Outlier-Aware Semi-Supervised Learning for Generalizable Non-Intrusive Load Monitoring

Keke Li[1], Bo Liu[1*], Wenpeng Luan[1], Yang Han[1], Zhixin Li[2], Jin Bao[2]

[1]*School of Electrical and Information Engineering, Tianjin University, Tianjin, China*
[2]*Marketing Service Center, State Grid Jiangsu Electric Power Co.Ltd., Nanjing, China*
Email: {kekeli, Boliu, wenpeng.luan}@tju.edu.cn, 920816hanyang@sina.com,
zxli_epc@163.com, baj8403@163.com

***Abstract*—Non-intrusive load monitoring (NILM) can estimate the power consumption and operational states of individual appliances from total measurements, eliminating the need for dedicated sub-meters. In practice, collecting large-scale labeled data for training NILM models is often costly and sometimes infeasible. While semi-supervised learning (SSL) techniques offer a promising solution by leveraging a small amount of labeled data along with abundant unlabeled data to learn the feature distribution of appliances. However, conventional SSL methods assume that the labeled and unlabeled data originate from the same label space, which may not hold in NILM application scenarios, where the label space can be inconsistent due to the presence of unseen or unknown appliances in the unlabeled data, which can degrade the performance of SSL algorithms. To address this, an outlier-aware semi-supervised learning framework is proposed for NILM. The model integrates a classifier for known appliances with an outlier detection mechanism based on one-vs-all (OVA) schemes, enabling the model to learn discriminative representations of known appliances while effectively identifying and rejecting unknown ones. Specifically, the OVA classifiers provide inlier confidence scores for each class, which are used to determine whether a sample should be classified or rejected. To further enhance generalization under label space mismatch, a soft consistency regularization strategy is introduced to encourage prediction consistency across input augmentations, thereby improving the smoothness and robustness of the OVA classifiers in detecting unknown appliances. Experimental results on the NILM dataset show that the proposed method excels in accurately identifying known appliances and robustly detecting outliers.**

***Index Terms*—Non-intrusive load monitoring, semi-supervised learning, outlier detection, one-vs-all network**

I. Introduction

Achieving net-zero carbon emissions has become a global imperative, driving the transformation of power systems toward higher efficiency, sustainability, and intelligence. In this context, fine-grained visibility into end-user electricity consumption is essential for promoting energy-saving behavior, enhancing electricity safety, and enabling personalized energy services. These emerging demands highlight the critical need for accurate acquisition of appliance-level power consumption data [1]. However, deploying dedicated meters for each appliance is expensive and intrusive. To address this challenge, non-intrusive load monitoring (NILM) has emerged as a promising approach to estimate the operating states and consumption profiles of individual appliances using only the aggregate power measurements from a single point [2].

Deep learning has achieved remarkable success across various domains and has also been widely applied in the field of NILM [3]. Supervised learning methods have recently shown impressive NILM performance by leveraging deep neural networks to learn hierarchical features directly from raw power signals. However, these methods are heavily dependent on large amounts of labeled data and often lack generalization across households. To address the issue of cross-user deployment, transfer learning and domain adaptation techniques have been introduced to adapt a pre-trained NILM model to new user environments. Nonetheless, due to significant variations in user habits, device configurations, and load combinations, it remains challenging to learn transferable representations that are robust across different domains (users) [4]. Furthermore, to reduce the reliance on labeled data while maintaining satisfactory performance in NILM, semi-supervised learning (SSL) has been introduced into NILM, enabling the model to leverage both labeled and unlabeled data during training [5]. Nevertheless, conventional SSL methods typically operate under the assumption that the labeled and unlabeled data share the same label space, which does not hold in open-world NILM applications. In practice, the unlabeled data may include unknown appliances that are either absent from the labeled dataset, leading to a label space mismatch [6]. These outlier samples may mislead the model and degrade its performance on known appliances. Therefore, it is crucial to develop NILM methods that are not only label-efficient but also robust to unknown appliances in the wild.

To address these challenges, this paper proposes an outlier-aware SSL framework tailored for NILM under real-world, label-incomplete conditions. The proposed method introduces a dual-branch SSL architecture that combines standard appliance identification with one-vs-all classifiers, enabling the model to accurately identify known appliances while effectively detecting and rejecting unknown ones. To further improve learning stability and representation robustness, a consistency regularization strategy is employed between strongly and weakly augmented views of unlabeled data. This design allows the model to fully utilize unlabeled samples, even in the presence of outlier appliances. Extensive experiments conducted on publicly available NILM datasets demonstrate that the proposed method improves both appliance identification

979-8-3315-8138-1

accuracy and robustness against unknown appliances, offering a scalable and practical solution for NILM systems deployed in dynamic, open-world environments. The main contributions of this work are summarized as follows:

1) Propose an outlier-aware semi-supervised learning framework tailored for NILM, which unifies known appliance identification and unknown appliance rejection, while reducing the dependence of NILM models on labeled data and enabling robust identification under heterogeneous label space conditions.
2) Develop an appliance identification framework that integrates a closed-set classifier with an outlier detector, where the classifier predicts a candidate label and the corresponding one-vs-all sub-classifier verifies its validity. Soft consistency regularization is applied to enhance robustness and generalization in open-world scenarios.
3) Demonstrate competitive performance on public NILM datasets, showing improvements in both known appliance identification and unknown appliance rejection compared to existing SSL methods.

II. Proposed Method

A. Appliance V-I Trajectory Feature Extraction

In event-based NILM, appliance switching events are first detected from aggregate power or current waveforms. For each detected event, steady-state voltage and current segments before and after the transition are extracted [7]. By encoding the corresponding voltage and current values, a V–I trajectory is constructed to capture the appliance's electrical signature. This transformation converts appliance identification into a visual classification task, enabling the effective use of image-based deep learning frameworks within a SSL paradigm.

B. Problem Setting

This paper proposes a novel semi-supervised classification framework for event-based NILM, which effectively handles scenarios where only a small subset of appliance events are labeled, whereas the majority are unlabeled and may contain unknown or unseen appliances. For K-class appliance identification task, let $\mathcal{X} = \{(x_b, y_b) \mid b = 1, ..., B\}$ denote a batch of B labeled samples drawn from the labeled data S, where each x_b represents a V-I trajectory feature of the appliance and $y_b \in 1, ..., K$ is the corresponding appliance label. Let $\mathcal{U} = \{u_b \mid b = 1, ..., \mu B\}$ denote a batch of μB unlabeled samples drawn from the unlabeled data S_u, where μ is a hyperparameter defining the ratio of unlabeled to labeled data. Unlike traditional semi-supervised learning, which assumes that all unlabeled samples originate from known appliance classes, the unlabeled data here may include events produced by unknown appliances outside the labeled set. The objective of the paper is to accurately classify known appliances while learning to distinguish and reject unknown appliances.

C. Approach Overview

The proposed outlier-aware semi-supervised learning method consists of three main components, as shown in

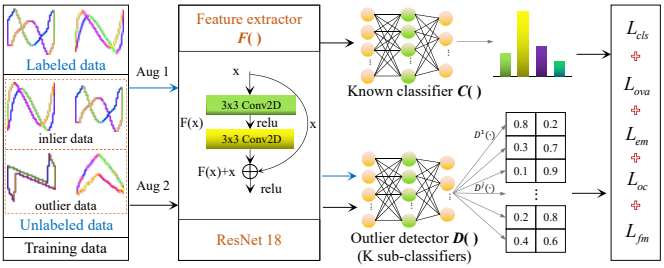

Fig. 1. The overall architecture of the proposed outliers-aware semi-supervised learning NILM method. The model comprises a feature extractor $F(\cdot)$, a known appliance classifier $C(\cdot)$, and an unknown appliance detector $D(\cdot)$ consisting of K sub-classifiers, each trained to determine whether a sample is an inlier of its corresponding class.

Fig. 1: (1) a shared feature extractor $F(\cdot)$ that maps each input V-I trajectory image to a latent embedding space, (2) an unknown appliance samples detection module composed of K one-vs-all (OVA) binary sub-classifiers $D^j(\cdot)$, where $j \in 1, ..., K$ corresponds to the known appliance classes, and (3) a multi-class classifier $C(\cdot)$ that produces a K-dimensional probability vector over known appliances.

During inference, the model first uses $C(\cdot)$ to generate the predicted appliance label $\hat{y}$ for a given input. The corresponding sub-classifier $D^{\hat{y}}(\cdot)$ then evaluates whether the input is a reliable instance of class $\hat{y}$. If the prediction is validated by $D^{\hat{y}}$, the final output is $\hat{y}$; otherwise, the input is identified as an unknown appliance. The key to this approach lies in using OVA sub-classifiers to reject unknown appliances, along with a soft consistency-based training strategy that enhances the model's ability to distinguish between known and unknown appliance behaviors in unlabeled event data.

1) One-vs-All Outlier Detector: In the context of NILM, unlabeled appliance events may stem from unknown appliances not present in the labeled training set. To detect these samples, the proposed method employs a set of one-vs-all sub-classifiers to verify whether an input event belongs to the predicted class or represents an unknown appliance. Specifically, for each known appliance category j, the corresponding sub-classifier $D^j(\cdot)$ receives the feature embedding $F(x_b)$ of an input V-I trajectory image and outputs a 2-dimensional vector $z_b^j = D^j(F(x_b)) \in \mathbb{R}^2$, where the two dimensions represent the scores for being an inlier and an outlier. These scores are converted into probabilities via a softmax function:

$$p^j(t = 0|x_b), p^j(t = 1|x_b) = \text{Softmax}(z_b^j) \tag{1}$$

with $t = 0$ indicating inlier and $t = 1$ indicating outlier, ensuring that $p^j(t = 0 \mid x_b) + p^j(t = 1 \mid x_b) = 1$. Each sub-classifier $D^j(\cdot)$ is trained to treat samples from appliance j as positives and samples from all other classes as negatives. The OVA training loss for a labeled batch $\mathcal{X} = (x_b, y_b)_{b=1}^B$ is defined as:

$$L_{ova}(X) = \frac{1}{B}\sum_{b=1}^{B} -\log(p^{y_b}(t=0|x_b)) - \min_{i\neq y_b} \log(p^{i}(t=1|x_b)) \tag{2}$$

This loss function enables each sub-classifier to learn compact intra-class representations and robustly reject unknown appliances. In this design, each sub-classifier outputs a confidence score that reflects the probability of the input to the corresponding appliance class. These scores serve as soft indicators of appliance classes and are effective for identifying events triggered by unknown appliances. The outlier detection process leverages both the multi-class classifier $C(\cdot)$ and the corresponding sub-classifier $D^{\hat{y}}(\cdot)$, where $\hat{y} = \arg\max_j C(F(u_b))$ is the predicted class. If the probability of sample u_b being an inlier for class $\hat{y}$ is below a fixed threshold, that is, $p^{\hat{y}}(t = 0 \mid u_b) < 0.5$, the sample is treated as an unknown appliance.

2) Open-set Entropy Minimization: To enhance the separation between known and unknown appliances, an entropy minimization objective is applied to unlabeled samples concerning the outputs of the sub-classifiers. This loss $L_{\mathrm{em}}(\mathcal{U})$, encourages confident predictions by reducing the uncertainty in OVA outputs. Specifically, for a batch of unlabeled samples $\mathcal{U} = \{u_b\}_{b=1}^{\mu B}$, minimizing $L_{em}(\mathcal{U})$ helps the model assign sharper inlier or outlier probabilities for each known appliance, thereby improving the model's ability to distinguish unknown appliances from these seen during training.

$$L_{em}(\mathcal{U}) = -\frac{1}{\mu B}\sum_{b=1}^{\mu B}\sum_{j=1}^{K} p^j(t=0|u_b)\ \log\ p^j(t=0|u_b) + p^j(t=1|u_b)\ \log\ p^j(t=1|u_b) \tag{3}$$

3) Soft Open-set Consistency Regularization: To address the challenge posed by unknown appliances in unlabeled data, the proposed method employs an entropy minimization objective that encourages confident predictions for known appliances while avoiding overconfident assignments for unknown ones. However, entropy minimization may not provide sufficiently reliable training signals for all unlabeled data. To further improve the model's ability to distinguish unknown appliances, a soft consistency regularization strategy is introduced that stabilizes the outlier detector. To smooth the decision boundary without relying on hard labels, the method minimizes the discrepancy between model outputs under different augmentations of the same input. Unlike conventional pseudo-labeling, it leverages unsharpened logits to preserve soft probabilistic guidance, reducing the risk of incorrect class assignments. This training objective, termed soft outlier consistency regularization (SOCR), improves the reliability of entropy-based supervision and promotes more robust separation between inlier and outlier samples.

Specifically, the proposed SOCR encourages the outlier detector to produce stable predictions in the presence of input perturbations. In implementation, the data augmentation strategy $\mathcal{T}(\cdot)$ consists of random cropping. For each unlabeled sample $u_b \in \mathcal{U}$, we generate two different views $\mathcal{T}_1(u_b)$ and $\mathcal{T}_2(u_b)$ by applying two independently sampled transformations from $\mathcal{T}(\cdot)$. Let $p^j(t \mid \mathcal{T}_1(u_b))$ and $p^j(t \mid \mathcal{T}_2(u_b))$ denote the predicted probability from the j-th OVA sub-classifier for the two augmented inputs. To promote local smoothness and suppress overconfident but unreliable predictions, we minimize the divergence between these outputs. The SOCR loss is defined as:

$$L_{oc}(\mathcal{U}, \mathcal{T}) = \frac{1}{\mu B}\sum_{b=1}^{\mu B}\sum_{j=1}^{K}\sum_{t\in(0,1)} |p^j(t|\mathcal{T}_1(u_b)) - p^j(t|\mathcal{T}_2(u_b))|^2 \tag{4}$$

This regularization objective minimizes the discrepancy between output distributions under data augmentation $\mathcal{T}(\cdot)$, encouraging consistent and stable responses from the outlier detector. To preserve uncertainty in ambiguous cases, output logits remain unsharpened. The regularization loss is computed directly from the outlier detector's outputs, enabling the model to better distinguish between known and unknown appliances and to reject unknown appliances while maintaining smooth decision boundaries.

D. Overall Framework

The overall training procedure is summarized in **Algorithm 1**. To enhance outlier detection performance, a FixMatch SSL loss is introduced to guide the classification of unlabeled instances from known appliances [8]. FixMatch generates pseudo-labels based on the model's predictions on weakly augmented input and then enforces consistency by training the model to predict the same labels under strong augmentations of the same input. This mechanism allows the model to refine its classification boundaries for known appliances while maintaining against unknown samples in the unlabeled data.

For labeled samples, we use the standard cross-entropy loss to train the closed-set classifier $C(\cdot)$, denoted as $L_{cls}(\mathcal{X})$, where $\mathcal{X}$ represents the labeled samples. In addition, the OVA outlier detectors are trained using $L_{ova}(\mathcal{X})$, which guides each sub-classifier to distinguish between inlier and non-inlier events for its corresponding class. The total supervised loss is defined as $L_{sup}(\mathcal{X}) = L_{cls}(\mathcal{X}) + L_{ova}(\mathcal{X})$. The model is first trained for E_{fix} epochs using $L_{sup}(\mathcal{X})$, along with the unsupervised outlier-aware objectives L_{em} (entropy minimization) and L_{oc} (soft consistency regularization). After this initial training phase, the model begins selecting pseudo-inlier samples from the unlabeled data at each subsequent epoch. As described in **Algorithm 1**, the function $\mathrm{Select}(w; \mathcal{D}_u)$ represents this selection process, where the current model parameters w including the feature extractor $F(\cdot)$, the outlier detector $D(\cdot)$, and the classifier $C(\cdot)$ are used to identify unlabeled samples that belong to known appliance classes.

Subsequently, the FixMatch loss $L_{fm}(\mathcal{I})$ is applied exclusively to the selected pseudo-inliers, where $\mathcal{I}$ denotes a batch of the unlabeled data identified as belonging to known

classes. This loss encourages the model to produce consistent predictions across different augmentations of the same sample, further refining its classification boundary. The complete training objective thus integrates supervised, unsupervised, and semi-supervised components, and is formulated as follows:

$$L_{all} = L_{sup} + \lambda_{em} L_{em} + \lambda_{oc} L_{oc} + \lambda_{fm} L_{fm} \tag{5}$$

where λ_{em}, λ_{oc} and λ_{fm} weight each objective.

Algorithm 1 Outlier-Aware SSL for Generalizable NILM

Require: The labeled samples are denoted as $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{N}$, and the unlabeled samples as $\mathcal{S}_u = \{u_i\}_{i=1}^{N_u}$. The pseudo-inlier collection is initialized as $\mathcal{K} = \emptyset$, and $\mathcal{T}$ represents the data augmentation transformation. Model parameters w, learning rate η, fixed and maximum epochs E_{fix}, and E_{max}, maximum iteration number I_{max}, and trade-off coefficients λ_{em}, λ_{oc}, λ_{fm}.

Ensure: Model parameters w

for epoch $= 1$ to E_{max} **do**
  for iteration $= 1$ to I_{max} **do**
    Sample a batch of labeled data $\mathcal{X} \subset \mathcal{S}$ and unlabeled data $\mathcal{U} \subset \mathcal{S}_u$
    Compute $\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{sup}}(\mathcal{X}) + \lambda_{\text{em}}\mathcal{L}_{\text{em}}(\mathcal{U}) + \lambda_{\text{oc}}\mathcal{L}_{\text{oc}}(\mathcal{U}, \mathcal{T})$
    if epoch $> E_{\text{fix}}$ **then**
      Sample pseudo-inliers $\mathcal{I} \subset \mathcal{K}$
      $\mathcal{L}_{\text{all}} \leftarrow \mathcal{L}_{\text{all}} + \lambda_{\text{fm}}\mathcal{L}_{\text{fm}}(\mathcal{I})$
    end if
    Update $w \leftarrow w - \eta \nabla_w \mathcal{L}_{\text{all}}$
  end for
  if epoch $\geq E_{\text{fix}}$ **then**
    $\mathcal{K} \leftarrow \text{Select}(w, \mathcal{D}_u)$
  end if
end for

III. Experiment

A. Dataset and Experimental Setup

In this study, experiments are conducted using the publicly available WHITED [9] dataset, which contains voltage and current waveform recordings of 55 appliance types. For each appliance, multiple waveform cycles during its operating process are extracted using a sliding window, and voltage–current trajectory features are generated (300 samples for each appliance). These features are then divided into the training, validation, and test sets with a ratio of 7:1:2, and are used to train and evaluate the model. In the experiment, 10 appliances are selected for evaluation: AC (AP1), AirPump (AP2), BenchGrinder (AP3), CableModem (AP4), CoffeeMachine (AP5), DeepFryer (AP6), GuitarAmp (AP7), HalogenFluter (AP8), Heater (AP9), and MultiTool (AP10).

During the experiments, we define a subset of appliances as known classes, while treating the remaining types as unknown outliers. For each known class, only a small portion of the samples are labeled, and the rest are unlabeled. All samples from unknown appliances are unlabeled and are mixed with the unlabeled samples of known classes. The model is trained jointly using all labeled samples and the mixed unlabeled dataset. During testing, the test set contains both known and unknown appliance samples. The trained model is designed to classify known appliances accurately while effectively identifying and rejecting unknown ones.

To quantitatively evaluate the performance of the proposed method, the macro-averaged F1 score metric is adopted. We adopt a consistent set of hyperparameters across all experiments. Specifically, λ_{oc} is fixed at 0.5, and λ_{em} is set to 0.1 throughout the training process. The λ_{fm} is initially set to 0 for the first E_{fix} epochs and then updated to 1 for the remainder of the training. In all experiments, E_{fix} is set to 10.

B. Experiment Results

The experimental results on the WHITED dataset validate the effectiveness of the proposed outlier-aware SSL framework in both known appliance identification and unknown appliance rejection. As shown in **Table I**, under the setting of 50 labeled samples per class, the proposed method consistently outperforms the conventional semi-supervised baseline in terms of macro-averaged F1 score. When two appliances (AP1 and AP2) are treated as unknown, the proposed method achieves an F1 score of 0.9259, compared to 0.8126 for the baseline. Similarly, when AP9 and AP10 are treated as unknown appliances, the proposed method again achieves higher performance, with an F_{macro} of 0.9407 compared to 0.8224 for the baseline. This improvement is attributed to the integration of the one-vs-all outlier detector and soft consistency regularization, which enables more reliable separation of known and unknown appliance events within the unlabeled data.

To further examine the model's generalization ability under more complex scenarios, **Fig. 4** presents results with four unknown appliances (AP1, AP2, AP9, and AP10). In this challenging setting, the proposed method maintains a high F1 score of 0.9510, significantly outperforming the baseline (0.7385), which suffers from substantial confusion between inlier and outlier classes. The confusion matrices in **Fig. 2–4** illustrate that the proposed method maintains clean and sharp diagonal patterns for known classes while minimizing incorrect predictions on unknown samples (relabeled as class 8 or 6). In contrast, the baseline method exhibits substantial misclassifications, especially in the unknown class, which confirms its vulnerability under label space mismatch. These results indicate that the proposed method not only improves appliance classification accuracy in scenarios with limited labeled data but also enhances the model's robustness in real-world open-set NILM scenarios.

IV. Conclusion and Future Work

In this paper, we propose an outlier-aware SSL approach for NILM that tackles two key challenges in real-world applications: the limited availability of labeled appliance data and unknown appliance classes concealed in the unlabeled

TABLE I

THE F_{score} PERFORMANCE OF APPLIANCE IDENTIFICATION WITH 50 LABELED DATA PER CLASS UNDER DIFFERENT UNKNOWN APPLIANCES.

Unknown	Methods	AP1	AP2	AP3	AP4	AP5	AP6	AP7	AP8	AP9	AP10	Unknown	F_{macro}
AP1 AP2	Semi-supervised	-	-	0.7895	0.9945	0.9549	0.9524	1	1	0.9836	0.6383	0	0.8126
	Proposed	-	-	1	1	1	1	1	1	1	0.6667	0.6667	0.9259
AP9 AP10	Semi-supervised	1	0.8257	0.7759	1	0.8	1	1	1	-	-	0	0.8224
	Proposed	1	1	1	1	0.8	1	1	1	-	-	0.6667	0.9407
AP1 AP2 AP9 AP10	Semi-supervised	-	-	0.5607	1	0.6250	0.9836	1	1	-	-	0	0.7385
	Proposed	-	-	1	1	0.8	1	1	1	-	-	0.8571	0.9510

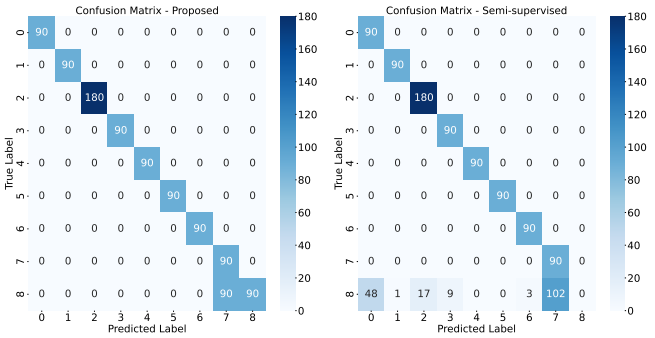

Fig. 2. Confusion matrices of different methods under 50 labeled samples with 2 unknown appliances (relabeled as class 8) (AP1 and AP2).

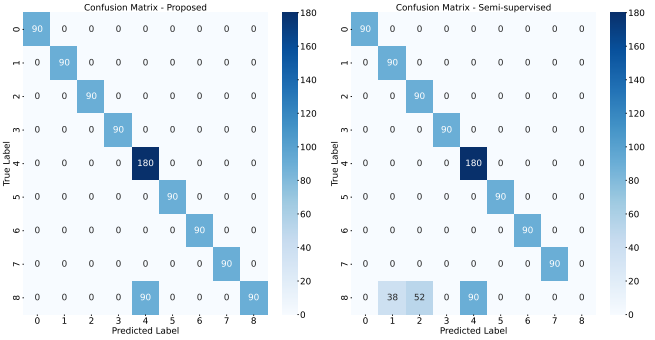

Fig. 3. Confusion matrices of different methods under 50 labeled samples with 2 unknown appliances (relabeled as class 8) (AP9, AP10).

samples. By extracting V-I trajectory images and using a dual-branch OVA classification architecture, the proposed method achieves accurate appliance identification and effective unknown appliance rejection. Experiments show that it improves identification accuracy and generalizes well across diverse appliances. Future research will aim to further improve the generalization capability of NILM models across diverse application scenarios.

ACKNOWLEDGMENTS

This work was supported in part by the Science and Technology Project of State Grid (No. 5700-202418277A-1-1-ZN).

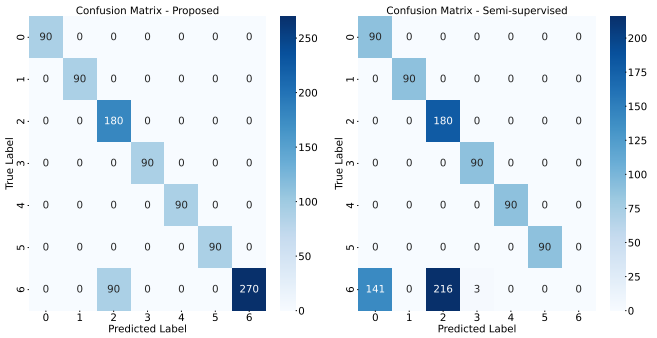

Fig. 4. Confusion matrices of different methods under 50 labeled samples with 4 unknown appliances (relabeled as class 6) (AP1, AP2, AP9, AP10).

REFERENCES

[1] S. Chotalia, C. Initiative, K. Lee, and D. Quispe, "Optimizing carbon emissions and cost reductions for household energy demand using machine learning," 2025.

[2] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[3] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*, 2015, pp. 55–64.

[4] G. Pusceddu, S. Manca, and L. Massidda, "Fine-tuning non-intrusive load monitoring model through user interaction: A practical approach to appliance recognition with limited labeled data," *Applied Energy*, vol. 391, p. 125943, 2025.

[5] Y. Han, K. Li, H. Feng, and Q. Zhao, "Non-intrusive load monitoring based on semi-supervised smooth teacher graph learning with voltage–current trajectory," *Neural Computing and Applications*, vol. 34, no. 21, pp. 19 147–19 160, 2022.

[6] B. Liu, K. Li, W. Luan, B. Zhao, J. Zhang, and Y. Wu, "Enhanced expandable representation learning for non-intrusive appliance incremental identification," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2024, pp. 1–5.

[7] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5609–5619, 2018.

[8] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[9] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "Whited-a worldwide household and industry transient energy data set," in *3rd International Workshop on Non-Intrusive Load Monitoring*, 2016, pp. 1–4.