Long-Term Energy Consumption Forecasting Using a Hybrid LSTM-XGBoost Approach

1st Nourhene Aouidi
ISITCOM, University of sousse
ENIM LASEE, University of Monastir
Sousse, Tunisia
nourheneaouidi@gmail.com

2nd Ferdaws Ben Naceur
LATIS, University of Sousse
Sousse, Tunisia
swadref87@hotmail.com

3th Chokri Ben Salah
Issat Sousse, University of Sousse
ENIM LASEE, University of Monastir
Sousse, Tunsia
chokribs@yahoo.fr

***Abstract*—Long-term energy consumption forecasting plays a crucial role in the effective management of smart grids , especially with the increasing integration of renewable energy sources and distributed energy resources (DERs). Accurate predictions are essential for optimizing grid operations, balancing supply and demand, and ensuring stability. This paper proposes a hybrid Long Short-Term Memory (LSTM) and XGBoost model for long-term energy consumption forecasting in smart grids. The LSTM component captures the temporal dependencies in energy consumption patterns, while the XGBoost component enhances forecasting accuracy through gradient-boosted decision trees. The hybrid model is evaluated using real-world data from smart grids, and its performance is assessed based on key forecasting accuracy metrics. Experimental results demonstrate that the hybrid LSTM-XGBoost model provides superior predictive performance, reducing forecast errors and offering a more reliable approach for long-term energy consumption predictions. This approach shows significant potential for improving energy management in smart grids, enabling better resource allocation and reducing energy consumption .**

***Index Terms*—forecasting, Long-term prediction, LSTM-XGBoost, Smart Grid.**

I. INTRODUCTION

The increasing complexity of modern power systems, driven by the integration of renewable energy sources and distributed energy resources (DERs), has underscored the need for accurate long-term energy consumption forecasting. Various AI techniques have been proposed to improve forecasting accuracy and optimize grid operations. Long Short-Term Memory (LSTM) networks, for example, are widely used in energy forecasting due to their ability to capture temporal dependencies. Xu et al. [1] introduced a Correlation Sorting-LSTM method to enhance short-term load forecasting by analyzing correlations between smart meters. Similarly, Kumar et al. [2] applied LSTM networks for demand response management in smart grids, demonstrating their computational efficiency.

In addition to LSTM, other methods like Derivative-Persistence [3] and neurocomputing [4] have been employed to improve forecasting accuracy in specific contexts, such as photovoltaic power forecasting and smart home energy management. Support Vector Regressors (SVR) and Random Forest (RF) models, combined in an ensemble approach, have also been shown to enhance regional energy demand forecasting [5].

Identify applicable funding agency here. If none, delete this.

Traditional models like ARIMA and SARIMA, while effective in short-term and seasonal load forecasting, struggle with non-linear and multivariate data, limiting their applicability in complex energy systems. In contrast, deep learning models such as Recurrent Neural Networks (RNNs) and LSTMs excel at capturing temporal patterns in energy consumption data, despite challenges like vanishing gradients. Recent hybrid deep learning architectures, such as the combination of gated recurrent units and temporal convolutional networks, have further improved model stability and reduced inference time [6]. Similarly, the PSGformer model uses multiscale decomposition with graph attention mechanisms to capture cyclical patterns and feature interdependencies [7].

Optimization algorithms like Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have been employed for load balancing and resource management, though they face challenges like convergence issues. Hybrid approaches, such as GA-PSO, combine the strengths of these algorithms for better performance in complex environments [8], [9]. Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) have also been adapted for time-series forecasting, with CNNs excelling in spatial feature extraction and DNNs offering scalability for big data applications [10]. These advancements highlight the growing trend of integrating AI techniques to address the complexities of load forecasting in modern energy systems [11].

A comprehensive long-term energy consumption forecasting solution is developed in this paper, which applies machine learning and deep learning (MDL) method for calculating daily forecast of energy consumption in smart grids. The proposed energy consumption forecast is tested and validated using historical real-world data : The primary dataset used in this research is sourced from the UK Power Networks' Low Carbon London project. Additional Data, including weather conditions such as temperature and humidity, are collected using the Dark Sky API also Economic indicators and demographic information related to the households' ACORN classification are also incorporated to enhance the predictive capabilities of the models. In this paper, we propose and apply a hybrid approach, as a single forecasting model may not provide satisfactory accuracy across all time periods and grid conditions. The proposed hybrid model combines the strengths of the LSTM network for capturing temporal dependencies and the gradient-boosted decision trees of XGBoost to enhance forecasting accuracy [17]. The rest of the paper is organized as follows: Section II describes the proposed long-term energy consumption forecasting methodology in detail. Section III presents the Results. the conclusions are drawn in Section IV.

II. Methodology

Methodology outlines the structured approach followed in the research. First, the Research approach is presented, The Data Collection Methods focus on the objectives and sources of data, providing a clear description of the data and the preprocessing steps, which include data cleaning, feature engineering, and splitting. Next, the AI Model Selection and Training Processes are discussed. This includes preliminary model exploration, the architecture of the chosen models, and the training and optimization processes necessary to refine the models for better accuracy and performance. This methodology ensures a systematic approach to building, training, and evaluating models as shown in this figure 1.

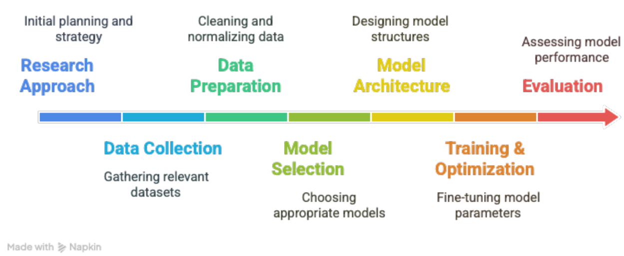

Fig. 1. Methodology Overview

A. Approach Overview

This study adopts a systematic and data-driven methodology to forecast long-term energy consumption in smart grids using a hybrid LSTM-XGBoost model. The proposed approach is designed to leverage the strengths of both deep learning and ensemble machine learning techniques: Long Short-Term Memory (LSTM) networks are used to capture temporal dependencies in sequential energy data, while the XGBoost model enhances the predictive accuracy through gradient-boosted decision trees.

The hybrid model is trained and tested using real-world energy consumption data. This includes preparing the dataset, handling missing values, scaling features, and constructing input sequences suitable for the LSTM model. The output from LSTM is then used as input for the XGBoost model to refine the predictions. Evaluation is performed using standard forecasting metrics such as MAE, RMSE allowing for a comprehensive assessment of model performance.

Overall, this approach is tailored to improve the reliability of long-term forecasts and support decision-making processes in smart grid energy management.

B. Dataset Description

1) Data Sources:: **Primary Data Source:** The primary dataset used in this research is sourced from the UK Power Networks' project. This dataset contains energy consumption readings from 5,567 London households between November 2011 and February 2013. The data reflects the rollout of smart meters, which were part of a broader European initiative to modernize energy systems and tackle climate change.
Additional Data: Supplementary data, including weather conditions such as temperature and humidity, are collected using the Dark Sky API. Economic indicators and demographic information related to the households' ACORN classification are also incorporated to enhance the predictive capabilities of the models.

C. Data Preparation

1) Data Loading: The energy consumption data was provided in multiple CSV files, with each file containing a portion of the data for various households over specific time periods.

2) Normalization Across Households: To account for varying daily household reporting, energy consumption was normalized by the number of reporting households. For each day d, the total energy consumption $E_{\text{sum},d} = \sum_{i=1}^{N_d} E_{i,d}$ was calculated, where N_d is the number of unique households (LCLid) reporting, and $E_{i,d}$ is the energy consumption of household i. The average energy consumption per household was then computed as $E_{\text{avg},d} = \frac{E_{\text{sum},d}}{N_d}$. This normalization ensures accurate per-household energy consumption despite daily fluctuations in reporting.

3) Weather Data Integration: Weather data, collected via the Dark Sky API, was merged with the energy consumption data to explore the relationship between weather conditions and energy usage.

- Loading Weather Data: The weather data included daily maximum and minimum temperatures, humidity, wind speed, and other atmospheric variables.
- Merging with Energy Data: The weather data was merged with the energy data on the day column.

4) Feature Engineering: To extract meaningful insights from the data, feature engineering was performed, including:

- Weather Clustering: Gaussian Mixture Model (GMM) clustering was used to categorize the weather data into distinct weather clusters as shown in figure 2 . These clusters were added as a new categorical variable to the dataset.
  - This scatter plot represents the Gaussian Mixture Model (GMM) clustering applied to the PCA components (PC1 and PC2).
  - The three clusters (purple, teal, and yellow) represent different regimes of weather patterns and their corresponding energy consumption levels.
  - The purple cluster seems to represent lower-energy consumption, while the yellow cluster represents higher energy consumption, with the teal cluster in between.

– The clustering appears relatively well-separated, meaning the GMM can distinguish between different weather-based energy consumption regimes.

- Bank Holiday Indicator: A bank holiday indicator was added to the dataset by merging with a list of UK bank holidays, with 1 indicating a holiday and 0 otherwise.

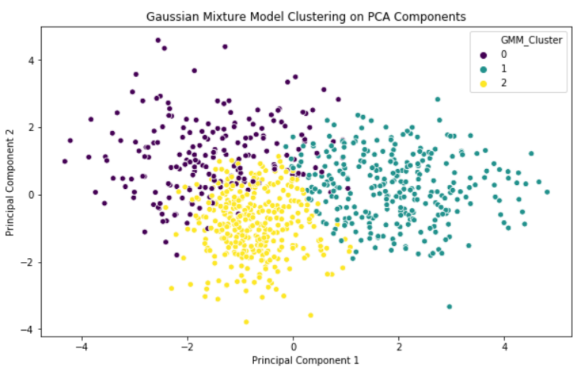

Fig. 2. Gaussian Mixture Model weather Clustering on PCA Components

5) Data Cleaning: Data cleaning is a critical step to ensure the reliability of the dataset before modeling. Outliers were identified using statistical methods such as the Z-score as shown in figure 3, which highlighted a few extreme values, particularly in the energy consumption data. These abnormal values, which may result from errors, unusual usage patterns, or sensor malfunctions, were either capped or removed depending on their severity. For weather-related features, robust scaling techniques were applied to mitigate the influence of extreme values while preserving the data structure. Missing values, commonly present in real-world time series data, were addressed differently depending on the dataset: energy consumption gaps were filled using the mean or median values based on similar days (e.g., weekdays or weekends), while weather-related gaps were filled using forward-filling for short sequences and linear interpolation for longer ones. Duplicate records were also removed by ensuring the uniqueness of each household's energy usage per day, eliminating any redundant entries that could skew the analysis.

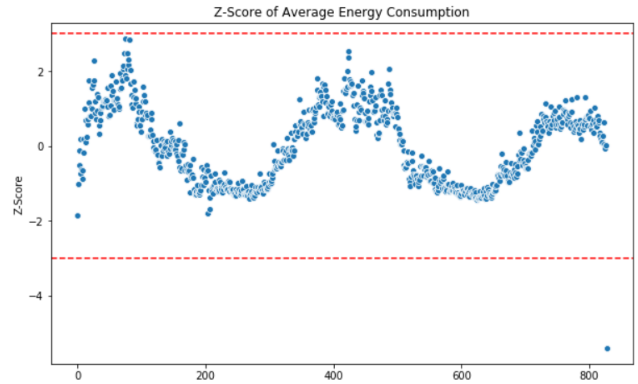

Fig. 3. Z-Score of Average Energy Consumption

6) Data Scaling: Data scaling is crucial for machine learning models like LSTM, which are sensitive to feature scales. Unscaled features with different units can skew model performance. We apply MinMax Scaling to transform features to a [0, 1] range using $x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$, where $x_{\min}$ and $x_{\max}$ are the feature's minimum and maximum values from the training data. This scaling, applied to both training and test data, prevents data leakage and ensures equal feature contribution, improving convergence and prediction accuracy in models like LSTM.

7) Data Splitting: We typically split the dataset into three key parts , we have energy consumption data from 2011 to 2014.

- **1-Training Set:** Used to train the model. Typically, this is the largest portion of the data, containing 60% of the dataset.
- **2-Validation Set:** This set is used to tune the model's hyperparameters and prevent overfitting. Around 20% of the dataset.
- **3-Test Set**: This is used to evaluate the final model performance on unseen data. It comprises 20% of the dataset.

D. Model Selection

Energy consumption data is highly sequential, with strong temporal dependencies influenced by factors such as time of day, seasonality, and external conditions like weather. Additionally, the data contains complex non-linear patterns that are not solely dependent on time, but also on external features such as temperature, humidity, and human behavior. This complexity led to the selection of a hybrid approach, combining LSTM and XGBoost.

1) Long Short-Term Memory (LSTM): LSTM, a Recurrent Neural Network variant, is designed for time series forecasting, capturing temporal dynamics via memory cells [15]. In this hybrid model, LSTM predicts future energy consumption from past patterns. As shown in Figure 4, each LSTM cell uses gates to manage information: the forget gate ($f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$) discards irrelevant data, the input gate ($i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$) and candidate cell state ($\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$) update the cell state ($C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$), and the output gate ($o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$) produces the hidden state ($h_t = o_t \cdot \tanh(C_t)$). This structure enables LSTM to retain long-term dependencies, enhancing prediction accuracy.

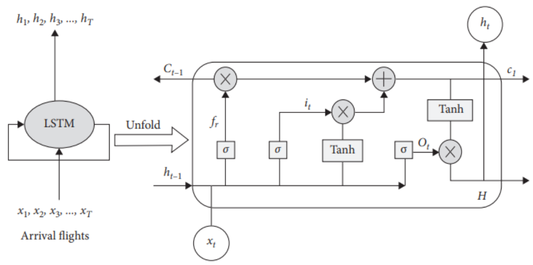

Fig. 4. The internal structure of an LSTM cell

2) XGBoost (Extreme Gradient Boosting)::

- **XGBoost's Role:** In this hybrid model, XGBoost is used to model the residuals from the LSTM predictions. By training XGBoost on the errors made by LSTM, the model can refine the predictions and enhance the overall accuracy of the system [16].

3) Extreme Gradient Boosting (XGBoost): XGBoost, an efficient gradient-boosted decision tree algorithm, builds an ensemble of trees sequentially, with each tree correcting residuals from prior trees to capture linear and non-linear patterns. It optimizes the objective function Obj $= \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$, where L is the loss function and Ω includes $L1$ and $L2$ regularization to prevent overfitting. A learning rate (η) shrinks tree contributions, enhancing model stability. In the hybrid model, LSTM captures temporal energy consumption trends, while XGBoost refines predictions by learning from residuals, improving forecast accuracy.

E. Hybrid Model Architecture

1) LSTM Architecture: The LSTM model, designed for sequential energy consumption data, comprises two LSTM layers to capture short- and long-term patterns. It takes 30 timesteps of past daily consumption as input. The first LSTM layer (50 units) processes the sequence, followed by a 20% dropout layer to prevent overfitting. The second LSTM layer (50 units) outputs a fixed-size vector, with another 20% dropout layer. A dense layer with one neuron predicts the next day's consumption. The model is trained using Mean Squared Error loss and the Adam optimizer for efficient regression.

2) XGBoost Architecture:: The XGBoost model operates in parallel with LSTM, focusing on modeling the residuals— the errors left by the LSTM predictions. By training XGBoost on these residuals, it corrects the LSTM's mistakes and provides a more accurate forecast. Input to XGBoost: The input to XGBoost is the same historical energy consumption data used by LSTM, but it is reshaped into a flat array. XGBoost does not operate on sequences like LSTM; instead, it treats each day's data as an independent observation.

3) Hybrid Model Approach: The hybrid model integrates LSTM and XGBoost for enhanced energy consumption forecasting. LSTM generates initial predictions from past temporal data, while XGBoost refines these by predicting residuals (errors). The final forecast combines LSTM's output with XGBoost's corrections, leveraging LSTM's temporal dependency capture and XGBoost's non-linear pattern recognition for improved accuracy. Hyperparameters are detailed in Table I.

TABLE I
SELECTION OF HYPERPARAMETERS FOR THE HYBRID LSTM + XGBOOST MODEL

Model	Hyperparameter	Value
LSTM	Number of hidden layers	2
LSTM	Neurons in each hidden layer	50
LSTM	Dropout rate	0.2
LSTM	Timesteps per sequence	30
LSTM	Batch size	32
LSTM	Optimizer	Adam
LSTM	Epochs	50
XGBoost	Number of estimators	100
XGBoost	Learning rate	0.1
XGBoost	Maximum tree depth	3

F. Hybrid Model Equations

The hybrid model combines LSTM and XGBoost to improve energy consumption prediction:

- **LSTM Prediction:**

$$\hat{y}_{\text{LSTM}}(t+1) = \text{LSTM}(X_t)$$

- **Residual Calculation:**

$$r_t = y_t - \hat{y}_{\text{LSTM}}(t)$$

- **XGBoost Prediction on Residuals:**

$$\hat{r}_{\text{XGB}}(t+1) = \text{XGBoost}(X_t)$$

- **Final Hybrid Prediction:**

$$\hat{y}_{\text{Hybrid}}(t+1) = \hat{y}_{\text{LSTM}}(t+1) + \hat{r}_{\text{XGB}}(t+1)$$

G. Training and Optimization

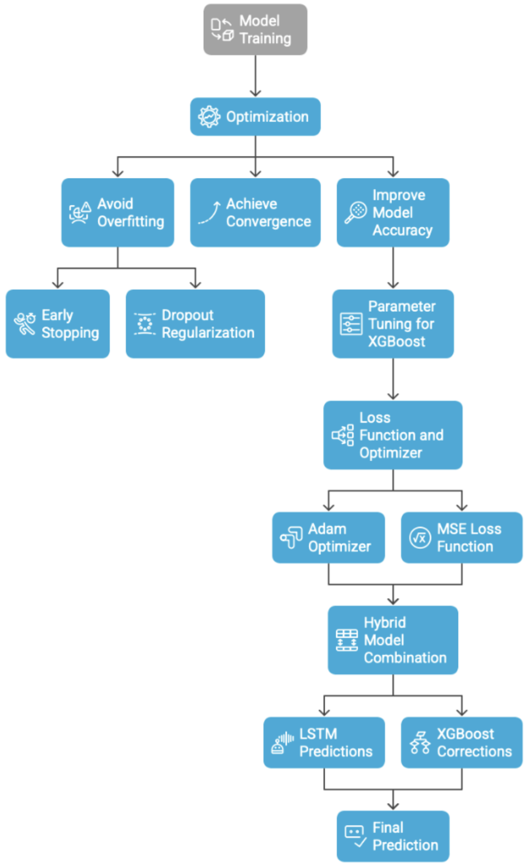

Fig. 5. Model Training and Optimization

The hybrid LSTM + XGBoost model was trained in two stages, with each stage undergoing optimization to ensure the model performed well on unseen data. The training and optimization strategies focused on avoiding overfitting, achieving convergence, and improving model accuracy through techniques such as early stopping, dropout regularization, and parameter tuning for XGBoost as shown in figure 5.

a) Loss Function and Optimizer: The model was compiled using the Adam optimizer and the Mean Squared Error (MSE) loss function to minimize the difference between predicted and actual values.

b) Hybrid Model Combination:: The hybrid model combines LSTM predictions with XGBoost corrections to improve overall prediction accuracy.

c) Final Prediction:: The final prediction is the sum of the LSTM output and the residual predicted by XGBoost, enhancing both temporal and non-linear modeling.

III. Results

A. Discussion of Metrics

The evaluation metrics used in this study include RMSE, MAE, and R^2. Below is a brief explanation of each metric:

R^2 (Coefficient of Determination) R^2 measures the proportion of variance in the dependent variable explained by the model. A value close to 1 indicates excellent performance:

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

RMSE (Root Mean Squared Error) RMSE measures the average magnitude of errors in the predictions, expressed in the same units as the target variable:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

MAE (Mean Absolute Error) MAE measures the average absolute difference between actual and predicted values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

As shown in Figure 6, the hybrid LSTM + XGBoost model predicted the energy consumption for the last 160 days with reasonable accuracy, closely following the true values. This demonstrates the model's robustness and capacity to generalize on unseen data.

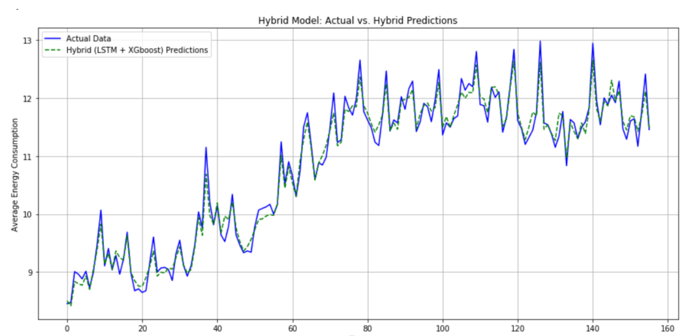

Fig. 6. LSTM-XGBoost Model: Test Set Predictions

- **LSTM-XGBoost 160 days Forecast RMSE**: 0.4339
- **LSTM-XGBoost 160 days Forecast MAE**: 0.3301
- **LSTM-XGBoost 160 days Forecast R²**: 0.9145

The graph 7 shows the residuals of the hybrid model (LSTM + XGBoost), representing the difference between actual and predicted values. Ideally, residuals should be randomly distributed around zero, indicating minimal, uncorrelated errors. Here, the residuals fluctuate around zero without a clear pattern, and their small magnitude suggests the model captures the data structure well with high accuracy and no significant bias. Thus, the hybrid model is effective and well-calibrated.

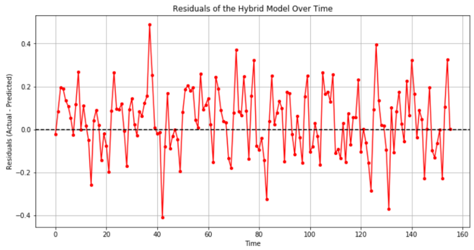

Fig. 7. Residuals in Megawatts of the Hybrid Model Over Time(days)

TABLE II: Comparative Performance of Forecasting Models

Model / Article Title	RMSE	MAE
Our Hybrid LSTM + XGBoost (160-Day)	**0.4339**	**0.3301**
Urban Daily Water Supply Forecasting Based on the LSTM-AM Model [12]	0.4939	N/A
Short-Term Load Forecasting Based on MI and Bi-LSTM [13]	0.5124	0.3912
Studies of Short-term Load Forecasting Model Based on LSTM-NBEATS [14]	0.4460	N/A

IV. Conclusion

This paper introduced a hybrid LSTM-XGBoost model for long-term energy consumption forecasting in smart grids, integrating sequential learning and gradient-boosted decision trees. The approach captures temporal dependencies and incorporates external factors like weather, enhancing predictive accuracy. Experimental results over a 160-day horizon show superior performance, with an RMSE of 0.4339, MAE of 0.3301, and R^2 of 0.9145, confirming its effectiveness for optimizing smart grid operations. This model supports utility companies in planning sustainable energy infrastructure. Its scalability offers potential for broader smart grid applications. The work advances sustainable energy management, a critical need in modern grids.

References

[1] C. Xu, W. Li, M. Yu, J. Xu, J. Liu, Y. Wang, and L. Zhu, “A correlation sorting-LSTM method for high accuracy short-term load forecasting based on smart meter data,” in *Proc. 7th Int. Conf. (Specific Conference Name)*, 2020, pp. 1–6, IEEE.

[2] J. Kumar and A. K. Singh, “A demand and response management model using load forecasting technique for smart grid,” in *Proc. 5th Int. Conf. (Specific Conference Name)*, 2023, pp. 1–6, IEEE.

[3] M. Bozorg, M. Carpita, P. De Falco, D. Lauria, F. Mottola, and D. Proto, “A derivative-persistence method for real time photovoltaic power forecasting,” in *Proc. Int. Conf. (Specific Conference Name)*, 2020, pp. 1–6, IEEE.

[4] Y.-H. Lin, H.-S. Tang, and T.-Y. Shen, “A smart home energy management system utilizing neurocomputing-based time-series load modeling and forecasting facilitated by energy decomposition for smart home automation,” *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2022, doi: 10.1109/ACCESS.2022.3219068.

[5] A. Muzumdar, C. Modi, and C. Vyjayanthi, “An efficient regional short-term load forecasting model for smart grid energy management,” in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, 2020, pp. 1–6, IEEE.

[6] W. Xu *et al.*, “Deep learning-driven hybrid model for short-term load forecasting and smart grid information management,” *Dental Sci. Rep.*, vol. 14, no. 1, Jun. 2024. [Online].

[7] Y. Zhang *et al.*, “PSGformer: A multiscale decomposition with graph attention mechanism for long-term energy forecasting,” in *Proc. Int. Conf. (Specific Conference Name)*, 2020, pp. 1–6, IEEE.

[8] M. Liu, C. Mo, and H. Wang, “Short-term power load forecasting model based on CNN-GRU and dual attention mechanism hybrid neural network model,” in *Proc. 8th Asia Conf. Power Elect. Eng. (ACPEE)*, 2023, pp. 1–6, IEEE.

[9] E. Vrettos and C. Gehbauer, “A hybrid approach for short-term PV power forecasting in predictive control applications,” in *Proc. IEEE Milan PowerTech*, 2019, pp. 1–6, IEEE.

[10] W. Wu, K. Chen, Y. Qiao, and Z. Lu, “Probabilistic short-term wind power forecasting based on deep neural networks,” in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, 2016, pp. 1–6, IEEE.

[11] “Demand-side load forecasting in smart grids using machine learning techniques,” *PeerJ*, vol. 10, pp. e1987–e1987, Apr. 2024. [Online].

[12] “Urban daily water supply forecasting based on the LSTM–AM model,” *2023 9th International Conference on Big Data and Information Analytics (BigDIA)*, Hengyang, China, 2023, IEEE. [Online]. Available: https://doi.org/10.1109/BigDIA60676.2023.10429385

[13] “Short-term load forecasting based on MI and BI-LSTM,” *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, IEEE. [Online]. Available: https://doi.org/10.1109/ICTC49870.2020.9289411

[14] “Studies of short-term load forecasting model based on LSTM-NBEATS,” *2022 IEEE International Conference on Data Science and Engineering (ICDSE)*, Wuhan, China, 2022, IEEE. [Online]. Available: https://doi.org/10.1109/ICDSE54319.2022.9848226

[15] J. Wu, R. Lin, M. Pan, B. Wang, O. Wang, and Q. Zhao, “Load forecasting method considering weather factors based on LSTM and XGBoost,” in *Proc. IEEE Int. Conf.*, 2024, pp. 1–6.

[16] Y. Wang and Y. Guo, “Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost,” *China Communications*, vol. 17, no. 3, pp. 260–272, 2020. IEEE.

[17] J. Song, L. Jin, Y. Xie, and C. Wei, “Optimized XGBoost based sparrow search algorithm for short-term load forecasting,” in *Proc. IEEE Int. Conf. Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)*.

[18] V. V. S. Lanka, M. Roy, S. Suman, and S. Prajapati, “Renewable energy and demand forecasting in an integrated smart grid,” in *Proc. 2021 Innovations in Energy Management and Renewable Resources (52042)*.

[19] L. A. Takara, L. F. R. Agottani, C. V. Züge, W. R. de Souza Chaves, L. dos Santos Coelho, and V. C. Mariani, “Deep learning-based probabilistic forecasting of household energy consumption in smart grids,” in *Proc. 2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*.

[20] M. Sakib, S. Mustajab, K. Anwar, and M. Alam, “BiLSTM based approach for energy demand forecasting and synchronization among smart grids,” in *Proc. 2024 IEEE Third Int. Conf. Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*.