FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 9, September, 2025, pp 327 – 338
DOI: https://doi.org/10.33003/fjs-2025-0909-3746

SHORT-TERM LOAD FORECASTING IN MICROGRIDS: A CLUSTERING-ENHANCED DEEP LEARNING APPROACH

***Ahmad Zakari Salihu and Munir Husein**

Department of Electrical and Computer Engineering, Faculty of Engineering, Baze University, Abuja, Nigeria.

*Corresponding authors' email: ahmad8728@bazeuniversity.edu.ng

ABSTRACT

Accurate short-term load forecasting is vital for efficient microgrid energy management, unit commitment, and renewable energy integration. Traditional and deep learning models often struggle with the complex, time-varying patterns in residential, commercial, and industrial loads. To address this, clustering algorithms are applied to group similar consumption patterns, enhancing forecasting accuracy. This study presents a clustering-enhanced long short-term memory (LSTM) framework that segments hourly load profiles using K-means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Models (GMM), and Hierarchical Clustering before training. Using 8,760-point synthetic load profiles per sector, baseline models and standalone LSTMs were compared against clustering-enhanced LSTMs. Results show that clustering reduces mean squared error (MSE) by up to 48% and mean absolute error (MAE) by up to 28% in residential forecasts (GMM), improves commercial forecasting by 18.7% (MSE) and 14.4% (MAE) with Hierarchical Clustering, and yields modest gains of up to 2.4% (MSE) and -0.026% (MAE) in stable industrial profiles with K-means. The proposed framework offers a scalable, sector-specific solution to improve microgrid forecasting and support renewable integration.

Keywords: STLF, Microgrids, Deep learning, Clustering, LSTM, Load forecasting

INTRODUCTION

The accelerating adoption of microgrids has brought about a paradigm shift in the way electricity is generated, distributed, and consumed. Microgrids which are localized energy systems capable of operating independently or in connection with the main power grid have emerged as vital components in modern energy infrastructure. They promote energy resilience, enable integration of renewable energy sources, and support sustainable power delivery for residential, commercial, and industrial consumers (Mallah et al., 2024; Saxena et al., 2024). However, effective microgrid operation hinges on the ability to forecast short-term electricity demand with high precision. Accurate short-term load forecasting (STLF) is essential for real-time decision-making in power dispatch, demand response, battery storage utilization, and peak load management (Mukhtiar et al., 2024).

The diversity and volatility of electricity consumption patterns within different microgrid types introduce significant forecasting challenges. Residential loads, for instance, exhibit high variability influenced by human activity, seasonality, and weather conditions, often peaking in the morning and evening (Karn & Kakran, 2022). Commercial loads are largely dependent on business hours, with notable drops during weekends and holidays (Glazunova et al., 2024), while industrial load profiles remain relatively stable but can fluctuate due to machinery cycles or production schedules. Traditional forecasting models such as autoregressive integrated moving average (ARIMA), linear regression, and persistence methods often fail to capture the complex, nonlinear, and time-dependent characteristics of these diverse load profiles. These models struggle particularly in dynamic environments that involve renewable energy sources and demand-side uncertainties (Cai et al., 2019; Makris et al., 2024).

Recent advancements in artificial intelligence (AI) and machine learning (ML) have opened new pathways for improving forecasting performance in microgrids. Among these, deep learning models, particularly Long Short-Term Memory (LSTM) networks, have gained prominence due to their strength in capturing temporal dependencies and handling large, nonlinear time-series datasets (Rafi et al., 2021). LSTMs have been widely applied in STLF due to their ability to retain long-term information and model sequential behaviour. They outperform traditional models and shallow networks like feed-forward neural networks (FNNs), especially when exogenous variables such as weather or calendar events are considered (Bashir et al., 2022; Husein & Chung, 2019) .

Several studies have demonstrated the efficacy of LSTM-based forecasting systems. Muzaffar and Afshari (2019) found that LSTM outperformed ARIMA and ARMAX models by up to 20% in forecasting error, while Aurangzeb et al. (2021) employed a CNN-LSTM framework enhanced with clustering techniques to forecast smart grid loads. Han et al. (2020) applied K-means clustering to segment residential load profiles, achieving improvements in forecasting accuracy by tailoring models to homogeneous clusters. However, these approaches often face limitations, such as the use of static clustering assumptions or their inability to adapt to evolving load patterns. Moreover, many studies focus on a single sector typically residential without generalizing the framework across commercial and industrial sectors. In some cases, clustering techniques are used without validating their effectiveness against other methods such as DBSCAN, Gaussian Mixture Models (GMMs), or hierarchical clustering (Han et al., 2020; Indralaksono et al., 2022).

In this research, we propose a hybrid short-term load forecasting framework that combines deep LSTM neural networks with unsupervised clustering techniques to improve prediction accuracy across residential, commercial, and industrial microgrids. While previous studies have explored LSTM or clustering independently for load forecasting, few have systematically combined both approaches (Han et al., 2020). Our results demonstrate that integrating clustering with LSTM yields significant improvements in forecasting accuracy for sectors with high consumption variability. Notably, most studies reviewed focused on single-sector analysis, often using residential data only, which limits the generalisability of their findings. In contrast, our approach is evaluated across three distinct microgrid types, each with unique consumption behaviour, providing a more comprehensive understanding of model performance. Additionally, prior studies have primarily emphasized prediction accuracy without evaluating how segmentation strategies contribute to model adaptability. In this work, we show that clustering enhances LSTM performance by allowing models to specialize in different consumption regimes, offering a scalable and sector-aware forecasting solution for smart microgrid management

In summary, the contributions of this work to the literature are threefold. First, we develop a hybrid forecasting framework that combines clustering algorithms with LSTM neural networks to improve short-term load forecasting across diverse microgrid sectors. This approach allows models to be trained on segmented data groups with similar consumption patterns, enhancing their ability to capture nuanced behaviours without requiring additional domain-specific feature engineering. Second, we demonstrate that LSTM consistently outperform traditional methods such as persistence and feed-forward neural networks (FFNN), with even greater accuracy achieved when clustering is integrated. The hybrid models achieved lower forecasting errors across all sectors, outperforming the benchmark models. Third, by evaluating the framework on synthetic hourly data representing residential, commercial, and industrial microgrids, we show the adaptability and scalability of the approach across different consumption environments. This multi-sector validation demonstrates not only the forecasting capability of the proposed method but also its practical value for microgrid energy management, where accurate demand prediction is critical for operational efficiency, cost optimization, and integration of renewable energy resources.

The remainder of this paper is structured as follows. Section 2 outlines the research methodology, including data collection, preprocessing, and the design of the forecasting models. Section 3 presents the results and discusses the performance of the benchmark models, LSTM model, and hybrid models across different microgrid . Finally, Section 4 summarizes the findings, discusses the limitations of the study, and provides recommendations for future research.

MATERIALS AND METHODS

Data Description

The dataset used in this study comprises of synthetic hourly load profiles generated using EnergyPlus simulations, as curated by Angizeh et al. (2020). These profiles span a full calendar year with 8,760 hourly observations per sector, representing typical consumption patterns in residential, commercial, and industrial microgrids.

The analysis of load profiles across residential, commercial, and industrial microgrids revealed distinct consumption patterns that significantly influenced forecasting model performance. Each sector exhibited unique temporal characteristics, variability profiles, and demand behaviours, as evidenced by statistical and visual diagnostics.

Residential Load Data

The residential microgrid dataset reflects electricity usage trends typical of household consumption. The hourly load time-series in Figure 1 illustrates demand across the year, segmented by months. Residential demand shows pronounced seasonal variation, with peaks occurring more frequently during colder months when heating is required or hotter months when cooling is necessary. Within each 24-hour cycle, demand also fluctuates significantly due to human behaviour, such as morning and evening household activities.

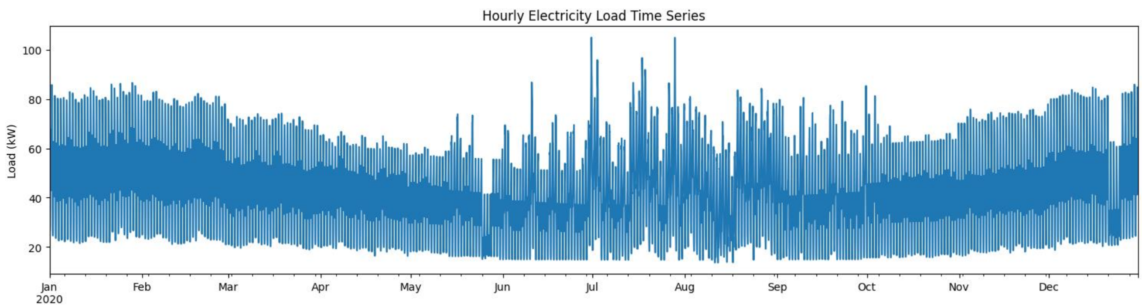


Figure 1: Hourly Electricity Load Time Series (Residential)

The histogram in Figure 2 indicates a right-skewed load distribution, as most demand values cluster near the average while occasional high-demand events extend the curve to the right. Demand surges are particularly noticeable on extremely hot or cold days when heating or air-conditioning use intensifies.

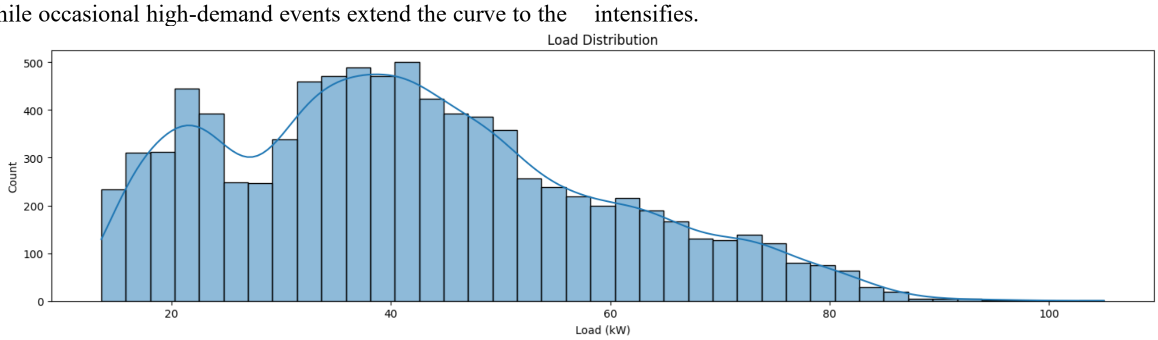


Figure 2: Load Distribution (Residential)

A strong positive correlation is observed at 24-hour intervals in the residential load, as shown by the autocorrelation function (ACF) in Figure 3. This pattern confirms daily periodicity, suggesting that load values at the same hour on consecutive days serve as reliable predictors of one another.

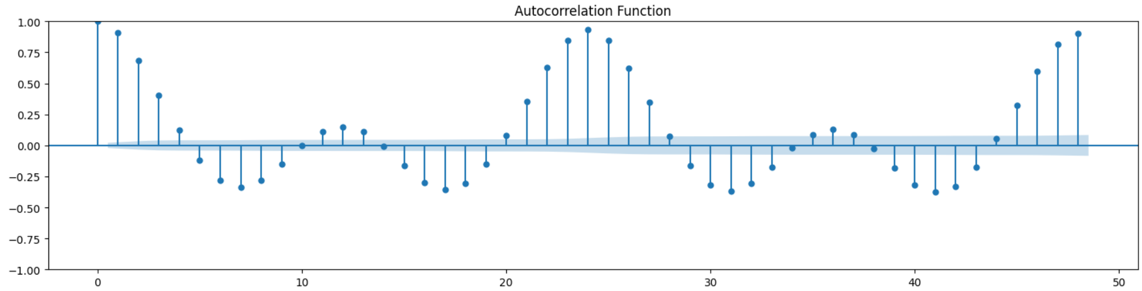


Figure 3: Autocorrelation Function (Residential)

The hourly average load curve in Figure 4 highlights the familiar double-peak profile of residential demand. Consumption rises in the morning, declines slightly around midday, and then climbs sharply in the evening when households engage in meal preparation, heating or cooling, and lighting.

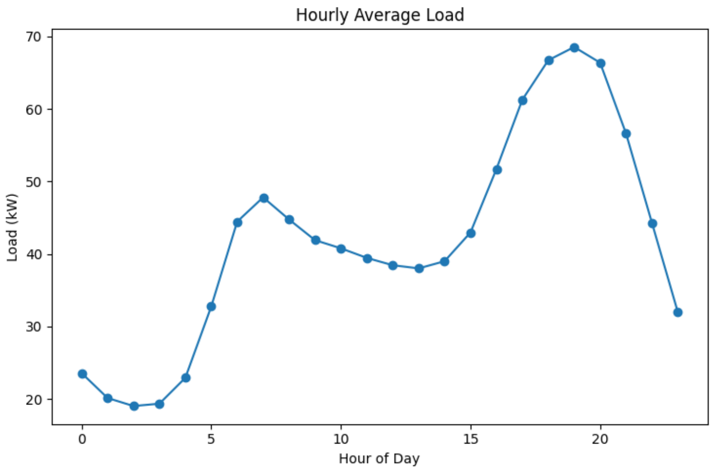


Figure 4: Hourly Average Load (Residential)

Commercial Load Data

The commercial microgrid dataset displays usage characteristics that differ from residential demand, showing more consistent patterns during working hours. The annual hourly time-series in Figure 5 reflects steady weekday profiles with marked reductions on weekends and holidays. Activity is higher during weekdays and falls consistently on Saturdays and Sundays.

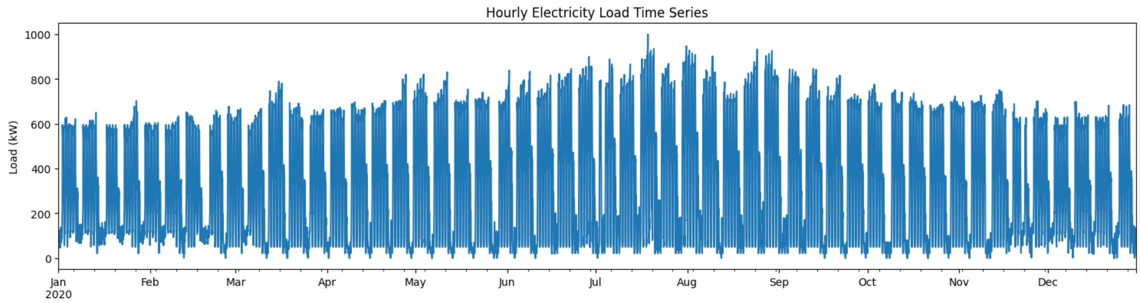


Figure 5: Hourly Electricity Load Time Series (Commercial)

The histogram in Figure 6 shows a symmetrical distribution centred around the mean, reflecting controlled and predictable usage in commercial buildings, unlike residential loads. The absence of a heavy tail indicates that extreme consumption spikes occur infrequently.

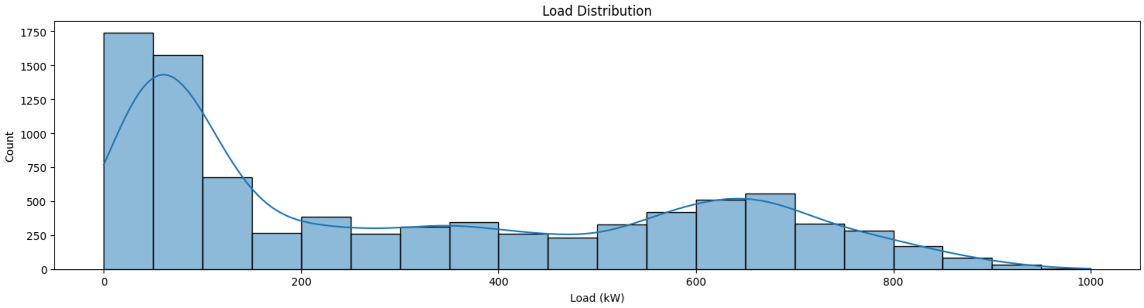


Figure 6: Load Distribution (Commercial)

As shown in Figure 7, the autocorrelation function displays 24-hour periodicity similar to residential demand but with quicker decay. This suggests that while commercial load follows structured daily cycles, its autocorrelation weakens over longer periods due to minor operational variations. Current demand is a useful indicator for same-hour usage on subsequent days, though fluctuations accumulate gradually.

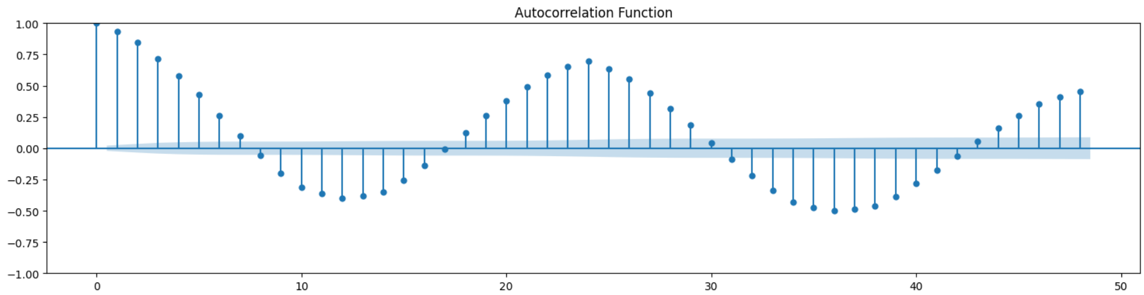


Figure 7: Autocorrelation Function (Commercial)

The hourly average load pattern in Figure 8 shows a sharp morning rise at the start of business hours, sustained high demand throughout the day, and a decline after evening closure. The profile is characterized by a single peak, consistent with standard commercial and retail working hours.

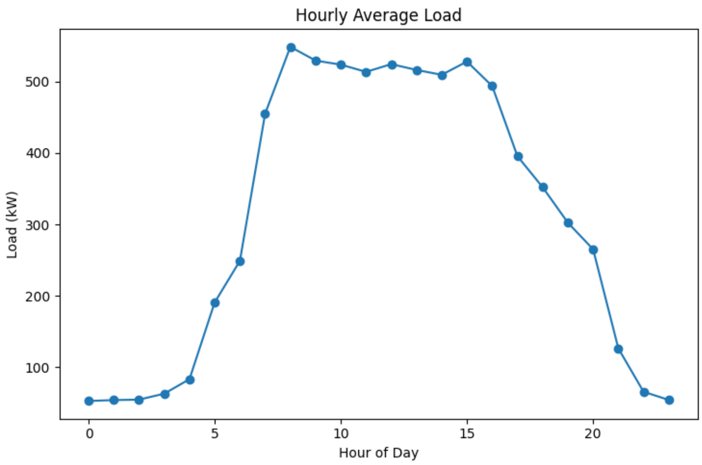


Figure 8: Hourly Average Load (Commercial)

Industrial Load Profile

The industrial microgrid dataset presents distinct characteristics compared to residential and commercial loads, with generally steady, high-level usage interrupted occasionally by operational changes. Figure 9 shows that industrial facilities maintain stable consumption across the year, reflecting their continuous operations and contribution to a strong base load.

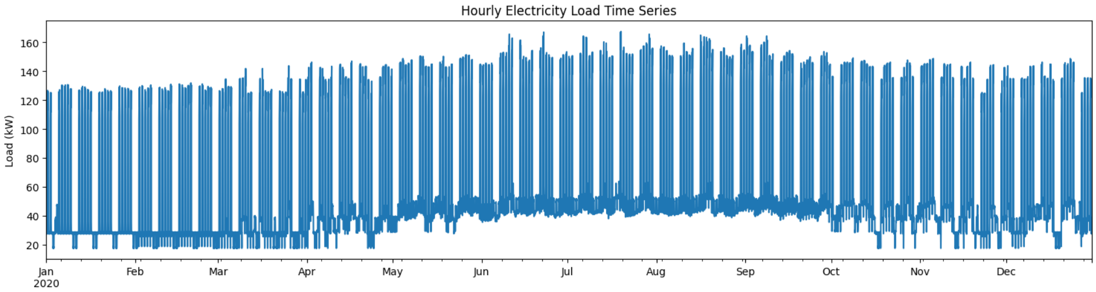


Figure 9: Hourly Electricity Load Time Series (Industrial)

The load distribution in Figure 10 demonstrates a bimodal pattern with uneven peaks. The dominant peak corresponds to baseline energy demand from regular production, while the smaller secondary peak captures intermittent high-load events such as machinery startups, batch cycles, or maintenance.

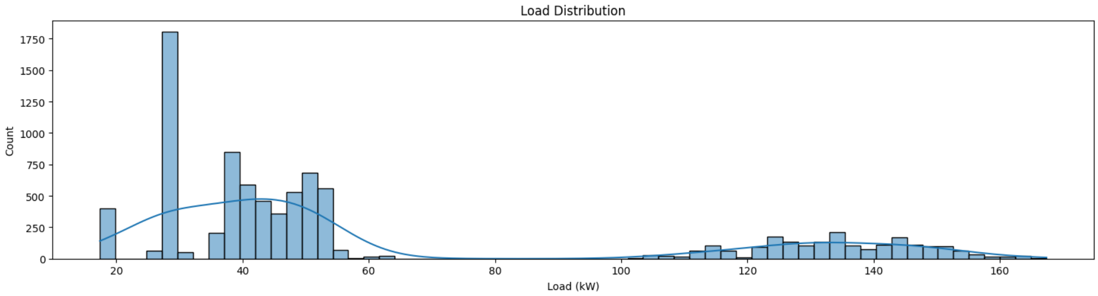


Figure 10: Load Distribution (Industrial)

The ACF in Figure 11 confirms 24-hour periodicity, though with lower peak intensity than residential or commercial loads. Industrial processes operate more continuously, with minimal influence from daily human behavioural routines.

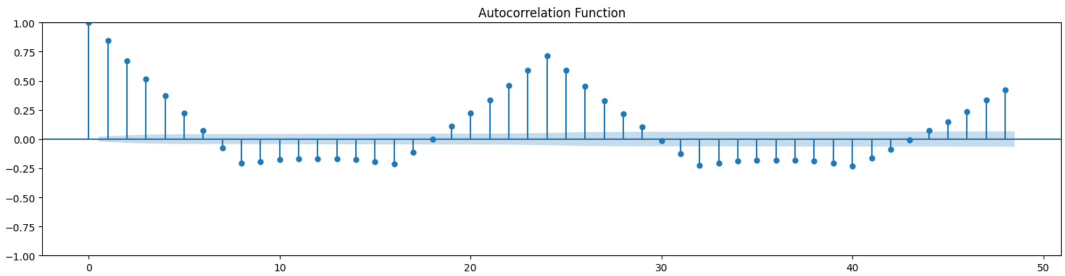


Figure 11: Autocorrelation Function (Industrial)

The hourly average load shown in Figure 12 reveals a sharp rise during morning work hours, sustained high demand throughout the day, and a drop-off after 4 p.m. This pattern aligns with the operational schedules of production and manufacturing facilities.

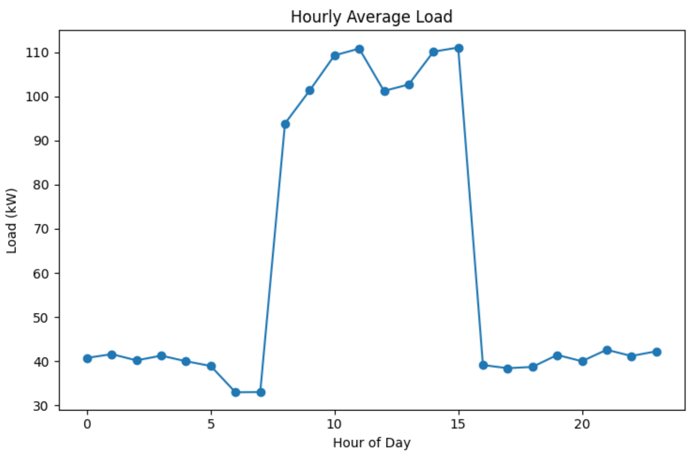


Figure 12: Hourly Average Load (Industrial)

Data Preprocessing

Upon loading the datasets, each time series was indexed by datetime, starting from January 1st, 2020. The data was cleaned to remove anomalies and missing values using linear interpolation, ensuring a continuous and reliable sequence for modelling. Exploratory Data Analysis (EDA) was then conducted to understand the temporal behaviour of each sector. This included plotting hourly load patterns, examining histograms to assess distribution, and calculating autocorrelation to determine temporal dependencies. These insights informed the selection of a 24-hour sliding window approach for supervised learning, where past daily consumption was used to forecast the next hour's load.

To prepare the data for neural network training, all values were normalized to the range [0, 1] using Min-Max scaling. This step ensures that the model training process is stable and that features contribute proportionally during gradient descent optimization.

Clustering Load Profiles

Clustering was employed to segment the dataset into homogeneous groups before applying LSTM models. The objective was to exploit similarities in consumption patterns to train specialized forecasting models tailored to each cluster. For clustering purposes, the 24-hour daily load profiles were extracted by reshaping the time series into 365 daily segments, each containing 24 hourly values.

Four clustering algorithms were applied: K-Means, Hierarchical Agglomerative Clustering, DBSCAN, and Gaussian Mixture Models (GMM). These are four widely used clustering algorithms, each with distinct theoretical foundations and applications in data segmentation.

K-Means is a partition-based algorithm that divides the dataset into k clusters by minimizing the within-cluster sum of squared distances. It iteratively assigns data points to the nearest centroid and then recalculates centroids until convergence. K-Means assumes that clusters are spherical and of roughly equal size, which makes it effective for well-separated, uniformly distributed datasets. However, it requires the number of clusters to be specified in advance and can be sensitive to the initial placement of centroids.

Hierarchical Agglomerative Clustering builds a hierarchy of clusters using a bottom-up approach. Initially, each data point is treated as an individual cluster, and pairs of clusters are merged step by step based on their similarity, often measured by linkage criteria such as single linkage, complete linkage, or average linkage. The process produces a dendrogram that visually represents the clustering structure. Unlike K-Means, it does not require specifying the number of clusters beforehand, but its computational cost increases significantly with larger datasets, and once clusters are merged, they cannot be undone.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies clusters as dense regions of points separated by areas of low density. It requires two parameters: ε (the neighbourhood radius) and minPts (minimum number of points in a neighbourhood to form a cluster). DBSCAN is capable of detecting clusters of arbitrary shapes and is robust to noise and outliers. However, it struggles with datasets where cluster densities vary greatly and can be computationally demanding in high-dimensional spaces.

Gaussian Mixture Models (GMM) provide a probabilistic approach to clustering by assuming that data points are generated from a mixture of several Gaussian distributions, each characterized by its mean and covariance. Unlike K-Means, which assigns points to the nearest cluster centroid, GMM performs soft clustering, assigning probabilities to each point's membership across clusters. This allows GMM to model elliptical clusters with different orientations and scales. However, it requires specifying the number of clusters in advance, is sensitive to initialization, and can become computationally intensive for large datasets.

Table 1 shows the clustering parameters used to segment residential, commercial, and industrial load profiles. These settings ensured effective grouping of load patterns to improve LSTM forecasting accuracy.

Table 1: Clustering Parameters Used in the Study

Algorithm	Key Parameters	Notes on Use
K-Means	Number of clusters (k) = 3–5, Initialization = k-means++	Applied to segment load profiles into distinct groups based on similarity.
Hierarchical Clustering	Linkage method = Ward, Distance metric = Euclidean	Generated dendrograms to visualize relationships and identify optimal splits.
DBSCAN	ϵ (neighbourhood radius) = 0.5, MinPts = 5	Captured arbitrary-shaped clusters and handled noise effectively.
Gaussian Mixture Model	Number of components = 3, Initialization = k-means, Covariance type = full	Allowed soft clustering with probabilistic assignments of load patterns.

Model Architecture

The study employed three categories of forecasting models: baseline models, Long Short-Term Memory (LSTM) network, and clustering-enhanced LSTM models.

Baseline Models

The baseline methods included the Persistence model and Feedforward Neural Networks (FFNN). The Persistence model assumes that the next load value is equal to the most recent observed value, serving as a simple benchmark. FFNNs consist of fully connected layers where data flows unidirectionally from input to output without feedback loops, making them suitable for nonlinear regression but limited in handling sequential dependencies.

Long Short-Term Memory (LSTM) Model

The LSTM is a type of Recurrent Neural Network (RNN) designed to address the vanishing gradient problem of standard RNNs. It achieves this by introducing memory cells and gating mechanisms (input, output, and forget gates) that regulate the flow of information over time. This allows LSTMs to capture long-term dependencies in sequential data, which is essential for electricity demand forecasting where past patterns strongly influence future load. The architecture in this study consisted of stacked LSTM layers followed by dense layers for final prediction.

Clustering-Enhanced LSTM Models

To improve forecasting accuracy in heterogeneous datasets, clustering algorithms were applied as a preprocessing step before LSTM training. K-Means clustering partitions data into k clusters based on minimizing intra-cluster variance. Hierarchical Agglomerative Clustering builds a hierarchy of clusters through successive merging, producing a dendrogram for structure visualization. DBSCAN identifies dense regions of data to form clusters, while Gaussian Mixture Models (GMMs) adopt a probabilistic approach, modelling the data as a mixture of Gaussian distributions and enabling soft assignments of points to multiple clusters. By grouping similar load profiles, these clustering methods ensured that the LSTM learned from more homogeneous patterns, thereby enhancing generalization and predictive performance.

Training Process

The dataset was transformed into supervised learning format using a sliding window approach, with the previous 24 hourly load values as input features and the next hour's load as the target output. The data was normalized to ensure stability during gradient descent optimization. The training and testing sets were split in an 80:20 ratio, and the LSTM models were trained using backpropagation through time (BPTT) with the Adam optimizer. Mean Squared Error (MSE) was used as the loss function, while performance was evaluated using both MSE and Mean Absolute Error (MAE).

Model Hyperparameters

The configuration of hyperparameters played a critical role in shaping the performance of the models. For the baseline Feedforward Neural Network (FFNN), the architecture was kept relatively simple, consisting of two hidden layers with 64 and 32 neurons respectively. The Rectified Linear Unit (ReLU) activation function was applied to capture nonlinear patterns, while the Adam optimizer with a learning rate of 0.001 was used to ensure stable and efficient convergence. A batch size of 32 was selected to balance training speed and gradient stability, and early stopping was implemented to avoid overfitting.

For the Long Short-Term Memory (LSTM) models, the hyperparameters were chosen to optimize their ability to capture long-term dependencies in the load data. Each LSTM layer contained 50 units, which provided sufficient memory capacity without leading to excessive computational cost. The learning rate of 0.001 was selected after empirical testing, as higher values led to unstable training, while lower values slowed convergence. A batch size of 64 and sequence length of 24 (corresponding to the 24-hour look-back window) were applied to reflect the daily periodicity of electricity consumption. Dropout with a rate of 0.2 was introduced to reduce overfitting by randomly deactivating neurons during training.

The clustering-enhanced LSTM models required additional hyperparameter selection for the clustering stage. For K-Means, the number of clusters was determined using the Elbow Method and Silhouette Score, ensuring a balance between compactness and separation of load profiles. DBSCAN parameters ε (neighbourhood radius) and minPts (minimum points per cluster) were selected through experimentation to capture meaningful consumption patterns while avoiding excessive noise. For Gaussian Mixture Models (GMM), the number of components was set in line with the optimal cluster count from K-Means for consistency. Hierarchical clustering applied Ward's linkage criterion to minimize intra-cluster variance, providing interpretable dendrogram-based grouping.

Overall, hyperparameters were carefully tuned through cross-validation and iterative testing to achieve robust forecasting performance while maintaining computational efficiency. Their selection reflects a balance between accuracy, generalizability, and practicality for real-world microgrid applications. Table 2 below summarizes the hyperparameters and configurations of the key forecasting models used in this study.

Table 2: Summary of Model Hyperparameters

Hyperparameter	Value/Range Used	Purpose in Training
Look-back Window	24 hours	Captures past daily consumption patterns to forecast the next hour load.
Forecast Horizon	1 hour ahead	Ensures short-term forecasting suitable for operational microgrid management.
Number of LSTM Units	64	Provides sufficient capacity to learn temporal dependencies without overfitting.
Hidden Layers	2	Balances model depth to capture nonlinear features while avoiding excess complexity.
Batch Size	32	Allows efficient training with stable gradient updates.
Epochs	100 (with early stopping)	Provides enough training cycles with a safeguard to prevent overfitting.
Optimizer	Adam	Chosen for adaptive learning rate adjustment and efficient convergence.
Learning Rate	0.001	Balances convergence speed and model stability.
Loss Function	Mean Squared Error (MSE)	Measures prediction errors, suitable for continuous regression tasks.
Regularization	Dropout (0.2)	Reduces overfitting by randomly disabling neurons during training.

Evaluation Metrics

To assess the performance of the forecasting models developed in this study, two standard error metrics were employed: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics were chosen for their widespread use in regression-based forecasting tasks and their ability to capture different aspects of prediction error (Terven et al., 2025).

Mean Squared Error (MSE) measures the average of the squared differences between actual and predicted values. It is especially useful for penalizing large errors more heavily, making it effective in detecting models that perform poorly on sudden demand changes or anomalies in the load profile. The mathematical formulation of MSE is given by:

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \widehat{y_i})^2 \qquad (1)$$

where $\mathrm{y_i}$ is the actual value, $\widehat{y_i}$ is the predicted value, and N is the total number of observations. The MSE provides a measure of the average error magnitude, with greater emphasis on larger errors due to the squaring of differences.

Mean Absolute Error (MAE) provides the average magnitude of the prediction errors, offering a more interpretable and less sensitive measure compared to MSE. It treats all errors equally, regardless of their direction or size, making it a reliable metric for evaluating the overall accuracy of the forecasting model. The MAE is calculated as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (2)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the total number of observations. The MAE provides a straightforward measure of the average error magnitude, making it easy to interpret.

Both metrics were computed using the test dataset, which consisted of 20% of the total data, withheld during training to ensure unbiased evaluation. For clustering-enhanced models, MSE and MAE were calculated for each cluster individually and then aggregated to determine overall performance.

Implementation Setup

All experiments were conducted using Google Colaboratory leveraging cloud-based GPU acceleration. The implementation utilized Python 3.10, with TensorFlow/Keras for deep learning, Scikit-learn for clustering, and Matplotlib and Seaborn for visualization. A fixed random seed was used throughout to ensure the reproducibility of results. Each model was trained and evaluated independently for the residential, commercial, and industrial datasets, enabling a sector-specific analysis of forecasting performance.

RESULTS AND DISCUSSION

Baseline Models vs LSTM

Table 3 presents the comparative performance of the baseline models, Persistence and Feedforward Neural Networks (FFNN), against the Long Short-Term Memory (LSTM) model across residential, commercial, and industrial microgrids. The results clearly demonstrate that LSTM consistently outperformed the baseline models in terms of both Mean Squared Error (MSE) and Mean Absolute Error (MAE). For the residential dataset, Persistence produced the highest error values with an MSE of 60.13 kW and MAE of 6.05 kW, while FFNN significantly reduced errors (MSE 5.83 kW, MAE 2.05 kW). However, the LSTM achieved superior accuracy with MSE 4.42 kW and MAE 1.53 kW, confirming its ability to capture temporal dependencies within household consumption patterns. Similarly, in the commercial sector, Persistence delivered very poor results with an MSE of 8007 and MAE of 46.85 kW, while FFNN reduced errors considerably (MSE 796.79 kW, MAE 20.56 kW). The LSTM again outperformed with MSE 621.92 kW and MAE 17.43 kW, reflecting its robustness in identifying consistent weekday–weekend and operational load variations. In the industrial sector, where load profiles were more stable, all models performed better compared to residential and commercial cases. Still, LSTM achieved the lowest errors with MSE 93.14 and MAE 4.53 kW, outperforming Persistence (MSE 550.85 kW, MAE 8.88 kW) and FFNN (MSE 98.91 kW, MAE 5.43 kW).

Table 3: Performance Comparison of Baseline and LSTM Models Across Load Profiles

Load Profile	Model	MSE (kW)	MAE (kW)
Residential	Persistence	60.1305	6.0525
	FFNN	5.8279	2.0503
	LSTM	4.4184	1.5312
Commercial	Persistence	8007.000	46.8516
	FFNN	796.789	20.5649
	LSTM	621.921	17.4252
Industrial	Persistence	550.850	8.8803
	FFNN	98.911	5.4288
	LSTM	93.143	4.5347

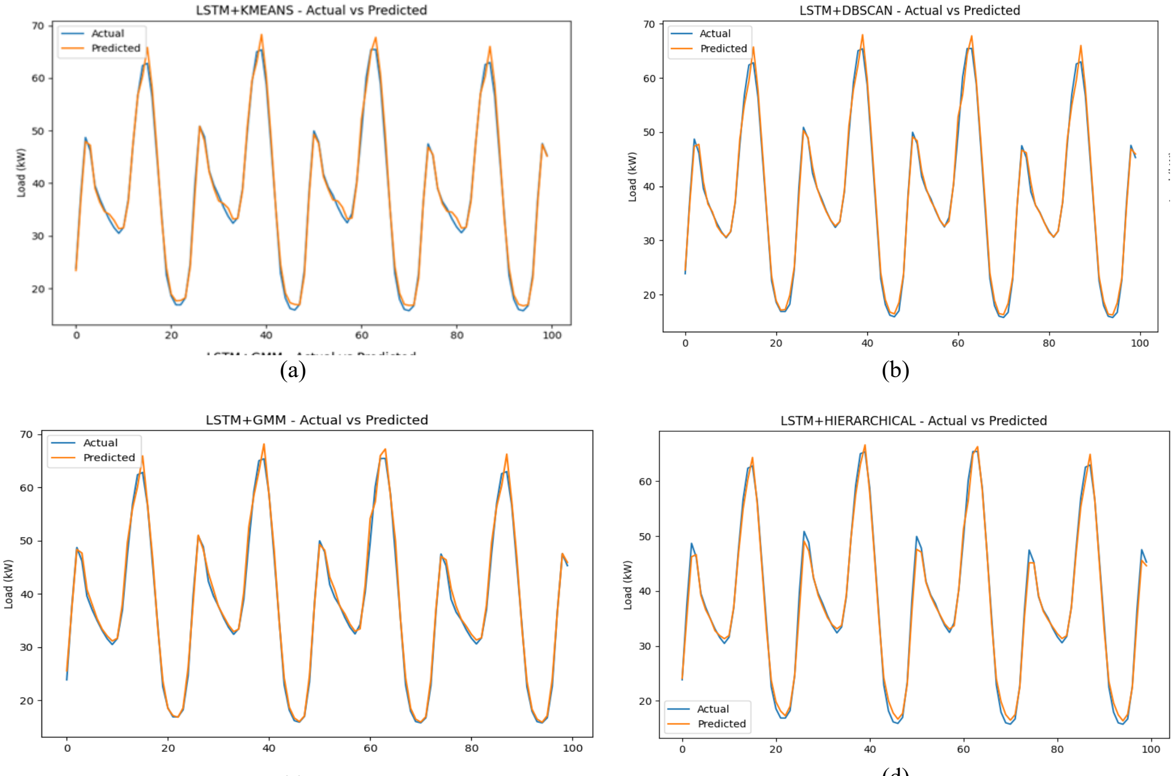
LSTM vs Clustering-Enhanced LSTM

Table 4 reports the performance of clustering-enhanced LSTM hybrids across residential, commercial, and industrial microgrids, with the baseline LSTM shown for reference. In the residential case, every clustered variant improves upon the standalone LSTM (MSE = 4.42 kW, MAE = 1.53 kW): the best performer is LSTM+GMM (MSE = 2.30 kW, MAE = 1.10 kW), delivering a 48.0% reduction in MSE and 28.3% reduction in MAE relative to LSTM, closely followed by LSTM+K-Means (MSE = 2.32 kW, MAE = 1.12 kW; 47.5% MSE and 26.8% MAE gains). LSTM+DBSCAN and LSTM+Hierarchical also lower errors substantially (MSE improvements of 36.2% and 37.5%, respectively), though they trail the probabilistic and centroid-based approaches. In the commercial sector, clustering yields a more nuanced picture: LSTM+Hierarchical is dominant (MSE = 505.44 kW, MAE = 14.92 kW), improving on LSTM by 18.7% in MSE and 14.4% in MAE, while LSTM+GMM offers modest MSE gains (8.3%) and roughly parity in MAE (0.8% better). In contrast, LSTM+K-Means and LSTM+DBSCAN underperform the baseline, indicating that commercial load regularities benefit from multi-scale structure discovery (hierarchical) more than from rigid centroid or density partitions. For the industrial profile, where load is comparatively stable, clustering has limited effect: LSTM+K-Means yields a small 2.4% MSE improvement with essentially unchanged MAE, and the other hybrids are neutral to slightly worse, underscoring that segmentation adds little when variability is low and patterns are already well captured by the sequence model.

Table 4: Performance of Clustered LSTM Models Across Load Profiles

Model	Residential (kW)		Commercial (kW)		Industrial (kW)	
Performance metric	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
LSTM (baseline)	4.418	1.531	621.921	17.425	93.143	4.535
LSTM + K-Means	2.322	1.121	1182.900	24.819	90.889	4.536
LSTM + DBSCAN	2.820	1.196	1025.590	23.134	90.916	5.007
LSTM + GMM	2.296	1.097	570.079	17.288	96.522	5.323
LSTM + Hierarchical	2.763	1.313	505.442	14.924	97.562	5.590

Visually, Figure 13 shows the actual versus predicted residential load for LSTM models enhanced with K-Means, DBSCAN, GMM, and Hierarchical clustering, illustrating how each clustering technique affects forecasting precision in highly variable household consumption patterns.

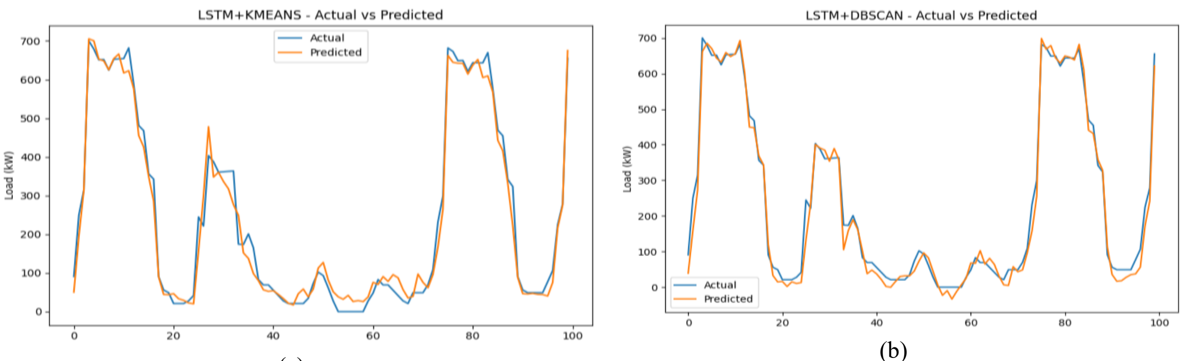


(a) (b)

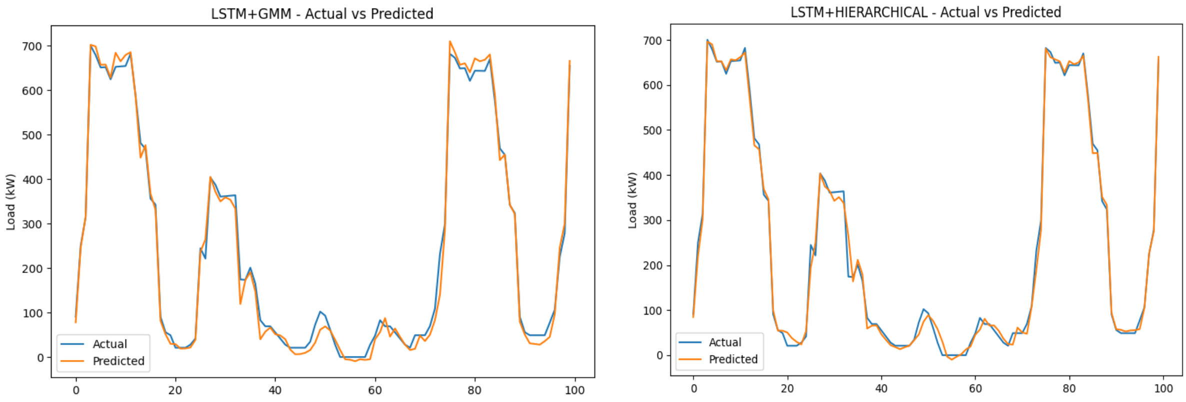
(c) (d)

Figure 13: Actual vs Predicted Graph of (a) K-means, (b) DBSCAN, (c) GMM and (d) Hierarchical (Residential)

Figure 14 compares the actual versus predicted load values for commercial microgrids using four clustering-enhanced LSTM models (K-means, DBSCAN, GMM, and Hierarchical), demonstrating their relative forecasting accuracy across different demand patterns.



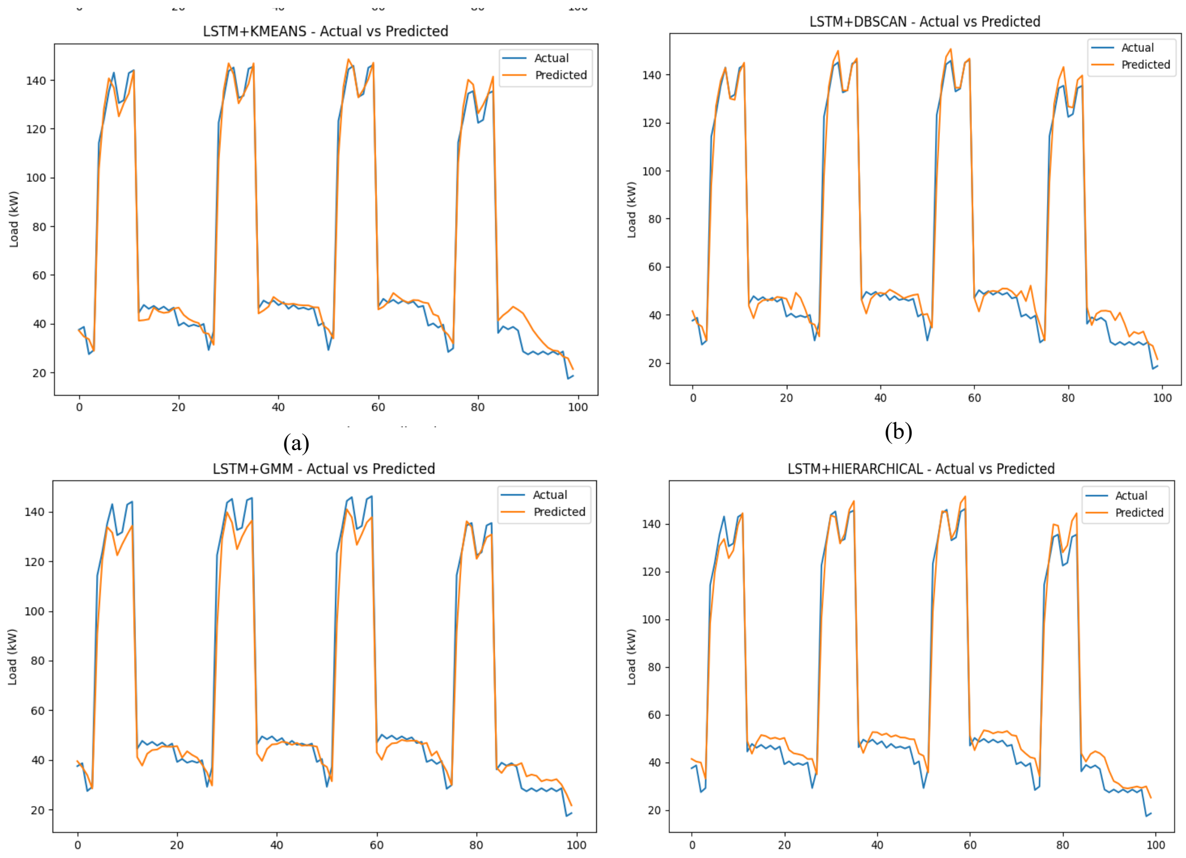
(a) (b)



(c) (d)

Figure 14: Actual vs Predicted Graph of (a) K-means, (b) DBSCAN, (c) GMM and (d) Hierarchical (Commercial)

Figure 15 contrasts the actual and predicted load values for industrial microgrids across four clustering-enhanced LSTM approaches (K-means, DBSCAN, GMM, and Hierarchical), revealing their comparative performance in handling stable industrial demand patterns with minimal fluctuations.



(a) (b) (c) (d)

Figure 15: Actual vs Predicted Graph of (a) K-means, (b) DBSCAN, (c) GMM and (d) Hierarchical (Industrial)

Collectively, these findings indicate that clustering is most valuable where intraclass heterogeneity is high (residential), selectively helpful where multi-regime structure exists (commercial), and least impactful where profiles are homogeneous (industrial).

Discussion

The findings of this study highlight important implications for the application of deep learning and clustering techniques in short-term load forecasting across different microgrid sectors. The comparative results show that while Long Short-Term Memory (LSTM) networks consistently outperform Feedforward Neural Networks (FFNN), the extent of this improvement diminishes across residential, commercial, and industrial load profiles. This reduction in relative performance suggests that the benefit of sequential modelling becomes less pronounced in environments with more stable and predictable consumption patterns. In particular, the industrial sector, characterized by steady operational demand, presents fewer temporal variations, thereby narrowing the advantage of LSTM over simpler models like FFNN.

The integration of clustering methods further demonstrates the potential of data segmentation in improving forecasting accuracy. Clustering enhanced LSTM models achieved notable performance gains, especially in the residential sector where load profiles are highly heterogeneous and influenced by diverse behavioural patterns. By grouping similar patterns, clustering reduced variability and enabled LSTM models to capture clearer relationships between past and future consumption. However, the improvements introduced by clustering were less substantial in the industrial sector, where consumption patterns were already uniform. This trend underlines the importance of tailoring forecasting methods to the specific characteristics of each load profile rather than applying a uniform approach across all sectors.

Despite these promising results, the study also has limitations that may influence its overall findings. The use of synthetic datasets, while providing control over data quality and availability, may not fully capture the irregularities and noise present in real-world microgrid operations. Additionally, clustering algorithms such as K-Means and GMM required predefined parameters, which may have constrained their adaptability to dynamic consumption behaviours. The study also focused on a fixed set of exogenous factors, and the exclusion of additional contextual variables such as socio-economic activities or real-time renewable energy fluctuations could have limited the models' ability to generalize. Finally, the computational overhead of clustering enhanced approaches may restrict their applicability in real-time microgrid management where speed and efficiency are critical.

CONCLUSION

This research proposed a hybrid short-term load forecasting framework that combines clustering algorithms with LSTM neural networks to improve prediction accuracy across residential, commercial, and industrial microgrids. The results demonstrated that baseline models, Persistence and Feedforward Neural Networks (FFNN), were consistently outperformed by Long Short-Term Memory (LSTM) networks, reflecting LSTM's ability to capture sequential dependencies in consumption patterns. In residential loads, Persistence exhibited an MSE of 60.13 kW and MAE of 6.05 kW, FFNN reduced these errors to 5.83 kW and 2.05 kW, while standalone LSTM further improved accuracy to an MSE of 4.42 kW and MAE of 1.53 kW. Similar trends were observed in commercial and industrial sectors; however, the relative improvement of LSTM over FFNN diminished for industrial loads due to their stable and predictable demand patterns.

Incorporating clustering prior to LSTM training further enhanced forecasting accuracy by segmenting load profiles into more homogeneous groups. Residential forecasts benefited most, with GMM reducing MSE by 48% and MAE by 28% compared to standalone LSTM, reflecting the high variability and behavioural heterogeneity in this sector. Commercial forecasting showed meaningful improvements, with Hierarchical Clustering yielding an 18.7% reduction in MSE and 14.4% reduction in MAE. Industrial profiles, due to their uniformity, exhibited modest gains, with K-Means delivering a 2.4% improvement in MSE and negligible change in MAE. These findings underscore the importance of tailoring forecasting strategies to sector-specific consumption characteristics rather than applying uniform methods across all microgrid contexts.

Despite the promising results, the study was based on synthetic datasets, which may not capture the full complexity of real-world consumption behaviours. Future work should focus on applying the framework to real-time microgrid data, incorporating external features such as weather, holidays, and socioeconomic variables to improve generalization. Additionally, adaptive or online clustering techniques and probabilistic forecasting should be explored to enhance responsiveness in dynamic environments.

REFERENCES

Angizeh, F., Ghofrani, A., & Jafari, M. A. (2020). Dataset on Hourly Load Profiles for a Set of 24 Facilities from Industrial, Commercial, and Residential End-use Sectors [Dataset]. Mendeley Data. https://doi.org/10.17632/rfnp2d3kjp.1

Aurangzeb, K., Alhussein, M., Javaid, K., & Haider, S. I. (2021). A pyramid-CNN based deep learning model for power load forecasting of similar-profile energy customers based on clustering. IEEE Access, 9, 14992–15003.

Bashir, T., Haoyong, C., Tahir, M. F., & Liqiang, Z. (2022). Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Reports, 8, 1678–1686.

Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. Traditional time-series techniques. Applied Energy, 236, 1078–1088.

Glazunova, A., Kovalchuk, D., Tomin, N., & Iskakov, A. (2024). Design of daily load profiles in commercial and industrial microgrids based on renewable energy sources. IFAC-PapersOnLine, 58(13), 272–277.

Han, F., Pu, T., Li, M., & Taylor, G. (2020). Short-term forecasting of individual residential load based on deep learning and K-means clustering. CSEE Journal of Power and Energy Systems, 7(2), 261–269.

Husein, M., & Chung, I.-Y. (2019). Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. Energies, 12(10), 1856.

Indralaksono, R., Wakhid, M. A., Wibowo, G. H., Abdillah, M., Rahardjo, A. B., & Purwitasari, D. (2022). Hierarchical Clustering and Deep Learning for Short-Term Load Forecasting with Influenced Factors. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 6(4), 692–701.

Karn, A. K. L., & Kakran, S. (2022). Operation Management of Microgrid Supplying to the Residential, Industrial and Commercial Community using Different Demand Response Techniques. 2022 4th International Conference on Energy, Power and Environment (ICEPE), 1–6. https://ieeexplore.ieee.org/abstract/document/9798064/

Makris, I., Moschos, N., Radoglou-Grammatikis, P., Andriopoulos, N., Tzanis, N., Malamaki, K.-N., Fotopoulou, M., Kaousias, K., Tompros, G. L., & Sarigiannidis, P. (2024). Exploring Load Forecasting: Bridging Statistical Methods to Deep Learning Techniques in Real-World Environment. 2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), 491–496. https://ieeexplore.ieee.org/abstract/document/10621486/

Mallah, M., Youssef, A., Ali, A., & Mohamed, E. E. (2024). Advancements in Microgrid Technologies: A Critical Review. SVU-International Journal of Engineering Sciences and Applications, 5(2), 241–251.

Mukhtiar, F., Javed, M., & Arbab, M. N. (2024). Comparing Microgrid Planning and Management Methods for Energy Resilience and Sustainability.

Muzaffar, S., & Afshari, A. (2019). Short-term load forecasts using LSTM networks. Energy Procedia, 158, 2922–2927.

Rafi, S. H., Deeba, S. R., & Hossain, E. (2021). A short-term load forecasting method using integrated CNN and LSTM network. IEEE Access, 9, 32436–32448.

Saxena, A., Shankar, R., El-Saadany, E., Kumar, M., Al Zaabi, O., Al Hosani, K., & Muduli, U. R. (2024). Intelligent load forecasting and renewable energy integration for enhanced grid reliability. IEEE Transactions on Industry Applications. https://ieeexplore.ieee.org/abstract/document/10616221/

Terven, J., Cordova-Esparza, D.-M., Romero-González, J.-A., Ramírez-Pedraza, A., & Chávez-Urbiola, E. A. (2025). A comprehensive survey of loss functions and metrics in deep learning. Artificial Intelligence Review, 58(7), 195. https://doi.org/10.1007/s10462-025-11198-7

©2025 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.