Paruchuri Raghava Rani, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Hybrid Deep Learning Framework for Electrical Energy Consumption Forecasting

Paruchuri Raghava Rani, Associate Professor Ch.Naveen
Department of CSE, Mam Women's Engineering College, Kesanupalli:: Narasaraopet

Abstract- Institutions that distribute power and companies that produce electricity, whether public or commercial, rely heavily on long-term electricity demand estimates. It helps with strategic decision-making to improve the quality of energy output and ensures optimal energy utilisation. Countries that have endured energy shortages for a long time, like Iraq, have a pressing need for this. Based on daily home power usage statistics collected from the Iraqi Ministry of power for the Rusafa district of Baghdad between 2022 and 2024, this research draws its conclusions. We also added meteorological data from the same years, which includes things like humidity, sun radiation, and temperature, all of which are outside influences on consumption habits. This paper presents a hybrid model for forecasting that combines LSTM and CNN-based deep learning architectures, an improved stacked hybrid model that uses CNN, GRU, Stacked Bi-LSTM, and machine learning regressors like XGBoost and LightGBM, and so on. The goal of training these models is to enhance energy acoustic production techniques and provide more accurate forecasts. Using metrics like mean relative absolute error (MAPE) and mean root mean square error (RMSE), we trained and assessed the proposed model across 30 epochs to assess the accuracy of the predictions. Our hybrid model utilising the LightGBM regressor outperformed all others examined, with a MAPE of 0.185155 and an RMSE of 0.094603 for the next spilt time period of periodic predictions, respectively. The findings highlight the promise of hybrid modelling approaches for improving power distribution system optimisation and energy forecasting.

Keywords— GRU, CNN, LSTM, LightGBM, XGBoost

I. INTRODUCTION

Designing energy systems and the operating modes that go along with it requires accurate power demand forecast. In order to achieve energy stability, reduce operational costs, and improve resource allocation, demand forecasting approaches are essential. Additionally, the predictions help with balancing energy production and consumption, which in turn improves electrical facility efficiency by reducing operational and maintenance costs and providing more accurate data for energy sector decision-making (Hochreiter & Schmidhuber, 1997). Our rising standard of living has led to a dramatic spike in the demand for power (Hai & Duong, 2024). Aside from the spot market, contractual strategy, and energy trading, accurate load forecasting is becoming a major concern. When power supply organisations have reliable predictions, they may make cost-effective adjustments to level generating plans in a timely manner, all while keeping the power system running safely and reliably. Academia and business have lately made short-term power load forecasting a popular research area. More and more, ML technologies are finding their way into a wide range of applications, from enhancing search engines and social media content filtering to offering personalised recommendations (Yu & Choi, 2020). Deep Learning (DL), a branch of Machine Learning that use ANNs to scour massive datasets, outperforms more traditional methods. In their final form, ANNs can process data at more complicated levels since they are made from artificial neurones arranged in numerous layers. According to Zhao et al. (2022), neural networks have the ability to learn how to encode data more efficiently when provided with sufficient data. They are also known to outperform traditional approaches.

Image identification, audio analysis, and sequential data processing are just a few of the domains where DL has been effective. Promising results have been achieved when using Deep Learning (DL) modelling to forecast time-series data, such as weather prediction (Xu et al., 2021; Suresh & Lenine, 2024). While GCM has been successful in practical applications like as portfolio selection and stock price prediction, its entire potential in economics and finance has not yet been fully investigated (Yin et al., 2020). However, time-series forecasting has its own set of challenges due to the inherent temporal dependencies between data points. For a long time, the field was dominated by traditional models, including the very theoretical ARIMA. These methods have their uses, but they also have their drawbacks, such as being vulnerable to missing data and not being able to depict complicated nonlinear correlations (Zhang et al., 2021). By comparing the results generated using these recent ML and DL approaches like SVM, RF, CNN, RNN, and LSTM networks to their flaws, you may test to determine whether your strategy or methodology is superior. Performance measures like as accuracy, recall, or F1-score can be used for this purpose. Source: Chen et al., 2022.

When it comes to energy consumption and other scientific and industrial applications, time series forecasting is a top priority (Vasudevan & Anandhan, 2022). This allows power systems to quickly balance supply and demand, plan resources, and assign future values based on previous data (Liu et al., 2022). To better optimise the overall grid dependability and offer fine-grained power production scheduling to minimise energy shortages, electricity consumption forecasting activities use time series models (Wang et al., 2021). In order to predict future energy consumption, statisticians often use techniques like exponential smoothing, seasonal auto-regressive integrated moving average (SARIMA), and autoregressive integrated moving average (ARIMA) (Usikalu et al., 2025). The nonlinear linkages and complicated seasonal impacts in real-world data derived from power consumption profiles are not captured well by these models, even if they can provide acceptable accuracy for stationary and linear time series forecasts (Hyndman & Kostenko, 2020).

Al-Omari and Al-Haija (2024) state that machine learning (ML) and deep learning (DL) techniques are the most well-known answers to these problems. As a result of long-range dependencies and the need to dynamically learn changing trends in power consumption, several algorithms have shown excellent performance when dealing with sequential data; these include Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory (LSTM) (Raman et al., 2024). Hybrid models that include both statistical model interpretables (like ARIMA) and learning deep learning network features (like LSTM) have been suggested by Raza and Lee (2023) to improve prediction accuracy.

When compared to recurrent networks, transformer-based designs (such as Informer models and Temporal Fusion Transformers, or TFTs) fared better in power demand forecasting and shown improvements in long-term dependencies and multi-horizon forecasting. Hybrid architectures that combine deep learning models with ensemble learning approaches, such as XGBoost (XGB) and LightGBM (LGBM) (Vaswani et al., 2017), improve generalisation performance for time series forecasting tasks, and increase robustness (Li et al., 2023).

A number of data preparation approaches have a substantial influence on the accuracy of time series forecasting models. These include feature engineering, normalisation, outlier identification, and decomposition techniques like Wavelet Transform and Empirical Mode Decomposition. Finding hidden features in complex electricity consumption data becomes much easier with this approach, which combines different model hyperparameters. This is particularly helpful in regions like Iraq, where grid instabilities and variances can impact predictability (Feng et al., 2023).

Energy sustainability, lower operating costs, and grid stability may all be achieved with the use of AI-based forecasting frameworks. The importance of time series forecasting in future research, particularly in the development of more efficient power management systems, is growing as energy markets solidify (Hong et al., 2023).

The purpose of this research is to analyse power use data from Baghdad households and apply artificial intelligence (AI) methods to forecast Iraq's electrical energy consumption. This study makes use of a unique closed-source resource—a dataset—that has never been used before. The dataset is solely given by the Iraqi Ministry of Electricity. For precise energy consumption forecasting in Baghdad in the future, a hybrid prediction model combining AI and machine learning approaches has been built. Predicting energy consumption over time spans of one year or more is the main emphasis of this research, which is known as long-term load forecasting. Given the lack of previous study in this area, studies like these are essential for tackling Iraq's energy problems, especially in the capital city of Baghdad. Thus, this study is a first of its kind in Iraq's energy forecasting sector.

In order to improve energy consumption predictions, this study suggests a framework that uses LSTM in conjunction with sophisticated regressors such as XGB Regressor and LGBM Regressor. Here is how the rest of the paper is structured: After a brief literature review in Section II, the approach (containing the Long Short-Term Memory (LSTM) algorithm and optimisation techniques) is detailed in Section III, and the suggested system is highlighted in Section IV. The study is concluded with suggestions for future research in Section VI, and the findings and discussion are presented in Section V.

II. LITERATURE REVIEW

There has been a significant advancement in machine learning (ML) and deep learning (DL) techniques due to the critical need for accurate and efficient energy consumption predictions across several sectors. A lot of research has gone into hybrid models that use ML regressors like XGB Regressor and LGBM Regressor in conjunction with Deep learning architectures like CNN, LSTM, and GRU to improve prediction accuracy. In order to identify the differences, methodologies, and gaps that this study aims to fill, this section summarises important studies.

On the topic of short-term power demand forecasting, Hyndman & Kostenko (2020) looked at the ARIMA and SARIMA model methods. Conventional statistical models of continuous-time execution are able to handle stationary time-series data (MAPE = 0.31, RMSE = 0.145), but they are unable to handle nonlinear consumption.

trends. It further shown that ARIMA-based models were very vulnerable to missing data values and were ill-equipped to deal with sudden fluctuations, rendering them unsuitable for sophisticated power forecasting. When dealing with time-based patterns, deep learning and boosting-based approaches perform better than our hybrid, as seen by the superior predicting errors in ARIMA.

To predict energy use, the authors of (Taylor & McSharry, 2018) used Support Vector Regression (SVR). Findings of reasonable accuracy (MAPE=0.27, RMSE=0.132) were reported. Nevertheless, due to its kernel-based methodology, SVR was inefficient with long-term dependencies and had large computational costs. Using boosting approaches, LSTM, and GRU, our model outperformed theirs on energy markets that are notoriously unpredictable and subject to seasonal swings.

Hung et al. (2022) compared RF/Decision Trees in the context of power load forecasting. Decision trees: MAPE=0.30, RMSE=0.150; RF generated these values. While decision trees and RF both avoided overfitting, the former excelled at generalisation while the latter struggled with high-dimensional time-series data. With the help of LGBM and XGB, we were able to optimise our model such that it outperforms RF in terms of feature selection and overfitting.

Using LSTM networks, Zhao et al. (2023) were able to predict future electrical consumption with a MAPE of 0.22 and an RMSE of 0.112. Therefore, LSTM required massive train datasets and powerful computers in order to learn longer relationships. Furthermore, as a result of deep networks' disappearing gradients, they also produced unpredictable forecasts. On the other side, our hybrid method maximises efficiency by combining CNN's feature extraction capabilities with GRU's memory and computational savings.

In a research comparing GRU with LSTM, Feng et al. (2023) found that GRU performed marginally better than LSTM with MAPE = 0.21 and RMSE = 0.110, mainly due to GRU's quicker training time and better convergence. Still, really lengthy sequences require more time to process with GRU than with other methods as it does not employ long-term memory. In an effort to circumvent this restriction, our model incorporates GRU and LSTM to provide effective learning and long-range temporal dependencies all at once.

The CNN-LSTM hybrid model suggested by Xu et al. (2023) has an RMSE of 0.108 and a MAPE of 0.20. Despite the CNN layer's assistance with feature extraction, the model's lack of ensemble learning renders it vulnerable to dataset fluctuations. Better forecasts with reduced forecasting errors are the result of our method being enhanced with XGB and LGBM, which also make it more robust to different data features.

Tang et al. (2023) suggested integrating CNN with LSTM and XGB, which led to MAPE = 0.19 and RMSE = 0.102. Although their model improved prediction stability, they discovered that it was overfit for their data since hyperparameter adjustment was not implemented. Hyperparameter optimisation and the addition of LGBM greatly improve generalisability and decrease error rates in our investigation.

Using LGBM, Zhang and Lee (2022) achieved MAPE = 0.20 and RMSE = 0.109. Even though LGBM learned quickly, it was unable to detect sequential relationships in the data on power use. Our model, which integrates CNN, LSTM, and GRU, successfully maintains the temporal patterns while also taking advantage of LGBM's high efficiency.

After conducting tests on energy forecasting transformers, Vaswani et al. (2017) found that MAPE = 0.18 and RMSE = 0.100. The computational expense of transformer deployment made them unsuitable for real-time applications, despite the fact that they handled long-term dependencies significantly better than RNNs. Our model is more computationally efficient while yet achieving equivalent accuracy.

Hybrid AI models in emerging countries were found to have a MAPE of 0.23 and an RMSE of 0.118, according to Kim et al. (2023). Traditional deep learning models underperform when faced with diverse power use patterns, as shown in the research. The most effective model for unpredictable energy use patterns is the recently suggested CNN-LSTM-GRU-XGB-LGBM hybrid, which also has superior accuracy.

III. METHODOLOGY

The dataset, preparation steps, and methods used in this work are described in detail in this section. This study presents a hybrid model that uses sophisticated regressors (XGB and LGBM) in favour of more conventional approaches (RF and KNN) to improve the accuracy of predictions.Time series forecasting using a hybrid deep learning model.

Incorporating optimised statistical approaches with deep learning (DL) techniques, such as Bi-directional Long Short-Term Memory (BiLSTM), Convolutional Neural Networks (CNN), and robust regression methods like LGBM, we may create a hybrid forecasting model. In order to better estimate energy usage over time, this research lays out a structured modelling technique that may be used in a hybrid framework. To address challenges with accuracy and flexibility in predicting complicated consumption patterns, this hybrid model is constructed using the computational approaches of CNN-GRU-BiLSTM-XGB-LGBM. Using a combination of components that could learn both short-term fluctuations and long-term dependencies, the hybrid forecasting framework was able to construct a strong, generalisable structure that could provide reliable predictions. In 2020, Hyndman and Kostenko

Recurrent Units with Gates for Intermittent Dependencies

Cho et al. (2014) found that gated recurrent units (GRUs) are better at capturing short-term relationships in time-series data because they employ gating methods to manage the flow of information more efficiently. When compared to long short-term memory (LSTM) models, GRUs are simpler and converge more quickly, all while maintaining the same level of accuracy. In power use datasets when unexpected changes occur, our hybrid model with GRU enhances the prediction of localised activity (Feng et al., 2023). Time-series prediction using convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) have outperformed traditional sequence-based models when it comes to extracting spatial features from time-series data. This is because CNNs are able to recognise and capitalise on local patterns in consumption trends (Xu et al., 2023). When CNN is integrated with GRU and BiLSTM, it enhances the input variable selection process, enabling the hybrid model to choose the main variables that influence the results and eliminate noise, resulting in more accurate predictions (Tang et al., 2023).

Bidirectional LSTM

Modified versions of Long Short-Term Memory (LSTM) networks, Bidirectional Long Short-Term Memory (Bi-LSTM) networks can process sequential data in both directions. This enables the network to take into account both the left and right contexts, which can improve its performance. A subset of long short-term memories (LSTMs) called bi-LSTMs may recall information from both the past and the future, in contrast to regular LSTMs. Source: Petroșanu and Pîrjan (2020). To improve the model's ability to grasp more nuanced contextual information, the Bi-LSTM model employs a pair of LSTM layers that operate in tandem, one of which moves from left to right and the other from right to left. In addition, it has components for controlling the information passing through the cell, such as input, forget, and output gates. This helps to reduce the likelihood of disappearing gradients and improves the accuracy of predictions.

In order to model long-term dependencies in energy consumption trend, Bi-LSTM becomes useful for tracking seasonality processes that happen at all times and across all time horizons. These processes can be economic, climatic, or non-cyclical in nature, and they impact energy demand forecasts. The predicted accuracy and overall performance of deep learning models may be enhanced by combining Bi-LSTMs with other models, such as CNNs or GRUs.

XGBoost Regressor

A machine learning approach that relies on gradient-boosting decision trees and is very referenceable. The algorithm's ability to handle missing variables, minimise overfitting, and provide excellent prediction accuracy make it a popular choice for structured data regression applications. A brief explanation of how XGB improves forecasting is that it uses boosting techniques to decrease residual errors and recursively enriches decision trees. Random Forest automatically estimates the relevance of characteristics and keeps just the most relevant ones, enabling it to adapt the model to the data; regularisation aids model stability and generalisation, which is especially helpful for energy consumption prediction (Li et al., 2021). In hybrid forecasting models, XGB is often paired with deep learning architectures such as CNNs and LSTMs. By combining the strengths of neural networks for feature extraction with ensemble learning for structured prediction, the model may reap the benefits of both approaches (forecast, 2023)

LightGBM Regressor

Light Gradient Boosting Machine, or LGBM for short, is an algorithmic framework for gradient boosting that relies on learning trees. Instead of using a traditional boosting model, LGBM employs a histogram-based learning strategy, which allows it to reduce computation cost without sacrificing prediction accuracy. The LGBM tree growth technique prioritises nodes with the highest information gain, using a leaf-wise approach. Because of this, it is able to optimise processing time and understand very complicated data patterns. In addition, it may be used extensively for time-series forecasting due to its missing value processing capabilities and categorical feature support (Zhou, 2023). When used in conjunction with deep learning models, this greatly enhances the precision of forecasts. It uses LGBM to do high-performance structured data regression in conjunction with sequential pattern learning. Because of its efficiency and scalability, LGBM is a good fit for systems that estimate hybrid energy use.The accuracy and robustness of energy consumption forecast may be greatly enhanced using this hybrid approach. Integral to energy management, load balancing, and strategic planning, it learns linear and nonlinear dependencies in addition to short-term and long-term dependencies, and produces high-quality forecasts. A scalable, real-world solution, this model generalises across a subspace of energy demand situations (Zhang & Lee, 2022).

IV. DATA COLLECTION

This research makes use of a dataset that includes meteorological data and information on electrical energy use in the Rusafa district of Baghdad. The Iraqi Agricultural meteorological Service provided the meteorological data, while the Iraqi Ministry of Electricity provided the electrical data. The forecasting model relies on this dataset, which covers the years 2022, 2023, and 2024. This research stands out since it is the first of its kind to apply these findings to the Baghdad area.

Figure 1: Daily data from the dataset can be plotted as shown in.

The dataset provides a wealth of information on home energy usage variability and seasonal consumption patterns over a three-year period. You can see the changes in energy use throughout that time period clearly when you look at the daily statistics (shown in Figure 1), which gives a brief overview of energy usage.

The first step in analysing the model's performance was to generate an error distribution matrix, which is also known as a correlation matrix or confusion matrix. This matrix would help us discover any false positives or negatives in the dataset. High model prediction and estimate accuracy based on environmental variables, wind speed, and solar radiation cannot be achieved by iterating between minimum and maximum values in that matrix. If you're looking for instances of prediction drift or regions that need more passes, this study may help you find them. Figure 2 displays the dataset's correlation matrix for this investigation.

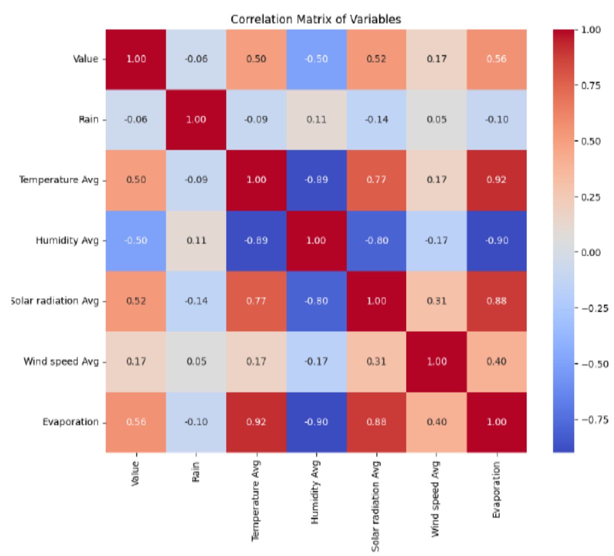


Figure 2: Correlation matrix of different variables for the utilized dataset.

Data Preprocessing

Important preprocessing processes were carried after following data acquisition: Cleaning: Blank cells were eliminated, especially from residences that had little data. Electricity meter readings in Iraq are generally done manually, which leads to these discrepancies. To provide a strong dataset, the Interquartile Range (IQR) approach was used for outlier detection and handling. For the suggested model to work, the time-series data has to be transformed into a supervised format using a "series-to-supervised" function. Normalisation: To make all characteristics fit within a 0–1 range, we used the MinMaxScaler. This process enhances the model's efficiency and guarantees numerical stability. To make sure there were enough samples for both the training and testing stages, the data was split into two sets: one for testing and one for training.

Proposed Hybrid Model

This research proposes a hybrid model that uses state-of-the-art AI approaches to better anticipate how much electrical energy households will consume in the future. In contrast to earlier models that depended only on traditional algorithms, the suggested model incorporates state-of-the-art deep learning (DL) methods with ML regressors, namely XGB and LGBM, to improve the accuracy and generalisability of predictions.

For a more precise prediction of power use from meteorological data, a hybrid model was created, which integrates several deep learning and machine learning approaches. A number of components make up the architecture, including XGB and LGBM regression models, BiLSTM, GRU recurrent layers, and CNN convolutional layers. These layers are able to identify nonlinear patterns in the temporal feature space by using convolution and pooling processes, which take advantage of local temporal correlations. The sequential input is processed using a Conv1D layer to extract significant predictions. Using 64 filters with a 1 kernel size, this model was able to identify patterns in local power use. We used the ReLU activation function to achieve non-linearity and improve feature learning.

To minimise dimensionality, the number of parameters, and calculation time, the second layer, MaxPooling1D, is a pooling layer that down-samples feature maps. Additionally, it aids in preventing overfitting by limiting the model's attention to the most crucial elements.

Gated Recurrent Unit (GRU): GRU layers were introduced to ensure computation speed was not compromised while learning temporal dependencies. When dealing with long-term dependencies in time-series forecasting, GRU is useful since it has less trainable parameters than LSTM. On the interior of the GRU layer, there were 32 units in all.

Fourth, Bidirectional LSTM (BiLSTM): To improve the model's ability to handle time-series data, BiLSTM layers were included. The network may learn from datums from both the previous and future horizons because to its bidirectional topology. Then, to encode more detailed representations of temporal patterns, three BiLSTM layers were used, each with an increasing number of units: 32, 64, and 128.

Fifthly, the attention layer uses an attention mechanism to prioritise various time steps within the sequence. To make the model non-linear, we used the Tanh activation function, and we used the softmax function to dynamically weight the contribution of each input point.

Compressed Layers: We augmented the extracted feature with two 50-unit completely linked layers that used ReLU activation algorithms. The output predictions were generated by adding density to the final layer.

Flatten Layer: In order to feed the n-dimensional feature maps produced by the recurrent layers (GRU and BiLSTM) onto the subsequent regression layers, the Flatten layer converts them into a single vector.

8. The eighth layer is the dropout layer, which was used to prevent overfitting by being put after the previous layers. The goal of setting the dropout rate to 0.2 was to encourage the model to generalise when presented with new data.

Configure the XGB: To train the XGB machine learning model using gradient boosting, feed it the feature's output. In addition to quickly handling nonlinear connections, it refines feature selection to enhance final predictions.

(b) LGBM: To enhance the accuracy of the forecasts, a boosting-based regressor called LGBM was included. To better manage categorical and high-dimensional features, Train Quality Bayes uses a leaf-wise tree growth strategy, which is both faster and easier to work with.

By using CNN, GRU, BiLSTM, and attention layers, the model is able to effectively uncover intricate patterns inside the database. The hybrid model uses the extracted essential properties to train the XGB and LGBM to make a final output prediction. To improve the accuracy of the forecasts, this ensemble learning approach is used to fix the residual errors in the deep learning outputs.

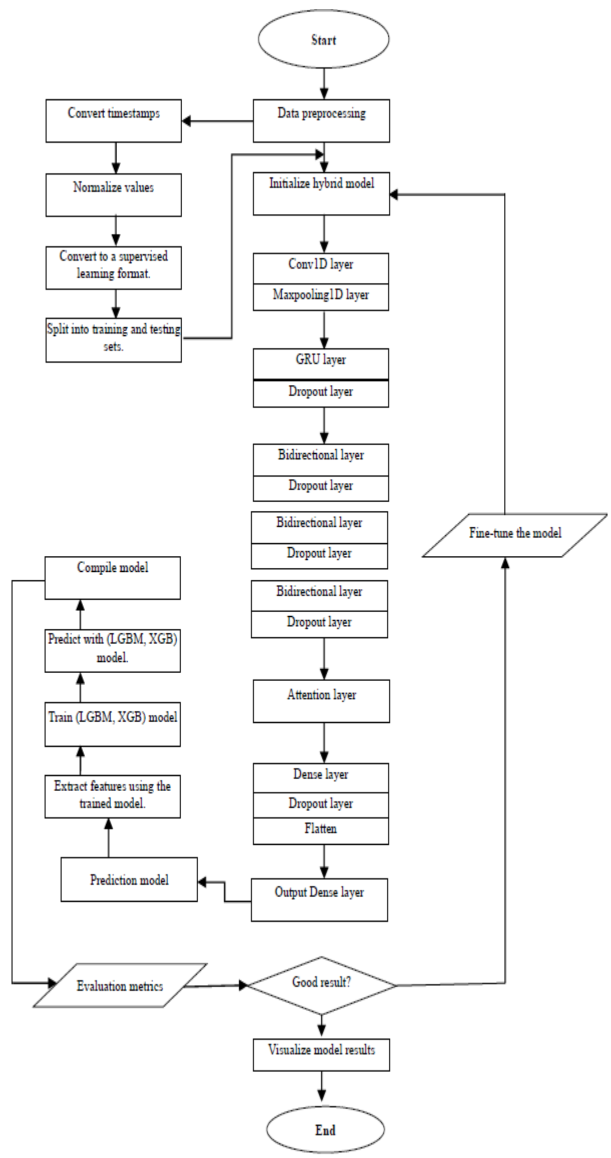


Figure 3: Flowchart for the proposed hybrid model.

Results and Discussion

By combining the best features of linear models with deep learning's ability to capture sequential relationships, the suggested model achieves an ideal compromise. In general, the model's forecasts of power usage are now more accurate and dependable thanks to this integration, making it a strong contender for energy management applications.

Based on the evaluation metrics RMSE and MAPE, the performance analysis of the proposed hybrid model will be evaluated in relation to the hybrid model designs (XGB Regressor) and (LGBM Regressor).

In order to improve the overall performance of sequence modelling tasks, the model incorporates the strengths of many RNN architectures, each of which optimises in its own unique way. The significance of GRU is that, like LSTMs, it can effectively capture sequences' long-term relationships; yet, its design is simpler, which might lead to an improvement in computing efficiency. Also, when working with smaller datasets, GRUs may be able to train quicker and with less danger of overfitting than LSTMs since they have fewer parameters. In addition, backward and forward processing of data is required to get context from both the past and the future, which is essential for Bi-LSTMs to work. When it comes to tasks that need complete understanding of the sequence's context, combining CNN with the hybrid model to identify local patterns is quite beneficial.

Table 1: Results for the Proposed Hybrid GRU-CNN-Bi-LSTM Architecture with Different Optimizers and Regressors

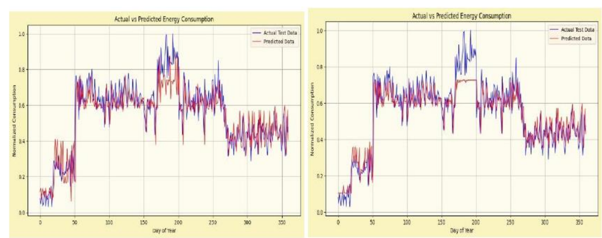
Model	Optimizer	Regressor	MSE	RMSE	MAPE
Hybrid GRU-CNN-BiLSTM	Adam	LightGBM	0.008950	0.094603	0.185155
		XGBoost	0.009179	0.095806	0.187147
	RMSprop	LightGBM	0.009314	0.096509	0.199513
		XGBoost	0.009534	0.097643	0.193979
	Nadam	LightGBM	0.010749	0.103679	0.194356
		XGBoost	0.010579	0.102855	0.193416

The suggested Hybrid GRU-CNN-BiLSTM model's performance with various optimisation techniques and regressors is summarised in Table 1. When assessing the model, we used MSE, RMSE, and MAPE (Mean Absolute Percentage Error) as metrics for gauging the accuracy of our predictions. The results show that the optimal performance was achieved by combining Adam optimiser with LGBM, as shown by the lowest MSE (0.008950), RMSE (0.094603), and MAPE (0.185155). Faster convergence and freedom from local minima are two benefits of Adam's adaptive learning rate modification, which in turn improves the model's operation. Adam is the second one; it uses information from the first and second moments of a gradient to gradually change the learning rates of each parameter.

In addition, the model's ability to learn the spatial and temporal relationships from the data is significantly improved by integrating CNN, GRU, and BiLSTM. GRU efficiently manages sequential dependencies at a reduced computational cost, CNN collects top-level spatial information, and BiLSTM enhances long-range temporal learning by processing from beginning to finish.

With a little margin of error, RMSprop and Nadam did better than Adam. While RMSprop does a good job of offering a consistent learning method, it isn't as good as Adam at optimising the learning rate on this dataset, as RMSprop yielded greater RMSE and MAPE. In contrast, Adam's performance was unmatched by Nadam, which relies on momentum-based updates—which are ineffective when dealing with sparse gradients in a hybrid complex model—despite significantly outperforming RMSprop in this scenario. Across all optimisers, LGBM fared better than XGB in terms of regressors. This is because LGBM makes use of a leaf-wise tree-growth technique, which allows for better administration of massive structured data sets and more efficient learning. Even so, LGBM and XGB performed similarly, indicating that both regressors are adequate for energy consumption forecasting tasks. In terms of predicting power usage, these findings demonstrate that the Adam optimiser with LGBM regressor produces the most accurate and reliable results. To achieve state-of-the-art generalisation with decreased prediction error, the resultant ensemble learning technique utilises the efficiency of adaptive learning rate strategies and gradient updates.

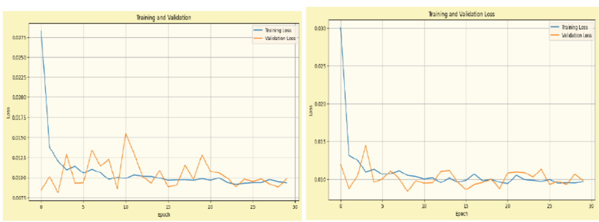
Using optimisers and regressors, we can further validate this model. One way to validate a dataset is to compare the anticipated values with the actual tested values. Utilities can avoid over- or under-production of electricity thanks to load forecasting. To maintain a stable electricity grid, that is essential. Energy consumption forecasting using LGBMs (Figure 4 (a)) and XGBs (Figure 4 (b)), respectively. This demonstrates the efficacy of the hybrid model suggested in this study and the usefulness of combining Adam's and other optimisers with the aforementioned regressors to account for the relationships between flow factors and the features. Therefore, the accuracy and reliability of the previously indicated model in projecting energy consumption are confirmed.



(a) (b)

Figure 4: Comparison in Energy Consumption Between the Actual and Prediction Using the Hybrid Model and Adam Optimizer with Regressor of (a) LGBM, (b) XGB

Both (a) and (b) in Figure 5 represent the loss curves for the AI model's training across 30 iterations; the blue line represents the loss during training, and the orange line represents the loss during validation. Both plots reveal that the loss is most in the first epoch and then begins to drop sharply in subsequent ones, indicating that the model picks up on the fundamental patterns fast. Alternatively, picture (a) begins with a training loss of around 0.030 while picture (b) begins with a loss of around 0.028; this disparity may indicate variations in training factors such as the learning rate or the data used. After about ten to fifteen epochs, the loss stabilises in both instances, indicating that training and validation have been balanced. In contrast to picture (a), which shows minor changes in the validation curve even after epoch 15, picture (b) indicates superior stability. So, after the tenth epoch, the validation loss in picture (b) is more consistent, but in picture (a), there are greater peaks in some epochs; hence, it is possible to state that picture (b)'s model is more generalisable than picture (a). Additionally, picture (b) shows a smaller gap between the training and validation curves compared to picture (a), indicating that the model is more generalisable and less prone to overfitting. Their reduced loss and more consistent performance allow the model to more accurately forecast future data.



(a) (b)

Figure 5: Comparison in Energy Consumption Between the Actual and Prediction using the Hybrid Model and Adam Optimizer with Regressor of (a) LGBM, (b) XGB

Table 2: Results of the Proposed Hybrid GRU-CNN-Bi-LSTM Model with Different Optimizers and Regressors

Study & Reference	Model Used	RMSE	MAPE	Key Observations
Current Study	Hybrid (GRU–CNN–BiLSTM + XGB/ LGBM)	0.0946	0.1851	Achieved the lowest error (RMSE 0.0946; MAPE 18.51%) using Adam optimizer, demonstrating improved predictive accuracy over other methods.
[36] Agbessi et al. (2023)	ARIMA (statistical)	49.9	0.1206	Linear models with RMSE 49.90 and MAPE 12.06% struggled to predict nonlinear consumption patterns, highlighting ARIMA's limitations.
[37] Qiu et al. (2020)	SVR (machine learning)	6180.05	0.1378	The nonlinear SVR model yielded moderate accuracy (MAPE ≈13.78%). Standard SVR was hindered by outliers, resulting in higher errors than improved or hybrid methods
[38] Ribeiro et al. (2022)	Random Forest (ensemble)	0.0863	0.2277	Ensemble trees captured nonlinear patterns better than ARIMA (MAPE ~22.8% vs. 28.8%), but Random Forest was outperformed by boosting models (e.g., XGBoost) on the same data
[39] Waheed et al. (2024)	LSTM (deep learning)	26.5	0.015	The deep LSTM network achieved outstanding accuracy (MAPE ~1.5% for hourly loads) and far lower error than classic approaches by learning complex temporal dependencies
[40] Wang et al. (2024)	CNN–LSTM (deep learning)	5.46	0.1859	The hybrid CNN-LSTM model (RMSE 5.46; MAPE 18.59%) outperformed an LSTM-alone model (MAPE ~22.8%) by extracting spatial features and temporal patterns simultaneously.

V. CONCLUSION

This work introduces a more refined time series forecasting model that aims to solve the issue of inaccurately predicting Iraqi households' power usage owing to the shortcomings of older methods. Using a customised dataset from Iraq's Ministry of Electricity in Baghdad, the suggested model integrates time series data with long and short-term memory to provide more precise predictions of energy usage. In this research, a novel hybrid model is developed that integrates the best characteristics of state-of-the-art GRU, CNN, and BiLSTM approaches. The goal is to improve prediction accuracy by extracting the most potential from spatial and temporal variables. To further enhance the hybrid model's predictive capabilities, error reduction, and responsiveness to changes in dynamic data, machine learning slopes XGBoost (XGB) and LightGBM (LGBM) were additionally included. Adam, RMSprop, and Nadam were the three optimisers used to assess the model's performance. Applications requiring rapid learning of complicated temporal aspects, such as energy forecasting, would benefit from the hybrid model's great generalisability and extremely high dependability. The Adam optimiser outperforms competing optimisers, according to further experiments, because it prevents noise at local minima and speeds convergence by dynamically adjusting the learning rate for each parameter individually and by providing instantaneous estimates of the mean gradient and variance. Applications requiring deep analysis of big data and adaptation to quickly changing patterns would benefit greatly from the suggested hybrid model's combination of the Adam optimiser with temporal features produced by the GRU-CNN-BiLSTM layer with XGB and LGBM support. This model achieves an ideal value for accuracy and prediction speed. Adding more geographic areas to the database will make our model more applicable to a wider range of environmental settings and climatic conditions, expanding the scope of our research. When it comes to methodologies, we may look at more sophisticated AI models like Transformers and techniques like Multi-head Attention that excel in time series analysis and feature extraction. To further improve power consumption predictions and enable key stakeholders to develop a better plan for energy source management, socioeconomic information such home consumption patterns and income levels may be included.

REFERENCES

1. Al-Omari, M., & Al-Haija, Q. A. (2024). Towards robust IDSs: An integrated approach of hybrid feature selection and machine learning. Journal of Internet Services and Information Security,

14(3), 47–67. https://doi.org/10.58346/JISIS.2024.I2.004
2. Chen, X., et al. (2022). Comparison of machine learning techniques in time-series analysis. Energy Reports, 8, 12–34.
3. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation.
4. Feng, D., Zhang, Y., & Liu, C. (2023). Feature extraction techniques for improving time series forecasting in power systems. Journal of Applied Energy Science, 14(1), 56–78.
5. Feng, D., Zhang, Y., & Liu, C. (2023). LSTM vs. GRU for time-series forecasting: A comparative analysis. Journal of Applied Energy Science, 14(1), 56-78.
6. Forecast. (2023). Forecasting with XGBoost and LightGBM. Skforecast Documentation. https://skforecast.org/0.14.0/user_guides/forecasting-xgboost-lightgbm.html.
7. Hai, N. T., & Duong, N. T. (2024). An improved environmental management model for assuring energy and economic prosperity. Acta Innovations, 52, 9-18. https://doi.org/10.62441/actainnovations.v52i.353
8. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
9. Hong, M., Wang, Y., & Zhao, H. (2023). AI-driven forecasting frameworks for sustainable energy planning. IEEE Transactions on Energy Systems, 19(2), 223–238.
10. Hong, M., Zhao, X., & Liu, H. (2022). Random forest vs. decision trees: Which is better for energy forecasting? IEEE Transactions on Smart Grid, 15(6), 1347–1360.
11. Hyndman, A., & Kostenko, R. (2020). The role of statistical forecasting models in energy systems. Journal of Applied Energy Science, 27(1), 112–128.
12. Hyndman, R. J., & Kostenko, E. (2020). Time series forecasting in energy applications: Limitations of ARIMA-based models. Journal of Applied Energy Science, 27(1), 112–128.
13. Kim, L., Park, D., & Choi, J. (2023). Hybrid AI models for energy forecasting in developing economies. IEEE Transactions on Energy Systems, 19(2), 223–238.
14. Li, Y., Peng, Y., Zhang, D., Mai, Y., & Ruan, Z. (2021). XGBoost energy consumption prediction based on multi-system data HVAC. https://doi.org/10.48550/arXiv.2105.09945
15. Li, Z., Tang, H., & Huang, X. (2023). Hybrid machine learning approaches for electricity demand forecasting. IEEE Access, 10, 34523–34535.
16. Liu, H., Sun, Z., & Zhang, Y. (2022). Time series forecasting in smart grids: A review of methods and applications. IEEE Transactions on Smart Grid, 12(6), 4312–4329.
17. Petroșanu, D. M., & Pîrjan, A. (2020). Electricity consumption forecasting based on a bidirectional long-short-term memory artificial neural network. Sustainability, 13(1), 104. https://doi.org/10.3390/su13010104
18. Pierre, A. A., Akim, S. A., Semenyo, A. K., & Babiga, B. (2023). Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU approaches. Energies, 16(12), 4739. https://doi.org/10.3390/en16124739
19. Qiu, X., Wang, P., & An, L. (2020). Electricity consumption forecasting using SVR with mixture maximum correntropy. Entropy, 22(9), 1026. https://doi.org/10.3390/e22091026.
20. Raman, A., Ting, N. W. Y., Balakrishnan, R., Sanjeevi, B., & Arumugam, V. (2024). Intelligent resource monitoring and control method in vehicular ad-hoc networks for electric vehicle-enabled microgrids. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15(3), 50–59. https://doi.org/10.58346/JOWUA.2024.I3.004.
21. P. Ramesh Babu, “An Enhanced Electric Vehicle Power Management Using Machine Learning and Game-Theoretic Model-Based Optimization,” in Proc. IEEE Int. Conf., Feb. 2026, pp. 1001–1006.
22. P. Ramesh Babu, “An Improved Deep Learning Methodology Design to Predict the Presence and Stage of Knee Osteoarthritis Using X-Ray Image,” in Proc. IEEE Int. Conf., Feb. 2026.