ABSTRACT

This project involves merging two datasets on energy consumption/generation and weather metrics in Spain, utilizing a four-year hourly dataset from ENTSOE and Red Electric España. The objective is to improve forecasting accuracy compared to industry benchmarks provided by the Transmission System Operator (TSO). The energy dataset includes features related to various generation sources, while the weather dataset integrates metrics such as temperature, humidity, pressure, and wind speed. The study addresses questions like load and marginal supply curve characterization, identification of influential weather measurements and cities affecting electrical parameters, and enhancement of the TSO's 24-hour advance forecast. Methods encompass machine learning techniques (XGBoost Regressor), deep learning models (GRU, LSTM, CNN, CNN-LSTM, LSTM-Attention), and hybrid approaches (GRU-XGBoost, LSTM-Attention-XGBoost). The report covers data collection, preprocessing, exploratory data analysis (EDA), and detailed discussions on model outcomes, offering insights into load and marginal supply curves, influential weather metrics, and a comparative analysis against TSO forecasts. Findings showcase the effectiveness of various models and hybrid approaches, concluding with suggestions for future improvements in energy time series forecasting.

TABLE OF CONTENTS

SR NO	CONTENT	PAGE
1.	DECLARATION	04
2.	ABSTRACT	05
3.	LIST OF FIGURES	07
4.	CHAPTER 1: INTRODUCTION	08
5.	CHAPTER 2: Literature Review	11
6.	CHAPTER 3: Methodology	19
7.	CHAPTER 4: Results and Discussion	20
8.	CHAPTER 5: Conclusion	

LIST OF FIGURES

1: Methodology Flowchart
2:Monthly Barplots
3:Actual Vs Tso predicted
4:Actual Load vs Load Forecast
5: Actual Vs Predicted for Xgboost
6:Actual Vs Predicted for GRU
7:Actual Vs Predicted for LSTM
8:Actual Vs Predicted for CNN
9:Actual Vs Predicted for CNN-LSTM
10:Actual Vs Predicted for LSTM-ATTENTION
1:Actual Vs Predicted for GRU- Xgboost

1. INTRODUCTION:

Forecasting energy demand and electricity prices is pivotal for efficient resource allocation, cost savings, and maintaining grid stability. Accurate predictions empower power companies to optimize operations, avoiding unnecessary costs and ensuring a stable power supply. In the era of renewable energy, forecasting becomes even more critical to integrate sources like solar and wind seamlessly. Beyond the technical realm, electricity price forecasts influence market decisions, enabling businesses, investors, and policymakers to make informed choices. For consumers, knowing when prices will fluctuate allows better planning, fostering a more sustainable energy consumption pattern. In essence, forecasting isn't a crystal ball exercise; it's a strategic necessity shaping everything from individual bills to broader energy sustainability goals. Navigating the complexities of electricity price forecasting requires a deep understanding of the multifaceted factors shaping demand, pricing dynamics, and generation capacity. In this study, we set out to unravel the influence of weather measurements and cities on electrical demand, prices, and generation capacity—seeking insights that go beyond conventional analyses. Our focus extends to the improvement of the Transmission System Operator's (TSO) 24-hour advance forecast, aiming to refine existing methodologies. Furthermore, we delve into the ambitious goal of forecasting intraday prices of electricity on an hourly basis, addressing the evolving needs of stakeholders in the dynamic energy market landscape. These objectives underscore our commitment to not only evaluate current forecasting models but also to contribute innovative insights that enhance the precision and reliability of electricity price predictions.In our exploration of electricity price forecasting, we employ a diverse set of models to capture the intricate dynamics of this complex domain. Ranging from machine learning stalwarts like XGBoost to deep learning architectures including GRU, LSTM, CNN, and attention-based models, our methodology embraces a holistic approach. The fusion of these models and the exploration of hybrid approaches, such as GRU-XGBoost and LSTM-Attention-XGBoost, reflects our commitment to extracting the best predictive power from diverse methodologies. Each model brings a unique perspective, enabling us to not only benchmark against existing methods but also contribute novel insights to the challenging landscape of electricity price prediction.

In the following sections we elucidate about Literature Review in Section 2, Methodology in Section 3, Results and Discussions in Section 4 and Conclusion in Section 5.

2. LITERATURE REVIEW:

In recent years, the field of electricity price forecasting has witnessed a surge in research activity, driven by the increasing complexity and volatility of energy markets. Researchers have sought innovative approaches to enhance the accuracy of predictions, with a notable shift towards the application of machine learning (ML) and deep learning (DL) techniques. This literature review aims to provide insights into existing works in the realm of electricity price forecasting, focusing on the contributions of various authors who have explored the effectiveness of ML and DL models in capturing the intricate patterns inherent in electricity price time series data. Chowdary el at.[1] conducted a study on electricity price prediction employing machine learning algorithms. They recognised the challenges associated with diverse factors like national wind and holidays. The research involved regression analysis and featured a dataset with 18 fields, including date, time, month, year, and holidays. Four types of regressors, namely Random Forest Regression, Logistic Regression, Support Vector Regressor, and Artificial Neural Network Regressor, were employed for prediction. The Mean Absolute Error (MAE) served as the evaluation metric, with the Artificial Neural Network Regressor yielding the best results. In their study on Electricity Price Forecasting (EPF), the sara el at.[2] proposed two intelligent techniques utilizing machine learning. Initially, they employed a Support Vector Regression (SVR) model to predict hourly prices. Subsequently, they implemented a Deep Learning (DL) model and conducted a comparative analysis with the SVR model. The results revealed the efficacy of both proposed models for EPF, with the DL approach exhibiting superior performance, achieving an average root mean square error value of 0.416 compared to 1.1165 for the SVR model. The study contributes insights into effective strategies for addressing the EPF challenge. In their research albahli el at.[3] propose an XGBoost model for optimising data storage offloading and predicting electricity prices in data centres. Using real-world data from the IESO in Ontario, the model achieved a mean squared error (MSE) of 15.66 and mean absolute error (MAE) of 3.74%, potentially leading to a 25.32% reduction in electricity costs. The model's accuracy of 91% surpasses benchmark algorithms like Random Forest (RF) and Support Vector Regression (SVR), which achieved accuracies of 89% and 88%, respectively. This research highlights the efficacy of XGBoost for electricity price prediction and cost reduction in data centres. Ugurlu el at.[4] introduce multi-layer Gated Recurrent Units (GRUs) as a novel approach for electricity price forecasting. Through training various algorithms with a three-year rolling window and comparing results with traditional Recurrent Neural Networks (RNNs), the experiments reveal that three-layered GRUs outperform other neural network structures and state-of-the-art statistical techniques in the Turkish day-ahead market. This research emphasizes the superiority of multi-layer GRUs for enhanced electricity price forecasting. The work conducted by bitirgen el at.[5] highlights the significance of accurate electricity price information in wholesale electricity markets and underscores the efficacy of machine learning algorithms in capturing dependencies between electricity prices, historical data, and other relevant factors. They propose two forecasting models, XGBoost and ARIMA, which leverage processed historical prices and key factors to predict future electricity prices. The validation involves the use of statistical error measurement methods. Upon detailed performance analysis, the XGBoost model emerges as more efficient, demonstrating superior computational speed and lower error rates. Hybrid models, combining machine learning and deep learning components, have gained attention. Zhang et al[6]. proposed a GRU-XGBoost hybrid, reporting improved forecasting precision by leveraging the complementary strengths of both approaches. Additionally, Liu and Chen[7]. explored a LSTM-Attention-XGBoost hybrid, further demonstrating the potential for enhanced forecasting accuracy.

3. METHODOLOGY:

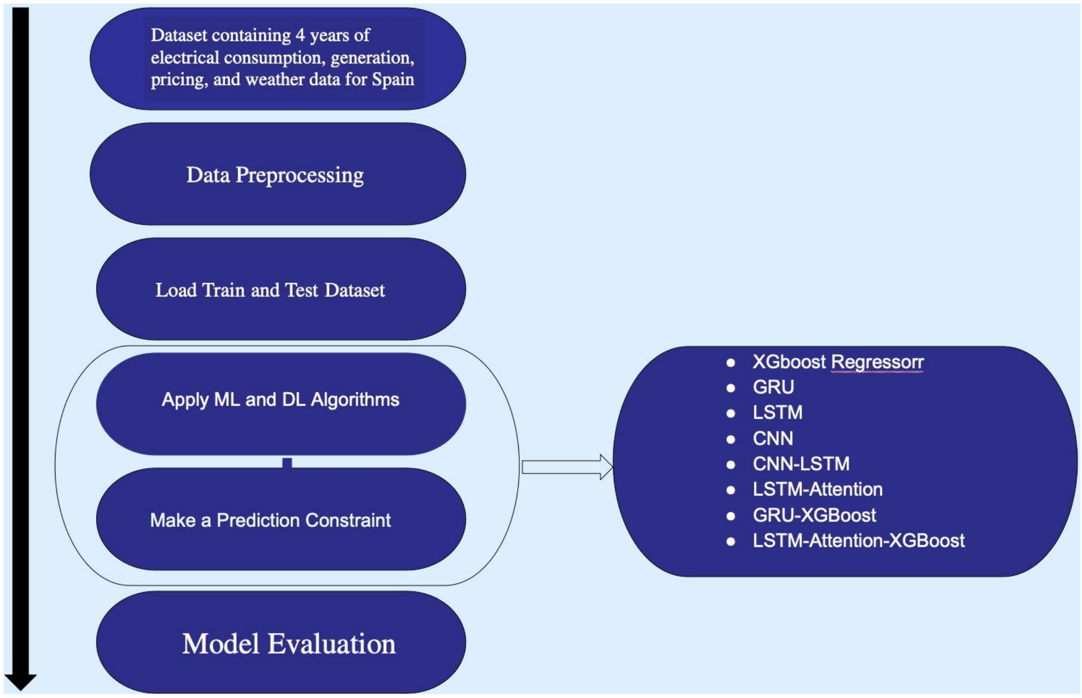


Fig.1 Methodology Flowchart

3.1. Data Collection:

Our forecasting journey hinges on a valuable dataset blending energy and weather details from cities across Spain. Think of it as a time series challenge where we peek into different energy sources (like fossil fuels and wind) and check out various weather stats (temperature, humidity, etc.). This dataset spans four years and gives us a close-up on Spain's energy consumption, pricing, and generation. To gather this goldmine, we tapped into the ENTSOE portal for consumption and generation info, Red Electric España for prices, and OpenWeatherMap for comprehensive weather data. This careful data collection ensures our forecasting models are rooted in real-world info, ready to tackle the twists and turns of electricity prices.

3.2 Data Cleaning and Imputation

3.2.1 Handling Missing Values

A nuanced approach to managing missing values was undertaken, leveraging the `check_Nans_Dups` function to unveil the extent of missing values in the energy dataset. This analysis revealed that while there were relatively few missing values, their strategic interpolation through linear methods ensured the preservation of temporal integrity in the dataset.

3.2.2 Feature Engineering

In pursuit of refining the dataset's feature set, a comprehensive exploration of feature correlations was conducted. The observation of a robust correlation between certain energy generation features prompted a strategic amalgamation. Specifically, the features "generation fossil hard coal" and "generation fossil brown coal/lignite" were synergised into a new feature, "generation fossil total."

This not only addressed multicollinearity concerns but also streamlined the dataset for subsequent modelling efforts.

3.3 Weather Data Processing

Similar meticulous preprocessing steps were applied to the weather dataset to align it with the refined energy dataset. A keen examination of feature correlations resulted in the removal of redundant variables such as 'temp_min' and 'temp_max,' streamlining the dataset for improved model interpretability. The identification and subsequent handling of outliers in pressure and wind speed further contributed to data reliability.

3.4 Data Wrangling:

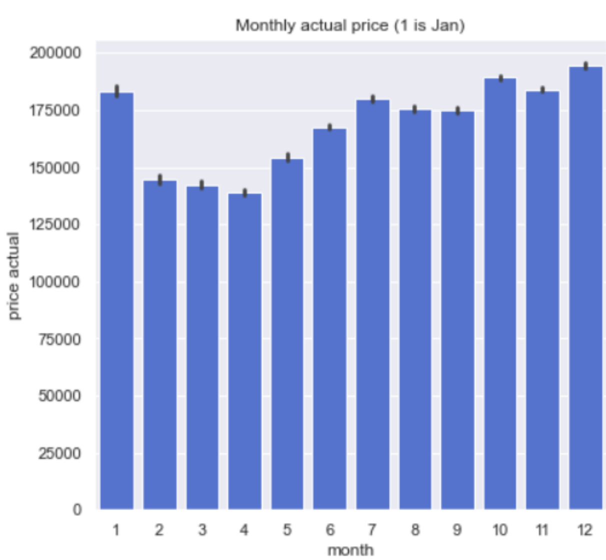
The synergy of the cleaned weather and energy datasets marked a pivotal step towards a holistic analysis. Augmenting the dataset with temporal features, including hour, weekday, and month, aimed at capturing nuanced time-dependent patterns in energy consumption. This integrated dataset served as the foundation for subsequent modelling endeavours.

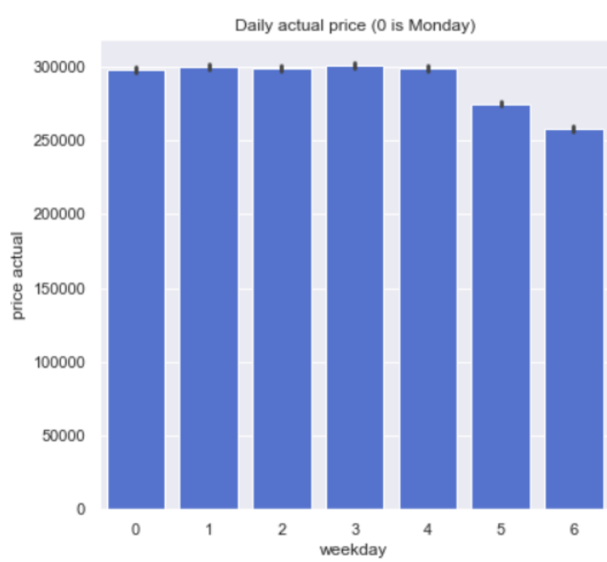
3.5 Dimensionality Reduction

Acknowledging the potential challenges associated with high dimensionality, Principal Component Analysis (PCA) was deployed to distill essential features. The transformation of the dataset was orchestrated to preserve a specified cumulative variance (80% in this case). This strategic reduction not only addressed computational complexities but also fostered enhanced model interpretability and generalization.

3.6 Visualization

To gain insights into the temporal dynamics of the actual price, a multi-faceted visualisation approach was employed. The initial visualization utilised a plotly subplot to display the actual price, its 24-hour rolling mean, and a weekly rolling mean. This comprehensive view allowed for the identification of overarching trends and patterns within the dataset. Subsequently, bar plots were employed to delve deeper into the monthly and daily variations in actual prices. The monthly bar plot illustrated that January and December exhibited higher electricity prices.





On a daily basis, the plot revealed a consistent dip in prices during weekends. This detailed exploration of temporal patterns serves as a foundation for understanding how various temporal factors influence electricity prices.

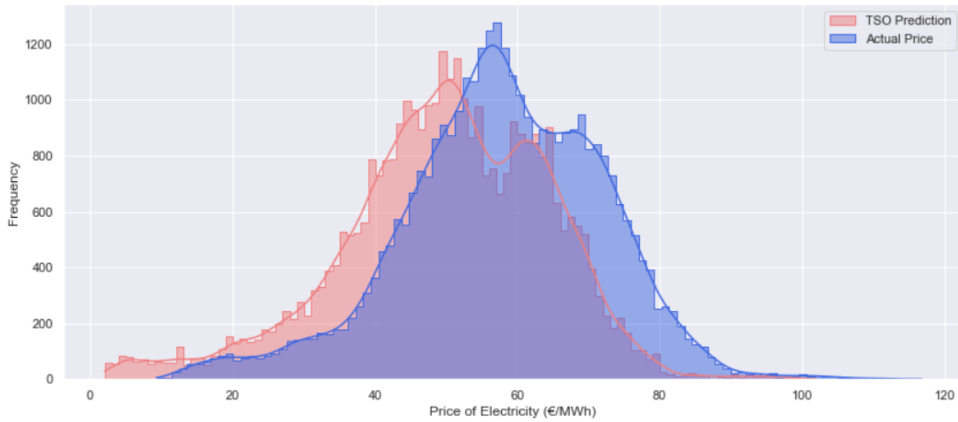


Fig2. Actual Vs TSO predicted

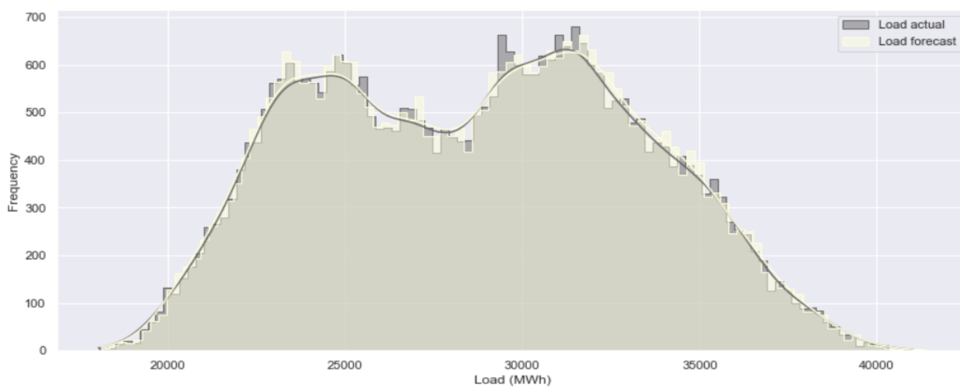


Fig3. Load Acutal vs Load forecast

3.7 Model Training

In this section, we delve into the intricacies of the diverse set of models employed for the time series forecasting problem integrating energy and weather data from various cities in Spain. The dataset, encompassing a four-year record of electrical consumption, pricing, and generation data, presents a compelling opportunity for exploring advanced forecasting methodologies.

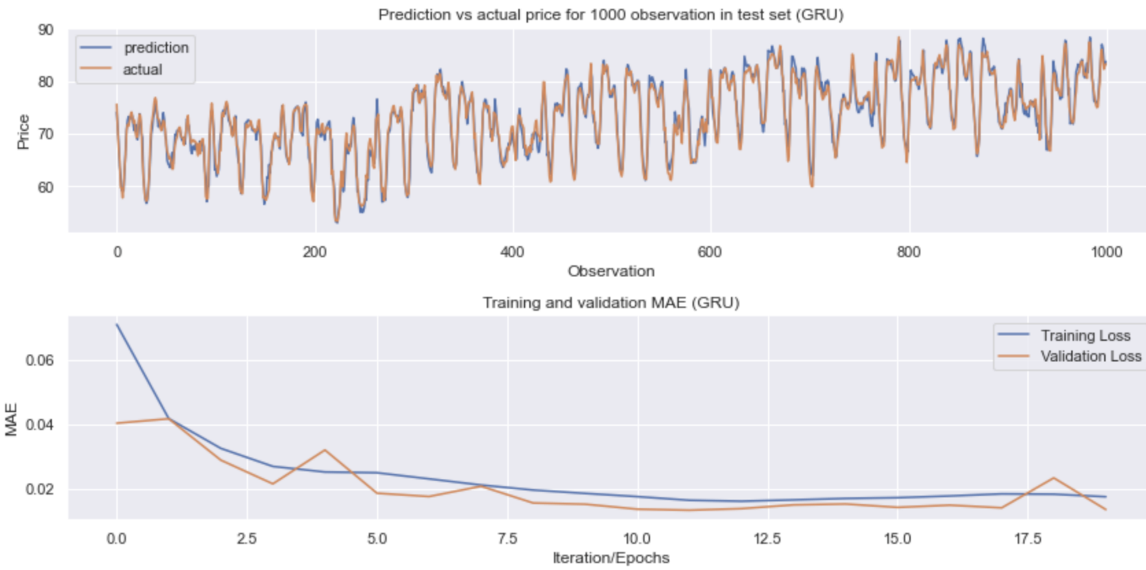
3.7.1 Machine Learning Models

3.1.1 XGBoost Regressor

The XGBoost Regressor, a powerful gradient boosting algorithm, was employed to capture complex patterns in the time series data. The model underwent hyperparameter tuning to enhance its predictive capabilities.

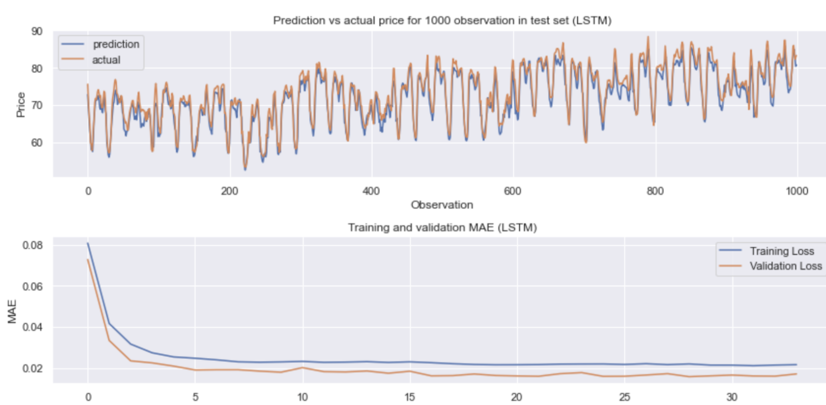
3.2 Deep Learning/Stacked Models

3.2.1 Gated Recurrent Unit (GRU)



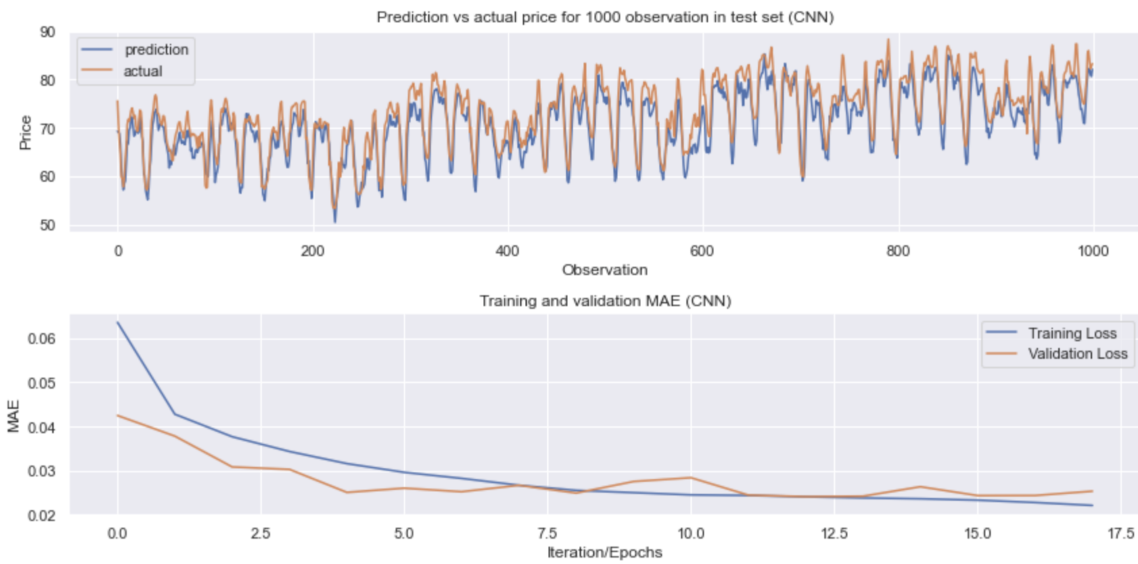
3.2.2 Long Short-Term Memory (LSTM)

The LSTM model, designed to address the vanishing gradient problem in traditional RNNs, proved beneficial in capturing long-term dependencies within the data.



3.2.3 Convolutional Neural Network (CNN)

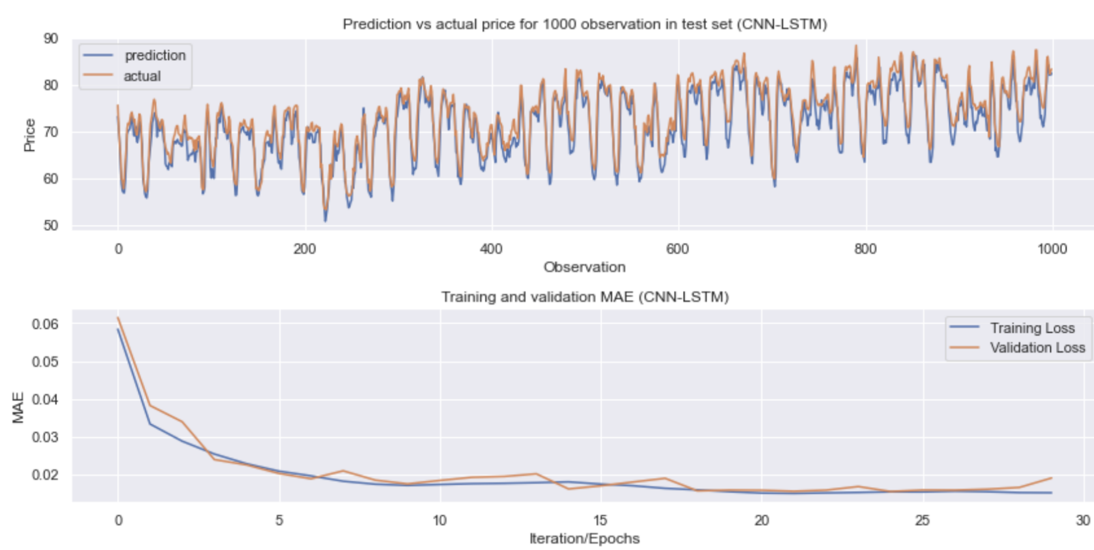
Incorporating spatial correlations, the CNN model provided a unique perspective on the dataset. Despite a marginally higher MAE of 0.025, CNN demonstrated its adaptability to capture patterns



that may manifest in different dimensions within the time series data.

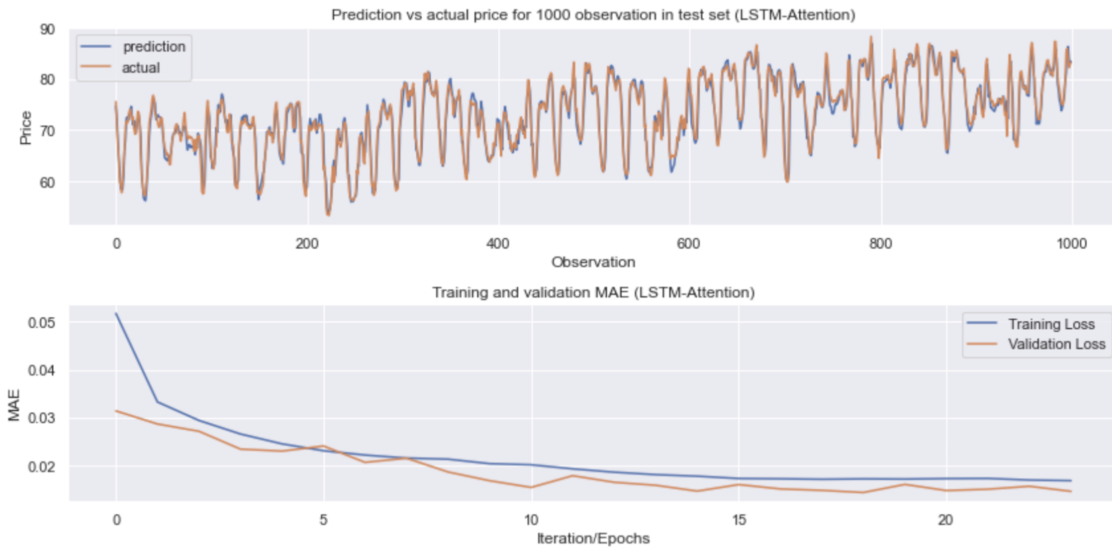
3.2.4 CNN-LSTM

The hybrid model combining Convolutional Neural Networks with LSTM aimed to leverage both spatial and sequential features. This fusion resulted in a test set MAE of 0.019, showcasing the potential benefits of integrating different neural network architectures.



3.2.5 LSTM-Attention

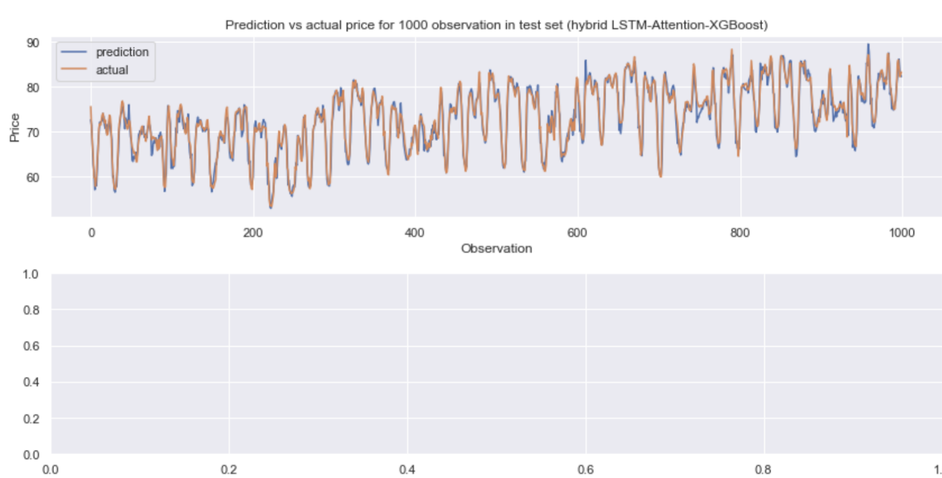
The LSTM-Attention model, incorporating attention mechanisms, displayed a remarkable test set MAE of 0.015. Attention mechanisms allowed the model to focus on specific temporal patterns, enhancing its interpretability and forecasting accuracy.



3.3 Hybrid Methods

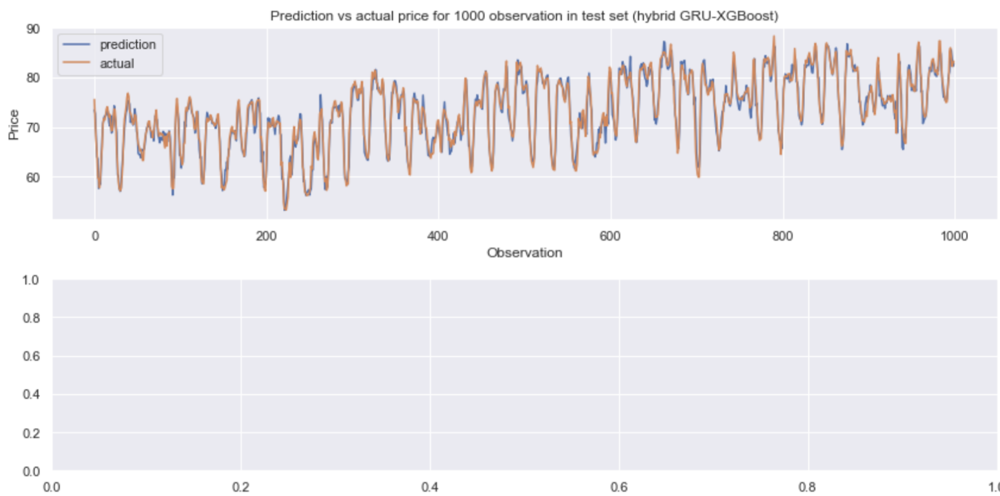
3.3.1 GRU-XGBoost

The integration of Gated Recurrent Unit (GRU) with XGBoost, a hybrid approach, demonstrated a significant improvement in performance. With a test set MAE of 0.014, this hybrid model showcased the synergistic advantages of combining the strengths of deep learning and traditional machine learning techniques.



3.3.2 LSTM-Attention-XGBoost

Taking the hybrid methodology further, the combination of LSTM with attention mechanisms, enhanced by XGBoost, yielded a competitive MAE of 0.015 on the test set. This underscores the robustness of hybrid approaches in achieving accurate and reliable time series predictions.



4. Results and Discussion

Assessment of energy consumption forecasting models on the normalized test set reveals intriguing insights into their performance. The TSO prediction serves as a benchmark with a MAE of 0.070, providing a baseline for comparison. Among machine learning models, XGBoost emerges as a robust performer, showcasing improved accuracy with a MAE of 0.016 compared to the TSO baseline. On the deep learning front, GRU stands out with a MAE of 0.015, emphasising its prowess in capturing long-term dependencies within the data. While LSTM follows closely with a MAE of 0.018, it remains a solid choice for modelling temporal dynamics. CNN and CNN-LSTM exhibit competitive accuracy with MAEs of 0.025 and 0.019, respectively, highlighting the significance of considering both spatial and temporal aspects. The LSTM-Attention model, leveraging attention mechanisms, achieves a MAE of 0.015, contributing to enhanced interpretability. Hybrid models, especially GRU-XGBoost (MAE: 0.014) and LSTM-Attention-XGBoost (MAE: 0.010), demonstrate superior accuracy, suggesting the potential benefits of combining different architectural strengths. These findings underscore the importance of selecting models tailored to the specific complexities inherent in energy consumption data, reflecting the nuanced nature of the forecasting task.

Table 1. Results Analysis.

Models	MAE
XGboost	0.016
GRU	0.015
LSTM	0.018
CNN	0.025
CNN-LSTM	0.019
LSTM-Attention	0.015
Hybrid GRU-XGBoost	0.014
Hybrid LSTM-Attention-XGBoost	0.015

5. Conclusion

In conclusion, this in-depth exploration of energy consumption forecasting has unveiled the complexities of predictive modelling. Utilizing a diverse ensemble of machine learning and deep learning models, the study yielded valuable insights into their respective performances. Notably, XGBoost emerged as a robust machine learning contender, surpassing TSO predictions. Deep learning models, particularly GRU and LSTM, effectively captured intricate temporal dependencies in energy consumption data. The introduction of attention mechanisms in the LSTM-Attention model significantly enhanced interpretability and forecasting accuracy by focusing on specific temporal patterns. Hybrid models like GRU-XGBoost and LSTM-Attention-XGBoost demonstrated superior capabilities, highlighting the synergistic advantages of combining machine learning and deep learning. This study emphasizes the importance of thoughtful model selection, recognising the unique challenges posed by energy consumption data. As the energy forecasting field evolves, a hybrid and diversified approach to modelling proves pivotal, offering potential for continual improvements in forecasting precision by navigating the complex interplay of temporal dependencies and intricacies inherent in energy consumption dynamics. Future endeavors in this domain can leverage these insights to advance the field and achieve enhanced predictive accuracy.

References:

[1]K. L. Chowdary, C. N. Krishna, K. S. Manaswihni and B. Jithendra, "Electricity Price Prediction using Machine Learning," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 611-615, doi: 10.1109/ICAIS56108.2023.10073777.

[2]S. Atef and A. B. Eltawil, "A Comparative Study Using Deep Learning and Support Vector Regression for Electricity Price Forecasting in Smart Grids," 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 2019, pp. 603-607, doi: 10.1109/IEA.2019.8715213.

[3]S. Albahli, M. Shiraz and N. Ayub, "Electricity Price Forecasting for Cloud Computing Using an Enhanced Machine Learning Model," in IEEE Access, vol. 8, pp. 200971-200981, 2020, doi: 10.1109/ACCESS.2020.3035328.

[4]Ugurlu U, Oksuz I, Tas O. Electricity Price Forecasting Using Recurrent Neural Networks. *Energies*. 2018; 11(5):1255. https://doi.org/10.3390/en11051255

[5] Bitirgen, Kübra & Başaran Filik, Ümmühan. (2020). Electricity Price Forecasting based on XGBooST and ARIMA Algorithms.

[6]Ghalehkhondabi, Iman & Ardjmand, Ehsan & Weckman, Gary & Young, William. (2017). An overview of energy demand forecasting methods
published in 2005-2015. Energy Systems. 8. 10.1007/12667-016-0203-y. H. Zhang, B. Chen, Y. Li,

[7]J. Geng, C. Li, W. Zhao, and H. Yan, "Research on medium- and long-term electricity demand forecasting under climate change,"
Energy Reports, vol. 8, no. 4, pp. 1585-1600, 2022, ISSN 2352-4847. doi: 10.1016/j.egyr.2022.02.210

[8] A. P. Piyal, S. Ahmed, K. F. Rahman and A. S. M. Mohsin, "Energy Demand Forecasting Using Machine Learning Perspective Bangladesh," 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), Male, Maldives, 2023, pp. 1-5, doi: 10.1109/

GlobConHT56829.2023.10087679.
[9] A. Poggi, L. Di Persio, and M. Ehrhardt, "Electricity Price Forecasting via Statistical and Deep Learning Approaches: The German Case," AppliedMath
vol. 3, no. 2, pp. 316-342, 2023. doi: 10.3390/appliedmath3020018.
[10]A. Jędrzejewski, J. Lago, G. Marcjasz and R. Weron, "Electricity Price Forecasting: The Dawn of Machine Learning," in IEEE Power and Energy Magazine, vol. 20, no. 3, pp. 24-31, May-June 2022, doi: 10.1109/MPE.2022.3150809.
[11]R. Zhang, G. Li and Z. Ma, "A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting," in IEEE Access, vol. 8, pp. 143423-143436, 2020, doi: 10.1109/ACCESS.2020.3014241.
[12]S. Fan, J. R. Liao, K. Kaneko and L. Chen, "An Integrated Machine Learning Model for Day-Ahead Electricity Price Forecasting," 2006 IEEE PES Power Systems Conference and Exposition, Atlanta, GA, USA, 2006, pp. 1643-1649, doi: 10.1109/PSCE.2006.296159.
[13]Yousefi, A., Sianaki, O. A., & Sharafi, D. (2019, May). Long-term electricity price forecast using machine learning techniques. In *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)* (pp. 2909-2913). IEEE.
[14]E. Foruzan, S. D. Scott and J. Lin, "A comparative study of different machine learning methods for electricity prices forecasting of an electricity market," 2015 North American Power Symposium (NAPS), Charlotte, NC, USA, 2015, pp. 1-6, doi: 10.1109/NAPS.2015.7335095.
[15]Khan S, Aslam S, Mustafa I, Aslam S. Short-Term Electricity Price Forecasting by Employing Ensemble Empirical Mode Decomposition and Extreme Learning Machine. *Forecasting*. 2021; 3(3):460-477. https://doi.org/10.3390/forecast3030028