Applied Soft Computing Journal 181 (2025) 113497

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A hyperparameter-space clustering methodology of residential electricity loads

Vasilis Michalakopoulos *, Ioannis Papias, Efstathios Sarantinopoulos, Elissaios Sarmas, Vangelis Marinakis, Dimitris Askounis

Decision Support Systems Laboratory, School of Electrical & Computer Engineering, National Technical University of Athens, Greece

ARTICLE INFO

Keywords:
Hyperparameter clustering
Unsupervised learning
Energy load analysis
Residential electricity consumption
Explainable AI (xAI)

ABSTRACT

Clustering residential electricity consumption patterns is a crucial step toward scalable and interpretable energy forecasting. Traditionally, clustering relies on direct analysis of load time series, which may obscure deeper behavioral or structural similarities between households. This study introduces a novel approach that shifts the focus from using raw consumption data to using the hyperparameters of stacked hour-ahead deep learning forecasting models—specifically based on LSTM, Bi-LSTM, and GRU architectures. By optimizing models independently for each household and clustering based on the resulting hyperparameter configurations, we reveal functional similarities in forecasting behavior that are not always evident in the original data. This method enables the formation of new, interpretable consumer segments, which can support the development of tailored forecasting models per cluster. Utilizing over three years of data from 200 households located in London, we evaluate the proposed hyperparameter-based clustering against traditional time-series clustering, applying standard metrics and explainable AI techniques (SHAP and LIME) to interpret the results. Findings demonstrate a strong alignment between the mean and median consumption patterns of clusters for both approaches, validating the effectiveness of the proposed method. Moreover, the analysis highlights that the feature transformation layers play the most critical role in shaping cluster formation, underscoring its significance in capturing underlying consumption behaviors. Beyond its analytical value, the method also offers a privacy-preserving advantage by requiring only the exchange of model hyperparameters rather than raw consumption data.

1. Introduction

In response to the growing urgency of climate change, international frameworks have set ambitious targets for reducing greenhouse gas (GHG) emissions. The Paris Agreement [1] aims to limit global temperature increases to 1.5 °C by cutting GHG emissions by 43% by 2030, a goal reaffirmed in the 2023 UN Climate Change Conference [2]. At the regional level, the European Union has committed to achieving climate neutrality by 2050 [3]. These commitments have placed energy efficiency at the center of decarbonization strategies across sectors. Among all sectors, the building sector plays a particularly critical role, accounting for approximately 40% of global primary energy consumption [4] and 30% of CO_2 emissions [5], with residential buildings representing nearly 70% of the sector's demand [6]. Urbanization and population growth continue to increase this demand [7], while leaving ample space for optimization [8]. In this context, advanced energy data analytics tools have become essential for identifying consumption patterns, informing policy, and enabling more efficient grid operations.

The widespread deployment of smart meters and Internet of Things (IoT) enabled energy systems has generated large-scale, fine-grained electricity consumption data [9,10]. One of the main challenges in utilizing this data lies in identifying meaningful groupings of consumers that capture underlying behavioral or structural patterns. Such segmentation enables a wide range of downstream applications, including tariff design, demand-side management (DSM), distributed learning strategies, and tailored energy services.

Traditional approaches to customer segmentation often rely on direct clustering of consumption time series [11]. While effective in some cases, these methods are sensitive to temporal misalignment and may struggle to capture deeper similarities that are not evident in raw usage profiles. Moreover, they offer limited interpretability, making them difficult to reconcile with expert knowledge or policy needs [12]. These challenges are particularly salient in settings such as Federated Learning (FL), where privacy constraints prevent central access to raw data [13,14].

* Corresponding author.
E-mail address: vmichalakopoulos@epu.ntua.gr (V. Michalakopoulos).

https://doi.org/10.1016/j.asoc.2025.113497
Received 22 October 2024; Received in revised form 31 May 2025; Accepted 15 June 2025
Available online 28 June 2025

Nomenclature

AI	Artificial Intelligence
Bi-LSTM	Bi-directional Long Short-Term Memory
CHI	Calinski–Harabasz Index
DBI	Davies–Bouldin Index
DL	Deep Learning
DP	Differential Privacy
FL	Federated Learning
FNN	Feed Forward Neural Networks
GMM	Gaussian Mixture Model
GRU	Gated Recurrent Units
HAC	Hierarchical Agglomerative Clustering
HBC	Hyperparameter-Based Clustering
LF	Load Forecasting
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
ML	Machine Learning
PV	Photovoltaic
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SHAP	SHapley Additive exPlanations
SIL	Silhouette Score
SMPC	Secure Multi-Party Computing
t-SNE	t-Distributed Stochastic Neighbor Embedding
xAI	Explainable AI

In this work, we introduce a fundamentally different perspective on the clustering of residential electricity loads. Instead of grouping households based on their direct consumption patterns, we propose to cluster them based on the hyperparameters of ML models trained to predict their energy usage. This approach is motivated by the observation that the optimal hyperparameters of forecasting models implicitly encode structural properties of the time series—such as regularity, seasonality, volatility, and predictability—which are often difficult to extract or compare directly. By treating these hyperparameter configurations as meta-representations of each household's load behavior, we enable a new form of clustering that is robust, interpretable, and privacy-preserving.

To implement this method, we optimize several sequence-based stacked Deep Learning (DL) models—based on Long Short-Term Memory (LSTM), Bidirectional (BiLSTM), and Gated Recurrent Unit (GRU) architectures—for each household using the Optuna hyperparameter tuning framework [15]. The resulting hyperparameter vectors are used as inputs to clustering algorithms, including K-means, Hierarchical Agglomerative Clustering (HAC) and Gaussian Mixture Model (GMM). We evaluate the quality of the clusters using standard metrics such as the Silhouette Score (SIL) [16], the Calinski–Harabasz Index (CHI) [17], and the Davies–Bouldin Index (DBI) [18]. To enhance interpretability and validate the contribution of individual hyperparameters, we further apply explainable AI (xAI) techniques, specifically SHAP (SHapley Additive exPlanations) [19] and LIME (Local Interpretable Model-agnostic Explanations) [20].

Using three years of smart meter data from 200 London households, we demonstrate that clustering in the hyperparameter space yields segments that are distinct, explainable, and aligned with differences in load behavior. Moreover, since hyperparameter vectors are compact and non-sensitive, they can be shared across distributed systems without compromising user privacy—making this approach suitable for real-world deployments at scale.

The key contributions of this paper are as follows:

- We introduce a novel methodology for clustering residential electricity loads based on the hyperparameters of their corresponding hour-ahead Load Forecasting (LF) models, offering a new abstraction layer for understanding load behavior. For this, distinct stacked DL models using three different Recurrent Neural Networks (RNN) variants were employed.
- We compare this method to conventional time-series clustering through three unsupervised clustering algorithms using multiple evaluation metrics and visualize the cluster structures with t-SNE to assess separability.
- xAI tools (SHAP and LIME) are applied to interpret the clustering process, identifying which hyperparameters contribute most to segmentation, thereby connecting model behavior with household characteristics.
- Our methodology results in a privacy-preserving clustering scheme in which only hyperparameter vectors, not raw consumption data, are shared—enabling scalable and secure deployment in distributed learning and smart grid environments.

The remainder of this paper is structured as follows. Section 2 reviews related literature. Section 3 outlines the proposed methodology. Section 4 presents the dataset and experimental setup, with results discussed in Section 5. Finally, Section 6 concludes the paper and suggests directions for future research.

2. Literature review

The emergence of ML, and particularly DL, has revolutionized the field of timeseries analysis, shifting from model-driven to data-driven approaches [21]. Neural networks, and more specifically, RNNs, excel in capturing complex, nonlinear patterns in data without the need for stringent statistical assumptions. RNNs are well-suited for sequential data due to their ability to process inputs over time, making them ideal for timeseries forecasting where temporal dynamics are critical [22–24]. Studies such as [25,26] have further validated the superiority of DL over traditional methods, whether for predicting financial metrics like the S&P 500 index or for energy LF. However, standard RNNs face challenges with long-term dependencies due to the vanishing gradient problem [27], leading to advancements like LSTM networks. LSTMs utilize gate mechanisms to regulate information flow, thereby effectively capturing long-range data dependencies. Liu et al. [28] utilized LSTM models for electricity LF, achieving higher precision than traditional RNNs. Similarly, Bouktif et al. [29] improved electricity LF accuracy by using LSTM models with optimally selected time-lagged features. Additionally, Sarmas et al. [30] developed and optimized four LSTM models to forecast the production of three rooftop PV plants in Portugal, further demonstrating the efficacy of LSTMs for energy forecasting.

To enhance the predictive capabilities of LSTMs further, Bi-LSTM models are used in [31,32] for photovoltaic power and solar radiation forecasting respectively in comparison with other models verifying their effectiveness. Alongside LSTM and Bi-LSTM, the GRU is another significant advancement in recurrent neural network architectures. Wu et al. [33] use a GRU network to forecast short-term energy load achieving better performance than traditional methods. The authors in [34] also report that their GRU model achieved superior results to the LSTM in a residential LF application. In this study [35], Ribeiro et al. evaluate the efficacy of various models, including RNN, LSTM, Support Vector Regression (SVR), Random Forest and GRU to forecast the energy load of a manufacturing plant in Brazil. Their findings indicate that the GRU model significantly outperforms the rest of the models, highlighting its superior capability in handling high-dimensional data.

Despite the superior capabilities of LSTM, Bi-LSTM and GRU models in handling complex timeseries data, their performance critically depends on hyperparameter configuration [29,36]. Recent research

has evolved beyond traditional RNN architectures, integrating hybrid approaches with sophisticated hyperparameter optimization techniques for renewable energy forecasting. While conventional methods like Bayesian optimization, genetic algorithms, and gradient-based approaches have been widely explored [37–39], recent innovations demonstrate significant performance improvements through customized optimization strategies. Zhang et al. [40] achieved over 70% accuracy improvement on wind and Photovoltaic (PV) predictions by combining decomposition algorithms with Convolutional Neural Networks (CNNs) and employing fuzzy entropy with particle swarm optimization to automatically fine-tune network parameters. Similarly, Neshat et al. [41] developed a hybrid framework merging deep residual learning with gated LSTM networks, using a differential covariance matrix adaptation for hyperparameter optimization. These studies collectively demonstrate that the integration of advanced model architectures with tailored hyperparameter optimization techniques substantially outperforms traditional approaches in renewable energy forecasting. The overall effectiveness of these methods varies by application context [42], with increasing importance in big data environments where managing computational complexity becomes a necessity.

Furthermore, clustering techniques are increasingly utilized in energy LF to categorize households or time intervals with similar consumption characteristics. This methodological shift allows for more accurate predictions by customizing model parameters to specific groups, rather than applying a one-size-fits-all approach across diverse datasets. Various clustering techniques have been used to reveal intricate consumption patterns in large datasets. Methods such as HAC, K-means, Fuzzy C-means (FCM), and Self-Organizing Maps (SOM) are especially recognized for their effectiveness in managing different data types and sizes [43]. McLoughlin et al. [44] assess the effectiveness of different clustering methods, including K-means, K-medoids, and SOM, using smart metering data to characterize domestic electricity load profiles. Their analysis, evaluated based on the DBI—a statistical measure for clustering algorithm evaluation—reveals varying degrees of effectiveness among the methods. Czetany et al. [45] performed an extensive study on electricity consumption in Hungarian households, utilizing K-means, FCM, and HAC to uncover daily and yearly consumption trends. Their results highlight K-means as particularly effective in segmenting the data into coherent groups, thus supporting its preferred use in clustering large-scale energy data. Devijver et al. [46] explore an advanced approach by utilizing high-dimensional regression mixture models for clustering residential electricity consumers. This method not only categorizes consumers into clusters but also fits predictive models specifically tailored to each cluster's characteristics, enhancing both the accuracy and applicability of the forecasts. In our previous studies [47,48], four clustering algorithms — K-means, K-medoids, HAC, and DBSCAN were compared — evaluating their effectiveness for energy LF. Multiple metrics including the DBI, CHI, and SIL were utilized in order to assess cluster quality and the distinctiveness of group separations. xAI was also utilized to delve further into the features that impacted the cluster selection the most. Lopez et al. [49] proposed a hybrid Hopfield–K-means algorithm to avoid random initialization in customer segmentation. Zarabie et al. [50] showed that Affinity Propagation outperforms K-means and K-medoids for residential load profiling. Similarly, Eskandarnia et al. [51] employed deep autoencoder-based clustering methods (e.g., DEC, IDEC, DynAE) on real-world smart meter datasets from the UK and Ireland, finding superior results compared to traditional algorithms. In addition, Wang et al. [52] and Okereke et al. [53] explored optimal cluster number selection and cluster validation metrics such as CHI, DBI, and SIL on energy datasets.

While these approaches involve tuning within the clustering process, they still rely on time-series or feature-based input spaces. A more recent line of work explores a fundamentally different approach, hyperparameter space clustering, in which clustering is performed not on raw data or features, but on the optimized hyperparameters of forecasting models. This technique is especially prominent in privacy-preserving FL settings, where raw consumption data cannot be shared. In [54], the authors applied a Hyperparameter-Based Clustering (HBC) methodology within an FL setup for residential short-term LF. Their comparison between clustered and unclustered datasets showed that the former led to faster convergence times and lower error metrics. Clustering based on hyperparameters is particularly useful for FL as it mitigates the issue of non-IID data across edge devices, without directly accessing the raw data. Similarly, in [55], the authors introduced a genetic clustered federated learning algorithm (Genetic CFL), which clusters edge devices in the FL architecture and genetically adjusts the hyperparameters of each cluster. This approach significantly improves performance while reducing both communication and computation costs. In our previous work [56], we implemented HBC for FL generation forecasting to enhance our privacy-preserving approach and tackle the inherent non-IID nature of the data. However, unlike these studies, our current work is not focused on FL applications. To illustrate the evolution of clustering techniques in the energy forecasting literature—and the growing role of hyperparameter space clustering—Table 1 presents a comparative overview of representative studies. As far as we are aware, this is the first study to incorporate xAI into the HBC process, aiming to improve human–computer interaction in the energy sector.

Adding to the previous, the increasing volume and sensitivity of personal data have underscored the importance of privacy-preserving data handling. For that reason, several methods have been developed to extract valuable knowledge from data while simultaneously safeguarding the privacy of the individuals or entities to whom the data pertain. Methods such as k-anonymity [57] and l-diversity [58] are privacy preservation techniques that aim to protect individual privacy by ensuring that each record in a dataset is indistinguishable from at least k-1 other records based on certain identifying attributes (k-anonymity) and that sensitive attributes within each such group have at least l well-represented values (l-diversity). While these methods can be effective in preventing the re-identification of individuals, they might not offer the same level of formal privacy guarantees other methods and can sometimes lead to a loss of data utility. More advanced methods include Differential Privacy (DP) and Secure Multi-Party Computing (SMPC). DP offers a mathematical framework for quantifying privacy loss [59], ensuring that the outcome of an analysis is not significantly affected by the presence or absence of any single individual's data. This is achieved by adding a controllable amount of noise. The level of privacy is also controlled by the privacy budget ϵ, with smaller values indicating stronger privacy but potentially lower accuracy. While DP provides a formal guarantee against information leakage, the added noise often distorts the data [60]. Secure multi-party computation, on the other hand, employs cryptographic protocols to enable multiple parties to jointly compute a result on their private data without revealing their individual inputs to each other. Techniques like homomorphic encryption allow computations on encrypted data, while secret sharing distributes data across multiple parties such that no single party can access the original data. SMPC offers strong privacy guarantees based on the underlying cryptographic assumptions, and ideally does not impact the results accuracy, since no noise is added. However, SMPC protocols can be computationally intensive, especially for complex algorithms or large datasets, and might require specialized hardware or significant communication overhead [61]. The privacy methods mentioned above, especially DP have been applied in clustering workloads extensively [62–64]. The difference of these methods compared to our implementation is that the privacy aspect we provide comes from sharing the hyperparameters of the data instead of the actual data or the data in an encrypted or transformed form.

Table 1
Comparison of clustering studies with alternate input spaces.

Study	Clustering Method	Tuning/Input Space	Application	Key Findings
[43]	HAC, K-means, SOM	No tuning/Smart meter data	Household Energy LF	Grouped households by patterns, no tuning used.
[44]	K-means, K-medoids, SOM	No tuning/Load curves	Load Profiling	Compared clustering methods using DBI, found performance varies.
[45]	K-means, FCM, HAC	No tuning/Residential profiles	Daily/Yearly LF	K-means captured daily/seasonal trends effectively.
[46]	Regression Mixture Models	Regression tuning/High-dimensional load features	Residential LF	Tailored regression models improved forecast accuracy.
[47]	K-means, K-medoids, HAC, DBSCAN	No tuning/Time-series input + xAI analysis	Residential Load Forecasting/DR	Combined clustering with xAI to interpret segments, no tuning applied.
[49]	Hopfield NN + K-means	Custom init/Static features	Customer Segmentation	Hopfield NN avoids random K-means init, improves clustering stability.
[50]	Affinity Propagation	No tuning/Load profile data	Residential Load Profiling	Outperformed K-means and K-medoids in consistency and clustering quality.
[51]	DEC, IDEC, DynAE	DL-based/Autoencoder features	Residential LF	Deep models better captured behavioral segmentation than traditional methods.
[52]	Adaptive DBSCAN + K-means	Optimal cluster selection using SIL, CHI, DBI/Retailer behavior	Electricity Market Segmentation	Combined DBSCAN to remove outliers and K-means to cluster valid points, optimal cluster number selected using multiple indices.
[53]	K-means	Elbow + Silhouette + DBI/Time-domain smart meter features	Residential Load Profiling	Used time-domain features from smart meters, evaluated clusters using Elbow method, SIL and DBI to select optimal k.
[54]	Custom (HBC)	Hyperparameter-space input/Tuned in FL setup	Residential LF (FL)	HBC improved convergence and accuracy under non-IID FL settings.
[55]	Genetic CFL (HBC)	Evolutionary tuning/Hyperparameters in FL	Federated LF	Genetic HBC enhanced efficiency while reducing communication cost.
[56]	K-means, HAC	HBC/Hyperparameter input	PV Generation Forecasting (FL)	Clustering on model hyperparameters improved privacy and performance under non-IID conditions.

3. Methodology

The diagram depicted in Fig. 1 outlines the systematic workflow of the proposed methodology. Commencing from the top, historical household data are collected from smart devices. Subsequently, the data for each of the n households undergoes segregation, followed by the training of n distinct ML models for each household individually. This stage includes a hyperparameter tuning process performed to each ML model. Once the optimal hyperparameters are determined, they serve as the input for various unsupervised clustering algorithms. The performance of these algorithms is assessed using cluster validity indices. Finally, the xAI part of our methodology is applied to interpret the clustering outcomes, providing insights into the underlying patterns and characteristics that define each cluster. This helps to clarify the reasoning behind the clustering decisions and ensures that the clusters are not only statistically sound but also meaningful and actionable in the selected set of hyperparameters. Both the xAI frameworks and algorithms employed are among the most widely used in the field. Specifically, we utilize LIME [20] and SHAP [19], which are renowned for their effectiveness in providing interpretability and insights into model predictions. At the same time, raw timeseries data are subjected to analogous procedures, leading to contrasting results.

3.1. Model selection

As detailed in Section 2, literature suggests that the prevailing practices in LF for residential settings predominantly revolve around RNNs. This is due to their proven ability to capture temporal dependencies in time-series data, which is essential for accurate forecasting. Additionally, RNNs are relatively lightweight in terms of computational requirements, making them well-suited for this context. While transformer-based architectures have shown promise in time-series forecasting, we neglect them in this methodology due to their high computational complexity and resource requirements.

Hence, we have created a stacked model architecture grounded in prominent RNN variants, including LSTMs, Bi-LSTMs, and GRUs, augmented by a Dropout Layer and two Feed Forward Neural Networks (FNNs). The Dropout Layer is incorporated as a regularization technique to mitigate overfitting, improving generalization by randomly deactivating a fraction of neurons during training. This is particularly important given the variability and noise inherent in residential settings. Also, the inclusion of FNNs enhances the model's predictive capability by transforming the extracted temporal features into more abstract representations, enabling the network to capture complex nonlinear relationships between input features and future load values. The specific configurations of the RNN models are described below in detail.

3.1.1. Long-short term memory (LSTM)

The LSTM model, developed by Sepp Hochreiter and Jürgen Schmidhuber in 1997, is an advanced form of an RNN, designed to capture long-term dependencies [65]. An LSTM cell consists of three main components: the input gate, the forget gate, and the output gate. These gates manage the flow of information, allowing the cell to selectively store and discard data over time. The input gate determines the amount of new information to incorporate into the cell state, combining the

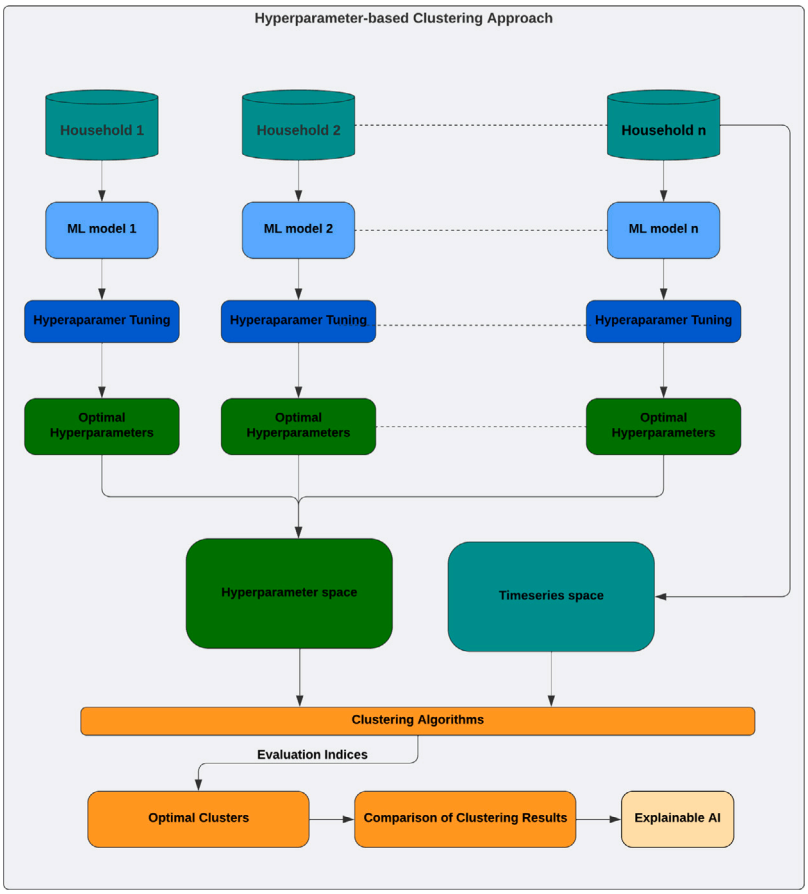


Fig. 1. Proposed methodology's workflow.

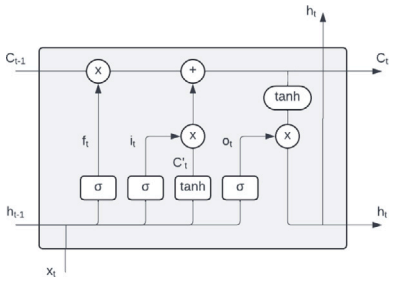


Fig. 2. Structure of the LSTM memory cell.

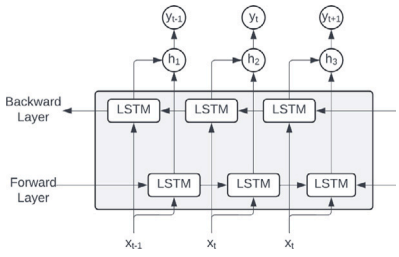


Fig. 3. Structure of the Bi-LSTM memory cell.

current input with the previous cell state and applying a sigmoid function to filter values. It then uses a hyperbolic tangent function to create a candidate vector that updates the cell state. The forget gate uses a sigmoid function to decide which parts of the existing cell state to keep or remove. Finally, the output gate uses another sigmoid function to select parts of the cell state to output, which is then modulated by the hyperbolic tangent of the updated cell state to produce the final output. Fig. 2 illustrates a basic LSTM cell structure.

3.1.2. Bi-directional long short-term memory (Bi-LSTM)

Bi-LSTM networks extend the traditional LSTM by adding a second layer, where the hidden-to-hidden linkages allow the information to flow in reverse temporal order [66]. This architecture displayed in Fig. 3 enables the network to process data in both forward and backward directions, capturing both past and future contexts at any point in time.

3.1.3. Gated recurrent unit (GRU)

The GRU, proposed by Cho et al. in 2014 [67], simplifies the LSTM architecture by merging the forget and input gates into a single "update gate". Additionally, it combines the cell state and hidden state, thereby reducing the model's complexity while maintaining similar performance levels to LSTM. Fig. 4 illustrates a basic GRU cell structure.

3.2. Hyperparameter tuning

After carefully selecting the stacked model components and architecture, such as the RNN, the FNN and Dropout Layers, the next step focuses on selecting the hyperparameters for the optimization process. The goal of this process is the minimization of the Mean Squared Error (MSE) of our predictions, in order to ensure a comprehensive fine-tuning of the model. To accomplish this, we employ the Optuna framework [15]. The framework enables the efficient exploration of various hyperparameter configurations to identify the one that best minimizes the objective function, whilst executing fifty trials. Optuna

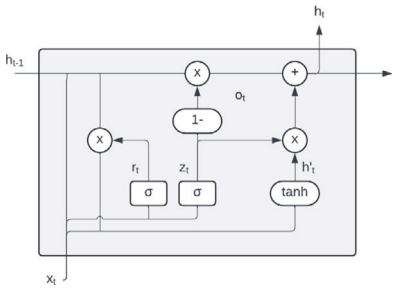


Fig. 4. Structure of the GRU memory cell.

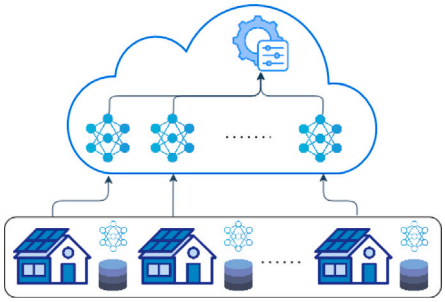


Fig. 5. Hyperparameter clustering schema.

employs a Tree-structured Parzen Estimator (TPE), a Bayesian optimization method, which is widely used in recent parameter tuning frameworks to iteratively refine the search space by modeling promising hyperparameter distributions [68]. Each trial consists of sampling a hyperparameter set, training the model, and evaluating its performance using the MSE loss function. To accelerate convergence, Optuna implements pruning, which terminates under performing trials early. After executing fifty trials, the configuration yielding the lowest MSE is selected as the optimal hyperparameter set. The representations of this process is displayed in Fig. 5 . As shown, the hyperparameter tuning process takes place locally for each house, and the hyperparameters are then collected from a central server where the clustering process will take place.

Furthermore, Table 2 presents a detailed account of the chosen hyperparameters and their descriptions. An additional column is included, indicating the range of possible values for each hyperparameter. This addition, helps with clarifying the extent of exploration during optimization. These ranges define the boundaries within which the framework searches for optimal hyperparameter values, ensuring that it explores feasible and meaningful parameter spaces for effective solutions. Subsequently, our approach explicitly identifies the key hyperparameters under investigation and outlines the expected ranges for each parameter during the optimization process. This clarity aids in replicating the optimization process and enhances the overall understanding of model fine-tuning. Furthermore, additional hyperparameters, such as the choice of activation function, validation method, number of layers, and dataset split ratios for training, testing, and validation, were investigated. However, these were ultimately not utilized due to their complexity and the desire for a simpler, more interpretable and compact solution.

3.3. Clustering algorithms & evaluation metrics

The clustering algorithms and the evaluation metrics that are utilized in this study will be reported here. Notably, the distance metric used for all the clustering algorithms is the Euclidean Distance (ED), as the hyperparameter input space does not leave room for time-series related distance metrics. The clustering algorithms utilized in this study are three of the most popular and effective clustering algorithms according to the literature, namely, the K-means [69], the HAC and the GMM [70]. Those three algorithms were chosen due to their different nature with the first being a centroid based algorithm, the second being a hierarchical based one and the third being a probabilistic model-based algorithm that assumes data is generated from a mixture of Gaussian distributions.

Utilizing evaluation metrics is crucial in determining the ideal number of clusters for the dataset. Commonly employed metrics in scientific research, including SIL [16], DBI [18], and CHI [17], are applied to assess the clustering results' quality. These metrics aid in identifying the optimal number of clusters that accurately depict the underlying data patterns. Leveraging such evaluation metrics empowers researchers and practitioners to make well-informed decisions regarding clustering outcomes, enabling them to refine their analyses and improve the overall quality of load profiling processes. It is worth noting that, the selection of the optimal number of clusters is based on choosing the clustering configuration that produces the highest SIL and CHI values, while aiming for a lower corresponding value of DBI.

- **Silhouette Score (SIL):** The silhouette score for a single sample is given by:

$$SI(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{1}$$

  Where $a(i)$ represents the mean distance between the ith object and all other objects within the same cluster, while $b(i)$ denotes the smallest mean distance from the ith object to any objects in a different cluster, indicating the closest cluster to the ith object.
- **Calinski–Harabasz Index (CHI):** The score can be defined as :

$$CH = \frac{BSS/(k-1)}{WSS/(N-k)} \tag{2}$$

  Here, BSS refers to the between-cluster sum of squares, quantifying the variance between different clusters. WSS denotes the within-cluster sum of squares, measuring the variance within each cluster. N is the total count of samples, and k represents the number of clusters.
- **Davies–Bouldin Index (DBI):** The score is defined as:

$$DB = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \tag{3}$$

  In this formulation, n denotes the total number of clusters. The term σ_i represents the average distance from all points within cluster i to its centroid c_i. The function $d(c_i, c_j)$ measures the distance between the centroids of cluster i and a different cluster j.

  Furthermore, we employ t-SNE [71] as a key tool for visualizing and interpreting our clustering results. t-SNE is a dimensionality reduction technique that projects high-dimensional data into a lower-dimensional space while preserving pairwise similarities, thus maintaining the relative distances between data points. This transformation simplifies the data representation, making it easier to discern distinct clusters and patterns. By converting complex multi-dimensional data into a two-dimensional space, t-SNE allows for clear visualization of clusters, enhancing our understanding of clustering outcomes.

4. Case study

4.1. Dataset

The research utilized energy consumption data from 5567 London households collected between November 2011 and February 2014.[1]

[1] https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

Table 2
Hyperparameters selected for model optimization and their corresponding value ranges.

Hyperparameter	Description	Range
n_{past}	Number of time steps considered in the input sequence	[12, 48]
RNN_units	Number of units in the recurrent layer	[32, 128]
dropout_rate	Dropout rate applied to the recurrent layer	[0.1, 0.5]
dense_units	Number of units in the first dense layer	[32, 64]
dense_2_units	Number of units in the second dense layer	[32, 64]
learning_rate	Learning rate used during model training	[0.00001, 0.1]
batch_size	Number of samples processed per training iteration	[16, 128]
epochs	Total number of training epochs	[20, 60]

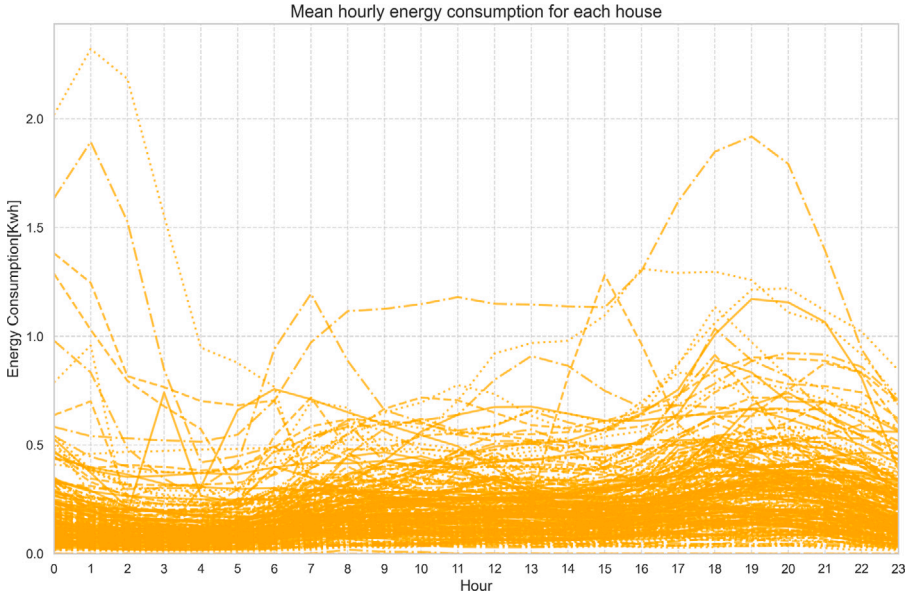


Fig. 6. Mean hourly energy profiles for each house.

This dataset includes energy usage measurements recorded every half-hour, along with timestamps and household identifiers. During 2013, approximately 1100 of these households were subjected to dynamic Time of Use (ToU) pricing, receiving advance notifications of price changes through smart meter in-home displays or mobile messages. The objective was to create scenarios for managing high renewable energy production and using elevated prices to reduce pressure on local distribution networks during peak times. On the other hand, about 4500 households not participating in the ToU programs provided more unfiltered data, making them a more representative sample for analyzing real-world conditions.

In order to manage computational resources effectively and ensure timely completion of our experiments, we have implemented an approach by concentrating on a subset of the larger dataset. Specifically, we have chosen to examine the initial 200 households from the pool of 4500 households not enrolled in Time of Use (ToU) tariff schemes. This random selection ensures that our sample is representative and free from selection bias. The mean hourly energy profiles for each house throughout a day can be seen in Fig. 6.

Prior to the analysis of the data, we conducted a data cleaning process to enhance their quality. This process can be separated into two steps, prior to forecasting and before the clustering procedure. The first one included the elimination of entries with missing values from the timeseries and application of the Min–max normalization technique, which adjusted all values to fall within the range of 0.0 to 1.0. Following that, after the hyperparameters were obtained the next step was to perform the same normalization technique to the new input-space, for the clustering algorithms. This normalization process is crucial for reducing the influence of differing data scales on the clustering analysis, thereby minimizing potential biases.

4.2. Results

After cleaning the dataset, hyperparameter tuning was performed for all the selected ML models. These experiments were quite time-intensive. However, in a real-world scenario, they could be distributed across different devices in each household. Each tuning cycle, involving the necessary training ones, took approximately ten minutes to complete. The average hyperparameter features outlined by the proposed framework for all 200 houses can be seen in Fig. 7. More specifically, the features outlined in Table 2 are divided into two Figures to enhance visualization and comprehension. The first Figure depicts the average values for the dropout rate and the learning rate, with the latter being multiplied with 100 as a different scale due to the differing ranges of the numbers. While, the second Figure presents the remaining features. The Bi-LSTM architecture notably exhibits greater complexity in terms of RNN units and past steps compared to other models. However, it excels in efficiency across all other aspects, resulting in superior running times and accuracy.

Following that, the careful extraction of the hyperparameter features were brought as input into the clustering algorithms and the results, for cluster numbers spanning from three to twenty were obtained. In Fig. 8 the SIL scores for each of the clustering algorithms and for all the selected approaches are visualized. As seen in the Figures, the SIL scores for plain time-series clustering surpass those for HBC in all three algorithms. Additionally, the optimal number of clusters is consistently four across all approaches. Notably, the Bi-LSTM method achieves the highest performance for this specific cluster number, in the hyperparameter domain.

The results shown in the plot can be confirmed in Table 3, which represents the most effective number of clusters for each algorithm and

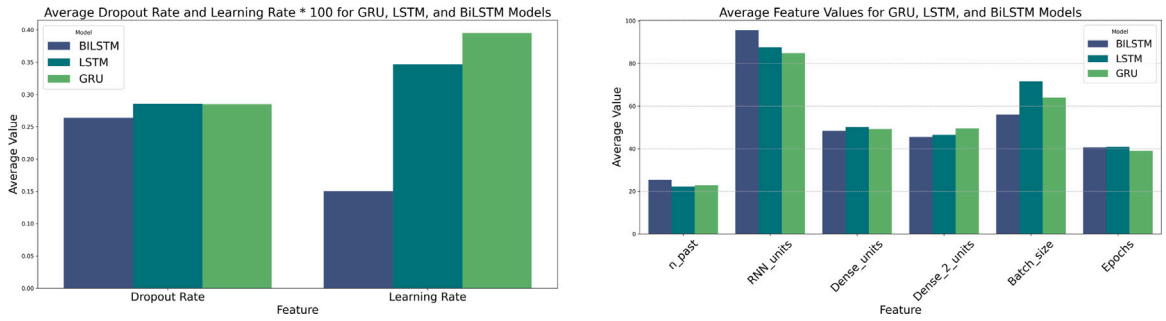
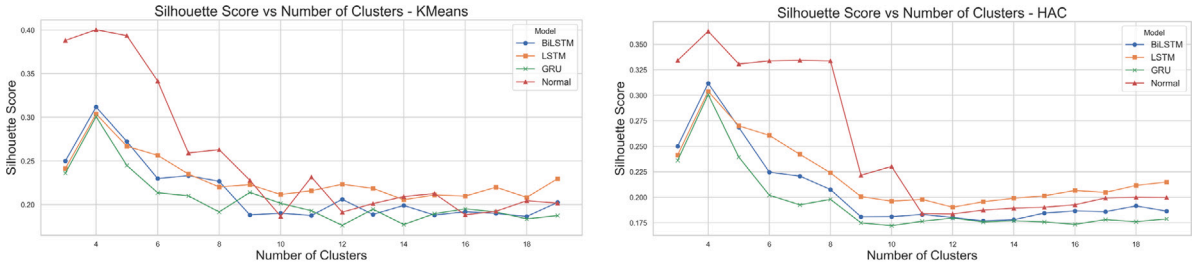
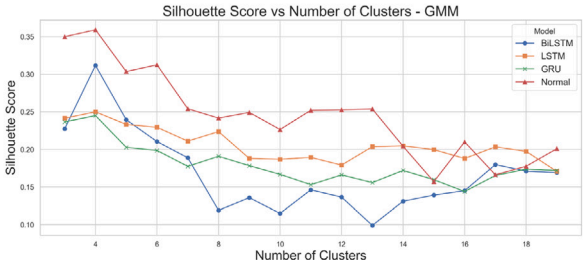


Fig. 7. Average hyperparameter values for all 200 houses and different model architectures.



(a) K-means (b) HAC



(c) GMM

Fig. 8. SIL scores of every approach utilized in this study for K-means, HAC, and GMM.

approach determined based on the selected evaluation metrics including SIL that was illustrated in Fig. 8. As observed, the optimal number of clusters consistently remains at four across all evaluation metrics. However, DBI shows inconsistencies in the results, often identifying nineteen as the optimal number of clusters. This suggests its limited reliability for this analysis, as the value of nineteen arises from single households or pairs forming clusters as outliers. Another significant observation not presented in the table is that once more, the Bi-LSTM approach outperformed other hyperparameter-based approaches in terms of DBI and CHI scores for a cluster number of four. Therefore, it will be considered as the ideal hyperparameter-based approach for the remainder of this study. Furthermore, all clustering algorithms produced comparable and effective results. However, when considering evaluation metric scores, which highlight cluster similarity and dissimilarity, K-means outperformed both GMM and HAC. Consequently, the labels obtained from the K-means algorithm will be utilized in the subsequent analysis.

Now that the best hyperparameter-based approach has been identified as the Bi-LSTM one, we can compare the clusters created by this approach and the plain-timeseries one. The clustering results are also visualized in Fig. 9 using a two-dimensional coordinate system with named t-SNE. Specifically, t-SNE is a ML-bases algorithm designed to project high-dimensional data into a lower-dimensional space while maintaining the relative distances between points. This allows the viewer to witness dependencies between clusters as t-SNE, can be considered as a merging of the original feature dimensions in the dataset for visualizing the data distribution in two-dimensional space. In Fig. 9, it is evident that the differentiation of clusters in the hyperparameter space is clear and well-defined. Conversely, in the timeseries space, while the clusters are more intertwined, there is still a notable level of separation. This indicates that both visualizations effectively demonstrate distinct clustering, each with its own strengths.

Moving on, in Fig. 10 the electrical demand profiles incorporating the median and mean values for all clusters for the plain timeseries analysis and in Fig. 11 the corresponding ones for the Bi-LSTM analysis are presented. This comparative analysis is conducted across four distinct clusters, as discussed before, revealing several insightful observations about the clustering results derived from the two different methodologies. Firstly, a significant observation from this analysis is that the mean and median values for corresponding clusters identified by the plain timeseries analysis and the Bi-LSTM analysis are strikingly similar. This similarity indicates a level of consistency between the two approaches in identifying groups with comparable central tendencies and distributions in the electrical demand profiles. However, it is crucial to note that the actual timeseries values within each cluster are not identical between the two approaches, revealing important differences in the clustering results. Although, a notable exception is observed in Cluster 1 between the two approaches. Here, the mean and median values diverge more noticeably than in the other clusters. This discrepancy suggests that Cluster 1 in Fig. 10 contains a higher degree

Table 3
Optimal number of clusters for each algorithm and approach based on the selected evaluation metrics.

Model	Algorithm	CHI	DBI	SIL
Timeseries	HAC	4	4	4
	K-means	4	18	4
	GMM	4	18	4
LSTM	HAC	4	19	4
	K-means	4	19	4
	GMM	4	4	4
Bi-LSTM	HAC	4	4	4
	K-means	4	18	4
	GMM	4	19	4
GRU	HAC	4	4	4
	K-means	4	18	4
	GMM	4	18	4

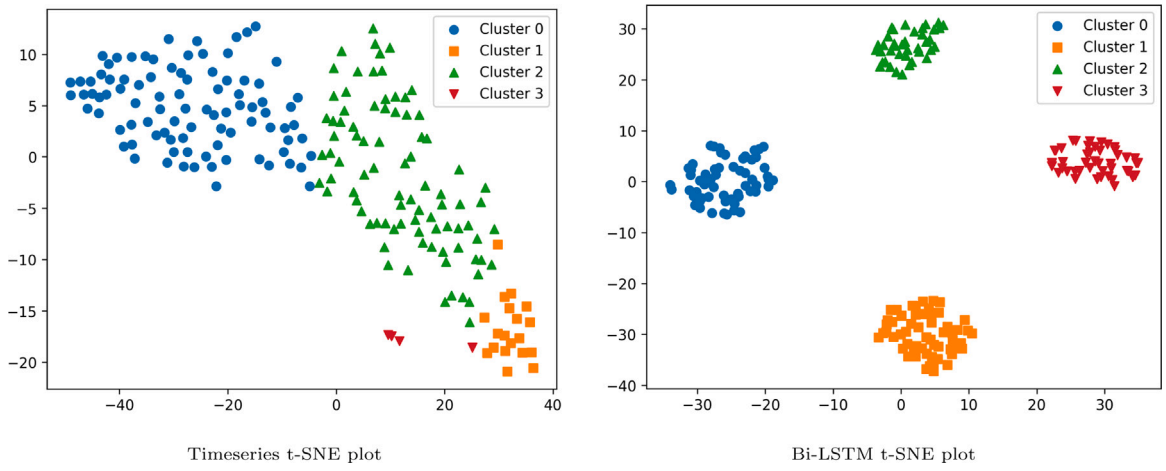


Fig. 9. t-SNE plots for hyperparameter-based and timeseries-based clustering.

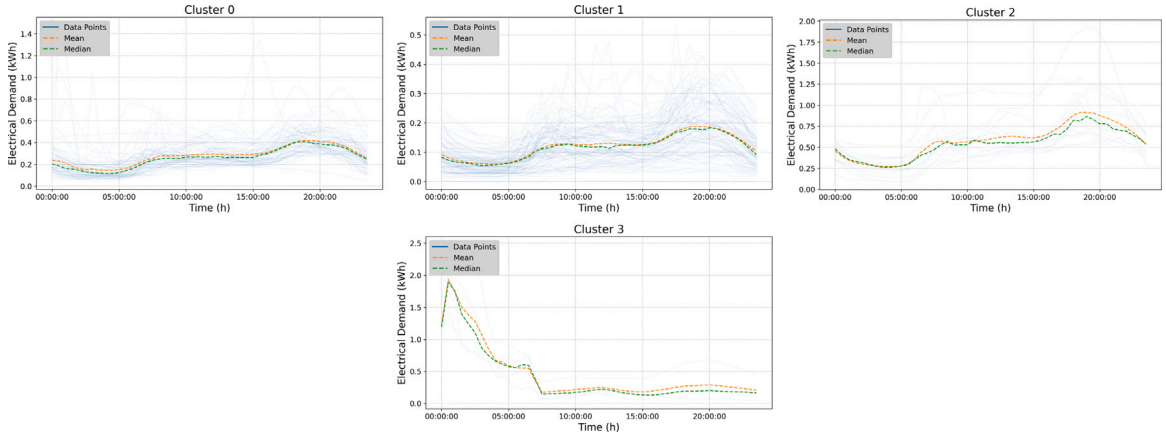


Fig. 10. Electrical demand profiles incorporating the median and mean values for every cluster for the plain timeseries analysis.

of variability or skewness in the data compared to its corresponding one in Fig. 11. In practical terms, this could imply that the electrical demand patterns within Cluster 1 are more diverse, possibly due to a mix of different consumer behaviors or operational characteristics that are not as uniformly distributed as in the other clusters.

When comparing the actual timeseries values within each cluster, further differences emerge between the timeseries analysis and the Bi-LSTM one. Although the mean and median profiles may appear similar at a glance, the individual timeseries data points that comprise these clusters are not identical. This distinction underscores the impact of the clustering algorithm and the analysis method on the formation of clusters. The plain timeseries analysis relies on traditional statistical measures to group similar patterns, which may not fully capture the temporal dependencies and intricate patterns present in the electrical demand data. In contrast, the Bi-LSTM analysis, leveraging the power of DL and sequential modeling, is likely to identify and group timeseries with more complex, non-linear relationships. The Bi-LSTM model is capable of capturing long-term dependencies and subtle variations in timeseries data.

Transitioning to the interoperable (xAI) part of our methodology, Fig. 12 illustrates the average impact of all features across each cluster, while Fig. 13 presents the same analysis conducted for a randomly

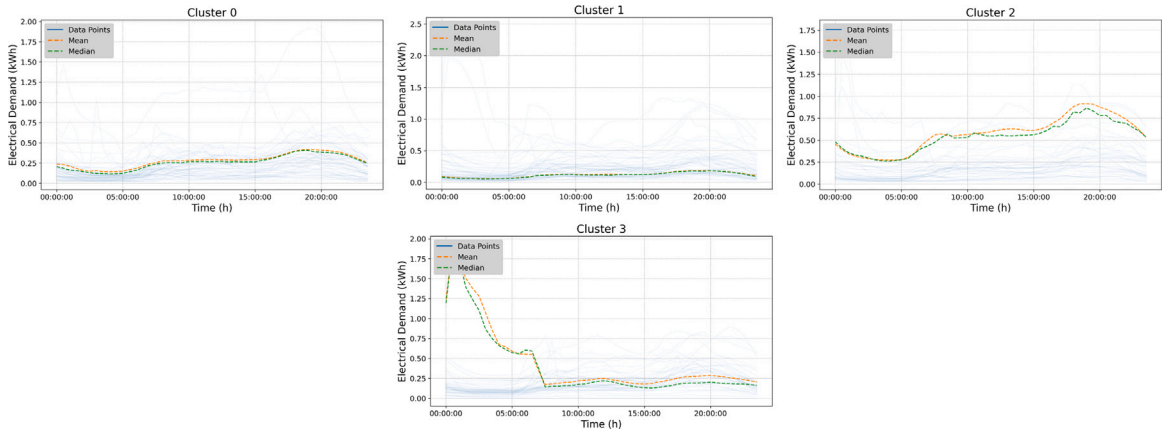


Fig. 11. Electrical demand profiles including the median and mean values for every cluster for the Bi-LSTM analysis.

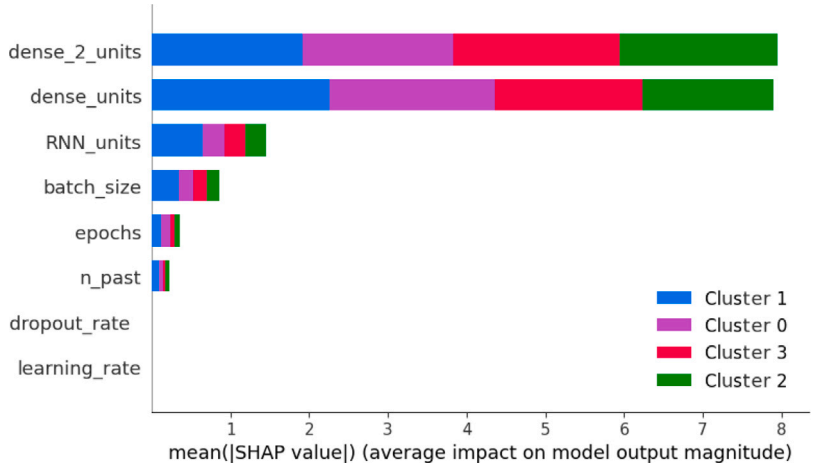


Fig. 12. Average impact of every feature for all classes.

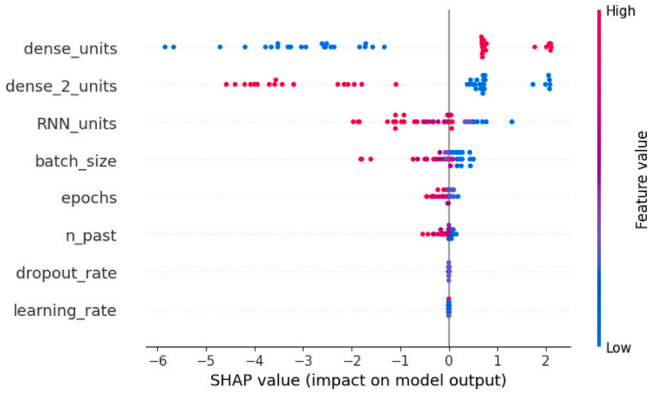


Fig. 13. Average impact of every feature for a random household.

selected household. In simpler words, the study examines how the various selected hyperparameters influence the clustering process. Interestingly, SHAP analysis indicate that the first and second feature extraction and representation layers are the most important factors in defining clusters. This suggests that the way the model transforms and represents features after the BiLSTM layers plays a critical role in distinguishing different hyperparameter configurations. This makes sense because the dense (FNN) units serve as the final layers that integrate and synthesize information from the preceding BiLSTM layers. They are responsible for feature extraction and representation, capturing complex patterns and relationships in the data, which directly influence the final clustering results. Therefore, the dense units play a crucial role in determining how data points are grouped, making them particularly important in the context of clustering analysis. The BiLSTM hidden units also contribute significantly but to a lesser extent, implying that while learning temporal dependencies is crucial, the transformation of these learned features is even more influential in differentiating model behaviors. Meanwhile, batch size, epochs, and number of past steps have a moderate impact on clustering, meaning that while they affect how the model learns, they do not drastically alter how different hyperparameter settings form distinct groups. Lastly, dropout rate and learning rate have minimal impact, indicating that regularization and optimization settings do not strongly dictate how models cluster based on their hyperparameters.

The insights derived from the SHAP analysis are further substantiated by the comprehensive LIME explanations presented in Fig. 14. These figures highlight the range of values for each feature and their influence on the selection across the four distinct clusters. Consistent with the SHAP findings, the number of dense (FNN) normalization layers emerges as the most influential factor in cluster selection. Specifically, for clusters 0 and 1, the first normalization layer holds the most significance, whereas for clusters 2 and 3, the second layer takes precedence. Beyond the normalization layers, slight variations can be observed in the importance of other features, such as the number of epochs and RNN units, both of which play prominent roles. Additionally, the dropout rate exhibited variable significance, being crucial in some cases and negligible in others. Similar patterns were noted for the batch size, learning rate, and the loop back window which, like

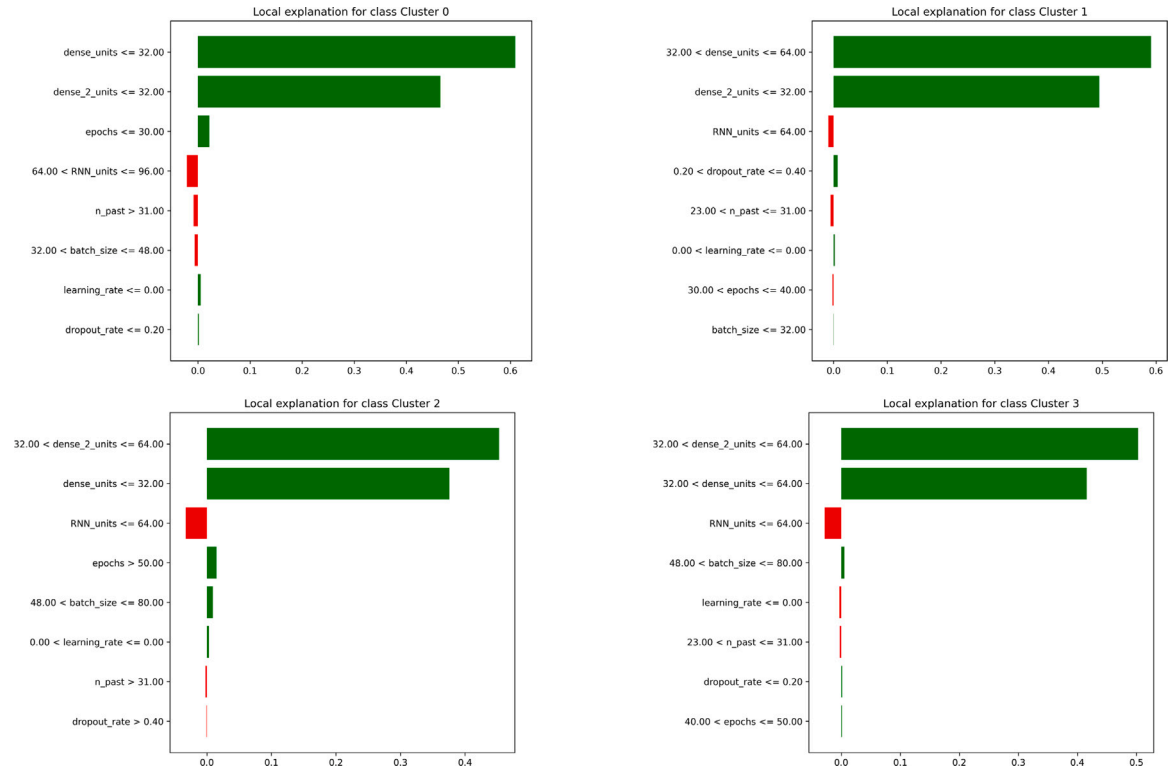


Fig. 14. LIME explanations for clusters 0, 1, 2, and 3.

the dropout rate, displayed varying levels of impact depending on the specific cluster. This analysis highlights that the architectural choices in feature extraction and representation layers are the primary drivers of differentiation, guiding future optimizations toward tuning these layers for better clustering and performance.

5. Discussion

Our hyperparameter-space based clustering approach presented in this study shows promising results as an alternative residential electricity consumption pattern analysis. However, several key considerations must be addressed when considering its extension to larger-scale deployments. When thousands of users are involved, the computational burden of per-household hyperparameter optimization and presents a significant challenge. Distributing the computational burden across devices within the households themselves represents an elegant solution to scaling concerns. Smart meters or home energy management systems capabilities can be leveraged, each household could independently perform its own hyperparameter optimization locally. This approach not only addresses the computational bottleneck but also enhances the privacy-preserving aspects of our method, as raw consumption data never leaves the premises. The central system would only need to collect the optimized hyperparameter configurations for clustering purposes, dramatically reducing both computational burden and data transmission requirements. Furthermore, once clusters are established and global models developed for each cluster, these optimized models can be deployed back to individual households, creating an efficient feedback loop that continuously improves while minimizing computational overhead. Implementation would benefit from a phased approach, beginning with pilot programs targeting specific geographic regions before scaling to wider deployment. Energy providers could initially implement this methodology alongside existing forecasting systems, gradually transitioning as benefits are demonstrated. While initial setup costs may be substantial, the long-term operational savings present a compelling business case for adoption in increasingly complex and distributed energy markets.

Moving on, our xAI analysis indicates that focusing optimization efforts on a subset of critical hyperparameters could significantly reduce computational requirements. It clearly demonstrated that hyperparameters related to feature transformation layers dominated cluster formation, with other hyperparameters showing minimal significance in determining cluster boundaries. Thus, in real-life applications reducing the number of hyperparameters can improve computational efficiency without compromising clustering results. Additionally, regarding cluster stability and transferability, while our current study focused on residential data from a single city, the robust methodological framework suggests promising potential for wider application. Although we have not explicitly tested the approach on expanded datasets from different geographic regions, the fundamental principles underlying our HBC approach are not inherently location-dependent. The method captures functional similarities in forecasting behavior that likely transcend specific geographic contexts, suggesting it could be effectively applied to different cities or countries with minimal adaptation.

6. Conclusions & future work

Concluding, this paper presents a new clustering method based on the hyperparameter-space retrieved from hour-ahead stacked ML models based LSTM, Bi-LSTM and GRU architectures, tailored for analyzing residential household electrical demand timeseries. The performance of K-means, HAC and GMM clustering algorithms is compared using a subset of a large public dataset containing data from 200 houses located in London. Our evaluation of the clustering outcomes, using established evaluation metrics, namely SIL, CHI and DBI, and t-SNE illustrates the reliability and precision of our approach. Utilizing the model's hyperparameters as the input space to the clustering algorithms our methodology gives another dimension to the problem which has been analyzed in comparison to the corresponding analysis using only timeseries data. That way, an important advantage of our method is showcased as its ability to conduct clustering without direct access to raw data, ensuring data privacy protection. By conducting hyperparameter tuning at individual households and transmitting only the results

to a central server, sensitive information is safeguarded while enabling accurate clustering.

Moreover, the integration of xAI in our method adds further transparency to the results. That way, this study is able to interpret the influence of different hyperparameters on the clustering outcomes, providing deeper insights into how specific model configurations contribute to the overall performance. This interpretability is crucial for validating the clustering decisions and ensuring that stakeholders can trust the process, particularly when making data-driven decisions for energy management and optimization. Further enhancing the human computer interaction of the proposed solution, the explainability aspect also allows for fine-tuning the hyperparameter selection process in a more informed manner, ultimately leading to better and more interpretable models.

Looking ahead, our research is committed to addressing current limitations and expanding the capabilities of our approach. A key focus will be on improving computational efficiency and developing real-world implementations using devices at individual households that can support the training of such models. To this end, given the privacy-preserving nature of the proposed approach, it aligns well with online learning methods and FL techniques, which enable decentralized model training without sharing raw data. Online learning could enhance adaptability by continuously updating the hyperparameters with new ones in real time reducing computational costs. As part of this advancement, we also aim to strengthen the role of xAI within our framework, ensuring that both the predictive models and clustering outcomes remain transparent and interpretable in real-world scenarios, incorporating more interpretable results of a more wide selection of hyperparameters and models. Additionally, future work will explore how energy providers and policymakers could adopt this approach in practice, considering the challenge of integrating with existing energy management systems. We also aim to test and validate our methodology in different cities and countries with larger and more diverse consumer bases, to assess its robustness across various socio-economic and energy consumption contexts. Addressing these challenges while incorporating xAI for clearer insights will refine our hyperparameter clustering method and broaden its applications leading to more effective, secure and comprehensive energy management strategies.

CRediT authorship contribution statement

Vasilis Michalakopoulos: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Ioannis Papias:** Writing – review & editing, Writing – original draft. **Efstathios Sarantinopoulos:** Writing – review & editing, Writing – original draft, Conceptualization. **Elissaios Sarmas:** Writing – review & editing, Writing – original draft, Project administration. **Vangelis Marinakis:** Writing – review & editing, Writing – original draft, Supervision. **Dimitris Askounis:** Writing – review & editing, Writing – original draft, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work presented is based on research conducted within the framework of the Horizon Europe European Commission projects DEDALUS (Grant Agreement No. 101103998) and ENPOWER (Grant Agreement No. 101096354). The content of the paper is the sole responsibility of its authors and does not necessary reflect the views of the EC.

Data availability

The link to the dataset is within the manuscript.

References

[1] The Paris Agreement - Publication | UNFCCC. https://unfccc.int/documents/184656,
[2] COP28 Agreement Signals “Beginning of the End” of the Fossil Fuel Era | UNFCCC. https://unfccc.int/news/cop28-agreement-signals-beginning-of-the-end-of-the-fossil-fuel-era,
[3] Directorate-General for Climate Action (European Commission), Going Climate-Neutral by 2050: A Strategic Long Term Vision for a Prosperous, Modern, Competitive and Climate Neutral EU Economy, Publications Office of the European Union, 2019.
[4] V. Michalakopoulos, S. Pelekis, G. Kormpakis, V. Karakolis, S. Mouzakitis, D. Askounis, Data-driven building energy efficiency prediction using physics-informed neural networks, in: 2024 IEEE Conference on Technologies for Sustainability, SusTech, IEEE, 2024, pp. 84–91.
[5] L. Yang, H. Yan, J.C. Lam, Thermal comfort and building energy consumption implications – A review, Appl. Energy 115 (2014) 164–173.
[6] C. Milan, C. Bojesen, M.P. Nielsen, A cost optimization model for 100% renewable residential energy supply systems, Energy 48 (1) (2012) 118–127.
[7] K. Dong, G. Hochman, Y. Zhang, R. Sun, H. Li, H. Liao, CO2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions, Energy Econ. 75 (2018) 180–192.
[8] O. Yelisieieva, Y. Lyzhnyk, I. Stolietova, N. Kutova, Study of best practices of green energy development in the EU Countries Based on correlation and bagatofactor autoregressive forecasting, Economics 11 (2) (2023) 183–197.
[9] D. Kolokotsa, The role of smart grids in the building sector, Energy Build. 116 (2016) 703–708.
[10] V. Marinakis, H. Doukas, J. Tsapelas, S. Mouzakitis, Á. Sicilia, L. Madrazo, S. Sgouridis, From big data to smart energy services: An application for intelligent energy management, Future Gener. Comput. Syst. 110 (2020) 572–586.
[11] L. Toderean, T. Cioara, I. Anghel, E. Sarmas, V. Michalakopoulos, V. Marinakis, Demand response optimization for smart grid integrated buildings: Review of technology enablers landscape and innovation challenges, Energy Build. (2024) 115067.
[12] S. Bobek, M. Kuk, M. Sze, G.J. Nalepa, Enhancing cluster analysis with explainable AI and multidimensional cluster prototypes, IEEE Access 10 (2022) 101556–101574.
[13] C. Briggs, Z. Fan, P. Andras, Federated learning for short-term residential load forecasting, IEEE Open Access J. Power Energy 9 (2022) 573–583.
[14] M. Savi, F. Olivadese, Short-term energy consumption forecasting at the edge: A Federated Learning Approach, IEEE Access 9 (2021) 95949–95969.
[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2623–2631.
[16] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.
[17] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Comm. Statist. Theory Methods 3 (1) (1974) 1–27.
[18] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. (2) (1979) 224–227.
[19] S. Lundberg, A unified approach to interpreting model predictions, 2017, arXiv preprint arXiv:1705.07874.
[20] M.T. Ribeiro, S. Singh, C. Guestrin, Why should i trust you? explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[21] G. Antonesi, T. Cioara, I. Anghel, I. Papias, V. Michalakopoulos, E. Sarmas, Hybrid transformer model with liquid neural networks and learnable encodings for buildings’ energy forecasting, Energy AI 20 (2025) 100489.
[22] S.-Y. Shih, F.-K. Sun, H.-y. Lee, Temporal pattern attention for multivariate time series forecasting, Mach. Learn. 108 (8–9) (2019) 1421–1441, Cited by: 484; All Open Access, Bronze Open Access, Green Open Access.
[23] H. Apaydin, H. Feizi, M.T. Sattari, M.S. Colak, S. Shamshirband, K.-W. Chau, Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting, Water (Switzerland) 12 (5) (2020) Cited by: 179; All Open Access, Gold Open Access.
[24] I. Papias, V. Michalakopoulos, E. Sarmas, V. Marinakis, Aggregated flexibility dynamic forecasting of residential energy prosumers for DR programs, in: 2024 15th International Conference on Information, Intelligence, Systems & Applications, IISA, IEEE, 2024, pp. 1–9.
[25] C. Krauss, X.A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500, European J. Oper. Res. 259 (2) (2017) 689–702.

[26] M.N. Fekri, H. Patel, K. Grolinger, V. Sharma, Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network, Appl. Energy 282 (2021) Cited by: 163; All Open Access, Green Open Access.

[27] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, 2013.

[28] C. Liu, Z. Jin, J. Gu, C. Qiu, Short-Term Load Forecasting using a Long Short-Term Memory Network, vol. 2018-January, 2017, pp. 1–6, Cited by: 66.

[29] S. Bouktif, A. Fiaz, A. Ouni, M.A. Serhani, Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches, Energies 11 (7) (2018) Cited by: 578; All Open Access, Gold Open Access, Green Open Access.

[30] E. Sarmas, E. Spiliotis, E. Stamatopoulos, V. Marinakis, H. Doukas, Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models, Renew. Energy 216 (2023) 118997.

[31] H. Zhen, D. Niu, K. Wang, Y. Shi, Z. Ji, X. Xu, Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information, Energy 231 (2021) Cited by: 82.

[32] T. Peng, C. Zhang, J. Zhou, M.S. Nazir, An integrated framework of bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting, Energy 221 (2021) Cited by: 180.

[33] W. Wu, W. Liao, J. Miao, G. Du, Using gated recurrent unit network to forecast short-term load considering impact of electricity price, Energy Procedia 158 (2019) 3369–3374, Cited by: 50; All Open Access, Gold Open Access.

[34] S. Kumar, L. Hussain, S. Banarjee, M. Reza, Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster, in: 2018 Fifth International Conference on Emerging Applications of Information Technology, EAIT, 2018, pp. 1–4.

[35] A.M.N.C. Ribeiro, P.R.X. Do Carmo, I.R. Rodrigues, D. Sadok, T. Lynn, P.T. Endo, Short-term firm-level energy-consumption forecasting for energy-intensive manufacturing: A comparison of machine learning and deep learning models, Algorithms 13 (11) (2020) 1–19, Cited by: 21; All Open Access, Gold Open Access, Green Open Access.

[36] H. Abbasimehr, M. Shabani, M. Yousefi, An optimized model using LSTM network for demand forecasting, Comput. Ind. Eng. 143 (2020) Cited by: 255.

[37] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, W. Rhee, Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks, IEEE Access 8 (2020) 52588–52608, Cited by: 126; All Open Access, Gold Open Access, Green Open Access.

[38] H. Alibrahim, S.A. Ludwig, Hyperparameter optimization: Comparing genetic algorithm against grid search and Bayesian optimization, 2021, pp. 1551–1559, http://dx.doi.org/10.1109/CEC45853.2021.9504761, Cited by: 113.

[39] M. MacKay, P. Vicol, J. Lorraine, D. Duvenaud, R. Grosse, Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions, 2019, Cited by: 60.

[40] Y. Zhang, Z. Pan, H. Wang, J. Wang, Z. Zhao, F. Wang, Achieving wind power and photovoltaic power prediction: An intelligent prediction system based on a deep learning approach, Energy 283 (2023) 129005.

[41] M. Neshat, M.M. Nezhad, S. Mirjalili, D.A. Garcia, E. Dahlquist, A.H. Gandomi, Short-term solar radiation forecasting using hybrid deep residual learning and gated LSTM recurrent network with differential covariance matrix adaptation evolution strategy, Energy 278 (2023) 127701.

[42] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing 415 (2020) 295–316.

[43] A. Rajabi, M. Eskandari, M.J. Ghadi, L. Li, J. Zhang, P. Siano, A comparative study of clustering techniques for electrical load pattern segmentation, Renew. Sustain. Energy Rev. 120 (2020) 109628.

[44] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Appl. Energy 141 (2015) 190–199.

[45] L. Czétány, V. Vámos, M. Horváth, Z. Szalay, A. Mota-Babiloni, Z. Deme-Bélafi, T. Csoknyai, Development of electricity consumption profiles of residential buildings based on smart meter data clustering, Energy Build. 252 (2021) 111376.

[46] E. Devijver, Y. Goude, J.-M. Poggi, Clustering electricity consumers using high-dimensional regression mixture models, Appl. Stoch. Models Bus. Ind. 36 (1) (2020) 159–177.

[47] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, H. Doukas, A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs, Appl. Energy 361 (2024) 122943.

[48] L. Toderean, M. Daian, T. Cioara, I. Anghel, V. Michalakopoulos, E. Sarantinopoulos, E. Sarmas, Heuristic based federated learning with adaptive hyperparameter tuning for households energy prediction, Sci. Rep. 15 (1) (2025) 12564.

[49] J.J. López, J.A. Aguado, F. Martín, F. Muñoz, A. Rodríguez, J.E. Ruiz, Hopfield–K-means clustering algorithm: A proposal for the segmentation of electricity customers, Electr. Power Syst. Res. 81 (2) (2011) 716–724.

[50] A.K. Zarabie, S. Lashkarbolooki, S. Das, K. Jhala, A. Pahwa, Load profile based electricity consumer clustering using affinity propagation, in: 2019 IEEE International Conference on Electro Information Technology, EIT, 2019, pp. 474–478.

[51] E. Eskandarnia, H.M. Al-Ammal, R. Ksantini, An embedded deep-clustering-based load profiling framework, Sustain. Cities Soc. 78 (2022) 103618.

[52] X. Wang, C. Zhou, Y. Yang, Y. Yang, T. Ji, J. Wang, J. Chen, Y. Zheng, Electricity market customer segmentation based on DBSCAN and k-means : —A case on yunnan electricity market, in: 2020 Asia Energy and Electrical Engineering Symposium, AEEES, 2020, pp. 869–874.

[53] G.E. Okereke, M.C. Bali, C.N. Okwueze, E.C. Ukekwe, S.C. Echezona, C.I. Ugwu, K-means clustering of electricity consumers using time-domain features from smart meter data, J. Electr. Syst. Inf. Technol. 10 (1) (2023) 1–18.

[54] N. Gholizadeh, P. Musilek, Federated learning with hyperparameter-based clustering for electrical load forecasting, Internet Things 17 (2022) 100470.

[55] S. Agrawal, S. Sarkar, M. Alazab, P.K.R. Maddikunta, T.R. Gadekallu, Q.-V. Pham, Genetic CFL: hyperparameter optimization in clustered federated learning, Comput. Intell. Neurosci. 2021 (1) (2021) 7156420.

[56] V. Michalakopoulos, E. Sarantinopoulos, E. Sarmas, V. Marinakis, Empowering federated learning techniques for privacy-preserving pv forecasting, Energy Rep. 12 (2024) 2244–2256.

[57] L. Sweeney, K-Anonymity: a Model for protecting privacy, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 10 (05) (2002) 557–570.

[58] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, L-diversity: Privacy beyond k-anonymity, ACM Trans. Knowl. Discov. Data 1 (1) (2007) 3–es.

[59] C. Dwork, Differential privacy, in: M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (Eds.), Automata, Languages and Programming, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 1–12.

[60] Y. Li, G. Xu, X. Meng, W. Du, X. Ren, LF3pfl: A practical privacy-preserving federated learning algorithm based on local federalization scheme, Entropy 26 (5) (2024) 353.

[61] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, X. Zheng, Privacy-preserving federated learning framework based on chained secure multiparty computing, IEEE Internet Things J. 8 (8) (2021) 6178–6186.

[62] M. Yang, L. Huang, C. Tang, *K*-means clustering with local distance privacy, Big Data Min. Anal. 6 (4) (2023) 433–442.

[63] A. Chang, B. Ghazi, R. Kumar, P. Manurangsi, Locally private K-means in one round, in: Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021, pp. 1441–1451.

[64] Y. Zhang, N. Liu, S. Wang, A differential privacy protecting K-means clustering algorithm based on contour coefficients, PLoS One 13 (11) (2018) e0206832.

[65] R. Wadawadagi, V. Pagi, Sentiment analysis with deep neural networks: comparative study and performance assessment, Artif. Intell. Rev. 53 (2020).

[66] B. Bohara, R.I. Fernandez, V. Gollapudi, X. Li, Short-term aggregated residential load forecasting using BiLSTM and CNN-BiLSTM, in: 2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT, IEEE, 2022.

[67] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, in: NIPS 2014 Workshop on Deep Learning, December 2014, 2014.

[68] S. Watanabe, Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance, 2023, arXiv preprint arXiv:2304.11127.

[69] D. Arthur, S. Vassilvitskii, K-means++ the advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.

[70] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), vol. 2, 1999, pp. 246–252.

[71] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).