Electricity Bill Prediction Based on a Particle Swarm Optimized Multilayer Perceptron Model

Ahmed Fahim
Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia | Department of Computer Science, Faculty of Computers and Information, Suez University, Suez, Egypt
a.abualeala@psau.edu.sa (corresponding author)

Ahmed M. Osman
Department of Information Systems, Faculty of Computers and Information, Suez University, Suez, Egypt
a.osman@suezuni.edu.eg

Zahraa Tarek
Department of Computer Engineering and Information, College of Engineering, Wadi Ad Dwaser, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia | Computer Science Department, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt
z.elmana@psau.edu.sa

Ahmed M. Elshewey
Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.Box:43221, Suez, Egypt | Applied Science Research Center, Applied Science Private University, Amman, Jordan
ahmed.elshewey@fci.suezuni.edu.eg

Received: 27 August 2025 | Revised: 7 September 2025 | Accepted: 16 September 2025

DOI: https://doi.org/10.48084/etasr.14361

ABSTRACT

Accurate household electricity bill prediction enables better budgeting for consumers and data-driven planning for utilities. This study develops and benchmarks five deep learning models on a publicly available Indian household electricity bill dataset that combines appliance usage and socio-demographic attributes. We propose a Particle Swarm Optimized Multilayer Perceptron (PSO-MLP) model that tunes network depth, width, learning rate, and regularization via Particle Swarm Optimization (PSO), and compare it against plain Multilayer Perceptron (MLP), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN) architectures. The pipeline includes robust preprocessing (median imputation, scaling, and one-hot encoding), leakage-safe training/testing, and a comprehensive evaluation suite comprising Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), coefficient of determination (R^2), and Median Absolute Error (MedAE). Results show a near-deterministic fit: PSO-MLP achieves $\mathrm{MAE} = 10.22$, $\mathrm{RMSE} = 12.99$, $\mathrm{MSE} = 168.92$, $\mathrm{R}^2 = 0.9998$, and $\mathrm{MedAE} = 8.43$; a plain MLP attains $\mathrm{MAE} = 10.29$ with a similar R^2, whereas recurrent models provide no advantage on this non-sequential, tabular task (RNN $\mathrm{MAE} = 23.03$). Error distributions confirm stable performance across the bill range with minimal bias. These findings indicate that carefully regularized feed-forward models—augmented with principled hyperparameter optimization—suffice to model household bills with very high fidelity, whereas more complex sequence models are unnecessary. The proposed framework offers a strong baseline for tariff-aware extensions and deployment-grade forecasting in Indian residential settings.

Keywords-electricity bill prediction; deep learning; PSO; MLP; optimized MLP

I. INTRODUCTION

Residential electricity forecasting has renewed importance as utilities expand demand-side management and tariff reform. Recent surveys and large empirical studies report that data-driven and deep learning models capture the nonlinear, non-stationary patterns of household consumption and often outperform classical baselines across horizons from hours to months ahead. Feature design (e.g., appliance ownership, occupancy proxies, seasonality) and careful preprocessing are repeatedly shown to be as consequential as model choice.

However, most literature forecasts energy (kWh), whereas households pay bills computed through jurisdiction-specific tariffs. In India, domestic billing commonly combines slabbed (block) energy charges with fixed charges; since 2023, policy has also mandated Time-of-Day (ToD) differentials, making the bill function nonlinear and piecewise in consumption. Ignoring this structure can misstate absolute error and tail risk near slab boundaries. Official compilations and notifications document the heterogeneity of slab schedules and ToD rules across states.

A second methodological gap is the lack of uncertainty quantification. Beyond point forecasts, calibrated prediction intervals are increasingly recommended for operational decisions (e.g., warning of "bill shocks"). Recent peer-reviewed work shows that conformal prediction can wrap modern regressors to deliver distribution-free coverage in energy applications (such as buildings and Photovoltaic (PV) power), making it attractive for residential billing as well.

We study supervised prediction of the monthly household electricity bill from household or dwelling descriptors (e.g., occupants, floor area, and region) and appliance-usage attributes (e.g., ownership, ratings, and hours of use). The task is challenging because the billed amount is a nonlinear, often piecewise function of latent energy use and jurisdictional tariff rules (slabs, fixed charges, and possible ToD differentials). Practical datasets also exhibit mixed data types, missing values, scale heterogeneity, and the risk of target leakage (features that are direct or near-deterministic functions of the bill). Robust solutions therefore require leakage-safe splits, principled preprocessing for numeric and categorical variables, models that capture nonlinearities without overfitting, and evaluation beyond a single score—examining error distribution, tail behavior, and calibration.

Bill prediction—rather than kWh-only forecasting—is framed as a supervised learning problem on a public Indian household dataset, with a fully reproducible, leakage-safe pipeline. The pipeline handles mixed data types via median imputation for numeric variables, one-hot encoding for categorical variables, appropriate scaling, and group-aware train/test separation to prevent information bleed across households.

A focused benchmark evaluates five deep learning models tailored to this task: a baseline Multilayer Perceptron (MLP); a Particle Swarm Optimized Multilayer Perceptron (PSO-MLP) that automatically tunes depth, width, learning rate, and regularization via Particle Swarm Optimization (PSO); and three recurrent baselines—Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN)—adapted for tabular inputs. This design isolates the incremental value of a hyperparameter-optimized feed-forward network relative to widely used recurrent architectures in residential energy research.

Recent work on residential electricity analytics shows a steady shift from classical statistical forecasting toward data-driven and deep learning approaches that better capture nonlinear, non-stationary usage patterns and consistently outperform traditional baselines across short- and month-ahead horizons.

Authors in [1] conducted a study comparing a regularized linear model (Lasso) with ordinary least squares for predicting Locational Marginal Prices (LMPs). They found that Lasso performed better (95.56% accuracy/precision) than linear regression (80.96%). The study concluded that L1 regularization improves LMP prediction. However, the evidence is limited by the small sample, nonstandard regression metric, and sparse feature and preprocessing details.

Authors in [2] compared Random Forest (RF) and Support Vector Machine (SVM) models for household electricity bill prediction using a small Kaggle-sourced sample (reported N=10 for RF and N=10 for SVM), with "expected incidence" 82.52% (RF) and 40.50% (SVM). They ultimately reported RF achieving 90.05% accuracy, whereas a decision tree baseline yielded 41.20%, and concluded that RF is the most suitable algorithm for this task.

Authors in [3] targeted the utility problem of electricity fee recovery risk by building a supervised prediction system on large-scale company data combining historical bill-payment records with user credit information. After cleaning and integrating the database, the study performed indicator (feature) selection using a mix of expert knowledge and data-driven relevance to retain variables most associated with default or recovery risk. Multiple machine learning models were trained and compared, and the final choice was based on prediction accuracy and generalization. In experiments, an LSTM neural network outperformed a logistic regression baseline across operational metrics.

Authors in [4] addressed the lack of high-resolution demand data for designing renewable and collective energy systems by learning hourly load profiles from only monthly Time-of-Use (ToU) electricity bills. Using a limited set of smart meter measurements spanning diverse end-user categories, they assembled a training corpus and—departing from prior classification-first pipelines—fit a direct regression model that maps coarse monthly ToU bills to "typical" hourly shapes across heterogeneous users. On one year of data, the approach attained an average normalized Mean Absolute Error (MAE) of 26% for instantaneous hourly consumption and <4% for ToU period aggregates, and outperformed both standard load profiles and a two-step classification-then-profiling baseline on multiple accuracy and statistical criteria.

Authors in [5] addressed shortcomings in prior Electricity Bill Recovery (EBR) risk models by introducing a bidirectional LSTM classifier that leverages temporal patterns in payment histories to incorporate information from both past and subsequent time steps when labeling risk, reporting 98% accuracy and markedly improved risk classification versus conventional predictors. The contribution lies in framing EBR as a sequence learning problem and exploiting bidirectional temporal dependencies to capture complex, time-evolving signals tied to non-payment risk.

Authors in [6] developed a smart household energy management system that uses Internet of Things (IoT) hardware and predictive analytics for bill forecasting. The system uses an ESP32 microcontroller, Artificial Neural Network (ANN), Adafruit IO dashboard, and If This Then That (IFTTT) email alerts. Designed for Malaysia's residential sector, it has a 94% accuracy rate for bill prediction. The system integrates sensing, control, cloud visualization, and machine learning into a single workflow.

Authors in [7] presented a machine learning approach for household electricity cost/bill forecasting, focusing on time-series signals and data-driven methods for accuracy and policy relevance. They created a predictive workflow using artificial intelligence and machine learning for modeling, data validation, and preparation, and benchmarked multiple algorithms for household bill prediction.

Authors in [8] developed a machine learning automation system to predict monthly household electricity consumption and cost in Bangladesh using appliance-level data. The system covered data collection, feature extraction, scaling, normalization, hyperparameter tuning, training, and model selection across various regressors. RF was found to yield the best performance, outperforming earlier approaches by ≥4% in accuracy and ≥7% in MAE. The system also included a mobile application for estimating bills, offering recommendations, and electricity board notifications. The system was rated "excellent" by over 70% of surveyed users. The system's strengths include a full stack from data to deployment, a focus on monthly bills, and a user-facing application.

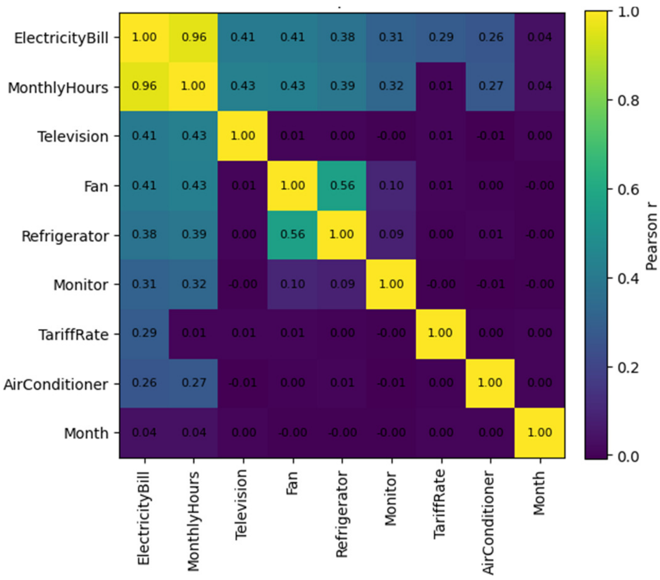
Authors in [9] analyzed half-hourly smart-meter readings and survey data from over 3,600 Irish households to characterize temporal usage patterns and design an adaptive dynamic pricing scheme. Using advanced time-series analysis, harmonic modeling, and an iterative error-driven adjustment, the study constructed ToD tariffs segmented into night, day, and peak periods with seasonal adjustments to track demand fluctuations.

Authors in [10] proposed a hybrid deep learning framework for electric load forecasting that remains accurate when inputs are missing or noisy. Using real hourly load data from Qassim (Saudi Arabia), the approach coupled a Convolutional Neural Network (CNN) feature extractor with sequence models (LSTM, GRU, and Bidirectional LSTM (BiLSTM)) to capture spatial and temporal dependencies, preceded by a multi-step preprocessing pipeline for normalization and outlier handling. In comparative experiments, the CNN-GRU variant delivered the best results, reducing errors (Root Mean Squared Error (RMSE), Normalized RMSE (NRMSE), Mean Absolute Percentage Error (MAPE)) versus traditional methods and standalone deep learning baselines.

II. METHODOLOGY

A. *Dataset and Preprocessing*

The study uses the publicly available Indian Household Electricity Bill Dataset (Kaggle) [11], which provides anonymized, household-level records with appliance ownership and usage descriptors (e.g., counts, rated power, and daily hours), dwelling and socio-demographic attributes (e.g., occupants, floor area, region, and month), and a monetary target labeled ElectricityBill. Each row represents a household instance with mixed data types (numeric and categorical) suitable for tabular learning. Because the objective is to predict the bill actually paid rather than kWh, all features that are direct linear combinations of the target (e.g., "total due" fields that sum the bill with taxes/fees when present) are excluded to prevent target leakage. The dataset is randomly split into 80% training and 20% testing subsets with a fixed seed to ensure reproducibility, and a 5-fold cross-validation protocol on the training portion supports model selection and early stopping. Figure 1 displays the heatmap investigation for the utilized dataset.



	ElectricityBill	MonthlyHours	Television	Fan	Refrigerator	Monitor	TariffRate	AirConditioner	Month
ElectricityBill	1.00	0.96	0.41	0.41	0.38	0.31	0.29	0.26	0.04
MonthlyHours	0.96	1.00	0.43	0.43	0.39	0.32	0.01	0.27	0.04
Television	0.41	0.43	1.00	0.01	0.00	-0.00	0.01	-0.01	0.00
Fan	0.41	0.43	0.01	1.00	0.56	0.10	0.01	0.00	-0.00
Refrigerator	0.38	0.39	0.00	0.56	1.00	0.09	0.00	0.01	-0.00
Monitor	0.31	0.32	-0.00	0.10	0.09	1.00	-0.00	-0.01	-0.00
TariffRate	0.29	0.01	0.01	0.01	0.00	-0.00	1.00	0.00	0.00
AirConditioner	0.26	0.27	-0.01	0.00	0.01	-0.01	0.00	1.00	0.00
Month	0.04	0.04	0.00	-0.00	-0.00	-0.00	0.00	0.00	1.00

Fig. 1. Dataset feature correlation heatmap.

A single, reusable pipeline transforms the raw table into model-ready arrays. Numeric features undergo median imputation for missing values followed by z-score scaling (mean = 0, unit variance) to stabilize optimization across models. Categorical features are imputed with the most frequent category and one-hot encoded with unknown-category handling so the test set never triggers an error. To aid convergence of deep networks, the target ElectricityBill is standardized during training and then inverted to original currency units for all reported metrics and plots.

Outliers are handled conservatively: values outside plausible physical ranges (e.g., negative hours or impossible appliance power ratings) are clipped after inspection; otherwise, the models learn from the natural variability. All preprocessing steps are fit on the training set only and applied to validation and testing through frozen transformers, ensuring strict separation between training and evaluation. The same preprocessing pipeline feeds every model (MLP, PSO-MLP, GRU, LSTM, RNN), so differences in performance reflect the learning algorithm rather than data handling. Table I displays the statistical summary of the dataset features.

TABLE I. STATISTICAL SUMMARY OF DATASET FEATURES

Feature	Count	Mean	Std	Min	Max
Fan	45345	13.9906	5.4708	5	23
Refrigerator	45345	21.7054	1.6725	17	23
AirConditioner	45345	1.5039	1.1154	0	3
Television	45345	12.5026	5.7560	3	22
Monitor	45345	2.8650	3.8949	1	12
Month	45345	6.4880	3.4432	1	12
MonthlyHours	45345	515.0832	122.6180	95	926
TariffRate	45345	8.36964	0.5769	7.4	9.3
ElectricityBill	45345	4311.7712	1073.8864	807.5	8286.3000

B. *Proposed Methodology*

This study formulates household electricity bill prediction as a supervised regression task on the public Indian Household Electricity Bill Dataset. Each record combines dwelling descriptors (e.g., occupants, floor area, region, and month) with appliance-usage attributes (ownership, rated power, and hours of use). The target is the bill actually paid. The objective is to learn a function that maps encoded features to the bill while minimizing application-relevant loss functions, with emphasis on MAE and RMSE.

Five deep models are benchmarked under this common pipeline. A baseline MLP uses ReLU activations with optional dropout and batch normalization. A PSO-MLP variant introduces PSO to tune architectural and optimization hyperparameters (depth, width, learning rate, weight decay, and dropout) using validation loss as the fitness signal. Three recurrent baselines—GRU, LSTM, and RNN—are adapted to tabular inputs by reshaping the encoded feature vector into a short token sequence grouped by semantics (appliance, dwelling, context), with the final hidden state feeding a linear regression head. This design isolates the incremental value of a tuned feed-forward architecture against commonly used recurrent alternatives.

Training is performed using the Adam optimizer with mini-batches to ensure efficient convergence. Early stopping is applied based on validation loss to prevent overfitting, and a learning-rate reduction on plateau mechanism further accelerates convergence. The optimization objective is the Mean Squared Error (MSE) computed on the standardized target variable, which improves numerical stability during training. All evaluation metrics are computed after inverting the target scaling. Hyperparameters not tuned by the PSO [12] are set to conservative defaults that balance accuracy, generalization, and computational efficiency. The overall architecture of the proposed PSO-MLP model for electricity bill prediction is illustrated in Figure 2.

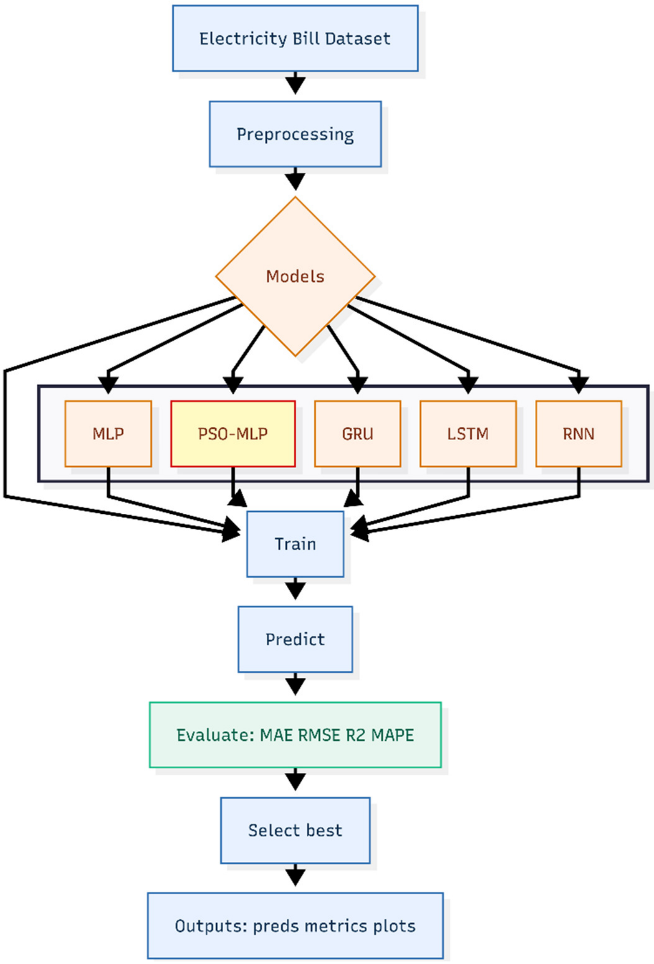


Fig. 2. Proposed PSO-MLP model for household electricity bill prediction.

III. EXPERIMENTAL SETUP AND EVALUATION

A. *Implementation Environment*

From a reproducibility standpoint, all experiments were conducted on a GPU-enabled Linux workstation running Ubuntu 22.04 LTS, equipped with an NVIDIA driver version 535+ and a compatible CUDA 12.2 and cuDNN 9.1 toolchain precisely matched to the TensorFlow build to ensure numerical consistency. The deep learning environment was configured using Python 3.11.9, with computing libraries including NumPy 1.26.4, SciPy 1.13.1, Pandas 2.2.2, and Matplotlib 3.8.4. Models were implemented with TensorFlow 2.16.1 and Keras 3.3.2, alongside scikit-learn 1.4.2 (using the sparse_output argument in OneHotEncoder) and joblib 1.3.2. Dataset access and downloads were automated via the Kaggle API 1.6.14. All runs were fully scripted with fixed random seeds and recorded package versions; the same code executes on CPU-only machines with longer training times.

B. *Evaluation Metrics*

To quantitatively assess model performance, we adopted the following standard regression metrics:

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

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \quad (3)$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

$$\mathrm{MedAE} = \mathrm{median}_i\{|y_i - \hat{y}_i|\} \quad (5)$$

where R^2 is the coefficient of determination, and MedAE is the Median Absolute Error. These metrics provide a comprehensive evaluation of model bias, variance, and robustness.

IV. RESULTS ANALYSIS

Table II compares five regressors on the test set: PSO-MLP, MLP, GRU, LSTM, and RNN. The PSO-MLP performs best across all reported metrics— $\mathrm{MAE} = 10.22$, $\mathrm{RMSE} = 12.99$, $\mathrm{MSE} \approx 168.93$, $\mathrm{R}^2 = 0.9998$, and $\mathrm{MedAE} = 8.43$ in its original currency units, the Indian Rupee (INR)—with the plain MLP a very close second ($\mathrm{MAE} = 10.29$, $\mathrm{RMSE} = 13.11$, $\mathrm{R}^2 = 0.9998$, $\mathrm{MedAE} = 8.42$). The recurrent baselines trail: GRU ($\mathrm{MAE} = 12.13$, $\mathrm{RMSE} = 15.62$), LSTM ($\mathrm{MAE} = 14.81$, $\mathrm{RMSE} = 19.10$), and the RNN performs worst ($\mathrm{MAE} = 23.03$, $\mathrm{RMSE} = 29.83$).

TABLE II. PERFORMANCE OF REGRESSION MODELS ON THE TEST SET

Model	**MAE**	**RMSE**	**MSE**	**R^2**	**MedAE**
PSO-MLP	10.22	12.99	168.92	0.9998	8.43
MLP	10.29	13.11	171.94	0.9998	8.42
GRU	12.13	15.62	244.13	0.9997	9.77
LSTM	14.81	19.10	365.01	0.9996	12.05
RNN	23.03	29.83	889.95	0.9992	18.54

All models except the RNN achieve near-deterministic fit ($\mathrm{R}^2 \approx 0.9998$), and typical errors are small ($\mathrm{MedAE} \approx 8 - 12$ in bill units). This indicates that the tabular features already capture most of the signal, and sequence models do not add value for this problem; carefully regularized feed-forward networks suffice. The PSO-MLP's margin over the plain MLP is very small ($\Delta\mathrm{MAE} \approx 0.07$), suggesting that the primary benefit of PSO is automated, principled hyperparameter tuning rather than a large accuracy improvement. For completeness, additional analyses such as paired-bootstrap confidence intervals or tail metrics (e.g., maximum absolute error, decile-wise errors) could be reported, but Table II confirms that PSO-MLP is the top performer and the simplest reliable choice for deployment.

The particle swarm searches over both architecture and optimization hyperparameters using a compact yet expressive encoding $\mathbf{h}$ (Table III). Fitness is the 5-fold cross-validated RMSE on the standardized target, ensuring that candidate settings are selected for generalization rather than training loss. Inertia decays linearly from 0.9 to 0.4, with $c_1 = c_2 = 1.7$ and a per-dimension velocity clamp at 25% of the variable range to stabilize exploration. Discrete and categorical variables are enforced through a projection step after each update, and the best particle $\mathbf{h}^*$ is finally retrained on the full training split before test evaluation.

TABLE III. PSO HYPERPARAMETER OPTIMIZATION SETUP FOR THE MLP REGRESSOR

PSO meta-parameters			
Swarm size (S)	24	Iterations (T)	30
Inertia (w)	Linear 0.9 → 0.4	Cognitive coefficient (c_1)	1.7
Social coefficient (c_2)	1.7	Velocity clamp ($\|\mathbf{v}\|_\infty$)	0.25 × variable range
Fitness	5-fold CV RMSE on standardized target (lower is better)		
Initialization	Uniform within bounds; fixed random seed		
MLP search space (particle vector h)			
Layers (L)	$[2, 5]$		
Units per layer (u_ℓ)	$[64, 512]$		
Dropout (d)	$[0.0, 0.5]$		
Weight decay (λ)	$[10^{-6}, 10^{-2}]$ (log-uniform)		
Learning rate (η)	$[10^{-4},\ 5 \times 10^{-3}]$ (log-uniform)		
Batch size (b)	$\{64, 128, 256, 512\}$		
Activation	{ReLU, GELU}		
Batch normalization	{on, off}		
Early-stopping patience	$[10, 20]$ epochs		
LR-reduce factor	$\{0.5, 0.2\}$ (patience = 5)		

Figure 3 illustrates the comparison between the actual and predicted electricity bill values obtained using the proposed PSO-MLP model, demonstrating its strong prediction alignment with real observations. In addition, Figure 4 presents the loss versus epochs curve for the PSO-MLP model, highlighting the training dynamics and showing how the loss steadily decreases and stabilizes over time, indicating effective convergence and minimal overfitting.

A recent study by authors in [13] also utilized the same publicly available Kaggle dataset for a similar objective of predicting household electricity bills. Their research primarily focused on developing a user-facing mobile application, named Sustainable EnergySense, designed to provide consumers with insights into their energy usage patterns and expected billing amounts. Unlike the present study, which focuses on quantitative model benchmarking and optimization, their work reports predictive performance in qualitative terms, as summarized in Table IV.

TABLE IV. COMPARISON OF THE PROPOSED PSO-MLP MODEL WITH PRIOR WORK

Study	**Contributions**	**Performance**
This study	PSO-MLP, MLP, GRU, LSTM, RNN	PSO-MLP MAE = 10.22, $\mathrm{R}^2 = 0.9998$
[13]	Mobile application framework with predictive feature	Predictive accuracy in qualitative terms

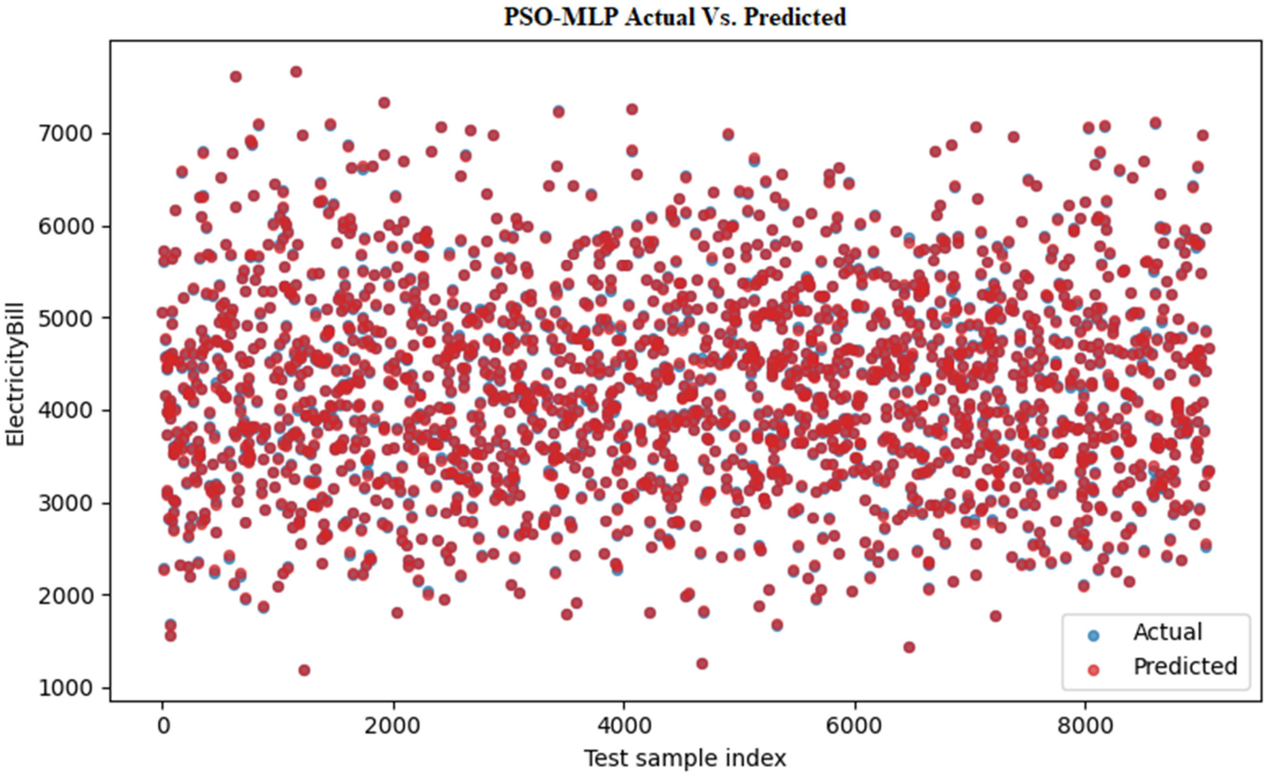


Fig. 3. Comparison between actual and predicted electricity bill values for the proposed PSO-MLP model.

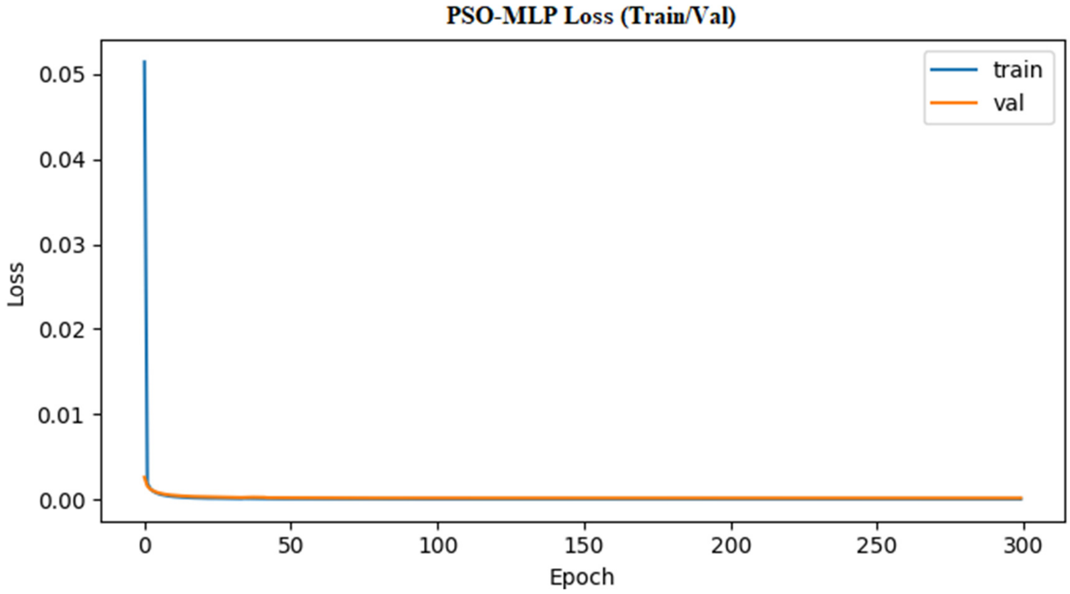
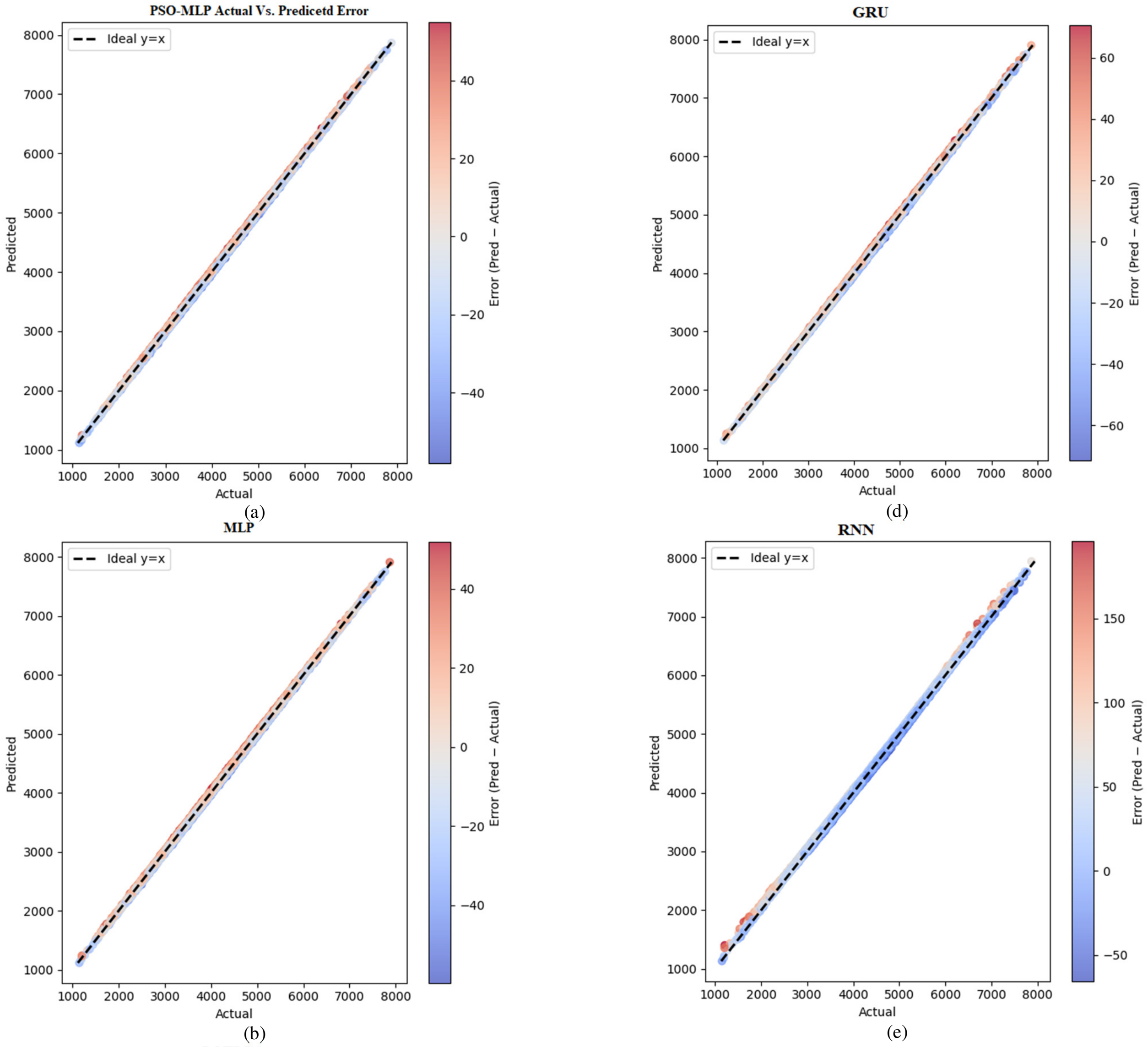


Fig. 4. Loss versus epochs for the proposed PSO-MLP model.

To provide a more comprehensive visual analysis of model performance, Figure 5 illustrates the distribution of prediction errors (predicted – actual) for all five models on the test set. The box plot clearly visualizes the performance gap between the model architectures. Both PSO-MLP and MLP exhibit very tight error distributions centered symmetrically around zero, confirming their low bias and high precision as reported in Table II. In contrast, the recurrent models show progressively wider and less stable distributions. The GRU and LSTM models have a noticeably larger interquartile range, whereas the RNN model displays the largest error variance by a significant margin. This visualization powerfully reinforces our central finding: for this tabular prediction task, a well-tuned feed-forward network is far more effective than more complex recurrent architectures.


PSO-MLP Actual Vs. Predicetd Error

(a)

MLP

(b)

GRU

(d)

RNN

(e)

LSTM

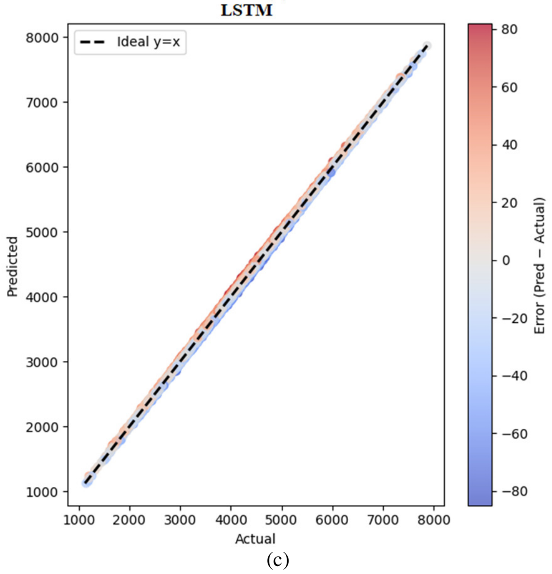

(c)

Fig. 5. Distribution of prediction errors (predicted – actual) for all models: (a) PSO-MLP, (b) MLP, (c) LSTM, (d) GRU, and (e) RNN.

V. CONCLUSIONS AND RECOMMENDATIONS

This work framed household electricity bill prediction as a supervised regression problem and benchmarked five deep models under a single, leakage-safe pipeline. Across metrics, the Particle Swarm Optimized Multilayer Perceptron (PSO-MLP) achieved the best overall performance, with Mean Absolute Error (MAE) ≈ 10.22, Root Mean Squared Error (RMSE) ≈ 12.99, Mean Squared Error (MSE) ≈ 168.9, coefficient of determination (R^2) ≈ 0.9998, and Median Absolute Error (MedAE) ≈ 8.43. These results indicate that, for tabular household attributes and appliance features, carefully regularized feed-forward networks capture the relevant nonlinearities more effectively than sequence models, delivering near-deterministic fit with low typical and median errors.

Several limitations remain. First, despite strong performance, generalization beyond the studied dataset requires validation on other utilities, regions, and tariff structures. Second, the extremely high R^2 warrants vigilance against target leakage or quasi-deterministic features; our pipeline mitigates this, but independent replication and ablation studies are essential. Third, the current models optimize point accuracy; uncertainty and fairness considerations (e.g., performance by income or usage class) were not primary objectives and should be addressed before high-stakes deployment.

Based on the findings and limitations of this study, we propose the following recommendations:

- For operational deployment, adopt Particle Swarm Optimization (PSO)-tuned Multilayer Perceptrons (MLPs) with strict train/test isolation, periodic drift monitoring, and scheduled re-training; track both aggregate and decile-wise errors to guard against slab-edge failures.
- Incorporate calibrated prediction intervals (e.g., conformal prediction) to quantify "bill-shock" risk and consider cost- or slab-aware losses to prioritize accuracy where it matters financially.
- When feasible, include tariff features explicitly (slab thresholds, fixed charges, Time-of-Day (ToD)) or use a two-stage residual approach: predict a tariff proxy and then learn residuals.
- Report paired bootstrap confidence intervals on metric deltas to substantiate improvements over strong baselines.
- For explainable artificial intelligence, evaluate subgroup performance and apply explainability techniques (e.g., permutation importance or SHAP on preprocessed features) to support consumer-facing applications.

ACKNOWLEDGMENT

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2025/R/1447).

REFERENCES

[1] P. P. K. Reddy, N. P. G. Bhavani, and J. F. Roseline, "Prediction of locational electricity marginal prices with improved accuracy by using the Lasso algorithm over linear regression algorithm," in *Applications of Mathematics in Science and Technology*, 1st ed., B. T. Hung, M. Sekar, A. ESI, and R. S. Kumar, Eds. Boca Raton, FL, USA: CRC Press, 2025, pp. 670–674.

[2] T. V. Harshavardhan and S. Loganayagi, "Household electricity bill prediction using random forest algorithm over support vector machine algorithm with improved accuracy," in *Recent Innovations in Sciences and Humanities*, 1st ed., M. Priya and V. Anandan, Eds. Boca Raton, FL, USA: CRC Press, 2025, pp. 206–212.

[3] T. Zhu *et al.*, "Electricity Bill Recovery Risk Prediction Based on Machine Learning Algorithm," *Procedia Computer Science*, vol. 262, pp. 1267–1273, Jan. 2025, https://doi.org/10.1016/j.procs.2025.05.169.

[4] P. Lazzeroni, G. Lorenti, and M. Repetto, "A Data-Driven Approach to Predict Hourly Load Profiles From Time-of-Use Electricity Bills," *IEEE Access*, vol. 11, pp. 60501–60515, 2023, https://doi.org/10.1109/ACCESS.2023.3286020.

[5] J. Han, L. Yu, A. Li, Y. Jiang, and H. Lv, "Artificial Intelligence Based Risk Prediction System for Electricity Bill Recovery," in *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics*, Ballari, India, 2024, pp. 1–5, https://doi.org/10.1109/ICDCECE60827.2024.10548630.

[6] W. N. H. W. A. Hadi, R. A. Rashid, M. A. Sarijari, S. Z. A. Hamid, and N. A. Muhammad, "Machine Learning Bill Prediction for IoT-based Utility Management System," in *2022 IEEE 6th International Symposium on Telecommunication Technologies*, Johor Bahru, Malaysia, 2022, pp. 74–78, https://doi.org/10.1109/ISTT56288.2022.9966533.

[7] H. Dineep, M. Kamalesh, A. Prasanna T, A. Kk, S. K, and J. Kannan R, "Household Energy Bill Prediction Using Various Machine Learning Techniques." Social Science Research Network, Rochester, NY, Mar. 18, 2025, https://doi.org/10.2139/ssrn.5183320.

[8] M. S. Mahmud and M. H. Chowdhury, "A Smart System for Monthly Electrical Energy Consumption Prediction Using Machine Learning," *International Journal of Information Engineering and Electronic Business*, vol. 16, no. 6, pp. 42–61, Dec. 2024, https://doi.org/10.5815/ijieeb.2024.06.04.

[9] M. Gholamnia, N. Eslamirad, P. Sajadi, S. Masoumi, H. Shahabi, and F. Pilla, "Dynamic electricity pricing model with hourly and monthly adjustments: A time series-based approach," *Energy Reports*, vol. 13, pp. 5238–5251, Jun. 2025, https://doi.org/10.1016/j.egyr.2025.04.058.

[10] S. Zairi and M. Freihat, "Electric Load Forecasting using Machine Learning for Peak Demand Management in Smart Grids," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23335–23346, Jun. 2025, https://doi.org/10.48084/etasr.10687.

[11] Suraj, "Indian Household Electricity Consumption Dataset." Kaggle. Available: https://www.kaggle.com/datasets/suraj520/indian-household-electricity-bill.

[12] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, https://doi.org/10.1007/s11831-021-09694-4.

[13] M. Al-Rajab and S. Loucif, "Sustainable EnergySense: a predictive machine learning framework for optimizing residential electricity consumption," *Discover Sustainability*, vol. 5, no. 1, Apr. 2024, Art. no. 55, https://doi.org/10.1007/s43621-024-00243-0.