Confidence-Aware Federated Learning for Smart Load Characterization in Cyber-Physical DER Systems

Vishal Krishna Singh, Vins Patel, Neeraj Jain, Chhaya Singh, Rajkumar Singh Rathore and Weiwei Jiang

***Abstract*—The integration of smart meters into residential environments has allowed a smooth collection of electricity consumption data, which is critical for demand response and coordinated operation in cyber-physical distributed energy resource systems. However, existing centralized methods of consumer characteristic identification pose significant risks to data privacy and confidentiality. Furthermore, the inherent limitations imposed due to noisy data cause a steep degradation in the accuracy and system-level decision-making. Addressing the issues of accuracy and data confidentiality, this paper proposes a confidence-aware federated learning framework for privacy-preserving inference of electricity consumer characteristics from raw smart meter data. The proposed method employs decentralized retention of smart meter data, using federated learning to refine network performance while ensuring that raw data remain localized at the client level. By evaluating a combined distribution of noisy and clean labels, erroneous data points are identified and excluded, thereby enhancing the model's efficacy and robustness. The effectiveness of the approach is validated using the Irish Commission for Energy Regulation dataset. The proposed framework demonstrates notable improvements, achieving an average gain of 3.97% in accuracy and 3.68% in MCC score compared to state-of-the-art methods, while maintaining strong data privacy guarantees.**

***Index Terms*—Confidence learning, Energy consumption patterns, Federated learning, Socio-demographic information, Smart meter, Noisy data.**

I. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) techniques have demonstrated significant potential in the analysis of smart meter data for smart load characterization, including the inference of socio-demographic attributes and electricity consumption patterns [1]. In cyber-physical distributed energy resource (DER) systems, such data-driven load intelligence plays a critical role in enabling demand response, coordinated operation, and efficient energy management [2], [3], [4], [5]. Accurate analysis of smart meter data supports informed decision-making, personalized energy services, and improved energy efficiency across residential and industrial consumers. However, the accuracy and reliability of the smart meter data depends on the quality of the labels in the data set used for model training [6], [7]. It is pertinent to consider that labeling extensive data sets, is challenging, time-consuming and susceptible to errors and inconsistencies. Noisy or inaccurate labels can significantly degrade model performance, undermining reliable load inference and, consequently, the effectiveness of cyber-physical DER system operations [8], [9].

One possible approach to overcome the issues of noisy labels is the logical estimation of confidence levels, linked to data set labels. This mapping, known as confidence learning (CL), can not only increase the overall quality and reliability of the model's inference process but also can significantly impact the overall quality of service. CL refers to the process of quantifying the level of confidence or uncertainty associated with each labeled sample in a data set. By estimating the uncertainty in data set labels, ML models can make informed decisions, assign appropriate weights to different labeled samples, and mitigate the negative effects of noisy or incorrect labels. One of the prominent works in [10], presents a comprehensive framework for estimating uncertainty in data set labels, utilizing techniques such as rank pruning and robust classification. The authors propose a novel approach to identify confident examples that are more likely to have accurate labels, thereby improving the overall performance and reliability of ML models. Targeting the applications of smart meter data, several works have been proposed to further enhance the quality and effectiveness of ML models. The authors in [11] present a deep learning-based approach for socio-demographic information identification from smart meter data. Similarly, the work in [12] presents a DL based detection and mitigation method for EV-based attacks on wind-integrated power grids. The authors in [13] focus on identifying incorrect annotations in multi-label classification data, addressing the challenge of label noise in the training process.

Among the existing approaches, the PCA-FANN model proposed in [14] serves as a key baseline for electricity consumer characteristic identification. This method combines Principal Component Analysis (PCA) for dimensionality reduction with a Feedforward Artificial Neural Network (FANN) for classi-

V. K. Singh is with School of Computer Science and Electronics Engineering, University of Essex, Colchester, U.K.
E-mail: v.k.singh@essex.ac.uk
V. Patel is with Indian Institute of Information Technology, Lucknow, India. (e-mail: mcs21016@iiitl.ac.in).
Neeraj Jain is with School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India. (e-mail: neeraj.jain@bennett.edu.in).
Chhaya Singh is with School of Computer Application and Technology, Galgotias University, Greater Noida. India. (e-mail: chhaya.singh@galgotiasuniversity.edu.in).
Rajkumar Singh Rathore is with Department of Computer Science, School of Technologies, Cardiff Metropolitan University, United Kingdom. (e-mail: rsrathore@cardiffmet.ac.uk).
Weiwei Jiang is with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. (e-mail: jww@bupt.edu.cn).
Manuscript received April, 2026.

TABLE I: Comparative Analysis of Existing Methods With Proposed Approach

Models	Noise Robustness	Scalability	Privacy	Interpretability	Accuracy	Advantages of Proposed Approach
IP2FL [3]	Medium	High	✓	High	High	1- Improved robustness to noisy labels through confidence-aware learning while improving reliability of inference.
PCA-FANN [14]	Low	Medium	✓	Medium	Medium	2- Low reliance on labeled data and centralized data sources, improving generalization.
DDEEP [15]	Medium	Medium	–	Low	Medium	3- Improved scalability and adaptability in large-scale energy environments.
Proposed Method	**High**	**High**	✓	**Medium-High**	**Excellent**	4- Integrated privacy-preserving FL with confidence-based noise filtering, reducing reliance on centralized data with improved accuracy.

fication. PCA reduces the dimensionality of high-dimensional smart meter data to improve computational efficiency, while the FANN performs the final classification task. However, this approach relies on centralized data collection, is sensitive to noisy labels, and does not incorporate any explicit privacy-preserving or uncertainty estimation mechanisms, which limits its effectiveness in distributed and real-world environments.

In addition to the aforementioned works, various studies have investigated the application of smart meter data analytics in the energy sector as well. For instance, the work in [16] analyzes management practices and risks in the electricity market. Interestingly, household characteristic identification based on smart meter data remains vastly unexplored, with very limited research in the domain. One of the few published research is the work in [17], where a time-frequency feature combination approach for household characteristic identification is proposed. Another approach in [18] employs ML techniques to uncover residential energy consumption patterns based on socioeconomic factors and smart meter data. One of the seminal works on feature identification is the work in [19] which presents a cross-domain feature extraction-based approach for household characteristic identification using smart meter data. However, these methods rely on centralized data collection, which raises significant privacy and security concerns while limiting user data ownership. They are also highly vulnerable to noisy labels, leading to degraded model performance and unreliable predictions. Furthermore, these approaches lack scalability in distributed environments and do not incorporate robust privacy-preserving mechanisms. The effectiveness of these established methods is assessed across multiple parameters, emphasizing the significance of evaluation metrics in validating the algorithms. Various metrics can be employed; for instance, [20] investigates the influence of class imbalance on classification performance metrics utilizing the binary confusion matrix. The authors in [21] present a federated learning (FL) driven secure computing mechanism for ensuring reliable and efficient transportation in cyber-physical systems.

Addressing the challenges of data confidentiality and unreliable inference in cyber-physical DER systems, this work proposes a novel confidence-aware FL framework for privacy-preserving smart load characterization through household characteristic inference from raw smart meter data. The proposed method employs decentralized retention of smart meter data, using FL to refine network performance. The proposed method ensures data confidentiality, while the use of CL enhances the accuracy of the label. By evaluating a combined distribution of noisy and clean labels, erroneous data points are identified and excluded, thereby enhancing the model's efficacy. The idea is to develop a framework for accurate consumer characteristic identification from smart meter data, while addressing challenges such as noisy labels, data privacy and robustness. Feature estimation, extraction, and classification from smart meter data is performed in noisy conditions while prioritizing end-user privacy and accuracy through weighted average estimation and distributed learning. A comprehensive comparison and advantages of the proposed method, over some of the existing methods, is presented in Table I. The novel contributions of this work are summarized as follows:

1) **Privacy-Preserving Smart Load Intelligence**: A privacy-preserving learning framework is proposed to enable smart load characterization in cyber-physical energy systems by minimizing raw data exchange across smart meters. Through FL, the proposed approach ensures that household electricity data remain localized while supporting collaborative model training across the network.
2) **Uncertainty Estimation**: A confidence-based learning mechanism is integrated within the federated framework to distinguish confident examples with potentially accurate labels from noisy examples. The proposed confidence-FL approach ensures data confidentiality without compromising on the accuracy of the label values.
3) **Robust Classification**: A comprehensive analysis is presented to prove that the proposed confidence-FL approach improves the overall performance of ML models by enhancing classification robustness through the pruning of potentially unreliable examples.
4) **Resilience to Label Noise in Federated Settings**: A comprehensive study of continual learning mechanisms and their ability to assist FL models in effectively identifying and mitigating label noise across devices. This ensures consistent and reliable learning throughout the network.

This paper is structured as follows: Section II describes the problem description. Section III provides information on the system model and data set description. Section IV describes the details of the proposed methodology. Section V presents the experimental setup and results, and finally Section VI concludes the study.

II. Problem Description

A majority of the existing approaches have failed to address the privacy and accuracy concerns of characteristic inference from smart meter data. A trade-off between the two i.e privacy and accuracy, has been mostly considered by the researchers and thus, this work addresses the following challenges in the existing works:

1) **Privacy:** Extracting information from smart meter data poses a substantial risk of privacy breaches for user data. Non-intrusive methods have the potential to utilize high-resolution data to dis-aggregate the total energy consumption of end users, accurately obtaining per-appliance consumption, thereby presenting a significant privacy threat.
2) **Noisy Labels:** Model accuracy is jeopardized by inaccurate labels, which can be introduced into the data for various reasons. For instance, consider cases where the wrong pairing of devices and appliances occurs. Additionally, the potential for compromised labels due to targeted attacks on local datasets cannot be overlooked.
3) **Robust Classification:** Ensuring model resilience against noisy data is crucial for reliable predictions and verifying the model's performance.

This work is aimed at developing a privacy-preserving learning framework to enable smart load characterization in cyber-physical energy systems by minimizing raw data exchange across smart meters while addressing challenges such as noisy labels, privacy concerns, and the model's robustness.

III. System Model and Dataset Description

A. System Model

A residential smart grid scenario is considered for the realization of the proposed framework, where households equipped with smart meters act as primary data sources and are connected through a cyber-physical DER network. Each household is treated as a retailer (client) in the FL architecture and retains its raw smart meter data locally. In the intended deployment, all data processing steps - including feature extraction, label mapping, and confidence-based noisy label filtering - are performed locally at each retailer without sharing raw electricity consumption data with any central entity. A master server is responsible only for coordinating the federated training process by aggregating local model updates received from participating retailers and redistributing the updated global model. The proposed framework is therefore privacy-preserving by design, as only model parameters or gradients are exchanged during the federated optimization process, while raw data remain strictly localized. The master server performs only model aggregation and coordination, and does not store or access raw smart meter data at any stage.

While the proposed framework is generalizable to broader smart city environments involving commercial and industrial infrastructures, the present study focuses on residential consumers due to the availability of detailed smart meter and socio-demographic data in the Irish CER dataset. Accordingly, each retailer corresponds to a residential consumer characterized by occupant information, dwelling properties, and appliance usage patterns. For experimental validation, the Irish CER dataset, which is originally available in centralized form, is partitioned into multiple logical retailers to simulate decentralized clients in a proof-of-concept FL setting. This partitioning is used solely for evaluation purposes and does not imply centralized data sharing in the intended deployment.

B. Dataset Description

The proposed work is based on the Irish CER dataset [22], which was collected as part of the *Smart Metering Project* in Ireland. The dataset contains raw smart meter data from households with a 30-minute granularity and covers a period of 536 days from July 2009 to December 2010. A total of $4,232$ residential consumers were included in the study. The proposed work utilizes 15 socio-demographic characteristics from the pre-trial survey, selected from a total of 140 questions. For better visualization and interpretability, a restricted number of class labels were assigned to each characteristic. These characteristics are categorized into three main groups: *occupant information*, *dwelling properties*, and *domestic appliance properties*.

Due to the inherent nature of the dataset, several class labels exhibit significant imbalance, with considerable disparity in the number of consumers across different categories. The proposed algorithm is designed to account for this imbalance during both model training and performance evaluation. Although the Irish CER dataset is available as a centralized benchmark dataset, in this work it is partitioned into multiple logical client nodes to emulate a decentralized FL environment. This partitioning is performed solely for proof-of-concept evaluation of the proposed privacy-preserving framework and does not imply centralized data storage or sharing in the target deployment.

IV. Proposed Methodology

The proposed approach integrates the capabilities of federated learning (FL) and confident learning (CL) to address key challenges such as data privacy, noisy labels, and robustness in electricity consumer characteristic identification from smart meter data. Using the Irish CER electricity dataset, 68 relevant features are extracted and used as inputs to the proposed *Confidence Federated ANN* framework, termed as *C-FANN*. In the intended deployment, raw smart meter data remain localized at each participating retailer/client, where feature engineering, confidence-based noisy label filtering, and local model training are performed independently. Only model parameters or gradients are shared with the master server for aggregation, thereby ensuring data confidentiality and preventing raw data exchange. For proof-of-concept evaluation, the centralized Irish CER dataset is partitioned into multiple logical retailers to emulate decentralized clients in a federated setting. Accordingly, the architecture in Fig. 1 reflects the decentralized nature of the proposed framework and follows a standard federated learning paradigm, where a single master server coordinates model aggregation while all data processing and training remain decentralized at the retailer level.

A. Feature Engineering and Pre-Processing

Selecting reliable characteristics that can capture daily and weekly trends is crucial, given the dynamic nature of an

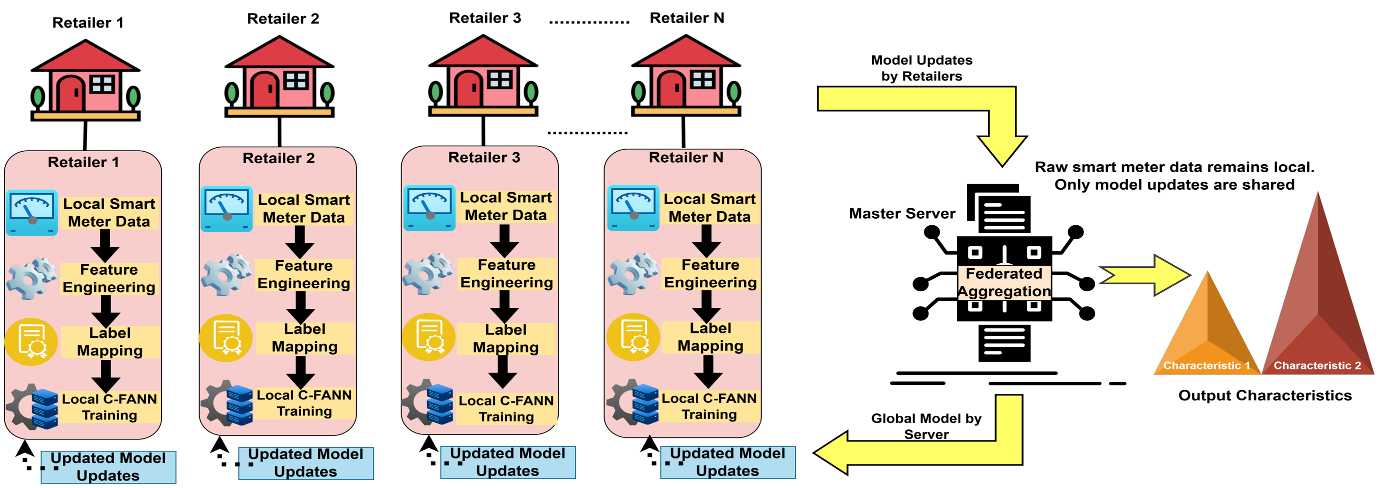


Fig. 1: Federated Architecture of the Proposed C-FANN Framework.

electricity user's load profile throughout a typical day. We establish typical weekly load profiles for each customer, drawing inspiration from previous work in [15] and [23] to manually define the feature set. This feature set is categorized into three main groups: *statistics* (including mean, minimum, maximum, and total consumption), *ratios* (such as the load factor), and *statistics* specific to time series data (including quantiles and entropy). We also consider the disparity in energy consumption patterns between weekdays (*Monday* to *Friday*) and weekends as part of the feature set. Additionally, we explore characteristics at distinct intervals throughout the day, categorizing these periods as *baseload* (2 am to 5 am), *morning* (6 am to 10 am), *midday* (11 am to 2 pm), and *evening* (6 pm to 10 pm). Throughout the trial period, these characteristics are determined for each customer by using weekly consumption data for *weekdays* as well as *weekends*. For the sake of clarity, a self-explanatory term is given to each feature, for instance, the term *'mean_morning_weekdays'* refers to the mean amount of electricity consumption during the early hours of a workday. In a similar vein, the term *'maximum_base_weekend'* refers to the maximum consumption that occurs on a weekend during baseload hours. The *'proportion_mean_GE'* shows the proportion of consumption that is more than the weekly mean, while the *'proportion_0.5kW_GE'* displays the proportion of energy consumption usage that is greater than or equivalent to 0.5 kW during the week. For the sake of simplicity, in the upcoming sections of this work, we will refer to the feature set created by the 'n^{th}' provider as '$\mathbf{SD}_n$' instead of the initial smart meter data.

In the intended federated deployment, the above feature engineering and preprocessing steps are executed locally at each retailer using its own smart meter records. Thus, the feature matrix corresponding to the n^{th} retailer, denoted by $\mathbf{SD}_n$, is generated without transferring raw smart meter data to any central entity. In the experimental proof-of-concept, the centralized Irish CER dataset is partitioned into multiple logical retailers, and the same local preprocessing pipeline is applied independently to each partition.

B. Data Processing Using Confidence Learning

Within the proposed federated framework, CL is applied locally at each retailer on its own feature set and labels before local model optimization. In this way, potentially unreliable samples are identified and filtered without exposing raw data or labels to the master server. The CL mechanism therefore serves as a local noise-mitigation step that improves the quality of client-side training data prior to federated aggregation. CL is employed to predict the collaborative distribution of the reported labels and hidden true labels while also considering the effect of noise. For the purpose of this estimation, two inputs are considered: *the out-of-sample predicted probabilities* and *the vector of noisy labels*. The step-wise execution of the algorithm *(Algorithm 1)*, is presented below:

1) Notation and Definitions: The training data set is denoted by $\mathbf{T_D} = \{(x_j, y_j)\}_{j=1}^{N}$, where N represents the total count of training samples, such that a label y_j is associated with true label input features x_j. Let x_j represent the attribute vector of the j^{th} instance, y_j be its corresponding observed label, and z_j denotes the unknown true label, which we aim to estimate using CL. It is empirical to note that the data set may contain mislabeled instances, where the label y_j does not match the true label z_j for instance x_j.
2) Model Training: Train a supervised learning model f by utilizing the labeled data set $\mathbf{SD}_p$. The model f can take the form of any classifier.
3) Model Confidence Score: For every instance (x_j) in the local data set, acquire the confidence scores ($s_j = (s_{j1}, s_{j2}, \ldots, s_{jC})$) from the model ($f$) for all potential classes. Here, C denotes the number of classes, and the confidence scores indicate the model's level of certainty regarding its predictions for each class.
4) Instance-Wise Confidence: Define instance-wise confidences (c_j), which indicate the classifier's (f) confidence in its predictions for each instance (x_j), using the margin-based approach as:

$$c_j = \max(s_j) - \text{average}(s_j) \tag{1}$$

where $\max(s_j)$ is the highest confidence score for the instance x_j and average(s_j) is the average confidence score for the same instance. A higher c_j indicates higher confidence in the model's prediction for x_j.

5) Ranking Instances: Sort the instances based on their instance-wise confidences (c_j) in descending order. Instances with higher c_j are considered more reliable, while those with lower c_j are more likely to be mislabeled.
6) Cleaning Mislabeled Instances: Start by examining the instance with the highest instance-wise confidence, then proceed iteratively to assess instances, identifying and rectifying mislabeled ones as necessary. This process can involve human inspection or domain-specific methods, with the underlying assumption that instances with higher confidences are more likely to possess accurate labels.
7) Re-training Classifier: After cleaning the data set, retrain the classifier using the cleaned data set. This helps the model to better adapt to the corrected labels and helps in the proliferation of output accuracy.
8) Uncertainty Estimation: Compute the uncertainty of the refined labels by contrasting the initial model's predictions for the cleaned data set with the retrained model's predictions. Instances where there is a discrepancy between the two models suggest increased uncertainty. Elevated uncertainty may signify that these instances remain difficult to classify or that they are potentially mislabeled.
9) Cleaning Iteration: The cleaning process may require multiple iterations. After one iteration of cleaning and retraining, the process may be repeated to further identify and correct mislabeled instances. The iterative approach continues until the data set's performance stabilizes or reaches a satisfactory level.

C. Federated C-FANN Training

A feed-forward ANN model with dense layers is trained for each characteristic. Assuming a fixed number of hidden layers, the transformation at the u^{th} hidden layer is expressed as:

$$e^{(u)} = W^{(u)}\mathbf{r}^{(u)} + \mathbf{b}^{(u)} \tag{2}$$

where $W^{(u)}$ and $\mathbf{b}^{(u)}$ denote the weights and biases of the u^{th} hidden layer, respectively. The input to the u^{th} layer, which is also the output of the $(u-1)^{th}$ layer, is given by $\mathbf{r}^{(u)} = \text{relu}(e^{(u-1)})$, where relu$(\cdot)$ is the activation function. The final classification layer uses the Softmax function. The overall identification model for the i^{th} characteristic is written as:

$$\hat{y}_{i,j} = CM_i(\mathbf{\Theta}_i, x_j) \tag{3}$$

where $\mathbf{\Theta}_i$ denotes the collection of all weights and biases used by the ANN to infer the i^{th} characteristic, and x_j is the j^{th} data sample. To train the model, categorical cross-entropy is used as the objective function:

$$L_i = \frac{1}{N}\sum_{j=1}^{N}[y_{i,j}\log\hat{y}_{i,j} - (1-y_{i,j})\log(1-\hat{y}_{i,j})] \tag{4}$$

Algorithm 1 Local Data Processing using CL

Require: Local tabular dataset $\mathbf{X}_p$ with K features at retailer p
Require: Local labels $\mathbf{y}_p$
Require: Threshold θ for filtering uncertain samples
Require: Maximum number of iterations *max_iter*
Ensure: Cleaned local dataset $\mathbf{X}_{p,\text{clean}}$ and labels $\mathbf{y}_{p,\text{clean}}$
Step 1: Preprocessing
Preprocess the local dataset $\mathbf{X}_p$ (handle missing values, feature scaling)
Step 2: Initial Classifier Training
Train an initial classifier $f(\mathbf{X}_p)$ on the local dataset $\mathbf{X}_p$ and obtain predictions $\hat{\mathbf{y}}_p$
Step 3: Confidence Score Calculation
for $i = 1$ to n_p **do**
  Calculate the class probabilities $p(c)$ for sample i
  Calculate the entropy $H_i = -\sum_{c=1}^{C} p(c)\log p(c)$
  Calculate the confidence score $s_i = 1 - H_i$
end for
Step 4: Data Filtering
Initialize $\mathbf{X}_{p,\text{clean}}$ and $\mathbf{y}_{p,\text{clean}}$ as empty lists
for $i = 1$ to n_p **do**
  if $s_i \geq \theta$ **then**
    Add $\mathbf{X}_{p,i}$ to $\mathbf{X}_{p,\text{clean}}$ and add $y_{p,i}$ to $\mathbf{y}_{p,\text{clean}}$
  end if
end for
Step 5: Classifier Re-training
Train a new classifier $f_{\text{clean}}(\mathbf{X}_{p,\text{clean}})$ on the cleaned local dataset $\mathbf{X}_{p,\text{clean}}$
Step 6: Iterative Confidence Estimation
Set $iter = 1$
while $iter \leq$ *max_iter* and confidence scores not converged **do**
  for $i = 1$ to n_p **do**
    Calculate the class probabilities $p(c)$ for sample i
    Calculate the entropy $H_i = -\sum_{c=1}^{C} p(c)\log p(c)$
    Calculate the confidence score $s_i = 1 - H_i$
  end for
  Filter out uncertain local samples as in Step 4
  Retrain the classifier $f_{\text{clean}}(\mathbf{X}_{p,\text{clean}})$ on the cleaned local dataset $\mathbf{X}_{p,\text{clean}}$ with labels $\mathbf{y}_{p,\text{clean}}$
  Increment $iter$ by 1
end while
Step 7: Final Prediction
Use the final local classifier f_{clean} to make predictions on new local data

Back-propagation is used to compute the gradient of the global objective as:

$$\nabla\Theta_i = \frac{\partial L_i(\Theta_i)}{\partial\Theta_i} \tag{5}$$

Within the proposed federated ANN architecture, the model is not trained on a centrally pooled dataset. Instead, each retailer p performs local training using its own cleaned feature set SD_p and corresponding labels SY_p, and computes the local gradient for the i^{th} characteristic as follows:

$$\nabla\Theta_{p,i} = \frac{\partial L_i(\Theta_{p,i})}{\partial\Theta_{p,i}} \tag{6}$$

The local gradients are then transmitted to the master server, where they are aggregated with other local gradients to determine the global gradient:

$$\nabla\Theta_i = \sum_{p=1}^{n} a_{p,i}\nabla_{\Theta_{p,i}} \tag{7}$$

where $a_{p,i}$ stands for the weight of the gradient contributed by the p^{th} retailer for the i^{th} characteristic. It is important to emphasize that only local model gradients/parameters are transmitted to the master server. Raw smart meter data, extracted features, and original labels remain local to each retailer throughout the training process. The master server performs only aggregation and redistributes the updated global model, thereby preserving consistency with the privacy-preserving federated setting.

The global parameter Θ_i is updated by the server using the Adadelta optimizer [24]. Specifically, at the q^{th} iteration, the biased first moment m_q and second moment v_q are initialized as follows:

$$m^q = \beta_1^q m^{q-1} + (1 - \beta_1^q)\nabla\Theta_i^q \tag{8}$$

$$v^q = \beta_2^q v^{q-1} + (1 - \beta_2^q)\nabla{\Theta_i^q}^2 \tag{9}$$

where β_1^q and β_2^q indicate the exponential moving-average coefficients for m^q and v^q, respectively. Considering this, the unbiased moments $\hat{m}_q$ and $\hat{v}_q$ are calculated as:

$$\hat{m}_q = \frac{m_q}{1 - \beta_1^q} \quad \hat{v}_q = \frac{v_q}{1 - \beta_2^q} \tag{10}$$

Therefore, the global parameter Θ_i is updated as:

$$\Theta_i^{q+1} = \Theta_i^q - \frac{\eta \hat{m}_q}{\sqrt{\hat{v}_q} + \epsilon} \tag{11}$$

where η is the learning rate, and ϵ is a very small constant introduced to prevent division by zero. The steps of the proposed federated training procedure are summarized in Algorithm 2. The proposed *C-FANN* framework combines confidence-based data cleaning, federated optimization, and ANN-based classification to achieve accurate consumer characteristic identification in a competitive retail environment while preserving privacy and data security.

Algorithm 2 Federated Training for Smart Load Characterization

Initialize hyperparameters and model architecture
Partition the benchmark dataset into n logical retailers for proof-of-concept federated simulation
At each retailer, prepare local training and testing partitions independently
for epoch $= 1$ to $t_{\max}$ **do**
    for retailer_id $= 1$ to n **do**
        Initialize local model model[retailer_id] from the current global model
        Apply local confidence-based filtering to the retailer-specific training data
        for $t = 1$ to n_{local} **do**
            for each batch in train_loader[retailer_id] **do**
                Compute model output and loss
                Compute gradients and accumulate local loss
            end for
        end for
    end for
    Reset gradients of the global model
    Transmit only local model updates from each retailer to the master server
    for retailer_id $= 1$ to n **do**
        Compute scaling factor a
        Aggregate gradients from model[retailer_id] into the global model
    end for
    Update global model parameters using Adadelta optimiser [24]
    Distribute the updated global model back to all retailers
    Evaluate the global model on the test dataset
    Compute Accuracy and MCC
    Update best Accuracy and MCC if improved
    Print evaluation metrics for current epoch
end for
Print best Accuracy and MCC achieved during training

V. Experimental Setup and Results

A. Experimental Setup

The identified dataset is partitioned into ten folds. In this analysis, K-fold cross-validation is performed, where K is set to 10. $K-1$ folds are evaluated for C-FANN training, and the remaining fold is used for testing. This process is repeated K times, such that each fold is used once for testing. The accuracy across the ten validation runs is averaged to obtain the final cross-validation performance of C-FANN. As mentioned before, the Irish CER dataset is originally available in a centralized form. For proof-of-concept evaluation of the proposed FL framework, the training portion of the dataset is partitioned into $N = 5$ logical retailers (clients), each containing approximately 16% of the total consumers. This partitioning is performed solely to emulate decentralized data ownership in a FL environment and does not imply centralized data sharing in the intended deployment. In the target system, each retailer independently retains its own smart meter data and performs local preprocessing, confidence-based filtering, and model training. Only model updates are exchanged with the master server for aggregation.

Due to the relatively small size of each local partition, the ANN model is configured with two hidden layers, along with dropout between the hidden and output layers. Each individual retailer trains a local model using its own data partition, thereby simulating decentralized training in a FL setting.

B. Results

The proposed *C-FANN* exhibits a substantial enhancement in *accuracy* and *MCC* score, as compared to the existing method of PCA-FANN in [14].

1) *Confusion Matrix Analysis:* The incorporation of weighted prediction during the training of the proposed C-FANN leads to a significant improvement in classification performance. This enhancement is reflected in increased true positives (TPs) and reduced false negatives (FNs) and false alarm rate (FAR). Fig. 2 presents the confusion matrices at different training iterations, specifically at 100, 500, and 1000. As observed in Fig. 2, the model demonstrates high TP rates and low FN rates, indicating its ability to correctly classify the majority of the samples. A low FN rate further reflects the model's strong sensitivity. This performance can be attributed

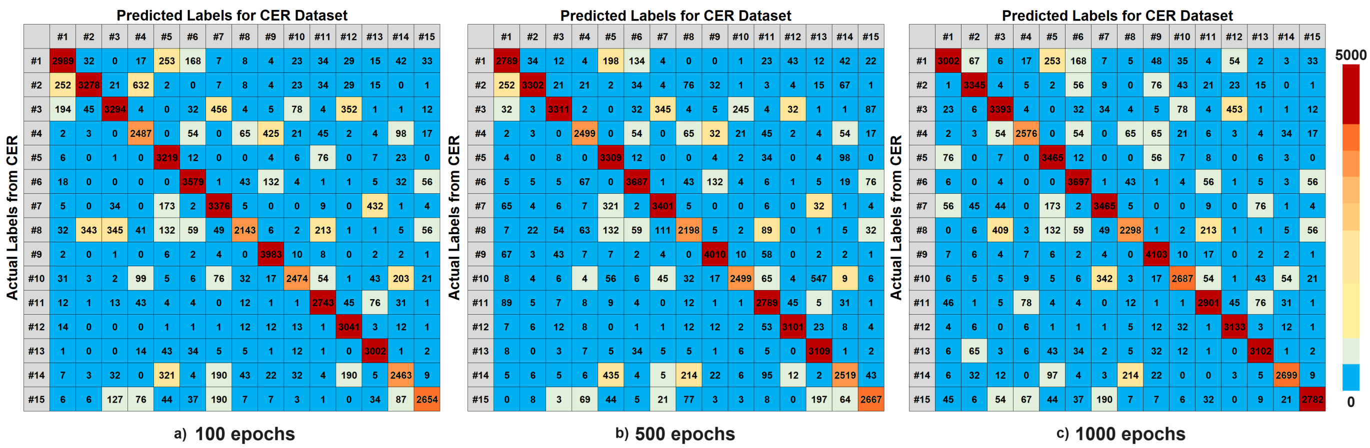


Fig. 2: Performance Analysis: Confusion Matrix of Proposed C-FANN at Different Iterations

to the model's ability to assign higher importance to reliable samples through confidence-based weighting. As discussed in Section IV-B, the data cleaning process followed by retraining enables the model to better adapt to corrected labels, thereby improving overall classification accuracy. The results indicate that the model effectively prioritizes informative and challenging instances, leading to enhanced predictive performance.

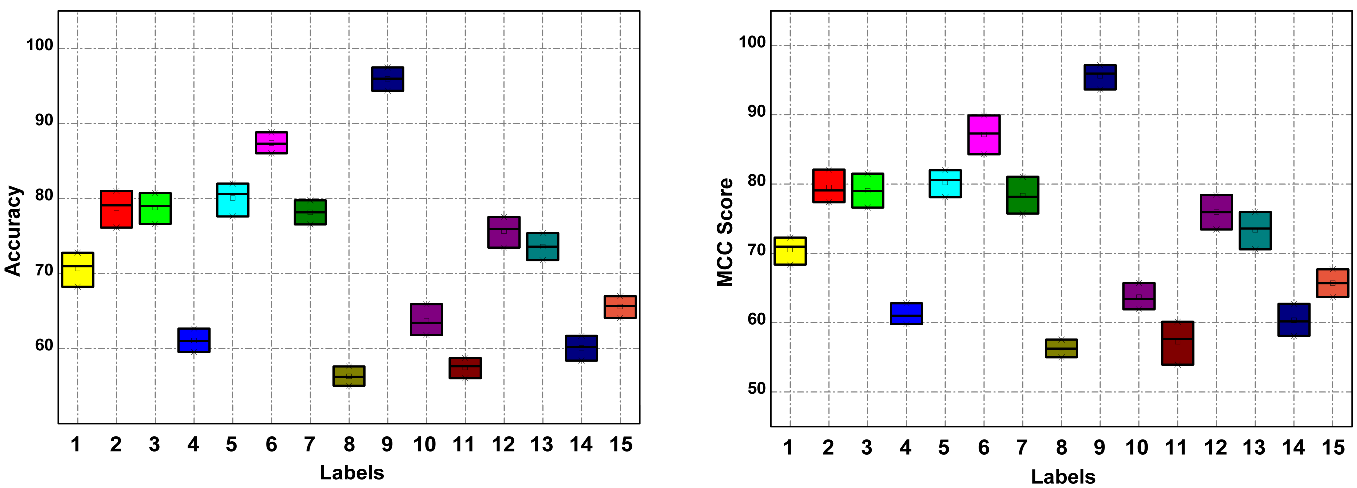
2) *Comparative Labeled Analysis:* Table II presents a comprehensive comparison of the accuracy and MCC scores for the proposed C-FANN and the PCA-FANN method [14] across various labels of the CER dataset. The performance is evaluated using two metrics: *Accuracy* and *MCC Score*. The evaluated characteristics, labeled from #1 to #15, correspond to attributes such as age, employment status, retirement status, social class, presence of children, living status, number of residents, age of the house, ownership status, number of bedrooms, cooking habits, number of home appliances, number of entertainment appliances, and proportion of light bulbs. An additional *Average* row summarizes the overall performance. It is evident from Table II that the proposed C-FANN consistently outperforms PCA-FANN across most characteristics in terms of both accuracy and MCC score. The key observations are summarized as follows:
   - The proposed C-FANN model generally exhibits higher accuracy and MCC scores compared to PCA-FANN.
   - The percentage improvement consistently favors the proposed C-FANN for both performance metrics.
   - Significant improvements are observed in attributes such as 'Age of House', 'Type of Home', 'Number of Home Appliances', and 'Number of Entertainment Appliances'.
   - Overall, the proposed model achieves an average accuracy improvement of approximately 3.97% and an average MCC improvement of approximately 3.67%.
3) *Accuracy Analysis:* The integration of CL plays a significant role in improving classification accuracy. By assigning confidence scores to individual data samples, the model prioritizes high-quality data while reducing the impact of noisy or uncertain samples. Fig. 3a illustrates the average accuracy achieved by the proposed C-FANN across different labels of the CER dataset. The results demonstrate that FL enables collaborative training across decentralized data sources, thereby improving model generalization and predictive performance.
4) *MCC Score Analysis:* Fig. 3b presents the average MCC score achieved by the proposed C-FANN across different labels. The MCC value of 0.3374 indicates a significant improvement in balanced classification performance. This improvement is attributed to the combined effect of CL and FL, where CL enhances robustness by filtering unreliable samples, and FL improves generalization by leveraging diverse data distributions. This synergy enables the model to learn more discriminative feature representations and achieve superior performance compared to existing methods.

VI. Conclusion and Future Work

In this work, we have proposed a novel confidence-aware FL framework, termed as C-FANN, for electricity consumer characteristic identification using raw smart meter data in a privacy-preserving manner. The proposed methodology integrates CL and FL to effectively address key challenges such as noisy labels, data privacy, and robustness in decentralized environments. The experimental evaluation, conducted on the Irish CER residential dataset, demonstrates that the proposed C-FANN framework achieves notable improvements in both classification accuracy and MCC score when compared with baseline approaches such as PCA-FANN and conventional FL models. These performance gains can be attributed to the effective integration of confidence-based noisy label filtering with federated optimization, along with carefully engineered domain-specific features. The results highlight the effectiveness of the proposed framework in accurately inferring household-level characteristics, including occupant information, dwelling properties, and appliance usage patterns, while ensuring that raw smart meter data remain localized at the client level.

TABLE II: Comparative Analysis of Average Accuracy and Average MCC Score for Different Labels of CER Dataset

Characteristics	**Average Accuracy**			**Average MCC Score**		
	Proposed C-FANN	**PCA-FANN [14]**	**Improvement**	**PCA-FANN [14]**	**Proposed C-FANN**	**Improvement**
#1 Age	70.97	67.43	**3.54**	0.2381	0.3474	**0.1093**
#2 Employment Status	79.09	74.20	**4.89**	0.4586	0.5852	**0.1266**
#3 Retirement Status	79.01	76.51	**2.50**	0.4255	0.5021	**0.0766**
#4 Social Class	61.01	56.47	**4.54**	0.2606	0.2707	**0.0101**
#5 Have Children	80.60	77.39	**3.21**	0.4104	0.5346	**0.1242**
#6 Living Status	87.31	84.65	**2.66**	0.4894	0.5216	**0.0322**
#7 Number of Residents	78.15	75.27	**2.88**	0.5044	0.5549	**0.0505**
#8 Age of House	56.25	51.21	**5.04**	0.1878	0.2472	**0.0594**
#9 Rent or Owned	95.96	92.98	**2.98**	0.1137	0.1452	**0.0315**
#10 Number of Bedroom	63.42	59.41	**4.01**	0.2679	0.3268	**0.0589**
#11 Type of Home	67.64	61.76	**5.88**	0.2303	0.2379	**0.0076**
#12 Cooking	75.95	73.11	**2.84**	0.2829	0.3217	**0.0388**
#13 Number of Home Appliances	73.58	67.06	**6.52**	0.3676	0.3977	**0.0301**
#14 Number of Entertainment Appliances	60.20	54.87	**5.33**	0.2575	0.3076	**0.0501**
#15 Proportion of Light Bulbs	65.70	62.84	**2.86**	0.0161	0.0389	**0.0228**
Average	**72.98**	**69.01**	**3.97**	**0.3007**	**0.3374**	**0.0367**



(a) Average Accuracy for Different Labels of CER

(b) Average MCC Score for Different Labels of CER

Fig. 3: Performance Analysis of Proposed C-FANN on CER Dataset

Although the present study focuses on residential consumers due to the availability of detailed benchmark data, the proposed framework is generalizable to broader cyber-physical energy systems, including commercial and industrial environments, subject to the availability of suitable datasets. As part of future work, the authors aim to explore ensemble learning strategies to further enhance the robustness and generalization capability of the proposed C-FANN framework. Additionally, the framework will be evaluated on diverse real-world datasets to assess its scalability and effectiveness in practical deployment scenarios.

REFERENCES

[1] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "Deepfed: Federated deep learning for intrusion detection in industrial cyber–physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5615–5624, 2021.

[2] M. K. Quan, P. N. Pathirana, M. Wijayasundara, S. Setunge, D. C. Nguyen, C. G. Brinton, D. J. Love, and H. V. Poor, "Federated learning for cyber physical systems: A comprehensive survey," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2025.

[3] D. Namakshenas, A. Yazdinejad, A. Dehghantanha, R. M. Parizi, and G. Srivastava, "Ip2fl: Interpretation-based privacy-preserving federated learning for industrial cyber-physical systems," *IEEE Transactions on Industrial Cyber-Physical Systems*, vol. 2, pp. 321–330, 2024.

[4] V. K. Singh, J. Singh, R. S. Rathore, and D. Nema, "Gradient-based compression and approximate computing for deep network optimization in multimedia data processing," *IEEE Transactions on Consumer Electronics*, 2026.

[5] V. K. Singh, N. Jain, G. Tripathi, and S. Sahani, "Correlated channeled spatio temporal graph attention network model for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[6] N. Jain, C. Singh, V. K. Singh, R. S. Rathore, N. S. Alghamdi, and C. Hewage, "Federated genetic optimization for secure and privacy-preserving sensor localization in consumer iot applications," *IEEE Transactions on Consumer Electronics*, 2025.

[7] V. K. Singh, C. Singh, and H. Raza, "Event classification and intensity discrimination for forest fire inference with iot," *IEEE sensors journal*, vol. 22, no. 9, pp. 8869–8880, 2022.

[8] Y. Liao, Y. Wang, Q. Cui, K.-C. Chen, G. Nan, and X. Tao, "Data-driven cyber-physical anomaly detection with gan in federated smart factories," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 4, pp. 3067–3076, 2025.

[9] S. Srivastava, D. Agarwal, B. K. Chaurasia, V. K. Singh, R. S. Rathore, and W. Jiang, "Federated blockchain-based resilient identity management for securing consumer vehicular networks," *IEEE Transactions on Consumer Electronics*, 2025.

[10] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *Journal of Artificial Intelligence Research*, vol. 70, pp. 1373–1411, 2021.

[11] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C. Kang, "Deep learning-based socio-demographic information identification from smart meter data," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2593–2602, 2018.

[12] A. Abazari, M. M. Soleymani, M. Ghafouri, D. Jafarigiv, R. Atallah, and C. Assi, "Deep learning detection and robust mpc mitigation for ev-based load-altering attacks on wind-integrated power grids," *IEEE Transactions on Industrial Cyber-Physical Systems*, vol. 2, pp. 244–263, 2024.

[13] A. Thyagarajan, E. Snorrason, C. Northcutt, and J. Mueller, "Identifying incorrect annotations in multi-label classification data," *arXiv preprint arXiv:2211.13895*, 2022.

[14] Y. Wang, I. L. Bennani, X. Liu, M. Sun, and Y. Zhou, "Electricity consumer characteristics identification: A federated learning approach," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3637–3647, 2021.

[15] H. Liang, J. Ma, R. Sun, and Y. Du, "A data-driven approach for targeting residential customers for energy efficiency programs," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1229–1238, 2019.

[16] Y. Zhilkina, "Risks in the energy sector: Management practices analysis in the electricity market," in *E3S Web of Conferences*, vol. 139. EDP Sciences, 2019, p. 01071.

[17] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, Z. Mi, H. Chen, and S. Chang, "Time–frequency feature combination based household characteristic identification approach using smart meter data," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2251–2262, 2020.

[18] W. Tang, H. Wang, X.-L. Lee, and H.-T. Yang, "Machine learning approach to uncovering residential energy consumption patterns based on socioeconomic and smart meter data," *Energy*, vol. 240, p. 122500, 2022.

[19] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, H. Chen, and S. Chang, "Cross-domain feature extraction-based household characteristics identification approach using smart meter data," in *2019 IEEE Industry Applications Society Annual Meeting*. IEEE, 2019, pp. 1–9.

[20] A. Luque, A. Carrasco, A. Martín, and A. de Las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019.

[21] N. K. Singh, S. Patni, S. Lim, and J. Lee, "Federated learning-based secure computing mechanism for consumer internet of vehicles-based transportation cyber-physical systems," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 2, pp. 4867–4882, 2025.

[22] Commission for Energy Regulation (CER), "CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010," [dataset] 1st Edition. Irish Social Science Data Archive. SN: 0012-00, 2012. [Online]. Available: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/

[23] S. AbdulRahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, "A survey on federated learning: The journey from centralized to distributed on-site learning and beyond," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5476–5497, 2020.

[24] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

Vishal Krishna Singh received his bachelor's degree in Information Technology, in 2010, the master's degree in Computer Technology and Application, in 2013, and PhD degree in Information Technology from Indian Institute of Information Technology, Allahabad, India in 2018. He is currently working as a Lecturer and is associated with the Networks and Communications Research Group at School of Computer Science and Electronics Engineering, University of Essex, Colchester, U.K.

Vins Patel received the bachelor's degree in Information Technology in 2016, the master's degree in Computer Science from Indian Institute of Information Technology, Lucknow, India in 2021. Currently he is working in Intel, Bengaluru, India. His research interests include Machine Learning, Internet of Things, and Data Analytics.

Neeraj Jain received his bachelor's degree in Information Technology, in 2007, the master's degree in Information Technology, in 2010, and PhD degree in Information Technology from Indian Institute of Information Technology, Allahabad, India in 2018. He is currently working as an Associate Professor at School of Computer Science Engineering and Technology, Bennet University, Greater Noida, India.

Chhaya Singh is an Associate Professor in the School of Computer Science and Engineering at Galgotias University, India. She received her Ph.D. from MANIT Bhopal and M.Tech. from IIIT Allahabad. Her research interests include Machine Learning, Data Science, and Fuzzy Logic. She has published several SCOPUS and SCI-indexed research articles and holds a design patent. Dr. Singh has also worked as a Research Assistant on a GCRF-funded project in collaboration with the University of Essex, UK. She has received multiple research fellowships and the Young Scientist Award for her contributions to computational biology.

Dr. Rajkumar Singh Rathore (Senior Member IEEE) is working as Head of Cyber Security of Connected and Autonomous Systems, CINC, Head of Cyber Physical and Networks Systems, CeRISS & Programme Director for MSc Computing and IT in Department of Computer Science at Cardiff Metropolitan University's School of Technologies, United Kingdom. He has gained doctorate degree, dual master's degrees, and bachelor's degree all in Computer Science and Engineering discipline. His research expertise are Wireless Communications, Internet of Things/Cyber Physical Systems, Cyber Security and Privacy, Connected and Autonomous Vehicles, EV Charging Infrastructure Management, Intelligent Networking of Drones and also use cases of AI/ML.

Prof. Weiwei Jiang (Senior Member IEEE) received the B.Sc. Degree of Electronic Engineering and Ph.D. Degree of Information and Communication Engineering from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an assistant professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, and Key Laboratory of Universal Wireless Communications, Ministry of Education. His current research interests include artificial intelligence, machine learning, big data, wireless communication and edge computing.