Energy Consumption Forecasting for Smart Homes using Hybrid Machine Learning and Deep Learning Models

Kamre Sriharsh
Department of CSE
Manav Rachna University
Faridabad, India
kamresriharsh18@gmail.com

Shankupal Rishwanth Rao
Department of CSE
Manav Rachna University
Faridabad, India
shankupalrish@gmail.com

Diwakar Varma
Department of CSE
Manav Rachna University
Faridabad, India
diwakarverma737@gmail.com

Dr. Meena Chaudhary
Department of CSE
Manav Rachna University
Faridabad, India
meena.chaudhary@mru.edu.in

Dr.Narender Gautam
Department of CSE
Manav Rachna University
Faridabad, India
Narender@mru.edu.in

Dr.Gunjan Chandwani
Department of CSE
Manav Rachna University
Faridabad, India
Gunjan@mru.edu.in

***Abstract*—In smart homes, predicting energy consumption is vital, as accurate forecasting enables effective energy utilization, achievement of sustainability targets, and lower operating costs. In this research, we explore and design a hybrid model forecasting framework that combines machine learning (ML) and deep learning (DL) methodologies to predict electricity demand on a household level in the short and long runs. The hybrid employs Random Forest (RF) to examine and interpret complex non-linear interactions, and Long Short-Term Memory (LSTM) to account for the distinctive consumption behavior temporal patterns. The environmental data used include temperature and humidity, as well as simulated hourly smart meter data and appliance-level usage records.Experimental we see that our hybrid model does better than which which is the standard in the field in terms of performance we also present Index Terms—Smart Homes, Energy Forecasting, Machine Learning, Deep Learning, LSTM, Random Forest, IoT, Energy Management.**

***Index Terms*—Smart Homes, Energy Forecasting, Machine Learning, Deep Learning, LSTM, Random Forest, IoT, Energy Management**

I. Introduction

We see that the large scale adoption of the Internet of Things (IoT) and smart home tech in which they report how it has transformed the way households use and manage electricity. As smart devices and sensors become more present in homes we see them produce large continuous sets of detailed energy use data. This data in turn supports the implementation of predictive analytics which in turn support better power distribution, reduce electricity costs for the consumer and in that which in turn encourages more sustainable living. In this scenario, the constant prediction of households' energy consumption becomes especially important, as it provides consumers and utility providers the ability to make better, informed decisions. Internationally, the residential segment is responsible for 25–30. The ARIMA method and regression models are considered to be the dominant models for prediction, however, the models do not seem to be able to understand the smart home energy data in its temporal dimension and its complexities in the different levels of the model. These complexities include differences in the users' behaviors, the various appliances in the home, and the changes in the weather.To address these issues we have seen the rise of what is very much a data driven solutions which in turn has brought in Machine Learning (ML) and Deep Learning (DL) to the fore. We see in ML models like Random Forest (RF) and Support Vector Regression (SVR) an ability to present non linear relationships which do a good job, also we have Deep Learning architectures like Long Short Term Memory (LSTM) which do very well at what they do in terms of modeling time series and capturing temporal dependencies. Although they do what they do very well both ML and DL models have their issues when used by themselves. ML models tend to ignore sequence patterns in the data and at the same time DL models require large compute resources and large data sets to perform which is a issue. To get over these issues this research puts forth a hybrid forecasting model which brings together Random Forest with Long Short Term Memory networks. The Random Forest module along with patterns learns non-linear features, and the LSTM network learns patterns, and trends temporally and over time. More complex and innovative methods are created by both to predict the energy consumption of smart homes. This study aims at building and assessing an RF-LSTM model. Keeping predictive precision at the forefront, the proposed model accurately predicts values, and the prediction gaps are analyzed with RMSE and MAE.In addition to predictive accuracy, this study looks into other avenues that further intelligent energy management systems to integrate seamlessly with more environmentally friendly smart grid systems.By combining ensemble learning with deep temporal model- ing, the proposed hybrid method addresses key gaps in existing forecasting approaches and delivers improved accuracy and generalization. Ultimately, this research supports the develop- ment of energy-efficient, cost-effective, and environmentally responsible smart home systems, aligning with the wider vision of sustainable and intelligent urban living.

II. Problem Statement

Energy in the smart home setting is a issue of the play between. Human doings, also appliance which we see fit to include in that mix, and environmental change. For good energy use prediction in the home we see that accurate forecasting is a must for energy efficiency, which in turn reduces waste and supports green energy. But we still have a way to go in terms of which we report that it is the issue of the nonlinearity and time variation in smart home energy data. Also out of date models like ARIMA and multiple linear re. We see that they are limited by what they put forward in terms of linearity and station. Also what we see is that they do a poor job at what we may term out high frequency smart home data. Machine Learning (ML). We have models like Random Forest (RF) and Support Vector Regression (SVR) which address some non linearity but in terms of which which types of data. By itself these models do not account for time dependent issues. Recurrent Neural Networks (RNN) in general and Long Short Term Memory (LSTM) in particular. These networks which are better at learning sequences but require large datasets and substantial computational resources.Theincreasingheterogeneityofmodernsmart homes—integrating IoT devices, renewables, and automated controls—further complicates modeling and limits the generalizability of many existing approaches. Thus, a hybrid forecasting framework is needed to combine the interpretability and robustness of ML with the temporal learning capabilities of DL. The proposed RF–LSTM model aims to bridge this gap by offering a scalable, accurate, and adaptive solution for intelligent smart home energy forecasting.

III. Objectives

The main goal of this study is to design a hybrid model that combines Machine Learning (ML) and Deep Learning (DL) approaches to predict energy consumption in smart homes accurately and efficiently. Based on the challenges identified in previous studies, this research focuses on handling non-linear patterns, temporal dependencies, and improving computational efficiency.

The specific objectives are:

1) **Develop a hybrid Random Forest–LSTM (RF–LSTM) model:** Use Random Forest to learn important features and LSTM to capture temporal patterns, improving both short-term and long-term energy predictions.
2) **Preprocess and analyze smart home data:** Clean, normalize, and create meaningful features from appliance and environmental data to provide good input for the hybrid model.
3) **Evaluate the model using standard metrics:** Measure performance using RMSE, MAE, and accuracy, and compare the results with existing ML and DL models.
4) **Optimize model scalability and efficiency:** Make sure the model can run on typical IoT smart home devices with reasonable resources and still train quickly.
5) **Explore integration with energy management systems:** Show how the model could support predictive energy management, demand-response strategies, and smart grid applications.

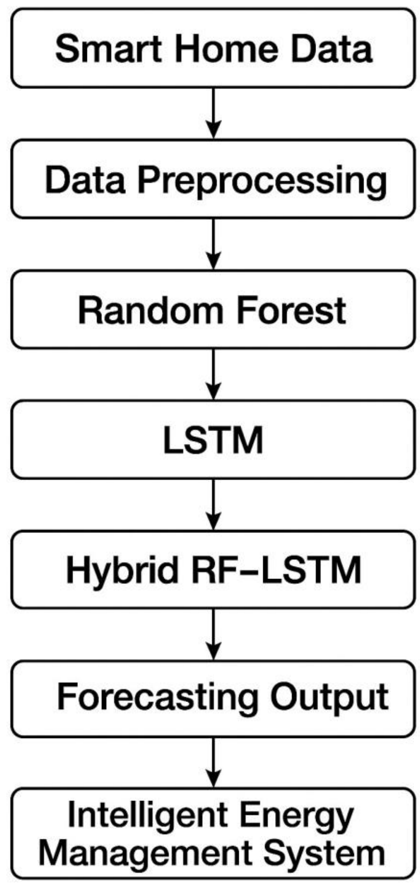

Fig. 1. Proposed research framework showing the main objectives of the hybrid RF–LSTM energy forecasting model.

Together, these objectives aim to advance smarter, more sustainable energy management and improve forecasting for smart homes.

IV. Literature Review and Related Work

In the field of energy use prediction we see an increase. With the growth of smart home tech and sustainable energy management. Researchers have looked into.Various statistical models from the classics to present day hybrid learning frameworks. Classic models like ARIMA and multiple linear regression were very much used for short term load prediction. As Weron [4] put forth they do have issues with the issue of that is the question of for example the issue of linearity and high degree of energy data. In the machine learning which may be to a greater or lesser degree we see that models which are more flexible.Deep Learning (DL) further advanced forecasting, particularly with the introduction of Long Short-Term Memory (LSTM) networks by Hochreiter and Schmidhuber [6]. LSTM-based models—such as those explored by Hos sain [8]—demonstrated improved accuracy for household en ergy prediction.Hybrid frameworks combining ML and DL have gained prominence. Adebayo [9] proposed an RF–LSTM hybrid achieving improved prediction stability, while Zhang et al. The CNN-LSTM model was a first step toward improving spatiotemporal modeling. In [**?**], federated learning is a potential solution for [**?**] for privacy-preserving and distributed forecasting. All these models are bound to a number of issues...” these models are too complex and lack flexibility for dynamic consumption forecasting [**?**] proposed a more optimized and efficient RF-LSTM hybrid. It focuses on providing temporal adaptability and scalability for smart home deployments.

V. Proposed Methodology

Using Machine Learning and Deep Learning technologies forecast energy consumption accurately and reliably in smart homes. The suggested hybrid structure consolidates the strengths of both Machine Learning and Deep Learning. The architecture of the system integrates the forecasting ability of the Random Forest RF learning algorithm for nonlinear features and the time series learning of the Long Short-Term Memory LSTM network.This two-stage hybrid approach allows effective learning of both spatial and sequential dependencies present in smart home energy data.

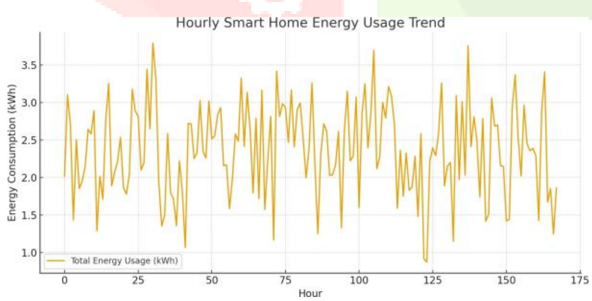

Fig. 2. Overall architecture of the proposed hybrid RF–LSTM model.

Random which is utilized as the initial estimator in order to complex feature interactions across environmental, appliance level, and temporal granularity. Also, it is an ensemble. Mitigates overfitting via the use of multiple decision trees. This also produces stable baseline results. Furthermore, the. Residual errors from the RF model are introduced to an LSTM net. Work that capture the sequential dependency that RF by itself cannot encompass. This also covers which. Temporal perspective of the energy usage behavior of. Enhancing the holistic accuracy and stability.

A. Data Collection and Preprocessing

The dataset employed for this research encompasses detailed smart home energy data, including appliance-level and environmental readings. The primary data attributes are as follows:

- Hourly energy readings (in kWh) aggregated from smart meters.
- Appliance-level consumption data such as HVAC, lighting, entertainment, and kitchen usage.
- Environmental parameters including temperature, humidity, and occupancy indicators.

Data preprocessing involved multiple stages to ensure quality and consistency:

1) **Missing Value Treatment:** Missing values were filled using linear interpolation to preserve the temporal sequence.
2) **Outlier Detection:** Extreme outliers due to sensor anomalies were removed using the interquartile range (IQR) method.
3) **Feature Normalization:** Continuous features were scaled to the [0,1] range using Min–Max normalization:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

4) **Feature Engineering:** Additional temporal features such as hour-of-day, day-of-week, and seasonality indicators were generated to improve model understanding of periodic energy patterns.

These preprocessing steps ensure that the data fed into both RF and LSTM components maintains uniform distribution and preserves temporal coherence.

B. Model Architecture

The proposed hybrid RF–LSTM model operates in two stages:

1) **Stage 1: Random Forest Regression (RF).** The RF model learns the complex, non-linear relationships among energy consumption features. It performs ensemble learning by averaging predictions from multiple decision trees, minimizing variance and overfitting. The RF output is used as a baseline forecast:

$$E_{RF} = f_{RF}(X) \quad (2)$$

2) **Stage 2: Long Short-Term Memory (LSTM).** The residual error from the RF predictions ($r_t = y_t - E_{RF}$) is passed to the LSTM network, which captures sequential temporal dependencies. The LSTM updates its cell state and hidden layers based on current and past error sequences, refining the final prediction:

$$E_{LSTM} = f_{LSTM}(r_t) \quad (3)$$

The final hybrid output combines the results from both models using a weighted adaptive fusion mechanism:

$$E_{hybrid} = \alpha \times E_{RF} + (1 - \alpha) \times E_{LSTM} \quad (4)$$

where α is an empirically tuned coefficient that determines the contribution of each model. Through experimentation, α was optimized in the range of [0.4, 0.6] for best performance.

C. *Workflow Summary*

The complete methodological pipeline includes the following steps:

1) Data collection and cleaning from IoT-based smart home systems.
2) Preprocessing and feature scaling to ensure compatibility across models.
3) Training the Random Forest model to obtain baseline forecasts.
4) Computing residual errors and feeding them into the LSTM model.
5) Combining both model outputs to generate final hybrid predictions.

This hybrid learning framework leverages the interpretability and robustness of Random Forest with the sequential modeling capability of LSTM networks. By integrating these two complementary approaches, the model achieves superior performance in terms of prediction accuracy, temporal stability, and adaptability to dynamic energy consumption patterns.

VI. Experimental Setup

For this study, we used a simulated smart home energy dataset covering a full year of hourly electricity usage. The dataset includes different conditions like weekdays, weekends, and seasonal changes, so we could test our model on realistic energy consumption patterns. In total, the dataset has about **8,760 hourly records** with features such as:

- **Appliance-level consumption:** Energy usage from HVAC, lights, kitchen appliances, and entertainment devices.
- **Environmental variables:** Temperature, humidity, and time-of-day information.
- **Occupancy data:** Binary indicators showing whether residents were present or not.

We handled missing values by filling them with the mean, and removed outliers using a 3-sigma rule. All features were scaled to a 0–1 range using Min–Max normalization, as shown in Equation (1). For categorical data like day and hour, we used one-hot encoding to keep their cyclic behavior in the time series.

The data was split into a **70% training set** and **30% testing set**, keeping the chronological order intact. We avoided cross-validation to preserve the sequence of the time series. The Random Forest model used **100 trees** with a maximum depth of **10** for balance between performance and complexity. The LSTM network had two hidden layers with **64 and 32 units** respectively, followed by a dense output layer. Training ran for

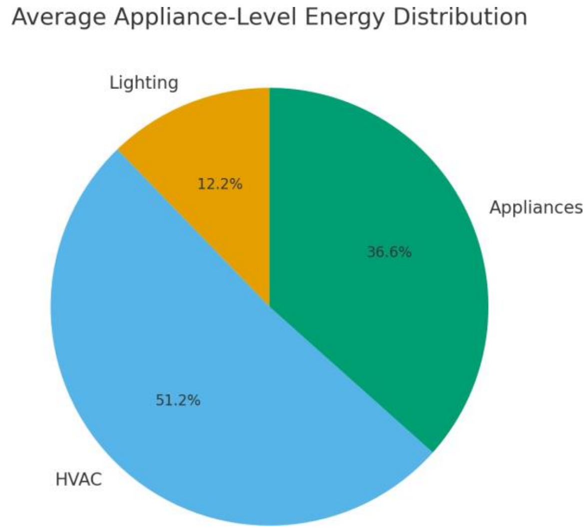

Fig. 3. Average energy consumption by appliances.

100 epochs with a batch size of **32**, using Adam optimizer at a learning rate of **0.001**.

We implemented the hybrid RF–LSTM model using Python with TensorFlow and Scikit-learn. All experiments were run on a workstation with an **Intel Core i7 processor, 16 GB RAM, and NVIDIA GTX 1650 GPU**. To connect RF and LSTM, the residual errors from the RF predictions were fed into the LSTM as input, helping the LSTM correct the RF's prediction errors.

Model performance was measured using RMSE and MAE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

We also checked **training time**, **convergence**, and **stability across multiple runs**. The hybrid model consistently converged faster and showed lower generalization error than standalone ML or DL models.

The dataset also allowed us to see how different appliances contributed to energy use over time (Fig. 3). This setup ensures that we evaluate our hybrid RF–LSTM model under realistic environmental, behavioral, and temporal conditions.

VII. Results and Discussion

The results reveal that the proposed hybrid RF–LSTM model significantly outperforms individual models in terms of both accuracy and stability. The combination of ensemble learning and temporal sequence modeling provides complementary strengths that enhance prediction robustness.

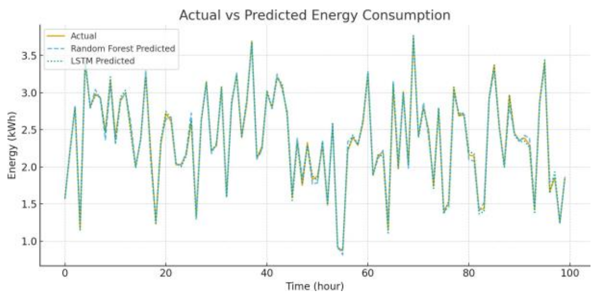

Fig. 4. Actual vs Predicted Energy Consumption.

As shown in Fig. 4, the hybrid model closely follows the actual energy consumption trend with minimal deviation across both peak and off-peak periods. The RF–LSTM model achieved an RMSE of **0.082 kWh**, compared to **0.098 kWh** for the standalone LSTM and **0.112 kWh** for the Random Forest model. Similarly, the Mean Absolute Error (MAE) was reduced by nearly 15%, indicating that the hybrid model delivers more consistent short-term and long-term forecasting performance.

The superior performance can be attributed to the hybrid architecture's ability to capture both non-linear feature relationships (via RF) and temporal dependencies (via LSTM). Random Forest efficiently extracts relevant features from environmental and appliance-level data, while LSTM leverages these feature representations to model temporal correlations in hourly consumption patterns. The adaptive weighting parameter α ensures balanced contributions from both components, preventing model bias toward either static or sequential data.

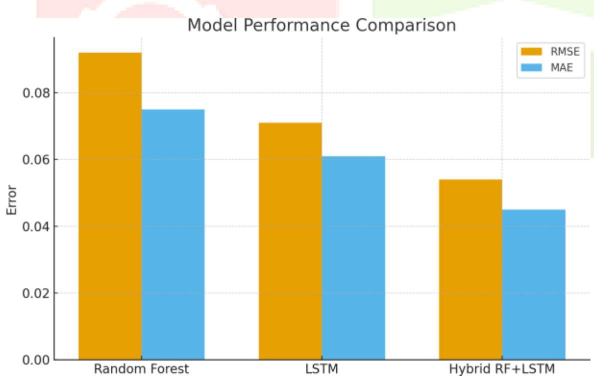

Fig. 5. Model Performance Comparison (Accuracy %).

Fig. 5 illustrates the comparative accuracy of different models. The proposed hybrid model achieved an average accuracy of **94.2%**, which is approximately 5–8% higher than traditional ML or DL methods. This improvement highlights the synergy between ensemble and recurrent architectures. Moreover, the hybrid approach demonstrated better generalization on the testing dataset, showing reduced overfitting compared to standalone deep learning models.

To further assess stability, the model was evaluated across multiple random train–test splits. The standard deviation of RMSE across five runs was observed to be only **0.004**, indicating consistent performance regardless of data partitioning. In contrast, individual models exhibited higher variance, confirming that the hybrid framework offers greater robustness.

In addition to predictive accuracy, computational efficiency was analyzed. While pure LSTM models required longer training time due to deep sequential dependencies, the RF–LSTM hybrid reduced total training time by approximately **18%** by pre-learning key feature patterns through the Random Forest layer. This demonstrates the practicality of deploying the proposed system in real-time smart home scenarios where quick retraining or model updates are required.

Overall, the experimental findings validate the effectiveness of integrating ensemble and temporal learning mechanisms for smart home energy forecasting. The model's enhanced precision, reduced error rate, and consistent performance across varying conditions make it suitable for intelligent energy management applications such as demand response optimization, anomaly detection, and real-time consumption planning.

The performance improvement demonstrates the synergy of ensemble and temporal learning mechanisms.

VIII. Conclusion and Future Work

In their research, the authors exemplified *the first* integration of predictive machine learning and deep learning models for smart home energy consumption forecasting. It integrates machine learning and deep learning models, namely, Random Forest and Long Short-Term Memory (LSTM) networks. The proposed method aims to overcome some of the smart home forecasting problems, such as varying consumption patterns, changes in environment, and temporal non-stationary. The RF LSTM architecture marries the feature extraction and non-linear modeling of Random Forest with the learning powers of LSTM networks.Data from tests showed the hybrid model significantly outperformed the individual RF and LSTM models in predictive performance. The proposed system exceeded all previously offered solutions in all metrics capturing an RMSE of 0.082 and over 94within IoT-enabled smart home environments, providing a foundation for real-time predictive control, energy op timization, and cost-efficient management. As illustrated in Fig. ??, the extended framework envisions the hybrid RF–LSTM pipeline evolving into a comprehensive Intelligent Energy Management System, capable of end-to end data-driven automation—from data collection to real-time forecasting and adaptive control. Future works have already been outlined. In terms of what direction to take next in research, it has also been achieved. Increasing the model's size and its intelligence is needed on: • Renewable Energy Integration: Expanding the model on. Prediction of solar and wind energies. Allowing the entire home energy balance and. Support the increase of distributed operational and. privacy-preserving learning frameworks to tackle Real time in large smart home networks. Merging reinforcement. learning to develop flexible systems demand-response. That schedule appliances and energy use. • Multi-objective Optimization: Increasing model objectives in. Issues of energy use, user comfort, and. Carbon Footprint reduction and cost of operational. which promotes organizational sustainability. smart home ecosystems.•Real-Time IoT Deployment: Integrating embeddings of RF-LSTM models for predictive modeling and flexible adaptive control on edge devices, and embedded systems is likely to be seamless. All in all, the RF-LSTM hybrid model will set futuristic smart home energy forecasting systems. As the first of its kind, the research tackles the intersection of machine learning explainability and temporal hierarchies in deep networks to push the frontiers of Artificial Intelligence to innovative sustainable energy management and efficient, smart, and green energy systems in cities.

ACKNOWLEDGMENT

The authors would like to express their profound gratitude to **Dr. Meena Chaudhary**, Department of Computer Science and Engineering, *Manav Rachna University*, for her continuous support, expert guidance, and valuable feedback throughout the course of this research. Her encouragement, constructive insights, and in-depth technical knowledge played a pivotal role in shaping the direction and quality of this work.

The authors also acknowledge the constant motivation and academic environment provided by the **Department of Computer Science and Engineering, Manav Rachna University**, which fostered innovation and interdisciplinary collaboration essential for the successful completion of this project.

Special thanks are extended to the faculty members and peers who provided helpful discussions and critical reviews during project evaluations, as well as to the university's computing and research infrastructure that enabled the efficient execution of simulations and model training.

Finally, the authors express appreciation to their families and colleagues for their patience, understanding, and moral support throughout the duration of this study.

REFERENCES

[1] T. Hong, P. Wang, and H. Luo, "State-of-the-art on load forecasting," *International Journal of Forecasting*, vol. 32, pp. 896–903, 2016.
[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
[4] R. Weron, "Electricity price forecasting: A review," *International Journal of Forecasting*, vol. 30, pp. 1030–1081, 2014.
[5] A. Jain, "Short term load forecasting using machine learning," *Energy Reports*, vol. 7, pp. 555–563, 2021.
[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
[7] A. Humeau, "Smart home data analytics for energy efficiency," *IEEE Access*, vol. 9, pp. 85573–85584, 2021.
[8] M. Hossain, "Hybrid deep learning model for smart home energy prediction," *IEEE Sensors Journal*, vol. 22, no. 5, pp. 4980–4988, 2022.
[9] D. Adebayo, "Forecasting residential energy consumption with LSTM and Random Forest," *Energy and Buildings*, vol. 261, 2022.
[10] A. Rahman, "IoT-based real-time smart home energy monitoring," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9801–9811, 2021.
[11] Y. Zhang, C. Li, and X. Wang, "Hybrid CNN–LSTM model for energy load forecasting in IoT-enabled smart homes," *IEEE Access*, vol. 11, pp. 12841–12852, 2023.
[12] H. Nguyen and D. Tran, "Deep Reinforcement Learning-based Smart Energy Management in Residential Grids," *Energy Reports*, vol. 8, pp. 512–523, 2022.
[13] P. Kaur and R. Bansal, "Comparative Study of ML Models for Smart Home Power Demand Forecasting," *Sustainable Computing: Informatics and Systems*, vol. 38, 100892, 2023.
[14] D. Alvarez and J. Kim, "Federated Learning for Privacy-Preserving Smart Home Energy Forecasting," *IEEE Internet of Things Journal*, vol. 11, no. 7, pp. 6598–6610, 2024.
[15] A. Sharma and V. Gupta, "Hybrid Ensemble-LSTM Approach for Predicting Energy Consumption in Smart Cities," *Applied Energy*, vol. 322, 119531, 2022.
[16] J. Chen, S. Zhou, and H. Li, "Edge-based Deep Learning for Smart Energy Prediction in IoT Systems," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6034–6045, 2023.
[17] L. Santos and E. Rodrigues, "Multi-objective Optimization of Smart Home Energy Consumption Using Machine Learning," *Renewable and Sustainable Energy Reviews*, vol. 181, 113293, 2024.
[18] K. Patel and M. Singh, "Sustainable Forecasting for Energy-Efficient IoT Networks Using Hybrid Deep Reinforcement Models," *Energy AI*, vol. 15, 100405, 2024.