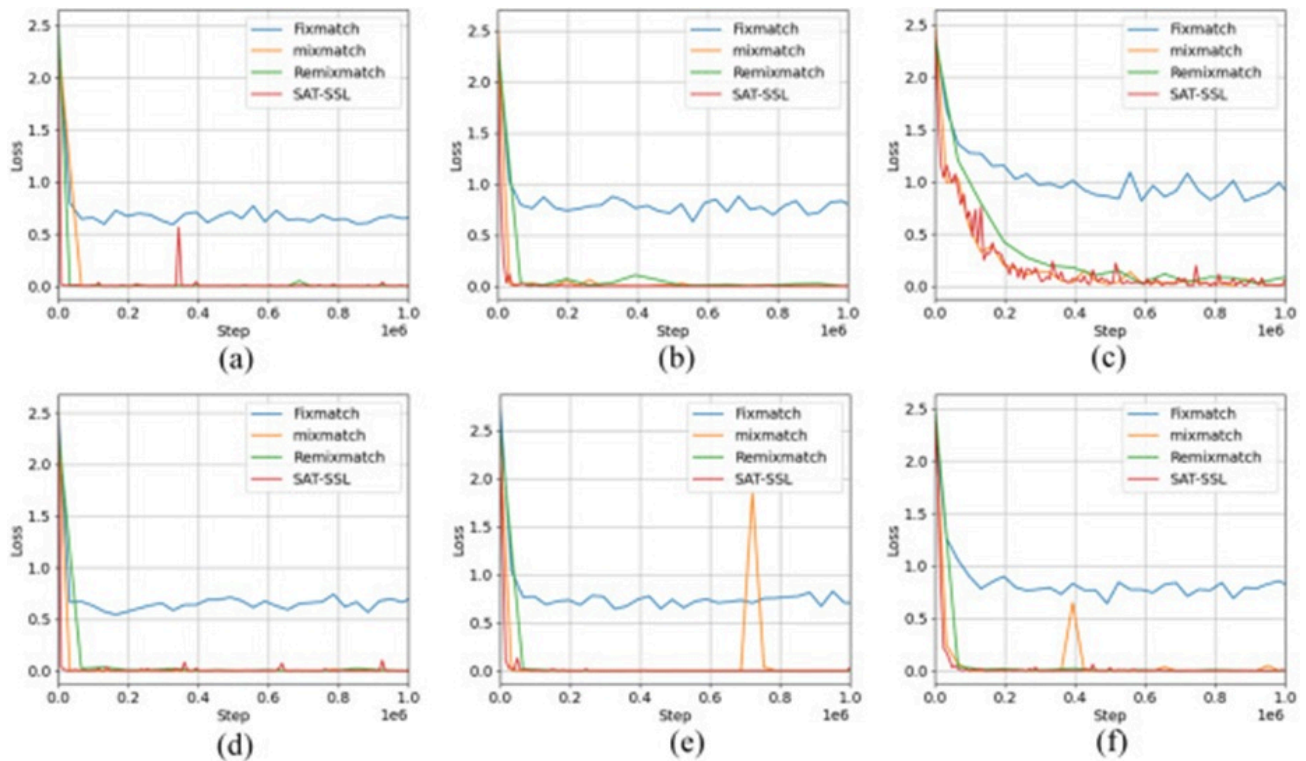


Figure 9. Loss variation on CIFAR-10 and SVHN datasets. (a) CIFAR-10,40 labels, (b) CIFAR-10,250 labels, (c) CIFAR-10,4000 labels, (d) SVHN,40 labels, (e) SVHN,40 labels, and (f) SVHN,40 labels.

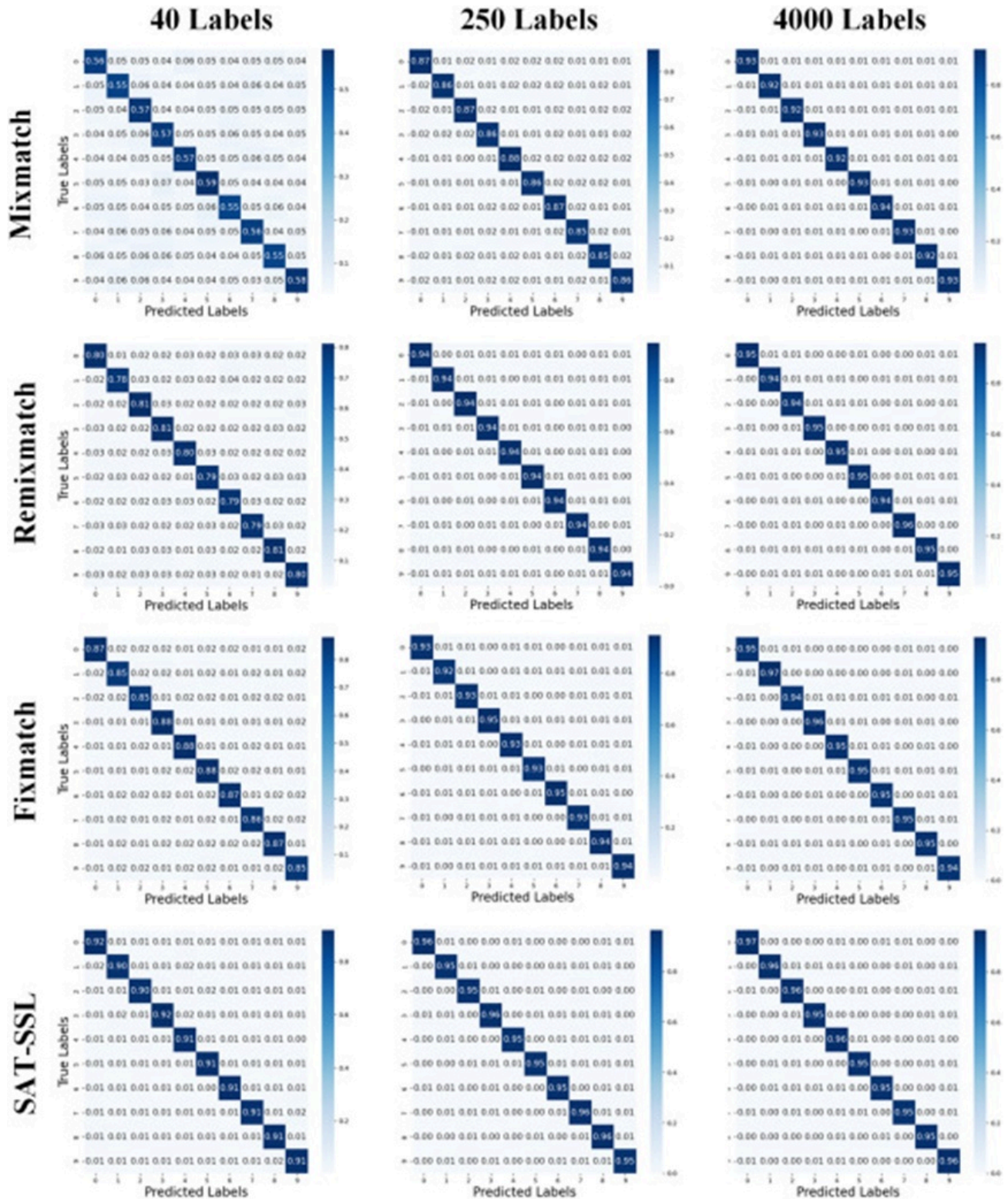


Figure 10. Confusion matrix of four algorithms under the CIFAR-10 dataset.

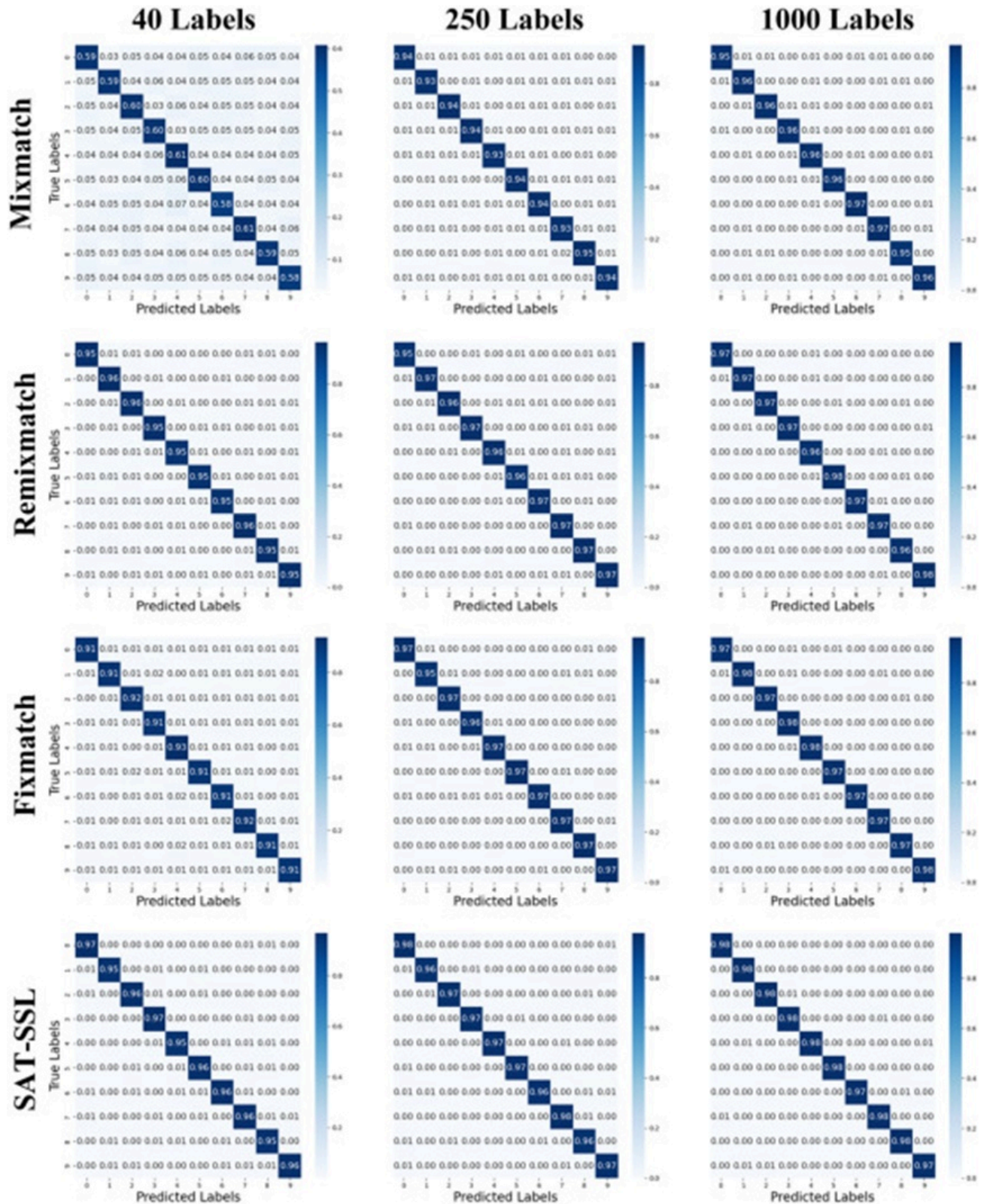


Figure 11. Confusion matrix of four algorithms under the SVHN dataset.

Power grid dataset

We next evaluated the proposed SAT-SSL method on the custom power grid frequency-event dataset. The experimental protocol mirrored that applied to the benchmark datasets: the full labeled set was subsampled to create five distinct label regimes—2%, 5%, 12.5%, 25%, and 50%. To ensure consistent, reproducible subsampling across runs, fixed random seeds were used. The procedure was repeated five times for each label ratio, and all reported metrics represent the average across these five independent trials.

Error rate comparisons across the four algorithms (MixMatch, ReMixMatch, FixMatch, and SAT-SSL) under the five label proportions are summarized in Table 4 and visualized in Figure 12. Training dynamics are depicted through the loss curves of all four methods in Figure 13. In contrast, classification behavior at different label budgets is detailed via confusion matrices in Figure 14.

The empirical results demonstrate that SAT-SSL delivers consistently strong performance on the power grid dataset. Superiority is evident across multiple evaluation dimensions, including test error rate, loss convergence behavior, and the structure of the confusion matrix. Collectively, these outcomes confirm the algorithm’s robust capability for accurate event detection and reliable discrimination among fault types in real-world grid frequency measurements.

Table 4. Error rates on power grid datasets.

Method	2% labeled	5% labeled	12.5% labeled	25% labeled	50% labeled
MixMatch	42.92 ± 10.67	40.14 ± 5.96	32.75 ± 1.89	21.14 ± 1.67	19.26 ± 0.89
ReMixMatch	21.25 ± 0.25	15.10 ± 0.13	10.59 ± 0.85	9.45 ± 0.76	6.21 ± 0.11
FixMatch	20.08 ± 3.65	12.64 ± 3.52	9.55 ± 0.56	6.08 ± 0.70	4.89 ± 0.12
SAT-SSL	13.56 ± 1.209	6.15 ± 1.13	4.30 ± 0.309	3.29 ± 0.227	1.21 ± 0.19

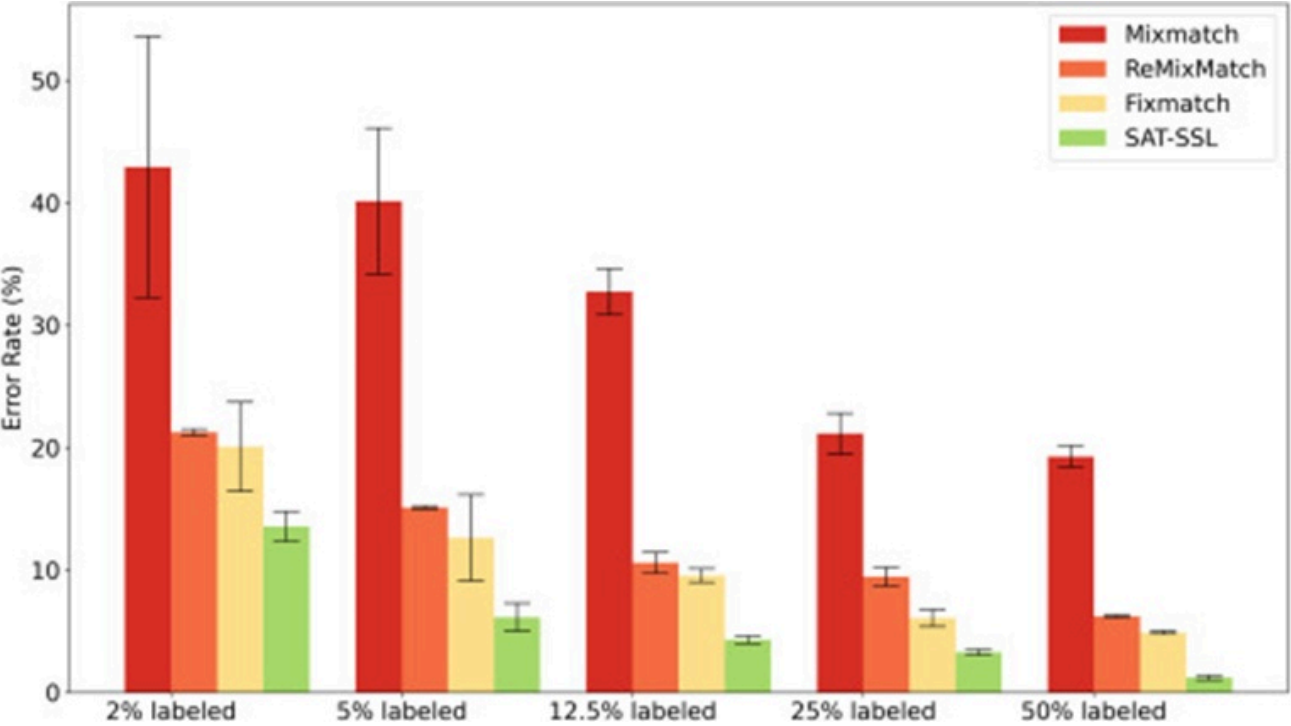


Figure 12. Comparative analysis of algorithm performance on power grid datasets.

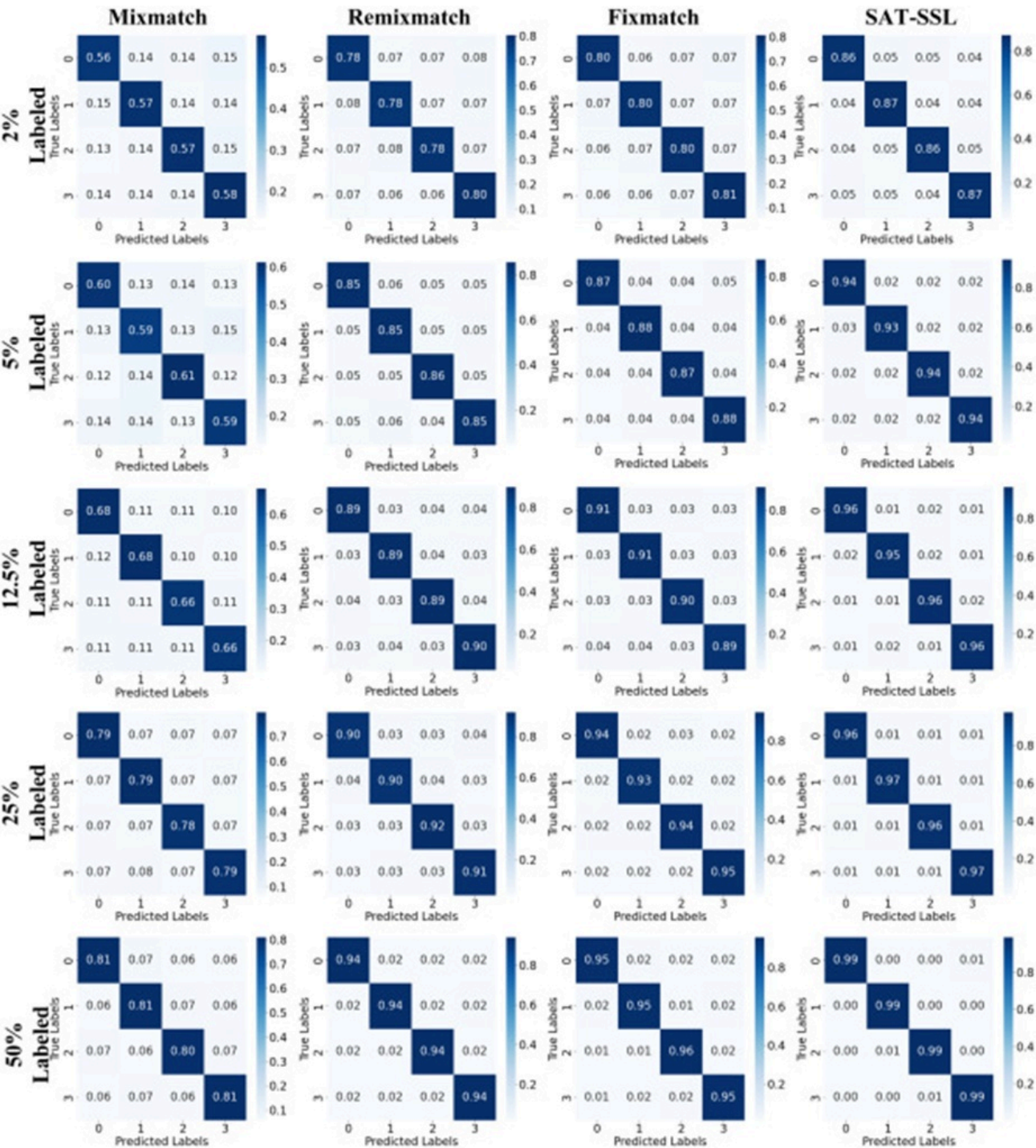


Figure 13. Confusion matrix for four algorithms on the power dataset.

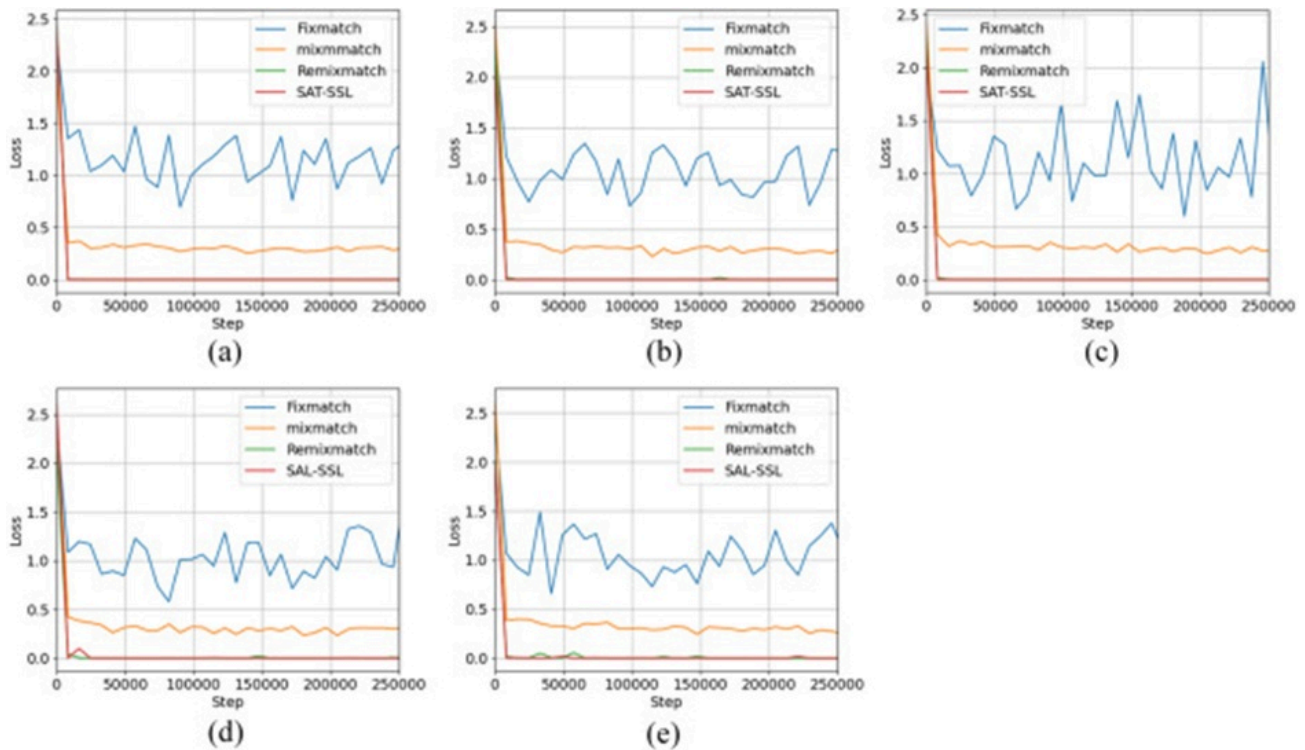


Figure 14. Comparative analysis of loss variations on power grid datasets: (a) 2% labeled, (b) 5% labeled, (c) 12.5% labeled, (d) 25% labeled, and (e) 50% labeled.

Conclusion

Given the intricate topology of modern urban power grids and the substantial challenges of acquiring sufficient labeled fault data, developing an algorithm capable of achieving high-accuracy fault detection and classification with only a small number of labeled examples holds considerable practical value. Such an approach can improve grid resilience while substantially lowering the cost and effort required for data annotation and model deployment.

In this study, we proposed the SAT-SSL algorithm, a semi-supervised learning framework tailored specifically for power grid fault identification and categorization in heavily label-scarce regimes. The work began with a detailed analysis of real grid-frequency signals recorded by phasor measurement units. Anomalous intervals were first isolated using a combination of Fast Fourier Transform (FFT) and Yule–Walker spectral estimation techniques. The detected events were then visually inspected and grouped into four categories based on characteristic waveform patterns, after which the corresponding segments were assigned class labels to construct a dedicated power grid fault dataset.

For feature representation, the Yule–Walker autoregressive method was applied to derive time-frequency characteristics from the frequency time series. To further strengthen the semi-supervised training pipeline, we incorporated a self-supervised auxiliary loss term together with dynamically adjusted confidence thresholds. These enhancements improved the quality of learned representations and accelerated overall convergence.

The SAT-SSL model was trained and evaluated on standard image classification benchmarks (CIFAR-10 and SVHN) and on a newly created power grid dataset, using controlled label-retention protocols to simulate varying degrees of supervision. Comparative experiments revealed marked performance gains over competing semi-supervised methods. Specifically, relative to FixMatch, SAT-SSL reduced the test error rate by 51.04% in the SVHN-40 label setting, by 34.21% on CIFAR-10 with only 40 labels, and by 51.34% on the power grid dataset when just 5% of the data were labeled. These

substantial improvements highlight the algorithm's ability to deliver excellent classification accuracy even under extremely limited supervision.

Future investigations will concentrate on deepening the understanding of underlying mechanisms driving different fault types in complex grid environments, expanding the dataset to include measurements associated with specific failure modes, and integrating additional operational parameters (e.g., voltage magnitude, phase angle, and active/reactive power flows) to enable more granular fault detection, precise localization, and potentially predictive maintenance capabilities.

Acknowledgements

None

Conflict of interest

None

Financial support

This research is supported by the National Natural Science Foundation China under Grants number [62073232](#), and the Science and Technology Project of Shenzhen, China ([KCXST20221021111402006](#), [JSGG20220831105800002](#)), and the “Nanling Team Project” of Shaoguan city, and the Science and Technology project of Tianjin, China ([22YFYSHZ00330](#)), and Shenzhen Excellent Innovative Talents [RCYX20221008093036022](#), Shenzhen-HongKong joint funding project (A) ([SGDX20230116092053005](#)), and the Shenzhen Undertaking the National Major Science and Technology Program, China ([CJGJZD20220517141405012](#)).

Ethics statement

None

Received: 18 Mar 2024

Revised: 29 Apr 2024

Accepted: 10 Jun 2024

Published online: 20 September 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Furse CM, Kafal M, Razzaghi R, Shin Y-J. Fault diagnosis for electrical systems and power networks: a review. *IEEE Sens J*. 2020;21(2):888–906.

Jayasree T, Devaraj D, Sukanesh R. Power quality disturbance classification using Hilbert transform and RBF networks. *Neurocomputing*. 2010;73(7–9):1451–6.

Buryanina N, Korolyuk Y, Koryakina M, Lesnykh E, Suslov K. Algorithm of current protection based on three instantaneous-value samples. In: 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). IEEE; 2019. p. 1–5.

Crossley P, McLaren P. Distance protection based on travelling waves. *IEEE Trans Power Appar Syst*. 1983;(9):2971–83.

Luo J, Yan R, Saha T, Sharma R. Waveform abnormality detection method for distribution system equipment condition monitoring. *Int J Electr Power Energy Syst*. 2023;152:109267.

Lin S, He Z, Li X, Qian Q. Travelling wave time–frequency characteristicbased fault location method for transmission lines. *IET Gener Transm Distrib*. 2012;6(8):764–72.

Fan C, Liu L, Tian Y. A fault-location method for 12-phase transmission lines based on twelve-sequence-component method. *IEEE Trans Power Deliv*. 2010;26(1):135–42.

Kalair A, Abas N, Kalair AR, Saleem Z, Khan N. Review of harmonic analysis, modeling and mitigation techniques. *Renew Sustain Energy Rev*. 2017;78:1152–87.

Primadianto A, Lu C-N. A review on distribution system state estimation. *IEEE Trans Power Syst*. 2016;32(5):3875–83.

Wang Z, Menke J-H, Schäfer F, Braun M, Scheidler A. Approximating multipurpose AC optimal power flow with reinforcement trained artificial neural network. *Energy AI*. 2022;7:100133.

Allen A, Santoso S, Muljadi E. Algorithm for screening phasor measurement unit data for power system events and categories and common characteristics for events seen in phasor measurement unit relative phase-angle differences and frequency signals. Golden, CO (United States): National Renewable Energy Lab. (NREL); 2013. Tech. Rep.

Phadke AG, Bi T. Phasor measurement units, WAMS, and their applications in protection and control of power systems. *J Mod Power Syst Clean Energy*. 2018;6(4):619–29.

Laverty DM, Best RJ, Brogan P, Al Khatib I, Vanfretti L, Morrow DJ. The OpenPMU platform for open-source phasor measurements. *IEEE Trans Instrum Meas*. 2013;62(4):701–9.

Guo Y, Li K, Laverty D, Xue Y. Synchrophasor-based islanding detection for distributed generation systems using systematic principal component analysis approaches. *IEEE Trans Power Deliv*. 2015;30(6):2544–52.

Shadi MR, Ameli M-T, Azad S. A real-time hierarchical framework for fault detection, classification, and location in power systems using PMUs data and deep learning. *Int J Electr Power Energy Syst*. 2022;134:107399.

Yu M, Du W, Li J, Zhang S, Hu J. Study on a comprehensive indicator and online classification of early warning of low frequency oscillation in power system. *Eng Appl Artif Intell*. 2023;124:106592.

Heymann F, Quest H, Garcia TL, Ballif C, Galus M. Reviewing 40 years of artificial intelligence applied to power systems—A taxonomic perspective. *Energy AI*. 2024;15:100322.

Vaish R, Dwivedi U, Tewari S, Tripathi SM. Machine learning applications in power system fault diagnosis: research advancements and perspectives. *Eng Appl Artif Intell*. 2021;106:104504.

Beinert D, Holzhüter C, Thomas JM, Vogt S. Power flow forecasts at transmission grid nodes using Graph Neural Networks. *Energy AI*. 2023;14:100262.

Husnoo MA, Anwar A, Reda HT, Hosseinzadeh N, Islam SN, Mahmood AN, et al. FedDiSC: a computation-efficient federated learning framework for power systems disturbance and cyber attack discrimination. *Energy AI*. 2023;14:100271.

Harish A, Asok P, Jayan M. A comparative evaluation of Stacked Auto-Encoder neural network and Multi-Layer Extreme Learning Machine for detection and classification of faults in transmission lines using WAMS data. *Energy AI*. 2023;14:100301.

Karić A, Konjić T, Jahić A. Power system fault detection, classification and location using artificial neural networks. In: *Advanced Technologies, Systems, and Applications II: Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT)*. Springer; 2018. p. 89–101.

Qu N, Zuo J, Chen J, Li Z. Series arc fault detection of indoor power distribution system based on LVQ-NN and PSO-SVM. *IEEE Access*. 2019;7:184020–8.

Fahim SR, Sarker Y, Sarker SK, Sheikh MRI, Das SK. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. *Electr Power Syst Res*. 2020;187:106437.

Rai P, Londhe ND, Raj R. Fault classification in power system distribution network integrated with distributed generators using CNN. *Electr Power Syst Res*. 2021;192:106914.

Belagoune S, Bali N, Bakdi A, Baadji B, Atif K. Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement*. 2021;177:109330.

Wang T, Wei X, Wang J, Huang T, Peng H, Song X, et al. A weighted corrective fuzzy reasoning spiking neural P system for fault diagnosis in power systems with variable topologies. *Eng Appl Artif Intell*. 2020;92:103680.

Yang X, Song Z, King I, Xu Z. A survey on deep semi-supervised learning. *IEEE Trans Knowl Data Eng*. 2022;35(9):8934–54.

Agrawala A. Learning with a probabilistic teacher. *IEEE Trans Inform Theory*. 1970;16(4):373–9.

Li C, Xu K, Zhu J, Liu J, Zhang B. Triple generative adversarial networks. *IEEE Trans Pattern Anal Mach Intell*. 2021;44(12):9629–40.

Wei X, Gong B, Liu Z, Lu W, Wang L. Improving the improved training of Wasserstein GANs: a consistency term and its dual effect. *arXiv preprint arXiv:1803.01541*. 2018.

Kingma DP, Mohamed S, Jimenez Rezende D, Welling M. Semi-supervised learning with deep generative models. *Adv Neural Inf Process Syst*. 2014;27.

Rasmus A, Berglund M, Honkala M, Valpola H, Raiko T. Semi-supervised learning with ladder networks. *Adv Neural Inf Process Syst*. 2015;28.

Sajjadi M, Javanmardi M, Tasdizen T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Adv Neural Inf Process Syst*. 2016;29.

Cao S, Lu W, Xu Q. Deep neural networks for learning graph representations. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30. 2016.

Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*. 2016.

Dong-DongChen W, WeiGao Z. Tri-net for semi-supervised deep learning. In: *Proceedings of Twenty-Seventh International Joint Conference on Artificial Intelligence*; 2018. p. 2014–20.

Lee D-H, Kim J, Park S, Choi M, Jung Y, Kang H, et al. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: *Workshop on Challenges in Representation Learning, ICML*. Vol. 3. Atlanta; 2013. p. 896.

Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA. Mixmatch: a holistic approach to semi-supervised learning. *Adv Neural Inf Process Syst*. 2019;32.

Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel CA, et al. Fixmatch: simplifying semi-supervised learning with consistency and confidence. *Adv Neural Inf Process Syst*. 2020;33:596–608.

Wu X, Zhang Y, Cheng C, Peng Z. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mech Syst Signal Process*. 2021;149:107327.

Long J, Chen Y, Yang Z, Huang Y, Li C. A novel self-training semisupervised deep learning approach for machinery fault diagnosis. *Int J Prod Res*. 2023;61(23):8238–51.

Mao W, Wei B, Xu X, Chen L, Wu T, Peng Z, et al. Fault diagnosis for power transformers through semi-supervised transfer learning. *Sensors*. 2022;22(12):4470.

Wallin E, Svensson L, Kahl F, Hammarstrand L. Doublematch: improving semisupervised learning with self-supervision. In: 2022 26th International Conference on Pattern Recognition (ICPR). IEEE; 2022. p. 2871–7.

Zhang B, Wang Y, Hou W, Wu H, Wang J, Okumura M, et al. Flexmatch: boosting semi-supervised learning with curriculum pseudo labeling. *Adv Neural Inf Process Syst*. 2021;34:18408–19.