Applied Machine Learning Approaches for Energy Consumption Prediction

Bhavya Kothari
Department of Computer Science
Birla Institute of Technology and Science Pilani, Dubai Campus
Dubai, United Arab Emirates
f20210103@dubai.bits-pilani.ac.in

Nethradevi Baskar
Department of Computer Science
Birla Institute of Technology and Science Pilani, Dubai Campus
Dubai, United Arab Emirates
f20210031@dubai.bits-pilani.ac.in

Maryam Sherrif
Department of Computer Science
Birla Institute of Technology and Science Pilani, Dubai Campus
Dubai, United Arab Emirates
f20210128@dubai.bits-pilani.ac.in

Abstract

Energy consumption forecasting plays a vital role in resource management and sustainability efforts.Many researchers have employed various traditional statistical and dynamic hybrid models and newer research has attempted to model energy consumption through machine learning as well. This study explores advanced machine learning techniques to improve the accuracy of energy consumption predictions. While previous research has employed various models for energy consumption prediction , our study introduces a novel approach by comparing and evaluating multiple methods, including XGBoost, Random Forest, and advanced architectures such as LSTMs, GRUs, and a hybrid RNN-CNN model. Our experiments demonstrate that among these machine learning models and architectures, the GRU significantly outperforms the others across five key metrics like RMSE (168.66), MAE (121.13), MAPE (1.04%), sMAPE (1.04%), and R^2 (0.99). These results highlight the superior performance of GRU in energy consumption prediction, making it a highly effective choice for time series analysis.

Keywords: *LSTM, GRU, Hybrid models, XGBoost, Random forest*

1. INTRODUCTION

Global energy demand is rising rapidly which is leading to the need for precisely forecasting energy consumption as it is becoming increasingly important for sustainable development and effective resource management. Predicting Energy consumption is essential in improving energy distribution which includes reducing operational costs and achieving a balanced energy supply and demand. Energy consumption has been predicted over a long time using traditional statistical methods like ARIMA and SARIMA . But most of these models often fail while capturing the complex, non-linear dynamic patterns in energy usage data. This is influenced by factors including human behaviour, changes in weather and seasonal variations which pose major challenges in accurately achieving precise predictions.Time series analysis which makes use of historical data to discover patterns and forecast future trends is a promising approach but still faces challenges with the non-stationary nature of energy data.

Recent developments in machine learning, especially deep learning models like Long Short-term Memory(LSTM) networks have shown considerable potential in tackling these challenges. These models are highly proficient in understanding temporal dependencies and complex patterns which makes them highly efficient for handling tasks which involves sequential data like energy consumption forecasting.Machine learning techniques achieve more predictive accuracy compared to traditional methods which has resulted in the integration of machine learning with time series analysis to boost prediction accuracy especially for short and medium term energy consumption forecasts in sectors such as commercial,residential and industrial applications.

In this paper ,we will be discussing and analyzing the patterns within the dataset which has been obtained by selecting a suitable model in the model selection phase with respect to the time series dataset .The content of the paper is presented as follows,Section 1 starts with an introduction to the topic chosen Section 2 comprises of the literature review which discusses the existing research on energy consumption forecasting highlighting the strengths,limitations and advancements of the models used, Section 3 provides a detailed description of the dataset used, Section 4 outlines the model selection process, data preprocessing techniques, and visualization methods used to extract insights from the data. It also discusses the training and testing of the selected model.Finally, Section 5 contains the references to the papers used in literature review.

2. LITERATURE REVIEW

We can discuss a few of the existing models used for predicting or forecasting energy/power consumption.

Back in a 2018 study **[1]**, Sat Qzturk 1 came up with a statistical ARIMA model for forecasting next 25 year's total renewable and energy consumption data of Turkey along with few others such as oil, natural gas, coal. The results revealed that the best fit model in comparison with others had obtained the lowest AIC values of -2.442.

Liu et al. **[2]** introduced the Fractional Polynomial Grey Model (FPGM(1,1,tα)) to improve electricity consumption forecasting by using time power terms and fractional accumulation , and it was made more efficient with quantum genetic algorithms. The model had good accuracy in case studies for Chongqing and Beijing, outperforming grey models using APE and RMSE. Improving this, Liu et al.**[3]** the same author with some others ,then used fractional improvements in their Adjacent Non-homogeneous Discrete Grey Model (ANDGM) for renewable energy forecasting in Europe, using accumulation generation and PSO -particle swarm optimization. Their model showed better accuracy, particularly in small-sample datasets, outperforming statistical methods in renewable energy prediction.

As years passed Researchers shifted from statistical models to machine learning and deep learning since they gave higher accurate results. Such as in the year of 2020 **[4]**, Jian Qi Wang had proposed the same LSTM model but for a long-term forecasting method to predict a time series data with strong periodicity, the model had achieved a lower RMSE compared to other models like BPNN, ARFIMA and ARMA discussed in the paper.Also in 2021 **[5]**, Ashutosh Kumar Dubey has discussed the use of a deep learning based LSTM model and a statistical based SARIMA models, where LSTM had proven to give better performance with Average MAE values of 0.27 and 0.23 in terms of lags and epochs.

Then, to find a better model than LSTM, in 2023 **[6]** It was compared with Bidirectional LSTM by Davi Guimar using a univariate time series data from various countries, was found that BLSTM performed better in terms of Normalized RMSE metric values.Then Ashutosh Kumar Dubey has discussed the use of a deep learning based LSTM model and a statistical based SARIMA models, where LSTM had proven to give better performance with Average MAE values of 0.27 and 0.23 in terms of lags and epochs.

The change in deep learning approaches for predicting residential energy consumption shows a progression of increasingly complex models over time. Kim and Cho **[7]** initially proposed a CNN-LSTM neural network that combined LSTM for temporal patterns with CNNs for spatial information extraction, achieving a MSE of 0.37 for minute-level predictions. Improving this, Khan et al.**[8]** developed a hybrid CNN-LSTM-AE model that improved performance, particularly for hourly predictions, with a MSE of 0.19 on the UCI dataset. Rui Gonçalves et al. **[9]**then introduced the Variable Split Convolutional Attention (VSCA) model, which incorporated a novel attention mechanism and outperformed previous models including ARIMA, LSTM, and WaveNet. Finally, Cascone et al.**[10]** proposed a hybrid ConvLSTM model that utilised Social Internet of Things (IoT) data for multi-fold time series analysis, achieving an RMSE of 367 kilowatts for weekly household consumption predictions, further advancing the field of energy consumption forecasting.

Then few of the studies also show the experiment with hybrid models performing better than traditional models for all kinds of time series datasets. In 2021 **[11]**, a study using 16 different time series datasets by Hossein Abbasimehr involved using a hybrid model combining LSTM and Multi-head attention, where it was noted that in comparison with other traditional models like ARIMA, ETS and MLL etc., The hybrid model proposed had outperformed in terms of SMAPE metric and achieved the best AR among all other models.Then in 2023 **[12]**, In case of data's with seasonal fluctuations, Yijue Sun has proposed a hybrid model on seasonal and trend decomposition using Loess(STL) algorithm and gray model, which ended by giving higher prediction accuracy. The final results gave the MAPE and RMSPE values as 1.77% and 2.37% for the model respectively. Then In the same year **[13]**,study on a advanced neuro-evolutionary NARX (NE-NARX) model was proposed by Nguyen Ngoc Son to predict a hourly energy consumption data, which resulted in a better performance than other models like LSTM, CNN, GRU etc..In terms of metrics like MAPE and MSE (0.0046 and 11.49). Then, Ullah et al. **[14]** introduced a hybrid CNN and Multi Layer Bi-Directional LSTM (M-BDLSTM) model applied to household power consumption which gave an MSE of 0.3738 and RMSE of 0.6114 which significantly improved accuracy over existing methods. Liu et al. **[15]** introduced an intelligent framework for forecasting quarterly and monthly energy consumptions by combining machine learning and grey system theory which gave a MAPE of 3.45%,RMSE of 12.34. Bhoj and Bhadoria **[16]** introduced a CNN-GRU hybrid model, which gave the best daily granularity results with an MSE of 0.041, while LSTM performed best for hourly predictions with an MSE of 0.114. Xu et al. **[17]** proposed a hybrid EMD-FRBF-AR model for nonlinear time series which achieved an MSE of 0.0521 and R^2 of 0.9835, showing strong forecast precision compared to other methods.Showing that hybrid models tend to give more good results.

A range of approaches have been suggested for forecasting short term energy consumption with each one tackling a different set of challenges in energy demand prediction. Shan et al. **[18]** developed an ensemble GRA_GRU model which was tested on building data and demonstrated a MAPE of 5.24%.Alghamdi et al. **[19]** developed a stacked LSTM snapshot ensemble with FFT for feature extraction, achieving an RMSE of 0.020, MAE of 0.013,MAPE of 0.017 and R^2 of 0.999. K.Yan et al. **[20]** integrated LSTM networks with Stationary Wavelet Transform (SWT) for individual household data, achieving an RMSE of 0.1234 and MAPE of 5.67%.This led to the effective addressing of unpredictable human behaviour in energy consumption. The gaps found in the research papers are:-

The dataset is limited and lack of vast coverage and lack of multiple data point coverage at the same time.Handling non stationary data is another challenge ,Also most of the models are more focused on short term energy consumptions as there are only few based on long term predictions. We will be addressing some of these gaps in our further research of the paper.

3. DATASET DESCRIPTION

The Dataset for our topic was found after extensive research, finalized with a dataset from kaggle for energy consumption in PJM Interconnection LLC (PJM) which is a regional transmission organization (RTO) in the United States and It is within Eastern Interconnection grid operating an electric transmission system serving different regions. The dataset covers the hourly energy consumption from 2004 to 2011.

So for our study we decided to choose a region for which there are not many predictions done so far, so opted for Northern Illinois Hub as the region.**[21]**

The dataset contained two columns:

- Date column containing the date and the hour of the day.
- NI_MW (Energy consumption measured in megawatts MW).

We have a total of 58450 rows in the dataset with zero null/missing values present.An example view of the dataset is shown below in **Figure 1.1.**

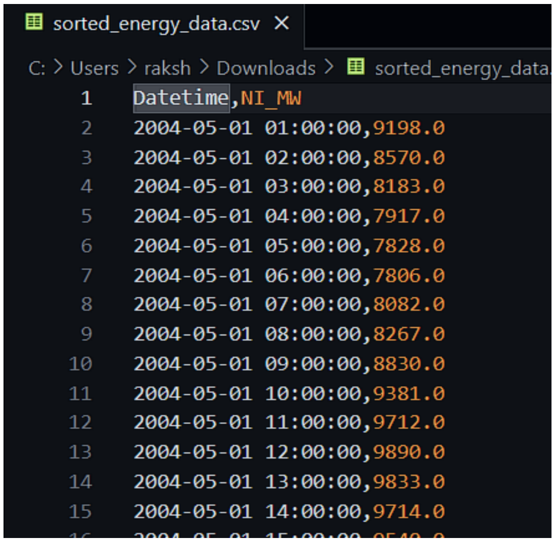

Figure 1.1: Represents the dataset in .csv format

4. METHODOLOGY

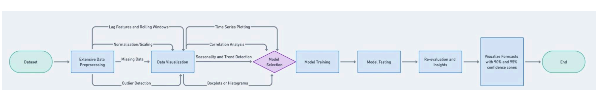

Figure 1.2 : Ideal Work-flow Diagram

This study employs advanced time series analysis techniques and machine learning approaches to forecast energy consumption. The methodology is structured into distinct phases to ensure a systematic approach to analyzing and modeling temporal patterns in electricity usage data.

4.1. Time Series Analysis

The exploratory data analysis (EDA) performed involves a variety of visualizations to understand the patterns, trends, and seasonality in power consumption data.

- Firstly , a raw time series plot provides the hour of consumption in a given series to visualize the fluctuations occurring during the progression of time. Furthermore, daily and monthly resampled averages smooth the variability in the short run, hence exposing the pattern that may have appeared in a long run. A heatmap has been used to summarize consumptions aggregated across different entities and temporal characteristics which maps consistent temporal dependencies. For identifying monthly seasonalities of the consumption series there has been several time series plotted for multiple years which has showed changes in month's consumption pattern to disclose their recurring annual patterns.

- Autocorrelation (ACF) and partial autocorrelation (PACF) plots quantified the relationship between data points over different lags, aiding in model selection for forecasting.

And to effectively model electricity consumption, the data is treated as a time series, and its key components are analyzed. The decomposition of the series includes:

- **Trend:** The long-term progression of the series.
- **Seasonality:** Regular patterns repeating at fixed intervals.
- **Residuals:** Random variations remaining after removing the trend and seasonality.

The **Seasonal-Trend decomposition using LOESS (STL)** is utilized to separate these components, providing better insight into the underlying patterns and simplifying subsequent modeling.

Lastly, a heatmap of hourly consumption by day of the week highlighted intraday and weekday-specific usage trends which provided insights into operational cycles. These analyses collectively identify critical consumption patterns.

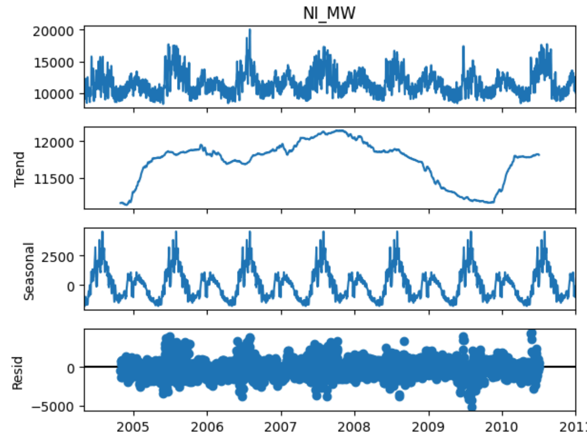

Figure 1.3: Decomposition plot to understand trends and seasonality

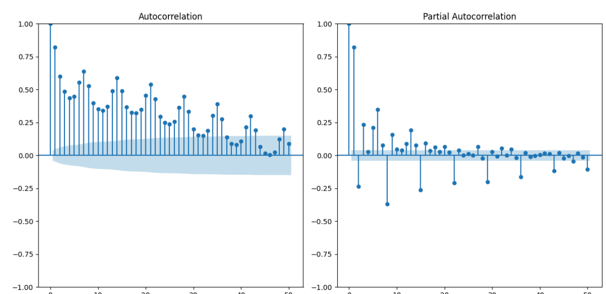

Figure 1.4: Autocorrelation and Partial Autocorrelation Plots

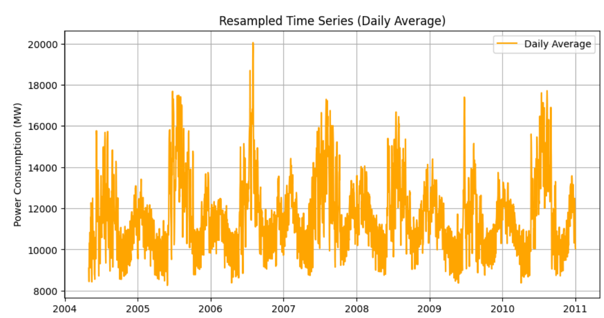

Figure 1.5: Daily average for clarity in long-term trends

4.2 Preprocessing

- **Creating lag variables:** The time series forecasting model incorporates comprehensive temporal feature engineering. For lag variables, the model utilizes the previous 24 hours of power demand (NI_MW) as predictors, creating features named 'lag_1' through 'lag_24'. These lag features capture recent historical patterns and autocorrelation in the power demand. The target variable is set as the power demand 24 hours into the future, enabling day-ahead forecasting. Temporal features are incorporated by extracting cyclical components from the datetime index, specifically the hour of the day (0-23) and day of the week (0-6), which helps capture daily and weekly seasonality patterns in power demand. The data is carefully preprocessed to maintain temporal order, with the most recent year reserved for testing (365 * 24 hours). This feature engineering approach combines autoregressive elements (through lag variables) with temporal patterns (through time-based features) to provide the model with a rich set of predictors for forecasting future power demand.

- **Creating Sequences:** First of all, the data was prepared by creating sequences of a specified length. 24 time steps in case of LSTM and the hybrid model discussed in later sections, and 72 time steps which were considered in the case of GRU. The target value serves as the label for each sequence, while the previous 24 time steps of data serve as features. This will allow the model to understand patterns over time, using the data from earlier time steps to predict future values.
- **Normalizing Data:** Normalization is a very important preprocessing step in time-series forecasting, especially when neural networks are used. The Min-Max scaling technique has been applied to the column NI_MW for normalization within the range of 0 to 1 in both training and test datasets, in a way that enhances model performance by keeping all input features on a similar scale. For this, we use the MinMaxScaler of sklearn.preprocessing. We call the fit transform method on the training set, and we create only a transform method for the dataset of tests to make absolutely sure these two datasets turn out properly scaled against its training partner.
- **Splitting Preprocessed data into a training-validation set and testing set:** This is quite a common practice to make sure that the model sees a lot of data while the test set remains unseen for the evaluation of the performance of the model. Another validation split is created from the training data, amounting to 20% of the training set, in order to make sure that the model can be evaluated on the validation set during training to monitor for overfitting.
- **Reshaping:** The neural network, more precisely a time-series model such as LSTM, takes a certain shape for input data. The required shape for this input data is: [samples, time steps, features]. After creating sequences, the data is reshaped to meet this requirement. The

reshaping of sequences by np.reshape means that each sequence now contains multiple time steps—24 in this case—and one variable. This is a necessary step to ensure that the data will be in a format that can be read by the LSTM model. Once reshaped, data is ready for use in the model, first for training and then for prediction.

4.3 Model Development

A diverse range of machine learning models is implemented and compared to achieve accurate energy consumption forecasting:

4.3.1 Machine Learning Models

We experimented with five different models that include two tree based models and three neural network based models on our dataset and thereby optimizing them through hyperparameter tuning in order to achieve the best possible results. The models utilized in our analysis are as follows:

4.3.1.1 Tree models:

- **XGBoost:** Leveraging time-based features for gradient-boosted tree models.

  The model implements a gradient boosting framework with 100 sequential trees, each having a maximum depth of 6 to control complexity. It employs a conservative learning rate of 0.1 to ensure stable convergence while building the ensemble. The model processes the same feature set as Random Forest but differs in its sequential learning approach, where each tree corrects the errors of previous trees. This architecture particularly excels at handling the non-linear relationships in power demand patterns while maintaining computational efficiency through its gradient-based optimization.

Hyperparameters	**Values**
n_estimators	100
max_depth	6
learning_rate	0.1
Forecast Horizon	24 hours ahead
Validation Period	365*24 hours

- **Random Forest:** Using ensemble learning to handle variability.

  The model is constructed with an ensemble of 100 decision trees, utilizing bootstrap aggregation (bagging) to reduce variance. The architecture processes 24 lag features representing hourly power demand values, along with temporal features (hour of day, day of week) to capture cyclical patterns. Each tree independently learns from a random subset of the data and features, with predictions aggregated through averaging to produce the final forecast. This architecture enables the model to capture non-linear relationships in the power demand data while being resistant to overfitting through its ensemble nature.

Hyperparameters	Values
n_estimators	100
Forecast Horizon	24 hours ahead
Validation Period	365*24 hours

4.3.1.2 Neural Networks:

- **LSTM (Long Short-Term Memory):**First, as seen in Figure 1.6, the model is made up of three LSTM layers , each unit being 50.The model starts with an input LSTM layer for the raw sequences of energy consumption input, followed by two successive LSTM layers to learn complicated temporal patterns.To handle the overfitting problem there are Dropout layers with a rate of 0.2 will be applied after each layer by randomly setting some fractions of links to zero when training.The last Dense layer gives us a single output which represents the energy consumption forecasted for the next time step.This architecture avoids long-term dependencies in time-series data and mitigates the vanishing gradient problem, which is faced by traditional neural networks during learning from long sequences.The LSTM architecture can grasp not only the short-term fluctuations but also the long-term trend in the use of energy, crucial for making appropriate predictions, given its ability to retain useful information over an extensive span of time.
- We have defined several important parameters during the training phase: includes epochs that determine how many times the model will iterate over this data and batch_size describes how many samples are fed in before updating the model's weights and hence balancing the speed of training or use of memory. It also utilizes an optimizer for effective adjustment of learning rates, while validation_data=(X_test, y_test) are utilized to assess model performance on unseen data and as such prevents overfitting.

Hyperparameters	Values
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)
Activation	tanh
Epochs	10, 50
Batch_size	64

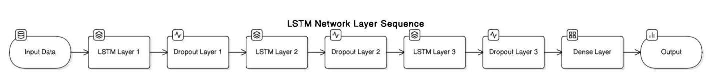

Figure 1.6 : LSTM’s Architecture diagram

- **GRU (Gated Recurrent Unit):** We have considered GRU in this paper, because it is a simplified form of LSTM which is used to capture long-term dependencies but with fewer parameters. This leads to a reduced computational complexity for the model while still considering both short-term fluctuations and long-term trends in the energy consumption data. The use of GRU here brings about the ability to handle sequential dependencies in an effective manner besides the strengths of the LSTM in modeling complex temporal relations.

  Also the model contains an architecture of four GRU layers as shown in Figure 1.7, each of them having 50 units with the tanh activation function that captures complex patterns within a time series. For this reason, the first three GRU layers in this implementation set return_sequences=True. Such a setting ensures the fact that the model passes output sequences coming from each time step to the next layer which allows it to capture temporal dependencies across multiple layers easily. To prevent over-fitting, dropout layers with a rate of 0.2 are attached .

  In this model, the optimizer used is SGD(Stochastic Gradient Descent) with a learning rate of 0.01, momentum of 0.9 and weight decay of 1e-7 for regularization. Batch size is set to 64, which regulates the number of samples that have to be processed before updating the weights of the model and balances training efficiency and memory usage. The model was trained up to 20 epochs with further experiments which were conducted at 30 and 60 epochs to show the influence of longer training on performance.

Hyperparameters	**Values**
Optimizer	SGD optimizer
Loss Function	mean_squared_error
Activation	tanh
Epochs	20, 30, 60
Batch_size	64

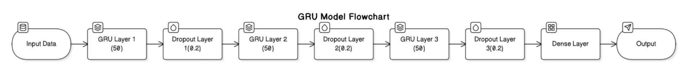

Figure 1.7: GRU’s Architecture diagram

- **Hybrid Model:** We have explored a hybrid model using the convolutional neural network(CNN) together with the recurrent neural network (RNN) for complex temporal pattern capture from the power consumption data. This approach leverages the strengths of both networks, the CNN acts as a feature extractor to identify the most critical patterns and trends in short windows of time, while the RNN, especially LSTM—a type of RNN—can grasp long-term sequential dependencies and dynamics. The CNN extracts key features at each time step, while the LSTM as seen previously learns from the sequential nature of energy consumption patterns and thus allowing it to predict future values by incorporating both local and long-term temporal relationships.

The architecture consists of two initial 1D convolutional layers followed by max pooling as shown in Figure 1.8, helping to capture local temporal patterns and reducing the complexity of retaining only the most vital features. Further, these are then processed through the LSTM layer, which is designed to handle such sequences and long-term dependencies, hence capturing the temporal structure of the data. Dropout after LSTM prevents overfitting, while the final dense layer provides a single prediction representing the next value in the series. The combination of CNN and LSTM ensures both the short-term fluctuations and the long-term trends in the power consumption data to be handled effectively.

The hybrid model uses the Adam Optimizer for efficiency during training. The loss function used here is mean_squared_error which is the standard for regression tasks .The learning rate is kept at 0.001 which provides a good balance between speed and stability. ReLU activation prevents vanishing gradients and speeds up the process of training. We have tested the model with epochs of 60, 80, and 100 to see how the model behaves at these different training durations. At the same time, it utilizes batch sizes of 64 and 32 respectively for better use of memory and to provide faster convergence.

Hyperparameters	**Values**
Optimizer	Adam
Loss Function	mean_squared_error
Learning Rate	0.001
Activation	Relu
Epochs	60, 80, 100
Batch_size	64, 32

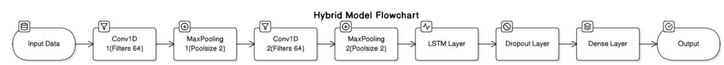

Figure 1.8 : Hybrid model's Architecture

5. Implementation Details

5.1 Software and Tools

The implementation was performed using the following tools and technologies:

- **Programming Language**: Python.
- **Libraries**:
  - Data manipulation and preprocessing: pandas, numpy, sklearn.
  - Visualization: matplotlib, seaborn, statsmodels.
  - Machine learning: xgboost, scikit-learn.
  - Deep learning: keras, tensorflow.
  - Time series analysis: statsmodels, prophet.

- **Development Environment**: Google Collab was used for executing scripts and experiments.
- **Hardware Specifications**:
  - Google Colab's virtual machine with access to NVIDIA Tesla T4 GPU for training deep learning models.

6. Results and Discussion

The performance of each model is evaluated using the following metrics:

- **Cross-validation with Time-Series Split:** To ensure reliable evaluation over time-dependent data.
- **RMSE (Root Mean Square Error):** For absolute error measurement with higher penalty on large errors.
- **MAE (Mean Absolute Error):** For direct average error measurement in original units.
- **MAPE (Mean Absolute Percentage Error):** For relative error measurement as a percentage.
- **R^2 (R-squared):** For measuring proportion of variance explained by the model.
- **SMAPE (Symmetric Mean Absolute Percentage Error):** For balanced percentage error measurement that handles zero values.

These metrics in Figure 1.9 provide a comprehensive assessment of model accuracy and robustness. And the rest of the Figures from Figure 2.0 to Figure 2.4 shows the visual representation of the results.

Model	RMSE	MAE	MAPE (%)	sMAPE (%)	R^2
GRU	168.66	121.13	1.04	1.04	0.99
HYBRID LSTM	353.75	247.46	2.11	2.11	0.98
LSTM	190.00	133.97	1.15	1.15	0.99
Random Forests	942.64	597.28	4.75	4.74	0.86
XGBoost	924.02	595.93	4.76	4.76	0.86

Figure 1.9: Metric comparison among all the models, where GRU model seems the best

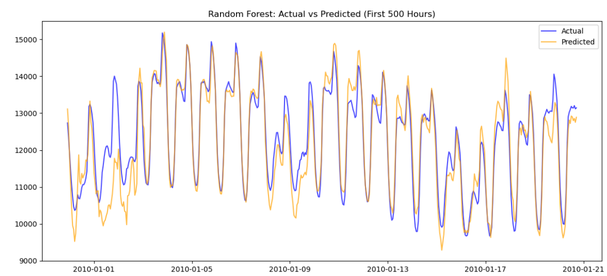

Figure 2.0: Random forest model's results

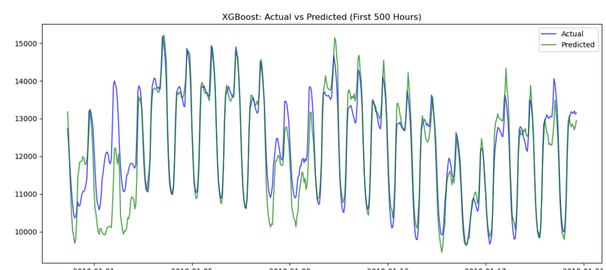

Figure 2.1: XG Boost model's results

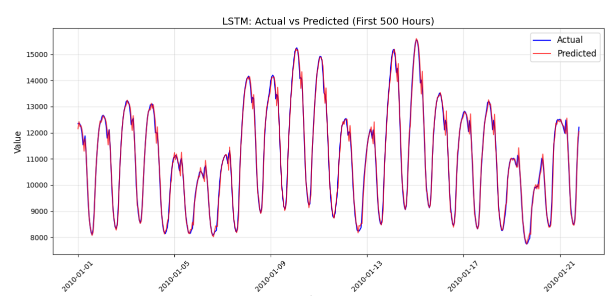

Figure 2.2: LSTM model's results

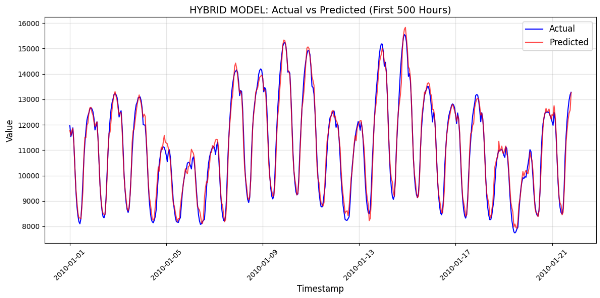

Figure 2.3: Hybrid model's results

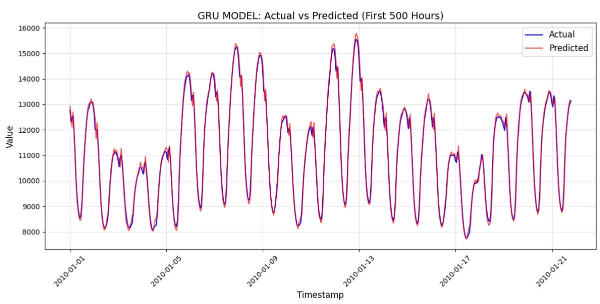

Figure 2.4: GRU model's results

7. CONCLUSIONS

In conclusion, our study demonstrates that non-statistical models, such as machine learning and deep learning approaches, outperform traditional statistical models for time series analysis in energy consumption prediction, as supported by our literature review. Building on this insight, we analyzed our dataset using five cutting-edge models: XGBoost, Random Forest, LSTM, a hybrid RNN-CNN model, and GRU.

We evaluated these models using five performance metrics like RMSE, MAE, MAPE, sMAPE, and R^2 and the results obtained highlight GRU as the best-performing model. The GRU model achieved exceptional accuracy with RMSE: 168.66, MAE: 121.13, MAPE: 1.04%, sMAPE: 1.04%, and R^2: 0.99 and thereby surpassing the performance of the other models. These findings clearly show the effectiveness of GRU for precise energy consumption prediction.

7. REFERENCES

[1] S. Ozturk, F. Ozturk, Forecasting energy consumption of Turkey by ARIMA model, Journal of Asian Scientific Research, Volume 8, 2018.

[2] Chong Liu, Wen-Ze Wu, Wanli Xie, Jun Zhang,Application of a novel fractional grey prediction model with time power term to predict the electricity consumption of India and China,Chaos, Solitons & Fractals,Volume 141,2020.

[3] Lianyi Liu, Lifeng Wu,Forecasting the renewable energy consumption of the European countries by an adjacent non-homogeneous grey model,Applied Mathematical Modelling,Volume 89, Part 2,2021.

[4] Jian Qi Wang, Yu Du, Jing Wang, LSTM based long-term energy consumption prediction with periodicity, Energy, Volume 197, 2020.

[5] Ashutosh Kumar Dubey, Abhishek Kumar, Vicente García-Díaz, Arpit Kumar Sharma, Kishan Kanhaiya, A study and analysis of SARIMA and LSTM in forecasting time series data, Sustainable Energy Technologies and Assessments, Volume 47, 2021.

[6] Davi Guimarães da Silva, Anderson Alvarenga de Moura Meneses, Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction, Energy Reports, Volume 10, 2023

[7] Tae-Young Kim, Sung-Bae Cho,Predicting residential energy consumption using CNN-LSTM neural networks,Energy,Volume 182,2019.

[8] Khan ZA, Hussain T, Ullah A, Rho S, Lee M, Baik SW, Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework, Sensors, 2020.

[9] Rui Gonçalves, Vitor Miguel Ribeiro, Fernando Lobo Pereira,Variable Split Convolutional Attention: A novel Deep Learning model applied to household electric power consumption,Energy,Volume 274,2023.

[10] Lucia Cascone, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Hafeez Ur Rehman Siddiqui, Muhammad Umer,Predicting Household Electric Power Consumption Using Multi-step Time Series with Convolutional LSTM,Big Data Research,Volume 31,2023.

[11] Hossein Abbasimehr, R. Paki, Improving time series forecasting using LSTM and attention models, Journal of Ambient Intelligence and Humanized Computing, Volume 13, 2022.

[12]Yijue Sun, F. Zhang, Quarterly electricity consumption prediction based on time series decomposition method and gray model, Environmental Science and Pollution Research, Volume 30, 2023.

[13] Nguyen Ngoc Son , N. Van Cuong, Neuro-evolutionary for time series forecasting and its application in hourly energy consumption prediction, Neural Computing and Applications, Volume 35, 2023.

[14] Fath U Min Ullah, Amin Ullah, Ijaz Ul Haq, Seungmin Rho, Sung Wook Baik, Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks, IEEE Access, Volume 8, 2020.

[15] Chong Liu, Hegui Zhu, Yuchen Ren, Zhimu Wang, A Novel Intelligent Forecasting Framework for Quarterly or Monthly Energy Consumption, IEEE Transactions on Industrial Informatics, Volume PP, 2023.

[16] Naman Bhoj, Robin Singh Bhadoria, Time-series based prediction for energy consumption of smart home data using hybrid convolution-recurrent neural network, Telematics and Informatics, Volume 75, 2022.

[17] Wenquan Xu, Hui Hu, Wei Yang, Energy Time Series Forecasting Based on Empirical Mode Decomposition and FRBF-AR Model, IEEE Access, Volume 7, 2019.

[18] Shubing Shan, Buyang Cao, Zhiqiang Wu, Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model, IEEE Access, Volume 7, 2019.

[19] Mona Ahamd Alghamdi, Abdullah S. AL–Malaise AL–Ghamdi, Mahmoud Ragab, Predicting Energy Consumption Using Stacked LSTM Snapshot Ensemble, Big Data Mining and Analytics, Volume 7, Issue 2, 2024.

[20] Ke Yan, Wei Li, Zhiwei Ji, Meng Qi, Yang Du, A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households, IEEE Access, Volume 7, 2019.

[21] PJM Interconnection LLC, Hourly Energy Consumption Data for over 10 years (2006–2018) in the PJM region (in Megawatts), Kaggle, 2018. Available at: https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption.