UNIVERSITÀ DEGLI STUDI DI BERGAMO
DEPARTMENT OF ENGINEERING AND APPLIED SCIENCE
DOCTORAL PROGRAMME IN ENGINEERING AND APPLIED SCIENCES

Methods for Analyzing Electricity Consumption Data and their Implementation in Multi-Platform Architectures

Doctoral Dissertation by:

Ayaz Hussain

Student ID: 1090373

Supervisor:

Prof. Giuseppe Franchini

Co-Supervisor:

Prof. Paolo Giangrande

XXXVIII Ph.D. Cycle

Academic Year 2024-2025

I dedicate this study to the memory of my late parents. Their prayers, sacrifice, and principles are like a bright light to me. Their knowledge and affection remain an unending source of power in my life. All of my successes are a tribute to the foundation that they built for me.

To my wife, whose patience, support, and understanding have been so useful to me in this journey in my absence for so long. The way she supported me and made silent sacrifices during the period I was studying away has given me the motivation to not lose track and continue my progress. The belief that she has placed in me has greatly encouraged me throughout the period of our struggle.

To my daughter, who, being thousands of miles away, never leaves me without comforting me with her love, cheerful heart, and innocence. Her presence in my life has consistently given me inspiration and purpose, even though she was far away while I worked towards this goal.

My siblings have been very helpful to me in my studies, with their constant encouragement, emotional support, and physical assistance. Through their mentorship, wisdom, prayers, and faith, my dreams have helped me stay focused and resilient. It has been their constant inspiration and support that have helped me through the challenges and guided me in achieving this milestone.

Acknowledgements

Firstly, I would like to acknowledge my supervisor, Prof. Giuseppe Franchini, who provided me with guidance, scientific support, and regular encouragement in the course of this PhD program, which was invaluable to me. His scientific knowledge, his patience, and commitment to the highest standards were priceless to my learning and accomplishment of this research piece. My co-supervisor, Professor Paolo Giangrande, with his professional advice, critical remarks, and continuous encouragement, also did a lot of work to improve the quality of this work. The mentoring has played a pivotal role in influencing my research, with a lot of emphasis on scientific research and technological development.

I would like to thank ABB SACE, Bergamo, for giving me the opportunity to work in an industrial research environment. I am especially grateful to Giulia Mandelli, Lorenzo Fenili, Enrico Ragaini, and Silvio Messi, and to the whole team of ABB Digital Department for their kind welcome and continuous support in my research activities. Their technical competence and collaborative focus created an environment that enabled me to quickly adapt to the corporate culture, and I have learned a lot about the industry standards in Italy. Working with them provided me with an understanding not only of the technical details of smart building analytics but also of the more general professional and cultural context. I am also very thankful to Simonetta Civera for her Italian language lessons. which considerably easier the quality of my time in Italy.

I also extend my gratitude to the ABB Corporate Research Center (DECRC) in Mannheim, Germany, where I had a six-month research internship. Working with AI and Optimisation teams on the Energy Management Project helped me to develop my technical and analytical skills in time-series forecasting, optimisation, and system framework. I am particularly thankful to Matthias Biskoping, Matthias Schloeder, and Arzam Kotriwala for their guidance, support, and invaluable suggestions during my stay at DECRC. I would like to thank the Department of Engineering and Applied Sciences (University of Bergamo) for providing me with the possibility of working in a productive and cooperative research environment. The support of the administrative team, laboratory members, and research fellows has all been very helpful in the joint research.

Finally, I wish to thank all those who helped me. Without them, I could not have completed my research activity.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
ARIMA	AutoRegressive Integrated Moving Average
BRR	Bayesian Ridge Regression
CNN	Convolutional Neural Network
CV	Cross-Validation
DL	Deep Learning
DR	Demand Response
DT	Decision Tree
EDA	Exploratory Data Analysis
GB	Gradient Boosting
GRU	Gated Recurrent Unit
HVAC	Heating, Ventilation and Air Conditioning
IoT	Internet of Things
kNN	k-Nearest Neighbours
kWh	Kilowatt-hour
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MCAR	Missing Completely At Random
MAR	Missing At Random
MICE	Multiple Imputation by Chained Equations
ML	Machine Learning
MNAR	Missing Not At Random
MSE	Mean Squared Error
PVGIS	Photovoltaic Geographical Information System
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
STLF	Short-Term Load Forecasting
SVR	Support Vector Regression
TPE	Tree-structured Parzen Estimator
UI	User Interface
XGBoost	Extreme Gradient Boosting

List of Publications

The following publications were produced during the course of the Ph.D. research:

1. Ayaz Hussain, Giuseppe Franchini, Muhammad Akram, Muhammad Ehtsham, Muhammad Hashim, Luca Fenili, Stefano Messi, Paolo Giangrande, "Hybrid ML/DL Approach to Optimize Mid-Term Electrical Load Forecasting for Smart Buildings," *Applied Sciences*, Accepted, 2025.

2. Ayaz Hussain, Paolo Giangrande, Giuseppe Franchini, Luca Fenili, Stefano Messi, "Machine Learning-Based Imputation Approaches for Efficient Electrical Load Forecasting," Accepted at *2025 IEEE 13th International Conference on Smart Energy Grid Engineering (SEGE 2025)*.

3. Ayaz Hussain, Paolo Giangrande, Giuseppe Franchini, Luca Fenili, Stefano Messi, "Analyzing the Effect of Error Estimation on Random Missing Data Patterns in Mid-Term Electrical Forecasting," *Electronics*, vol. 14, 2025, Art. no. 1383.

4. Ayaz Hussain, Giuseppe Franchini, Paolo Giangrande, Giovanni Mandelli, Luca Fenili, "A Comparative Analysis of Machine Learning Models for Medium-Term Load Forecasting in Smart Commercial Buildings," *2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE)*.

Abstract

The rapidly changing digital world of advanced infrastructure has allowed implementing Internet of Things (IoT) related metering services, robust monitoring architecture, and cloud-based analysis tools. The development of new technology creates substantial amounts of credible data that are essential to the forecast-based energy management of electric energy consumption. However, data-driven datasets often have limitations in data completeness due to sensor failures, interruptions in communication, and constraints associated with hardware, which considerably compromise the efficacy of Mid-Term Load Forecasting (MTLF) models. This doctoral research endeavours to address these challenges by proposing a comprehensive methodological framework that includes missing-data modelling, cutting-edge imputation methodologies, Machine Learning (ML), Deep Learning (DL), and hybrid ML-DL forecasting techniques specifically optimised for energy analytics in smart buildings.

The dissertation is based on four interrelated research analyses: two focused on analysing and improving MTLF models, while the other two focused on various evaluations of incomplete data patterns and the impact of imputation techniques on future forecasting accuracy. The empirical basis for this work is data from a 30,000 m^2 smart commercial building in northern Italy, as well as meteorological data from NASA POWER and the Photovoltaic Geographical Information System (PVGIS). Missing data mechanisms, such as linear block missingness and random point-wise missingness, were systematically introduced at different levels (5%-40%), and each result requires a thorough statistical, ML, DL, and hybrid imputation evaluation. The results indicated that hybrid imputation methods consistently produce lower reconstruction errors and improve subsequent forecasting stability in contrast to traditional approaches.

A detailed analysis of an extensive range of forecasting models is performed, incorporating Random Forest (RF), XGBoost, Support Vector Regression (SVR), Decision Trees (DT), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and hybrid frameworks such as FireNet–XGBoost. Among these methodologies, the combination of hybrid ML and DL frameworks illustrates the most significant predictive performance, particularly on imputed datasets. This analysis identifies a distinctive relationship between the quality of

the imputation, the representation of temporal variables, and the resilience of forecasting strategies, thereby offering critical findings for the realisation of reliable data-supported energy management initiatives.

Besides its methodological contributions, this thesis also presents a comprehensive software implementation, which is an interactive MTLF platform developed with Streamlit, Python, and modern ML/DL libraries. The proposed framework enables end-to-end forecasting, i.e., data upload, data preprocessing, outlier elimination, imputation, model training, and performance assessment, along with advanced forecasting modules-all conveniently accessible through its web user interface. In addition to simplifying model selection & hyperparameter optimisation in AutoML frameworks, an integrated CO_2 emissions analytics feature can provide crucial information about sustainability. This software application shows effectively the practical validity of the research in a way by bridging academic model design approaches with real-world decision support tools.

In summary, this dissertation contributes to improved knowledge of smart-building forecasting by offering a thoroughly validated methodological framework, emphasising the crucial role of high-quality imputation, and also making available a practical forecast platform for operational planning, resource allocation, optimisation, and sustainability-oriented decisions in modern intelligent buildings.

Chapter 1

Introduction

1.1 Background and Motivation

The global energy paradigm has been evolving in the past decade, making the creation of systems that are intelligent and data-driven more pressing, and able to assist in the efficient and sustainable management of energy. Advanced buildings and especially the commercial and institutional buildings have now become highly complex cyber-physical systems with advanced sensing and metering infrastructure, and automated controls. These systems are gathering massive volumes of data on power usage, environmental factors, system conditions, and user information and have presented a special chance to utilise data-driven analytics to enhance energy predictability and optimisation.

In this respect, MTLF, which lies between weeks and months, is determined as one of the essential energy management strategies at the building level. Compared to Short-Term Load Forecasting (STLF), which is more associated with real-time control or operational adjustments, MTLF could provide strategic planning for maintenance scheduling, tariff planning, and resource allocations of heating, ventilation, and air conditioning (HVAC) systems. Robust MTLF also enables building operators to actively participate in demand response, as greater generation of energy from renewable resources is added to both local and national grids. The high level of variability and uncertainty related to solar and wind generation intensifies the need for reliable mid-term planning techniques capable of adequately forecasting consumption trends under dynamic climatic and behavioural scenarios.

Despite its importance, MTLF faces numerous limitations. The energy consumption of buildings will be determined by a number of factors, such as the environment surrounding the building, the patterns of the occupants of the structure, the time of day when the occupants will carry out the activities, and how well the outer structure will be able to offer insulation.

Moreover, many of these factors exhibit non-linear interactions and often are unobservable or unpredictable. As a result, the forecasting model tends to be well-developed for one period but underperforms for others. Seasonal effects, consumers' behavioural disturbances, and irregularities are the most frequent problems encountered. Furthermore, more complex systems are assumed, including missing and noisy sensor data, which often produce poor experimental conditions typically observed in practical smart-building environments. The presence of missing data is especially challenging: it may be caused by network communication failure, sensor break-up, and data distortion, etc., which can decline the performance of forecasting models if not addressed properly.

However, such MTLF studies are typically over-simplistic in the context by focusing mainly on fully observed or almost fully observed data sets, ignoring the common problem of a large amount of missing data in real-world building practices. In recent times, ML and DL techniques have proved highly successful in the modelling of complex nonlinear systems; they are, however, sensitive to the quality of data. If ML or DL models are trained on incomplete data, or if imputing is performed poorly, the model will be biased when making predictions and undergeneralize to new samples with unstable performance in future prediction horizons.

The review shows that most of the MTLF work assumes complete and partially observed data, ignoring the missing-data cases frequently observed in actual building conditions. Moreover, although ML and DL methods have achieved success in modelling complex nonlinear systems, their performance is strongly affected by the quality of data. If trained on incomplete or poorly-imputed data, ML and DL models may generate biased predictions, insufficient generalisation, and show unstable performance over different forecasting horizons.

Another significant challenge lies in the practical implementation of forecasting systems. Many studies develop forecasting models in controlled environments, without integrating them into deployable platforms that building managers can use. Consequently, a gap persists between research and real-world application. Forecasting pipelines often require extensive preprocessing, feature engineering, imputation workflows, hyperparameter tuning, and evaluation procedures, yet few studies propose an end-to-end system that brings these components together in a cohesive and user-accessible manner.

The increasing interest in environmentally responsible energy strategies showcases the urgency for foresight methods that are not only accurate but also implementable in practice. This requires models that can ingest diverse data sources, such as historical load, high-resolution meteorological data, and calendar features, and produce reliable mid-term predictions even in the presence of missing or corrupted inputs. It also necessitates software systems capable of supporting decision-making for building managers, energy analysts, and stakeholders who may not have a technical background in ML.

Motivated by these challenges, the present dissertation investigates the role of missing-data mechanisms in smart-building load forecasting and proposes a comprehensive framework that integrates missing-data analysis, imputation methods, ML, and DL forecasting models, and a complete software platform for MTLF. This research seeks to align methodological progress with a practical software platform to integrate academic studies and practical uses in building-energy analysis.

1.2 Motivation for Addressing Missing Data

In smart-building datasets, the contribution of the missing value problem is still one of the most intuitive and challenging problems. Depending on the occurrence time and duration of such events, missing data might be presented under different patterns: there could be random single-point gaps, short duration of missing intervals, or block-type outage, long gaps for hours or days. These trends have a strong influence on the statistical characteristics of the smart building data, conceal time-based effects, and modify significant characteristics, including daily patterns, seasonal dynamics, or weather-load relationships.

Most past research assumes that forecasting algorithms require training data that accurately measure the actual dynamics of a system. On the other hand, if missing data are filled with simple methods, such as zero-filling or linear interpolation, in such a case, the dataset can deviate significantly from the physical properties of the building. These distortions will be amplified in ML and DL models, and hence will cause inaccurate predictions with reduced reliability. In the MTLF, such degradation is especially problematic as forecasting horizons are extended, and models are more vulnerable to error propagation.

The proposed research will be carried out with the purpose of providing a systematic analysis of how missing-data mechanisms impact the performance of different models in terms of forecasts. They are statistical imputers, ML-based imputers, DL-based imputers, as well as the hybrid imputers which combine different techniques. The aim of this research, using a sequence of simulated experiments with manipulated missing data, is to establish the imputation methods that can be used consistently and at the same time maintain forecasting accuracy under diverse conditions and missingness patterns.

1.3 Role of Meteorological Variables in Load Forecasting

Another reason is inspired by the necessity to develop more effective forecasting models that can leverage the improved feature space offered by the meteorological data. In addition to operational factors, weather conditions are crucial, such as temperature, relative humidity, solar radiation, and wind speed, which are very important parameters that affect building energy demand, particularly in regions with high seasonal ranges. Including these parameters from databases such as NASA POWER and PVGIS, this study improves the ability of forecasting models to represent environmental effects on load profiles.

Finally, a user-friendly forecasting tool has been developed that fulfils the needs in practice of building authorities and energy experts. A software package that can be deployed takes this research and transforms it into a tool that assists with energy forecasting, planning, and carbon emissions reduction.

1.4 Research Problem

Despite the remarkable progress in building-energy forecasting, there are a number of open issues that still prevent such MTLF models from being used reliably and effectively when applied in practice to real smart-building infrastructure. First is the problem of data quality. Smart meters and IoT sensors are now widely implemented; however, the online streaming data is commonly related to incomplete samples, skewed sampling intervals, and time gaps. However, the characteristics of these missing values, i.e., randomness, structuring, or due to external manipulation, have not been fully studied in the problem statement of MTLF. A very small fraction of the literature specifically addresses missing data, and even a smaller portion studies the impact of various missingness patterns on ML and DL based prediction models.

A second constraint is to determine the imputation approach and quantify its performance. Although there are many methods to address this, including statistical imputers, ML methods, or DL models, no standards have been reached about which techniques are particularly suitable for data retrieval on building-energy usage. Furthermore, imputation method interactions with forecasting models have been empirically reported but never systematically studied. The lack of experimental studies of missing data mechanisms under controlled conditions is a major practical challenge for energy analysts and building facility managers in selecting the correct systems.

The third issue is model interpretability and operational application. DL models, such as LSTM networks or convolution-based architectures like FireNet, have proven to extract long-range temporal dependence effectively, but are usually seen as "black-box" models. Their performance is often reliant on hyperparameter sensitivity and large quantities of data. However, in the absence of a clear and systematic assessment model, it could be difficult to determine whether such models indeed enable predictions to become more accurate or merely overfit the available training data under practical conditions.

Finally, even when research achieves high-performing models, these results rarely translate into operational tools. Many studies remain proof-of-concept prototypes without a usable interface or the capacity to run automated pipelines for data ingestion, preprocessing, imputation, model training, and forecasting. This disconnect between research and practical application limits the adoption of modern forecasting techniques in many building-management contexts.

1.5 Research Aim and Objectives

The overarching aim of this research is to develop a comprehensive, empirically validated framework that integrates missing-data analysis, imputation strategies, and advanced forecasting models, culminating in a functional software platform for MTLF in smart buildings.

To fulfil this aim, the dissertation pursues the following objectives:

1. **To investigate the nature, extent, and impact of missing-data patterns** in real smart-building datasets, and to evaluate a comprehensive set of statistical, ML, and hybrid imputation strategies under controlled missingness scenarios of varying severity.

2. **To develop and compare forecasting models** spanning classical ML, advanced DL architectures, and hybrid ML–DL ensembles, and to optimise these models through systematic hyperparameter tuning for fair and robust performance comparison.

3. **To design and implement a complete forecasting platform** capable of automating preprocessing, imputation, model training, validation, evaluation, and MTLF, and to assess its practical usability for smart-building energy management and load-planning applications.

These objectives are deliberately interconnected: analysing missing-data patterns informs imputation strategies; imputation quality affects forecasting accuracy; and the integrated pipeline ultimately enables a deployable forecasting tool.

1.6 Research Questions

Based on the research aim and the identified gaps in the literature, the following research questions guide the study:

- **RQ1.** What are the common patterns of the missing data in smart-building electricity datasets, and how do the various mechanisms of data absence affect the forecasting accuracy performance?

- **RQ2.** Which imputation strategies provide the greatest resilience across different missingness levels and patterns when used in conjunction with ML and DL forecasting models?

- **RQ3.** How accurate, robust, interpretable, and computationally efficient are ML, DL, and hybrid prediction models incorporated for forecasting of the mid-term building loads?

- **RQ4.** How can a unified forecasting platform be developed that integrates preprocessing, imputation, hyperparameter tuning, and forecasting into an operational and user-friendly system, and how can this platform support real-world energy-planning tasks, including demand response applications and the evaluation of CO_2 emissions associated with building-energy consumption?

These questions allow the thesis to systematically investigate both theoretical and practical aspects of mid-term building electrical load forecasting.

1.7 Scope of the Study

The scope of this dissertation is intentionally focused to ensure depth of analysis and methodological clarity:

- The study concentrates on forecasting horizons, typically spanning hourly horizons (12 hours ahead) up to 7 days.

- One real-world dataset from a European smart-building campus is used: the 2023 dataset, which incorporates meteorological information obtained from the NASA POWER and PVGIS platforms.

- Missing-data simulations include random uniform missingness and structured block-type gaps, evaluated at multiple severity levels.

- Methods of imputation vary from basic statistical techniques to sophisticated ML and hybrid methodologies that encompass Multiple Imputation by Chained Equations (MICE), Random Forest (RF), k-Nearest Neighbours (kNN), eXtreme Gradient Boosting (XGBoost), Gradient Boosting, and Autoencoders.

- The forecasting models include tree-based algorithms (DT, RF, XGBoost, GB), kernel-based models (SVR), distance-based models (kNN), LSTM networks, the FireNet convolutional architecture, and hybrid ML–DL combinations.

- The research questions explored are the importance of accurate prediction, reliability, and computing efficiency, along with other more practical factors like the preprocessing time and platform functionality.

The study does not address long-horizon forecasting, economic planning models, or real-time grid control applications, although the developed methods could be adapted for such purposes.

1.8 Significance of the Research

This research represents a significant step forward for the area of building-energy analytics in several important directions. Its first contribution is that it offers an extensive and systematic study of the missing data effect, which most forecasting literature excludes. Examining different mechanisms of missingness and imputation provides useful guidelines for selecting appropriate preprocessing methods for realistic data.

Second, the research involves a rigorous comparison of ML, DL, and hybrid forecasting models, including the adaptation of the FireNet architecture for time-series forecasting. This contributes to methodological innovation, particularly in applying hierarchical feature extraction to load patterns.

Third, the development of a comprehensive and functional forecasting platform proficiently aligns theoretical algorithmic research with its practical execution in real-world situations. By integrating imputation workflows, hyperparameter optimisation, forecasting pipelines, and a graphical interface, the platform provides building managers with a usable tool rather than isolated research results.

Finally, the results contribute to the overall objectives of energy and carbon sustainability by supporting a more efficient building-level energy planning. Reliable MTLF enables proactive planning, efficient operation, and better participation in demand response programs.

1.9 Novelty and Original Contributions

While the literature on load forecasting is extensive, especially in short-term and day-ahead horizons, MTLF within smart-building contexts remains comparatively underexplored. This dissertation work enhances the field by showcasing contributions aimed at fixing the practical, methodological, and analytical challenges acknowledged in current research.

The first major objective of this study is the comprehensive and systematic investigation of missing-data behaviour and its implications for forecasting performance. Unlike most existing work that only acknowledges missing values, this research rigorously simulates both random and block-type missingness at varying severity levels and evaluates their impact across multiple forecasting horizons. Within this controlled framework, a wide range of imputation strategies, spanning statistical approaches, ML-based imputers, deep autoencoders, and hybrid methods, is compared under identical conditions. This unified analysis not only clarifies how different missing-data patterns influence model behaviour but also identifies which imputation techniques remain robust when the dataset is significantly degraded.

The second contribution involves adapting and deploying advanced forecasting architectures specifically for MTLF. Despite the fact that LSTM networks are being used extensively in STLF, there is no systematic study of their performance over multi-week horizons where the available data is not fully collected. Moreover, none of the building energy prediction models employed the FireNet architecture that was initially suggested in image classification. Its hierarchical architecture allows the convolutional layers to extract features of time at varying granularities, and therefore is a natural alternative to recurrent networks.

The third contribution is the development and deployment of a forecasting system application. Unlike many studies that present models in isolation, in this thesis, an integrated platform is built, including data ingestion, preprocessing, missing data strategies handling, hyperparameter tuning, and training model evaluation and predictions. This working system demonstrates the potential to successfully transition research results into a usable tool for both energy analysts and researchers.

Finally, this work integrates several findings presented in four peer-reviewed papers carried out during the PhD research. These research publications analyse missing-data modelling, feature selection, imputation effectiveness, and hybrid forecasting methods. Their works directly contribute to the methodology section and validation frame employed in the dissertation, enhancing findings into the final forecasting platform.

1.10 Integration of Publications Within the Dissertation

This thesis will provide a thorough analysis, description, and combination of the results of four peer-reviewed articles that will be published throughout the PhD project. Every publication addresses a particular issue of the entire research problem, and the overall contributions they present are the basis of the proposed comprehensive forecasting framework.

The first paper investigates the effect of random missing-data on MTLF accuracy. It shows how different mechanisms of missingness correlate with the pattern of loads and performance of a model. These results are also used to design the missing-data simulation studies in this dissertation.

The second publication evaluates several classes of imputation techniques, highlighting the advantages of ML-imputers such as kNN and RF for datasets with moderate-to-high missingness. These results motivate the inclusion of these imputers in the comparative framework used in this thesis.

The third paper describes a combined DL and gradient boosting method for time series forecasting. This study shows that hybridisation can mitigate the drawbacks of pure DL models while preserving their capacity to learn complex nonlinear temporal dependencies. The hybrid experiments performed in this thesis generalise the idea by including a larger number of models.

The fourth publication focuses on hierarchical feature extraction for building-energy consumption. It provides early evidence that convolution-based architectures can capture multi-scale temporal patterns, motivating the adaptation of FireNet in this thesis.

By synthesising these publications, the dissertation provides a coherent and comprehensive understanding of how missing-data handling, model selection, and hybridisation strategies collectively influence MTLF in real smart-building environments.

1.11 Structure of the Dissertation

The remainder of this dissertation is organised into five main chapters:

- **Chapter 2: Literature Review** Presents an in-depth review of the state-of-the-art in electrical load forecasting, covering classical statistical models, ML techniques, DL architectures, hybrid models, missing-data theory, and imputation methodologies.

- **Chapter 3: Methodology** Describes the detailed methodological structure developed during this study. It includes comprehensive descriptions of the datasets employed, the methodologies for preprocessing, the strategies for modelling missing data, methodologies for imputation, forecasting models utilised, techniques for hyperparameter optimisation, and evaluation metrics for assessment.

- **Chapter 4: Results and Discussion I (Missing Data Analysis)** Explain the empirical results of simulations on missing data for random (uniform) and linear block-type missingness. A detailed comparison between statistical and ML-based imputation methods in different levels of missingness is made. The study highlights how different missing-data structures impact model inputs and provides the best imputation strategies under realistic conditions.

- **Chapter 5: Results and Discussion II (Forecasting Model Performance)** Focuses on the evaluation of forecasting models after imputation. This chapter investigates the prediction performance of conventional M, DL, and hybrid ML and DL methodologies over a specific time duration. It also includes the assessment of hyperparameter tuning results, model robustness, comparative error analysis, and the integration of the forecasting platform, linking outcomes to the research questions.

- **Chapter 6: Forecasting Platform Development** Details the design and implementation of the end-to-end forecasting platform, including system architecture, automation workflows, and user-interface components.

- **Chapter 7: Conclusion and Future Work** Summarises the primary outcomes, highlights the constraints of the study, and specifies the potential research directions, such as federated learning, edge AI, and real-time adaptive forecasting.

This structure ensures a logical progression from conceptual foundations and methodological development to empirical evaluation and practical implementation.

1.12 Scientific and Industrial Contributions

Beyond addressing methodological gaps, this dissertation contributes to both academic research and practical building-energy management in several meaningful ways.

From a scientific perspective, the thesis provides a structured analysis of missing-data mechanisms in smart-building datasets. By simulating both random and block-type missingness and systematically evaluating a wide range of imputation approaches, the study offers one of the most detailed examinations of missing-data impacts within the domain of MTLF.

The thesis includes a comparison of predictive models in ML and DL based techniques. This evaluation not only compares models' performance under various scenarios but also investigates model architectures on their robustness against biased data inputs. The optimisation of the FireNet architecture for multi-scale temporal feature extraction is a methodological contribution that may be useful in other domains beyond building energy prediction.

For the industrial sector, a forecasting platform is a significant step forward. In contrast to limiting the work to algorithmic experiments, the dissertation implements a complete end-to-end system including data ingestion, preprocessing, imputing, hyperparameter tuning, model training, and prediction. This tool simplifies complex analytical processes and provides a as well as researchers. By bridging the academic research and applied energy management, the approach becomes more practicable and feasible for industry application.

1.13 Broader Impacts

The findings of this study are not limited to smart buildings. Accurate MTLF predictions will allow more informed operational planning in the context of Micro-Grid, which can improve energy user comfort, reduce expenses, and promote sustainability planning. In the era of a decarbonised world, building load forecasting becomes highly coupled to grid flexibility, renewable energy integration, and demand-side management.

Moreover, the findings of this study on missing data handling strategies can be generalised to other domains where temporal datasets are incomplete (e.g., transportation, environment monitoring, health care, and industrial automation). The methodological framework presented in this research can be considered to be applicable to other time-series modelling problems.

Finally, the forecasting platform is an excellent tool that can help developers move beyond model development into actual applications. These platforms will be crucial in enhancing data-driven operational decision-making as companies endeavour to deliver on their commitment to digital evolution.

1.14 Closing Remarks

This introduction has covered the motivation, identified the research gap, specified the objectives of the study, and outlined the significance of this dissertation. It places the work in a wider context of energy analytics for smart buildings, where several challenges are interconnected, such as missing data, complex temporal dynamics, and methodological inconsistencies, which demand a unified approach.

The later chapters aim to develop such a framework methodically, starting with the analysis of forecasting methods and missing data in the review of previous literature. After that, the comprehensive methodological framework of the study is described in detail. Follow-up chapters introduce research results obtained from missing-data simulations, imputation method comparisons, and forecast model evaluations. Finally, an operational forecasting platform is designed in order to show how the results can be practically transformed into tools for building energy management.

Overall, this thesis aims to contribute with higher theoretical depth and a more practical solution to MTLF. With systematic experimentation, the formulation of a rigorous methodology, and the development of a viable system for building energy forecasting, it is our goal to work towards filling out this void. This research will aim to improve building energy management.

Chapter 2

Literature Review

2.1 Time Series Forecasting Techniques & Evaluation

A wide variety of time series forecasting methods of electricity consumption is presented in the literature that can be broadly classified as statistical, ML, DL, and hybrid models. The first layer is statistical techniques like Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing, which require the series to be stationary and linear, to define the trend and other seasonalities (Okanlawon, 2023). Indicatively, the Okanlawon 2023 developed a comparative analysis of the time series methods in the energy demand in Nigeria based on the ARIMA algorithm in order to analyse 10 years of data on electricity consumption in Nigeria. The chapter has highlighted the conciseness and effectiveness of these methods in recording trends, tendencies, and seasonalities in energy usage and the way that leads to the sustainability of the electricity industry over the long run. Non-linear and non-stationarity of the data are, however, a common problem in statistical methods in the case of electricity time series (Misiurek et al., 2025a).

ML and DL techniques overcome these limitations by dealing with complex and non-linear relationships. The LSTM networks are a type of RNN that can be used for time series tasks due to the network's ability to capture temporal dependencies, especially in the case of irregular patterns (Baek et al., 2025). (Baek et al., 2025) have proposed a hybrid model of SARIMAX-LSTM for monthly university electricity demand forecasting in Georgia, USA, using weather-related variables and academic schedules between 2019 and 2024 over the last 6 years. The model showed its superior performance in terms of lower error metrics and illustrated how one can integrate multiple predictors with DL for better accuracy in energy intensive environments. Similarly, Transfer Learning (TL) is beneficial for the already trained models of related time series for better predictions, as is done in the case study of (Tzortzis et al., 2024) on 27 European national electricity demand time series. By invoking a Multi-Layer Perceptron (MLP) with clustering to find similar patterns, TL outperformed the conventional Neural Network (NN) training, which was significantly better for day-ahead forecasting with potential for international-scale applications.

A combination of these paradigms that attempts to address the negative aspects of the individual models is called a hybrid model, such as ARIMA and artificial neural networks (ANNs) that are more adaptable to the linear and non-linear factors (Ali et al., 2024; Nti et al., 2020). Also, Ali et al. (2024) summarised the improvements on the time series traditionally to hybrid models in STLF, and heuristic optimisations have been made to improve the ANN performance with a positive effect on accuracy compared to the standalone time series models, including Particle Swarm optimisation (PSO) and Differential Evolution (DE) (Nti et al., 2020). In a systematic review on STLF for the residential sectors and more than 300 studies in 2012-2022, (Rodrigues et al., 2023a) found that the hybrids had lower error in hourly demand forecasts.

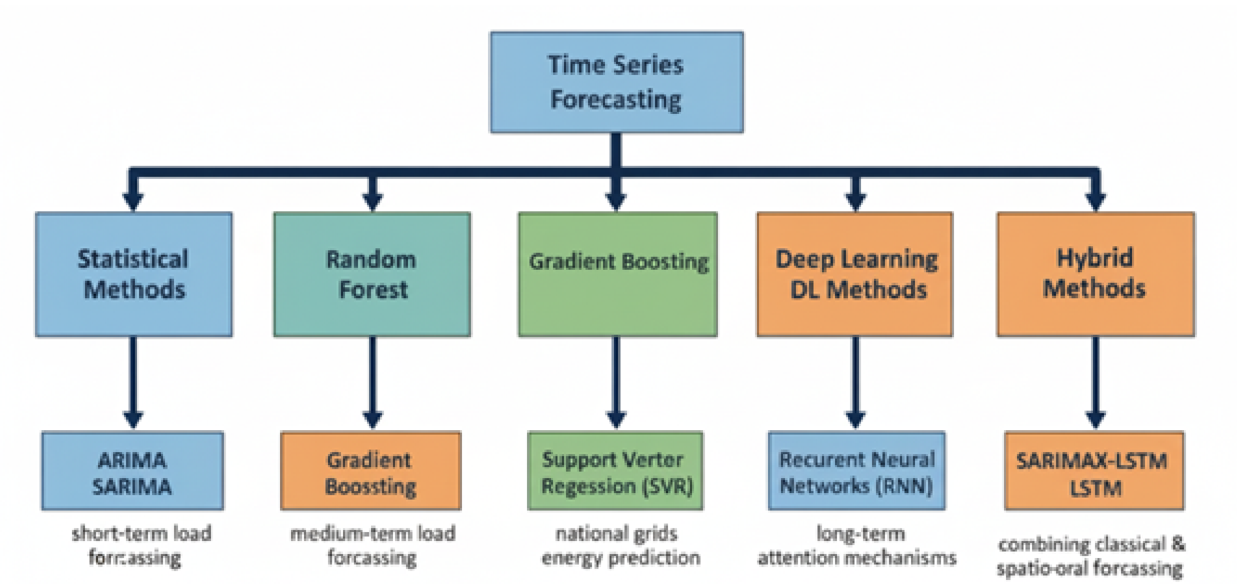


Figure 2.1: Classification of time-series forecasting techniques

In order to assess the precision and reliability of the predictions, these methodologies are analysed through the application of evaluation metrics. Common indicators include Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of variance (CV) for comparison of the forecasts with the actual consumption (Nti et al., 2020; Baek et al., 2025). For example, (Nti et al., 2020) presented that LSTM models provided lower MAPE than traditional time series in residential load forecasting, and (Okanlawon, 2023) presented these metrics to validate the efficacy of

ARIMA for Nigerian energy data. (Misiurek et al., 2025a) presented multiple forecasting horizons (from very short-term to long-term) and tested models such as fuzzy logic and evolutionary algorithms through MAPE and RMSE, highlighting the context-specific metrics in the context of energy integration of renewable energy and decentralisation.

Table 2.1: Summary of time-series forecasting techniques and evaluation for electricity consumption

Category	**Key Methods**	**Applications**	**Strengths**	**Limitations**	**Evaluation Metrics**	**References**
Statistical	ARIMA, Exponential Smoothing	Comparative analysis of energy demand in Nigeria, capturing trend and seasonality	Interpretability, low computational cost, suitable for stationary data	Struggles with non-linear/non-stationary behaviour	RMSE, MAPE, MAE, CV	(Okanlawon, 2023; Misiurek et al., 2025a)
Machine Learning	Support Vector Machines, Decision Trees	Applied for handling non-linear relationships in volatile demand	Captures complex patterns, improves accuracy with diverse features	Requires large datasets, reduced interpretability	RMSE, MAPE, MAE	(Nti et al., 2020; Ali et al., 2024)
Deep Learning	LSTM, MLP with Transfer Learning	Hybrid SARIMAX–LSTM for universities; TL for European national loads	Captures long-term temporal dependencies, performs well in irregular patterns	High computational cost, prone to overfitting	MAPE reduction, RMSE improvement	(Baek et al., 2025; Tzortzis et al., 2024)
Hybrid	ARIMA + ANN, PSO/DE optimised ANN	Short-term load forecasting with reduced residual errors	Mitigates individual weaknesses, synergistic performance	Integration complexity, tuning difficulties	Lower RMSE/ MAPE in hourly horizons	(Ali et al., 2024; Nti et al., 2020; Rodrigues et al., 2023a)
General Evaluation & Challenges	N/A	Exogenous variables, COVID-19 effects, smart grid scalability	Supports model comparison and reliability assessment	Non-stationarity, volatile renewable generation	MAPE, RMSE, MAE, CV	(Nti et al., 2020; Ali et al., 2024; Tzortzis et al., 2024; Rodrigues et al., 2023a)

Despite the advancements, challenges remain, such as non-stationarity of the data, exogenous variables integration (e.g., weather, impacts of Covid-19), and the scalability in smart grids (Ali et al., 2024; Tzortzis et al., 2024). Gaps in the literature include a limited amount of work on multi-platform implementations and the need for robust handling of volatile patterns from renewables (Rodrigues et al., 2023a). These issues highlight the importance of investigating specific forecasting modelling methods in more depth, e.g., statistical methods for baseline, ML for non-linearity, DL for temporal dependencies, and hybrid approaches for synergistic performance, which are explained in the following sections in more depth.

2.2 Techniques for Forecasting

Building on the evaluation frameworks presented in the above section, which highlight the importance of sound metrics, such as MAPE and RMSE, to evaluate model performance in the face of the complexities of electricity time series data, this section provides an overview of the main classifications of forecasting techniques. These methods are key to the prediction of electricity consumption, efficient grid management, demand response, and renewables integration, by capturing the patterns affected by factors such as weather, economics, and user behaviour (Misiurek et al., 2025a; Mystakidis et al., 2024).

Statistical techniques are based on mathematical models to analyse the historical data by assuming some underlying patterns, such as stationarity, linearity, etc., and are suitable for stable and well-understandable datasets in electricity forecasting (Ugbehe et al., 2025). ML approaches such as Support Vector Machines (SVM) and DT use data-driven ML algorithms to deal with non-linear relationships and huge feature sets to enhance accuracy in volatile consumption scenarios (Mamun et al., 2020). DL is a step forward of ML, where it is applied to neural architectures such as LSTMs to handle sequential data with a focus on long-term dependencies to make high-resolution predictions (Mystakidis et al., 2024). Hybrid models blend these paradigms to pick up on strengths, e.g., statistical stability coupled with ML adaptability to reduce errors and boost generalisation, as it has been demonstrated in smart grid applications (Almazrouee et al., 2025).

The choice of technique often rests on the nature of the data and the outcome of evaluation, with hybrids being the best performers in recent reviews. To explore these in depth, we begin with statistical forecasting approaches, which are a benchmark for the more advanced approaches in terms of interpretability and computational efficiency.

2.2.1 Statistical Forecasting Approaches

Statistical forecasting methods are the basic pillar of electricity consumption forecasting, based on mathematical models used to predict electricity consumption in the future based on historical patterns and variables. These methods, based on the theory of probability and time series, assume that the data of electricity loads has underlying structures such as trends, seasonality, and stationarity which can be modelled without consuming much computational resources (Kolkova et al., 2022; Nti et al., 2020). In the simple case of electricity consumption, statistical methods are particularly appreciated for their interpretability, which enables energy planners to make sense of the causal relations between some of the predictors, such as temperature, time of day, or economic indicators, and load variation (Rodrigues et al., 2023a; Mir et al., 2020). As outlined above, those methods serve as a reference to more elaborate approaches by offering simplicity and reliability in cases of linear or slightly seasonal information, and their evaluation using performance measures such as RMSE and MAPE indicates their baseline execution (Madin et al., 2025; Makris et al., 2024). A critical analysis, however, reveals that although these models are highly effective in controlled and linear contexts, their rigidity leads to poor performance in volatile and real-world scenarios, which motivates methodological innovation and highlights gaps in handling current grid complexities.

The ARIMA model (and its derivatives, e.g., exponential smoothing algorithms and regression-based models) is the basic statistical approach to electric load forecasting. ARIMA is a univariate model that decomposes time series dynamics into autoregressive (AR), differencing (I), and moving average (MA) components to handle non-stationarity and capture short-term dependencies in load data (Makris et al., 2024). The general form of the ARIMA (p, d, q) model is expressed as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d Y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t, \tag{2.1}$$

where Y_t is the load at time t, L is the lag operator, ϕ_i and θ_j are parameters, d is the differencing order, and ε_t is white noise. Autocorrelation functions, tuning parameters.

Variations Seasonal ARIMA (SARIMA) is a set of extensions that add seasonal differencing and seasonal terms (P, D, Q, s) to incorporate periodic trends such as demand on a day-to-day basis or week-to-week basis for electricity. The equation of the SARIMA (p, d, q) is extended to:

$$\Phi(L^s)(1 - L^s)^D(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d Y_t = \Theta(L^s)(1 + \sum_{j=1}^{q} \theta_j L^j)\varepsilon_t, \tag{2.2}$$

with ϕ_i and Θ as seasonal polynomials. SARIMAX also incorporates exogenous factors (e.g., weather or holidays), which further add to the predictive capacity by looking outside the box (Makris et al., 2024; Madin et al., 2025).

The Holt-Winters exponential smoothing is a commonly used method that uses weighted averages of the level, trend, and seasonal components, which can be expressed in additive or multiplicative form depending on whether or not statistical variance increases with time. The additive Holt-Winters equations are:

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \tag{2.3}$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \tag{2.4}$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}, \tag{2.5}$$

where L_t, T_t, and S_t are level, trend, and seasonal components, and α, β, γ are smoothing parameters between 0 and 1. This is computationally efficient for smoothing noisy data, and is commonly used for seasonal forecasting (Madin et al., 2025). State-space models, such as Kalman filters, are a way of providing a probabilistic framework for dynamic systems and allow for real-time updates, missing data by means of hidden states. Regression-based approaches, such as linear regression and generalised linear models, treat load as a dependent variable regressed on predictors such as temperature, Gross Domestic Product (GDP), or time dummies, and support simple hypothesis testing of variable significance (Kolkova et al., 2022; Karageorgou, 2023). Critically, despite these models sharing a common reliance on model parameters, each is more or less successful. The internal-dependent character of ARIMA compared to the external-dependent character of regression results in debates on whether hybrid forms of these models could be creatively advanced to connect such divides for improved multi-horizon forecast systems.

Applications of these methods to electricity forecasting range over a variety of different horizons and situations, indicating the versatility of the methods, but also inconsistencies in performance across different datasets. For STLF, the most commonly used models are ARIMA and SARIMA, which are used for hourly or daily predictions of load and can model intraday fluctuations caused by human activity. Makris et al. (Makris et al., 2024) used both ARIMA and SARIMAX on real data from the Ikaria islanded grid in Greece, achieving RMSE of about 0.13-0.15 MW for hourly load, demonstrating their suitability for isolated systems having a large seasonality. However, this success is in contrast with the results of Karageorgou (Karageorgou, 2023), who observed the loss of accuracy of SARIMA in the case of anomalies resulting from the Covid-19 pandemic in Spain, with MAPE to 10-15%, implying the loss of the stationarity assumptions of statistical models during disruptive events, which is an important gap for resilience planning in the grid. Similarly, Madin et al. (Madin et al., 2025) used Holt-Winters and SARIMA for monthly electricity consumption data from Kazakhstan (2002-2022), showing seasonal peaks in the winter months caused by heating demands. Although the MAPE was better for the aggregated national forecast (2.11%) using the Holt-Winters method, an important observation of this study, it is easy to overestimate the errors of urban subsets if nonlinear integration of EV loads is not considered.

For MTLF (weeks to months), economic variables are added in SARIMAX to plan demand, such as Karageorgou (Karageorgou, 2023), which suggests the ability of SARIMA in stable economic periods, but the under-forecasting ability in the case of anomalies such as Covid-19 disruptions. Additionally, Rai and De (Rai et al., 2021) demonstrated the superiority of support vector regression (SVR) compared to classical methods such as multiple linear regression and optimised Holt's method for MTLF in smart grid data with up to 3.6% Mean Absolute Percentage Error (MAPE) reduction using meteorological variables.

Similarly, for MTLF across utility systems, Jahan et al. (Jahan et al., 2020) synthesised hybrid models based on artificial neural networks, SVR, DT, and fuzzy sets for enhancing the outcomes through 10-20% of accuracy compared to single methods while being bulky in terms of heavy preprocessing. Further developments include Oreshkin et al. (Oreshkin et al., 2021), which introduced the N-BEATS neural network for MTLF, which achieved a 14% reduction in MAPE over baselines on European monthly demand series by capturing trends, seasonality, and stochastic elements well. Zimmermann and Ziel (Zimmermann et al., 2025) introduced Generalised Additive Models (GAMs) with interpretable P-splines for hourly MTLF using weather and socio-economic factors that showed better accuracy compared to state-of-the-art methods for 24 European countries and emphasised the computational efficiency and total interpretability. Zhang et al. (L. Zhang et al., 2023) proposed an error correction model using seasonal decomposition, which dealt with the error accumulation in MTLF and improved the precision of forecasting by using hybrid decomposition methods. These results highlight the viability of the use of more sophisticated techniques to improve the robustness of MTLF, specifically in the context of dynamic smart buildings.

Long-term applications frequently use regression models coupled with ARIMA residuals, such as ARIMAX, for example, in the work of Kolková and Ključnikov (Kolkova et al., 2022) who test statistical models for enterprise demand forecasting that are shown to be accurate in projecting annual trends related to industrial growth, but caution against overfitting in the case of SMEs with sparse data. Cai et al. (Cai et al., 2024) extended this to Electric Vehicle (EV) charging loads using Monte Carlo simulation, a probabilistic statistical approach to estimate spatial-temporal distributions, generating scenarios for urban grid planning with virtual parameter estimation from historical EV data; however, the approaches reliance on simulated variability is a creative way to address uncertainty, but tends to increase bias if initial distributions are improperly specified.

The strengths of statistical approaches include their transparency, low data requirements, and computational efficiency, which can make them accessible to resource-constrained utilities. For example, using regression models, the impacts of variables such as a 1℃ increase in temperature (℃ is the standard metric unit of temperature in scientific contexts) are a 2–5% increase in load in summer peaks (Karageorgou, 2023). They also perform well when there is an assumption of linearity, stationarity (established by Dickey–Fuller tests in Madin et al. (Madin et al., 2025) and do not need extensive tuning as do data-hungry models. Nevertheless, the nonlinearity cannot be fully treated, and volatility due to renewables or abrupt effects like a pandemic cannot be modelled (ARIMA model is more likely to overfit noise or fail to capture long-range dependence), (Razghandi et al., 2021; Widmark, 2022). According to Makris et al. (Makris et al., 2024), the upper MAPE of SARIMAX (up to 5 %)

was more pronounced in nonlinear island grids compared to DL baselines, which highlights rigidity in complex real-world environments, a major weakness that can further grid instability in systems that rely on renewable power. There are also emerging trends of hybrid forms with a few probabilistic elements, e.g. in Monte Carlo to quantify uncertainty of EV loads (Cai et al., 2024), though at this point there are no scalable solutions to big data and the flexibility to non-stationary series of smart grids (Mystakidis et al., 2024), although in this area some innovative solutions like adaptive parameter estimation may address the assumption breaches.

These shortcomings have resulted in the development of ML approaches that are more adaptable in modelling nonlinear trends and have the capacity to process a variety of characteristics, as examined below.

2.2.2 ML Forecasting Approaches

Although the use of statistical approaches, as identified above, has been at the forefront in offering a simplified structure relating to the electricity consumption by the use of models like ARIMA and exponential smoothing, its rigidity usually fails miserably in the attempt to represent the degree of nonlinearity and volatility in the current energy demand. These dynamics are constructed using dynamic aspects like the changing weather, consumer behaviours, and the incorporation of renewables, which have complexities that cannot be easily accommodated by linear assumptions. ML approaches have become an attractive solution to these shortcomings and are grounded in algorithms that discover multifaceted relationships without requiring the specification of functional forms. Emphasising pattern recognition and adaptivity, ML models have transformed forecasting, where more robust predictions of forecasters in general are possible in short-term and very short-term horizons, where statistical immobility leads to the poor performance of statistical models.

The introduction of ML in electricity load forecasting was initiated with basic algorithms such as SVM and kNN, which are good at dealing with nonlinearities using kernel tricks and distance-based similarity, respectively. For example, SVM has been initially used for STLF, and a higher accuracy than ARIMA was demonstrated in datasets with seasonal variations, after mapping the inputs into higher-dimensional spaces for a better separation of demand patterns. Regression trees and their extensions, such as decision trees, increased this even further by dividing data into hierarchical subsets, so that it could be intuitively handled, such as temperature and historical load. However, models using a single learner, such as these, are often plagued by overfitting or sensitivity to noise, inspiring a move to ensemble methods. RF, an ensemble of decision trees, has been widely adopted due to its capability of variance reduction with bootstrapping and randomness in features. Lotfi et al. (Lotfi et al., 2025) employed RF as a prediction method with Grid Search hyperparameter tuning for STLF, considering environmental factors such as temperature and previous day load, resulting in 0.894 R^2 in winter, which outperforms seasonal ARIMA and exponential smoothing, by reducing overfitting while incorporating diverse features.

Gradient Boosting Machines (GBM) are a culmination of the ensemble evolution, where weak learners are sequentially added to minimise the residual. The models may shine in flexibility and combination of features, which, in the case of the study of Muqtadir et al. (Muqtadir et al., 2025), is a combination of the three major models LightGBM, XGBoost, and CatBoost to nowcast residential load in 13 sites in North America and Europe. Their method showed a Root Mean Squared Logarithmic Error (RMSLE) of 0.1898 and an R^2 of 0.9745, which outperforms single ensembles thanks to the generalisation of the LightGBM, the prevention of overfitting of XGBoost, and the use of categorical variables such as weather indicators by CatBoost. These studies show the example of the ML strength of learning trends (nonlinear interaction between load, time-of-day, and externalities) as they adapt themselves to large sets of smart grid data. However, several criticisms remain, including the fact that ML needs large quantities of high-quality data to train, which is problematic in places with sparse metering, as noted in Zairi and Freihat (Zairi et al., 2025), who discovered that the noisy input in load data necessitated a multi-step preprocessing step prior to the use of a hybrid CNN-GRU.

There are still challenges, in particular, the problems of interpretability and computational cost. In comparison with transparent statistical models, black box ensemble models, such as GBM, cannot reveal decision pathways, which might be a problem for compliance with regulations in energy markets. Percuku et al. (Percuku et al., 2025) tackled this to some extent by comparing linear regression as a simpler ML baseline with LSTM in Kosovo substation data. Linear regression had been found to compete in terms of MAE for STLF but lacked in terms of nonlinear capture. The computational requirements are also proportional to the size of the ensemble; an example of this is Choubey et al. (Choubey et al., 2025), who proposed a stacking ensemble model that includes empirical mode decomposition and transfer learning to do robust load forecasting, which shows higher accuracy with less computational cost through effective feature decomposition and model transfer. In MTLF applications, Kotsiopoulos et al. (Kotsiopoulos et al., 2023) also used multivariate ensemble learning on data of a Greek energy system, where many variables are considered (such as injected renewables and economic indicators) to produce 15–20 % reductions in RMSE over base models, but at the cost of the increased preprocessing of variable synchronisation. Equally, Wang et al. (A. Wang et al., 2023) proposed a deep ensemble model that involves convolutional and recurrent networks as the training structure of MTLF reduced MAPE and RMSE by 4.96 and 12.31 per cent on historical data, respectively, at the expense of an intense consumption of GPU-resources to train the structural elements. Trade-off comparisons are made between models: Ensembles are the better of single learners in those aspects, but also raise complexity: e.g., RF vs SVM, robustness vs efficiency, etc. Nevertheless, with the addition of more paradigms to hybrid models, their combination requires further tuning and can be associated with a higher chance of overfitting.

This evolution highlights ML's transformative role, linking statistical foundations to more flexible paradigms. However, a structured analysis by learning paradigm and supervised techniques for labelled predictions and unsupervised techniques for pattern discovery provides a deeper insight.

2.2.2.1 Supervised ML Techniques

Supervised ML techniques have become more dominant in electricity load forecasting by using labelled datasets to model complex, non-linear relationships between inputs, such as historical demand, weather, and temporal features, and they frequently outperform traditional statistical methods in terms of accuracy, and are often interpreted as non-interpretable and dependent on data (Rubow, 2025; Satyanarayana et al., 2023). Ensemble techniques like RF and its variants of Gradient Boosting techniques (e.g., XGBoost, LightGBM) are a good example of this change, where multiple weak learners can be aggregated to counter overfitting and capture seasonal and disruptive patterns to offer a strong structure to tackle heterogeneous data from smart grids (Ibrahim et al., 2022; Jain et al., 2024). For example, in situations where data is scarce, or data variation is high, optimised ensembles expose both strengths and weaknesses: LightGBM and RF in German utility forecasts not only exposed economic aspects of bias, where low granularity caused higher costs despite R^2 values ranging from 0.57 to 0.81, challenging over-reliance on voluminous data without addressing quality (Rubow, 2025). This highlights a crucial argument that, although ensembles boost the predictive precision by regularization and feature integration, which can be seen in Indian power systems, where the DT classifiers achieved a near-perfect accuracy in predicting trends in electricity generation, outperforming SVM and Naive Bayes classifiers (Satyanarayana et al., 2023), they require a complex preprocessing to prevent the amplification of errors in noisy environments.

Comparatively speaking, the regression-based methods such as Support Vector Regression (SVR) and DT can provide computational efficiency, although different in terms of multivariate input, as the ensemble approaches are more resistant to noise in seasonal forecasting, as well as reducing Normalized Mean Squared Error (NMSE) by 1.35–13.51% when combined with neural variants such as LSTM (Jain et al., 2024). This complementarity can be seen in the case of the grid of Panama, for which DNN regression performed well for hourly predictions (R^2 0.93) compared to the limitations of AdaBoost for its weaknesses, but in multi-step horizons, boosted ensembles were preferred because of their regularisation at the expense of increased training times (Ibrahim et al., 2022). More crucially, these works indicate a significant conflict between supervised ML enhancing adaptability, i.e., the incorporation of humidity and past loads as the best predicting features, but with the black box predisposed to causal explanations and compelling the field to hybrid optimisations between precision and explainability of volatile energy markets (Rubow, 2025; Satyanarayana et al., 2023). The developments provide a good argument in support of supervision techniques in scalable forecasting, in which automated tuning is a part, to cut through the computational overheads and data limitations.

2.2.2.2 Un-Supervised ML Techniques

Self-directed ML techniques that learn the intrinsic properties in unlabeled electricity load data without the known results, provide a paradigm shift to the supervised techniques of learning latent properties that enhance forecasting resilience in the condition of data scarcity and nonlinearity (Bae et al., 2025; Tavakoli et al., 2020). The key areas here are clustering algorithms, including k-means and hierarchical algorithms, that group similar load profiles together to bypass the curse of dimensionality in large-scale data to enable more specific prediction. A detailed study by (Moradzadeh et al., 2022) focused on the application of Variational Autoencoders (VAEs) alongside bidirectional LSTMs aimed at predicting STLF. The researchers effectively enhanced the R^2 values by identifying hidden patterns with respect to unlabeled data. However, the substantial computational requirements make the approach challenging for real-time implementation, especially within dynamic power grid systems. This highlights an important point that while clustering is good at identifying patterns in segmentation (e.g., identifying peak vs. off-peak behaviours), it requires good initialisation to prevent poor clustering (Moradzadeh et al., 2022) for an example where VAEs were employed to improve clustering for bidirectional LSTMs forecasting, with great results in terms of R^2 values, but at the cost of computation power.

Dimensionality reduction techniques, Principal Component Analysis (PCA) and autoencoders, further boost unsupervised efficacy by compressing high-dimensional load time series into salient features, capturing variance while limiting noise. The study by (Tavakoli et al., 2020) suggested an autoencoder-based deep clustering for time series, converting unlabeled financial load proxies into supervised-like predictions with 87.5% accuracy, and suggested that such nonlinear embeddings were better than linear PCA at capturing temporal volatilities, while at the same time risking a loss of information from sparse datasets. Complementing this, (Li et al., 2023) used VAEs for renewable-integrated load forecasting, showing the power of latent space representations using Kullback–Leibler (KL) divergence optimisation in improving anomaly detection and prediction stability (with RMSE improvements of up to 15% compared with raw inputs), but the method's probabilistic assumptions break down in the presence of non-Gaussian noise, critiquing the generalizability in diverse climates. A trade-off is thus revealed: unsupervised methods support adaptability to unlabeled, heterogeneous data that is important for developing smart grids of the future, but which tends to lose interpretability due to reduced features that lose the causal relationship between variables such as weather and demand (Bae et al., 2025).

Hybrid unsupervised paradigms by combining clustering with autoencoders overcome these gaps by jointly optimising the representation and grouping tasks for scalable forecasting. According to the findings of (Varanasi et al., 2019), PV-integrated load forecasting through k-means on meteorological clusters obtained from VAE showed improvements of 10-12%, in research that shows synergies to alleviate overfitting in unsupervised regimes, but has exposed scalability issues for real-time applications. Critically, these advancements make the case for unsupervised techniques as basic preprocessors for progressive boost of downstream supervised models, while requiring hybrid precautions for cluster instability and dimensionality challenges, and therefore bridging to the hierarchical capabilities of deep learning in the following section.

2.2.3 DL Forecasting Approaches

DL methods have emerged as an advancement of ML paradigms in electricity consumption prediction that address the deficiencies of supervised and unsupervised methods in modelling hierarchical, nonlinear temporal dependencies at the scale of large, volatile load datasets (Dong et al., 2025; Percuku et al., 2025). Using the exploitation of architectures such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and transformers, DL models can be trained to capture the rich features of multivariate time series in an autonomous manner, capturing more factors such as weather and spikes in demand than classical ensembles. However, this excellence comes with the tradeoff of higher computational demands and non-interpretability that needs a critical evaluation of their relevance in a real-world environment where data sparsity or noise can increase overfitting (Bae et al., 2025; Yang et al., 2025).

Recurrent architectures like LSTM and GRU excel in STLF by capturing long-range dependencies but struggle with scalability in high-dimensional data (Percuku et al., 2025). For

MTLF, LSTM-based models reduce errors by 10–20% in smart grids, though computational demands increase (Jahan et al., 2020; A. Wang et al., 2023). The research work of (Percuku et al., 2025) applied LSTM to four-year substation data from Kosovo, yielding MAE of 6.231 and RMSE of 8.418 for winter predictions, but noted that it is inferior to the simpler linear regression in the smaller data sets, observing that LSTM's dependence on extensive datasets makes it less effective in grid contexts where computational and data resources are limited. This weakness is reflected in hybrid combinations of LSTMs and CNNs, in which LSTM is combined with CNN to achieve better spatial feature extraction, as (Zairi et al., 2025) showed that a CNN-GRU hybrid system lowered RMSE and MAPE by capturing the local and temporal pattern in Saudi Arabian load data, outperforming stand-alone DL models by as much as 20%. Such syntheses highlight an interesting case: although pure RNNs are good at sequential modelling, their combination with convolutional layers solves the gradient vanishing problem and produces good forecasts with nonlinearities, although with an additional training complexity (Muqtadir et al., 2025; Wen et al., 2024).

Transformers and attention mechanisms are another step towards advancing DL to enable parallel processing and the selective focus on the salient features, solving the inefficiencies of RNNs for long sequences. According to (Zhong, 2023) proposed an ANN–LSTM–Transformer model employing self-attention was proposed to capture spatiotemporal associations and achieved superior MAE and RMSE results on UCI Electric Load datasets from the University of California, Irvine ML Repository. Similarly, External-Convolution Attention in CNN-TCN-BiLSTM Hybrid showed an increase in accuracy by 2- 21% for different data sets, proving attention is an important aspect of prioritising weather-load interactions (Zare et al., 2025).

Recent optimisations, including self-adaptive evolutionary networks, have furthered DL's potential but exposed trade-offs with efficiency. The Self-Adaptive Differential Evolution (SADE)-optimised Kolmogorov-Arnold Network (KAN) achieved MAPE of 3.57–3.87% on ISO-NE data, with fewer parameters (132,000 vs. 394,000 for MLPs) in the trade-off between accuracy and cost, critiquing traditional DL for being parameter-inefficient (Abbas et al., 2025). Hybrid RNN-LSTM models, which were further optimised by the addition of meteorological features, produced N-RMSE of 0.16 while MAPE of 4.06% (Demir et al., 2025), which further highlighted the importance of feature engineering. However, (Yang et al., 2025) used CNN-BiLSTM in combination with SHapley Additive exPlanations (SHAP) for interpretability in very short-term forecasting, while achieving state-of-the-art results, the authors claimed that black box DL prevents causal explanations and therefore explainable AI integrations must be made, in view of regulatory requirements in energy markets.

These studies collectively underscore a compelling case for DL's hierarchical approach, which delivers superior precision in STLF and MTLF, as evidenced by consistent reductions in error metrics (Dong et al., 2025; Zairi et al., 2025; Jahan et al., 2020; Kotsiopoulos et al., 2023). That said, the computational heft and sensitivity to data quality pose real challenges, demanding robust hybrid safeguards and meticulous preprocessing to keep things on track. Even though certain models, such as GRU-TCN-Attention, deliver over 95% accuracy with 40–48% efficiency improvements (Wen et al., 2024), the existing weakness of extreme events needs to be tackled, and this is why future research on the creation of resilient architectures is required. This development preconditions the multi-platform implementations of the section above, where the data-driven capabilities of DL are combined with the deployment strategies of real-world analytics.

2.2.4 Hybrid Forecasting Approaches

Hybrid forecasting models have also taken a leading role in the electricity consumption analytics sector by innovatively integrating the statistical rigor of statistical forecasting with the flexibility of ML and DL to overcome the long-standing limitations of single-modality approaches, including the inability of ARIMA to account for the influence of nonlinear volatility or the challenge of disappearing gradients in LSTMs when using long sequences (Ashtar et al., 2025; Dong et al., 2025). This approach smartly blends the linear trend detection of statistical tools with the knack of neural networks for picking up residual nonlinear patterns, boosting accuracy for both STLF and MTLF as smart grid complexities keep ramping up (Ashtar et al., 2025; Hussain et al., 2025a; Jahan et al., 2020). Critically, the hybrids challenge the paradigm of model isolation by promoting architectures that are synergistic in the sense that they increase predictive resilience to data heterogeneity at the cost of often over-proportional computational overhead, which calls into question their viability in edge-computing scenarios, where latency is a prime concern (Ugbehe et al., 2025; Wen et al., 2024).

The basis of hybrid efficacy is decomposition strategies, which subdivide time series into parts that can be interpreted to minimise noise and seasonality. The study (Su et al., 2025) performs a similar practice with Hodrick–Prescott (HP) filtering to segregate trends for ARIMA modelling and cyclic residuals for RNNs, resulting in MAPE reductions between 9.70% and 35.66% over benchmarks such as Holt–Winters on Guangzhou data; but, its application is subject to criticism for possible over-smoothing in high variance renewables-integrated series, where the signal's fidelity is paramount to avoid forecast biases (Bae et al., 2025). Similarly, (Sinha et al., 2021) used Vector Auto Regression (VAR) to separate linear dependencies, sending nonlinear remains to CNN-LSTM hybrids, with RMSE reductions of 20–50% on UCI and Ontario data compared to ARIMAX with a persuasive claim that residual stacking limits the magnitude of error propagation at the cost of compounding assumptions from different paradigms, which could compromise robustness in non-stationary environments (Zairi et al., 2025).

DL infusions take hybrids to a higher level, particularly in the capture of spatiotemporal intricacies. In a recent study, (Wen et al., 2024) designed a GRU-TCN-Attention ensemble, where sequential memory processing is handled by GRU, convolutional patterns by TCN, and feature salience is further enhanced by attention, reducing MAPE by 39%. The benchmark dataset introduced in the Global Energy Forecasting Competition 2014, which provides real-world electricity load and weather data for model evaluation. This setup makes a compelling argument for the role of attention in exogenous variable prioritisation, surpassing the performance of vanilla RNNs, but suffers from criticisms of parametric density causing overfitting in sparse data and increased training costs (Yang et al., 2025). Echoing this, (Zairi et al., 2025) CNN with GRU/LSTM/BiLSTM hybridised, with noisy Saudi data filtered to minimise RMSE, showed hybrids' edge in peak demand management, but bidirectional LSTM variant's 2-way flow, while enriching context, doubles computational load, highlighting trade-offs between depth and efficiency hybrids evade. (Muqtadir et al., 2025).

In sustainable forecasting, hybrids offer a skillful incorporation of multifaceted inputs, as (Ashtar et al., 2025) combined SARIMAX with LSTM for demand in the Netherlands with 1.88% MAPE at 180 days using a blend of statistical seasonality and the adaptability of DL, while exogenous renewables improved short-term accuracy (MAPE from 2.13% to 1.09%) but offered marginal increases in accuracy at longer time horizons, while socio-economic factors sometimes increased errors due to multicollinearity, criticizing the over-reliance on Stacked innovations, such as (Elakkiya et al., 2025) transformer-ANN-fuzzy logic framework, employ self-attention for long dependency and fuzzy preprocessing for noise, achieving 15–20% reduction in RMSE over baselines using Texas/New York data, pushing another line in the argument that fuzzy logic moderates ANN's brittleness to irregularities, although interpretability is elusive, requiring tools such as SHAP to acquire regulatory compliance (Yang et al., 2025).

The conversation around clustering and dimensionality hybrids takes an intriguing turn with (Su et al., 2025), who paired Hodrick-Prescott filtering with RNNs to reduce MAPE by an impressive 9.70–35.66% in electricity demand forecasting. This setup enhances DL's ability to handle high-dimensional data with finesse. Still, there's a limitation in smoothing in volatile series, especially those tied to renewables, which could skew predictions if not carefully managed (Su et al., 2025). However, interpretability, as a trade-off, represents a defining characteristic of hybrid models compared to non-explainable counterparts, argues a transparency gap: a gap between scholarly innovation and field realisation, against black-box critiques that pervade stand-alone DL, (Yang et al., 2025) CNN-BiLSTM with SHAP (Percuku et al., 2025).

Finally, the superiority of the hybrids derives from their calibrated combination that exceeds monolithic models on accuracy and flexibility in the evolving markets, as (Ugbehe et al., 2025) review affirms their preference towards accuracy in reference to the Nigerian grid issues. However, the increasing complexity requires careful design to avoid performance saturation resulting from over-parameterisation, which calls for future focus on lightweight hybrid models for scalable multi-platform analytics in the above-mentioned section.

2.2.4.1 Bagging and Boosting Methods

Bagging and boosting, being ensemble methods, have been very helpful for the development of hybrid forecasting paradigms in the electricity load prediction problem by solving the variance-bias imbalance that is inherent to the performance of single models, which have resulted in more robust and accurate results in volatile energy systems (Khwaja et al., 2020; Muqtadir et al., 2025). Bagging is a variance reduction technique based on bootstrap aggregation that creates various subsets that are trained in parallel, including (Srivastava et al., 2020), which compared RF and Bagging regression trees in Australian market data, and the ensemble diversity enabled stabilisation of nonlinear load patterns. This reduction in variance is vital to STLF, in which stochastic swings of weather or consumer behaviour further inculcate the instability of forecasting, but the parallel character of bagging typically fails to factor in error-correction sequentially and thus, may instead perpetuate biases in high-dimensional time series against its effectiveness as a predictive standalone strategy in dynamic grids (Qinghe et al., 2022).

Boosting, on the other hand, optimises weak learners by calibrating them so that they reduce bias, boost performance, and the dominant variant is now XGBoost, which has regularisation and tree pruning that help prevent overfitting and make it compatible with sparse features such as temporal lags or renewables integration. According to (Qinghe et al., 2022), XGBoost using Tree-structured Parzen Estimator on the grid data from the region of Ile-de-France with a MAPE of 2.61%, an improvement of 14.14% compared with the baselines, arguing that boosting's adaptive weighting is very powerful to capture seasonal dependencies, but sequential dependency increases the computational cost, which may be non-scalable in real-time applications (Semmelmann et al., 2022). The integration of bagging and boosting techniques, as seen in (Khwaja et al., 2020) with their bagged-boosted ANNs, cleverly merges the strengths of both to cut down on bias and variance, delivering stronger results for STLF and MTLF. Their approach shines with lower MAPE on real-world datasets compared to standalone ANNs, pointing to better generalisation across the board. That said, the sensitivity to hyperparameters calls for automated tuning to keep things in check, as highlighted by (C. Wang et al., 2019; A. Wang et al., 2023) and (Jahan et al., 2020). However, one continuing problem with these black-box ensembles is their interpretability; knowing their internal decision-making processes is still a critical problem (Khwaja et al., 2020; C. Wang et al., 2019; A. Wang et al., 2023; Jahan et al., 2020).

Critically, boosting (LightGBM, CatBoost, etc.) integrated in (Muqtadir et al., 2025) can deal with categorical features and generalisation across multi-site residential loads, delivering RMSLE of 0.1898, a testament to the edge of boosting methods in noisy, multivariate data. However, ablation studies reveal diminishing performance gains of increasing complexity, so overfitting vulnerabilities are present without regularisation (Praca et al., 2020). In energy communities, (Semmelmann et al., 2022) LSTM-XGBoost hybrid uses boosting for peak refinement post-LSTM sequencing for better than standard profiles, but multicollinearity from smart meter inputs causes error inflation, critiquing unrefined feature engineering. Ultimately, although bagging stabilises and boosting intensifies precision, their hybrid implementations require particular validation for preventing computational inefficiencies, strengthening the cases for explainable ensembles in sustainable grid analytics.

2.2.4.2 Stacking and Blending Techniques

Stacking and blending as meta-ensemble methods have raised the level of hybrid forecasting by hierarchically incorporating the outputs of base learners, overcoming naive averaging methods regarding the ability to capture non-linear temporal dynamics and to reduce the individual model biases in electricity load prediction (Elakkiya et al., 2025; Sinha et al., 2021). Stacking uses a meta-learner to weight or regress upon base predictions, which promotes superior generalisation, as in the case of the transformer-ANN-fuzzy stack that improved RMSE by 15–20% on Texas/New York datasets by leveraging self-attention to capture dependencies and fuzzy to capture noise, arguing cogently that stacking's layered abstraction is superior in volatile grids, but its parametric escalation runs the risk of overfitting if not cross-validated, critiquing computational viability for real-time MTLF (Wen et al., 2024).

Blending, on the other hand, incorporates some prediction blending by using weighted fusion (and this is often simpler, but works well for bias-variance equilibrium). The study performs, (Sinha et al., 2021) VAR-CNN-LSTM residual blend isolates linear trends for VAR, nonlinearities for DL, resulting 20–50% RMSE improvements on UCI/Ontario data; this assumes blending's edge in residual decomposition for seasonal volatilities; however, unweighted schemes undervalue learner diversity, perhaps increasing errors for renewables-integrated systems, where multicollinearity from aggregated community data inflates variance (Semmelmann et al., 2022; Nespoli et al., 2025). In cases of grid optimisation, blending of global non-parametric models for Energy Hub (EH) and Heat Pump (HP) flexibility optimisation of control signals to reduce rebounds and costs was introduced by (Nespoli et al., 2025), illustrating the practicality of blending for sustainable integration; however, its dependence on simulation exposes sensitivity to input fidelity as the storage flexibility blending introduced in (Hale et al., 2016) revealed a decreased return from uncalibrated weights in high-RES scenarios.

More importantly, the meta-level sophistication of stacking, following the sophisticated 2022 New York integration analysis, which stacked decarbonization pathways, introduced accuracy in the face of policy uncertainties but must be carefully checked to prevent errors at the bottom, as proposed by (Gumerman et al., 2006) inter-regional congestion stacking as an example in the National Energy Modelling System (NEMS). Blending, as defined by (Pierpont et al., 2017), optimises low-carbon grid paths by combining a set of flexibility measures, but does not consider any interaction between heterogeneous loads, (Kiguchi, 2021). Smart meter blending to Demand Response (DR) was accused of ignoring temporal lags. The approaches appeal decisively to hybridisation in the sphere of efficiency and consumption of renewables in respect of automated tuning to deal with scalability, and in respect of analytics versus grid volatility.

2.3 Smart Buildings: MTLF Approaches

The central feature of smart buildings is MTLF, which is likely to be between weeks and months, and is critical to proactively manage energy and optimise energy resources and connect them with Demand Response (DR) processes as more and more renewable energy becomes integrated into the electricity, and consumption patterns change (Jahan et al., 2020; Hussain et al., 2025a). Unlike STLF, which is oriented at instant operation adjustment, MTLF is a strategic forecasting, which involves the organisation of the maintenance work or tariff negotiations, yet is more unpredictable due to the external factors, such as weather changes and human behaviour (Mir et al., 2020). Recent research in 5-6 years (2019 to 2025) indicates a tendency to de-emphasise classical statistical methods, and instead adopt hybrid ML and DL methods, more capable of describing non-linear relationships between smart building data. Nevertheless, this development reflects the continuing problems in data quality, model and computational practice generalisations, and their effectiveness and constraints must be critically assessed.

Classical techniques such as Linear Regression (LR) and ARIMA have been a foundation in MTLF and, in most cases, do not work in the case of smart buildings, due to the inability to handle stochastic elements such as missing data or seasonal anomalies (e.g., Dogoulis, 2022; Rai et al., 2021). For example, Rai and De (Rai et al., 2021) applied classical methods and ML models for smart grid data from an institutional campus and concluded that SVR is superior to Multiple Linear Regression (MLR) and optimised Holt's method when applied to MTLF with an MAPE reduction of up to 3.6% by incorporating the meteorological variables such as temperature and humidity. This highlights how the classical approaches, although being computationally light-weight, fail to synthesise the complex determinants of population growth, GDP, and historical load prevalent in low and middle-income contexts (Mir et al., 2020). In contrast (Jahan et al., 2020), ML ensembles integrate these factors more robustly synthesized Artificial Neural Networks (ANN), SVR, DT and fuzzy sets in a lightweight framework, fail to synthesize the complex determinants of population growth, GDP, (FS) across utility systems, arguing that hybrid models reduce overfitting and increase accuracy by 10-20% over singular methods, but require a lot of preprocessing.

DL techniques, especially the LSTM network, have proven to be better for MTLF in smart buildings, using sequential data from smart meters to solve temporal dependencies (Javedian, 2020) critically analysed the robustness of LSTM to random missing data patterns of commercial smart buildings and showed that ML-based imputation (e.g., k-Nearest Neighbours) outperforms statistical hybrid when missing values are more than 10%, and reduced the forecasting errors up to 15%. This is echoed in Javedian (Javedian, 2020), which applied LSTM and other variants of DL to institutional buildings, demonstrating their accuracy of prediction (RMSE below 5%) by combining weather and occupancy data, although critiquing them as black boxes and hence lacking in interpretability in energy policy applications. Hybrid DL-ML frameworks further enhance such advantages, e.g., Lin et al., 2022 proposed a Resistance-Capacitance (RC) thermal model combined with ensemble predictors (SVM, BPNN, GRNN) for residential loads with superior performance via genetic algorithm optimisation, but mostly validated for short-term horizons with mid-term extensions.

Integration with DR and prosumers management represents a promising frontier, as MTLF models increasingly incorporate renewable generation forecasts to balance supply-demand dynamics (Ruiz-Abellón et al., 2019). Recent advancements, such as ensemble voting regressors (e.g., LightGBM, XGBoost) for MTLF, demonstrate enhanced robustness by aggregating predictions, yielding R^2 scores above 0.95 in educational settings (Gupta et al., 2025). Nonetheless, these models often exhibit bias toward high-income contexts, under-representing determinants like economic volatility in low- and middle-income countries (Mir et al., 2020).

Chapter 3

Methodology

3.1 Overview of the Research Methodology

The methodological framework constructed in this research has been aimed at systematically addressing two interrelated problems of data-driven energy analytics for smart commercial buildings: (i) Mid-Term Electrical Load Forecasting (MTLF) and (ii) the robust imputation of missing data in multivariate time series as they come from intelligent metering infrastructures. Considering the integrated elements of smart buildings, such as the dynamic sensor measurements, communication networks, environmental factors, and operational timelines, these systems can be classified as complex Cyber-Physical Systems (CPS). Therefore, the input data, including quality, accuracy, and temporal continuity, makes a significant contribution to the forecast and prediction accuracy.

This research is based on a rigorous methodology that is divided into several stages involving data collection, data preprocessing, missing data modelling, missing data imputation strategies, forecasting techniques, ML, DL, and hybrid model architectures, hyperparameter optimisation techniques, and comprehensive evaluation metrics. The overall goal is to build an end-to-end methodological pipeline that will be able to provide resilient, scalable, and generalizable forecasting solutions that can be deployed in real-world smart building environments.

The methodological design draws from the experimental research principles and comparative model evaluation. Real operational data, as collected by a smart commercial building, is used to build a representative and real-time series forecasting problem. Because real-world buildings are associated with imperfections in the data, e.g., sensor malfunction, packet loss, environmental interference, etc. (fabricated missingness generation is one method), to ensure that this methodology combines the synthetic missingness generation and various imputation methods to test the robustness of the models in controlled environments. Statistical, ML-based, and hybrid imputation techniques are combined into the forecasting pipeline in order to test their impact on downstream MTLF. Furthermore, various families of forecasting models, including traditional ML regressors, RNN, Gated Recurrent Units (GRUs), LSTM, and two-stage hybrid ML-DL models (such as FireNet-XGBoost), are implemented in order to capture linear, non-linear, and temporal dependencies within the data. This diversity ensures that the methodology is not based on a single modelling paradigm but actually provides a holistic comparative evaluation.

The whole process is fully managed by a structured set of experiments, featuring the chronological data splitting and the five-fold time series cross-validation as well as standardised performance measures like MSE, RMSE, MAPE,R^2. Hyperparameter optimisation is carried out by using grid-search and Optuna-based Bayesian in order to make a fair and unbiased comparison among models. There exist considerations of ethics, standards of reproducibility, and computational transparency that are encompassed in the methodological design. Overall, this is a well-developed, rigorous empirical basis on which to analyse the interactions of data quality, imputation strategies, and modelling choices, which affect the MTLF performance. It enables the development of predictive systems that can facilitate energy-saving working, demand response, and sustainable urban energy management in advanced smart buildings .

3.1.1 High-Level Research Framework

The research presented in this dissertation is based on a well-planned and multi-stage research route with the aim of systematically mitigating the following dual objectives: (i) the development of accurate MTLF models for smart buildings, and (ii) the evaluation of the resilience of load forecasting models under different degrees of missing data and imputation approaches. The combination of several different interdependent methodological elements, including data collection, preprocessing, model development, and performance evaluation, is integrated into a unified operational pipeline in the high-level research framework. This makes it possible to ensure that every phase contributes significantly to the overall enhancement and evolution of the forecasting system.

At the basis of the research workflow is the data collection stage, which includes importing computational libraries that will be needed to run the calculations, extracting meteorological variables from credible online sources, and collecting electrical load data from smart circuit breakers and cloud-based monitoring platforms. This stage focuses on the importance of high-resolution and real-world datasets in order to maintain ecological validity and realistic system behaviour. The second stage, data processing, deals with the preparation of raw datasets for modelling. This involves setting up a structured database, outlier detection, fixing inconsistencies, a timestamp aligner, and missing region detection. Preprocessing is essential due to the high risk of poor data quality that can reduce the model accuracy and also propagate errors in the data through the forecasting pipeline.

The third phase focuses on model development, which includes the training and assessment phases of a wide range of algorithms, namely ML-based, DL-based, and also hybrid models. Traditional ML models like RF, XG Boost, DT, SVR, along with advanced DL techniques, are used as LSTM and FireNet. Hybrid models using feature extraction of DL and regression of ML (e.g., FireNet-XGBoost) are at the core of the proposed forecasting strategy. Feature selection based on analysis of correlation and hyperparameter optimisation using grid search and Optuna ensures that model configurations are efficient and unbiased. Finally, the fourth stage is performance evaluation, in which the forecasting accuracy, stability of the model, and the efficiency of the calculations are analysed using various performance measures such as MSE, RMSE, MAPE,R^2. Cross-validation procedures are further used to ensure that the generalisation of the models is well studied in multiple temporal partitions.

Figure 3.1 gives a visual representation of this high-level research frame and the sequential flow of the process from data acquisition to final evaluation. This block diagram is conceptual and is used to understand the interaction of the individual components to form a cohesive forecasting methodology.

3.2 Data Sources

This research uses two main data sources: (a) high-resolution building-level electricity demand from a smart commercial Building; and (b) meteorological data extracted from globally-trusted meteorological databases. Collectively, these datasets give the multivariate time series basis needed to assess MTLF models and missing-data imputation approaches in a realistic operating state. A combination of electrical and meteorological data simplifies the predictive models' use of both demand-side trends and external climatic variations.

3.2.1 Electrical Load Dataset

The electrical load data set was collected from a 30,000 m^2 smart commercial building in Bergamo, Italy. The whole facility is covered with advanced metering infrastructure, with IoT sensors, smart circuit breakers, and cloud-connected monitoring infrastructure, which records detailed electrical characteristics at hourly resolutions. The dataset includes 8,760

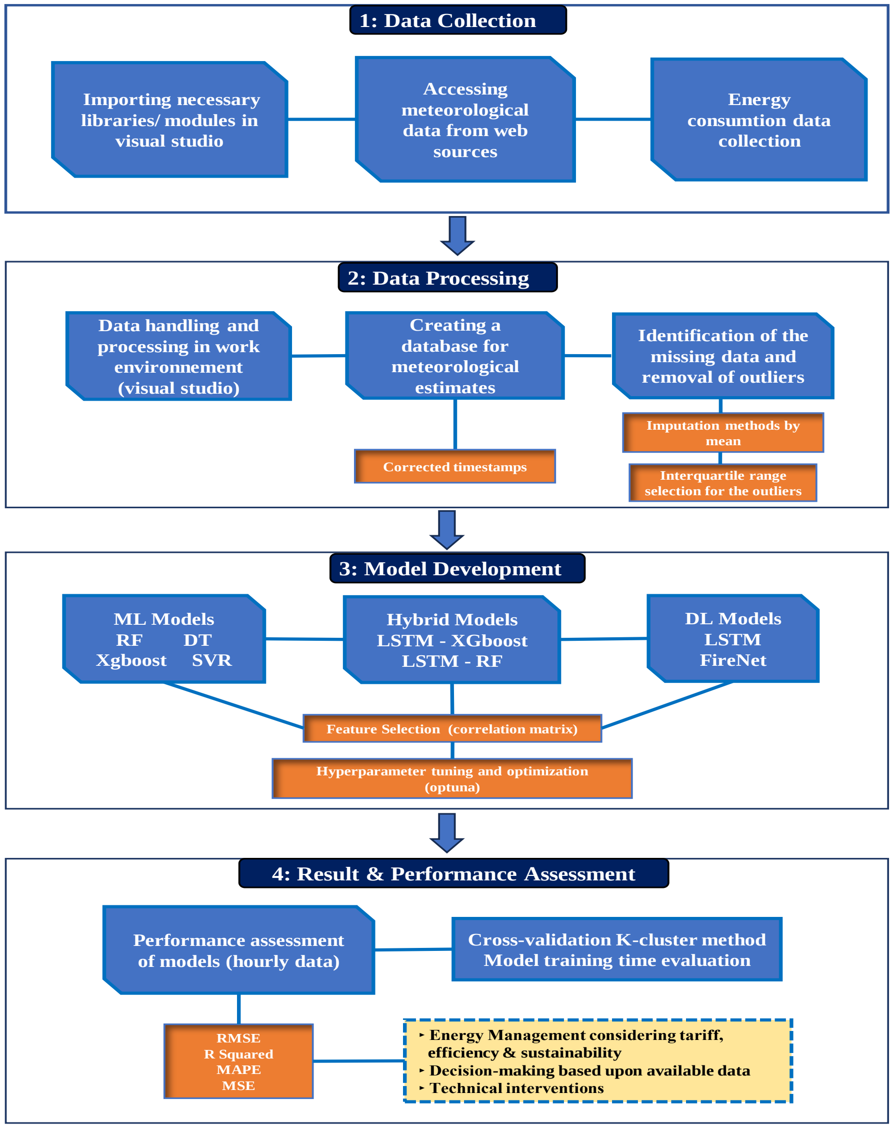


Figure 3.1: Proposed methodology for research framework

continuous hourly observations from January 1 to December 31, 2023 (Hussain et al., 2025b).

The dataset includes a comprehensive set of power-quality parameters:

- Active power: PAvg, PMin, PMax
- Reactive power: QAvg, QMin, QMax
- Apparent power: SAvg, SMin, SMax

These measurements provide reflections of the operation of equipment, HVAC operating patterns, occupant patterns, and seasonal behaviours, providing the right variability to train, validate, and perform robustness forecasting models. To give an initial visual characterisation of the load behaviour, the hourly active power for the entire year and aggregated daily, weekly, and monthly averages are shown in Figures 3.2 and 3.3. These plots are important in showing short-term variations, business hour peaks, declines on weekends, and seasonal demand changes, all patterns that are critical to the development of MTLF models.

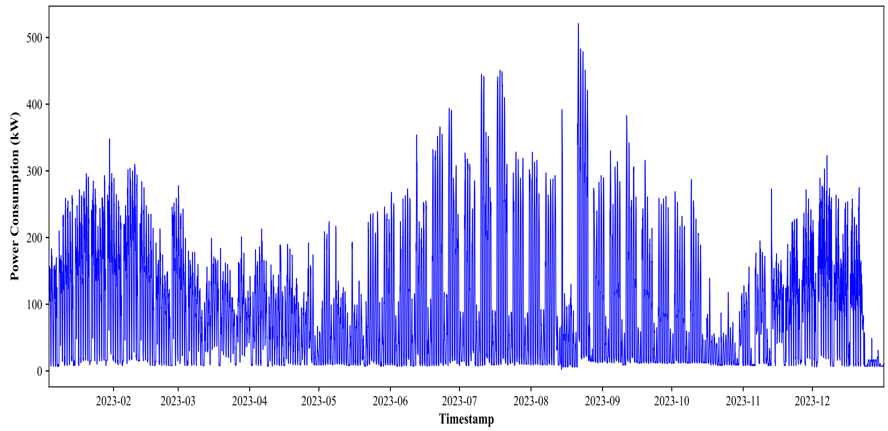


Figure 3.2: Time-based profiling of hourly active power consumption

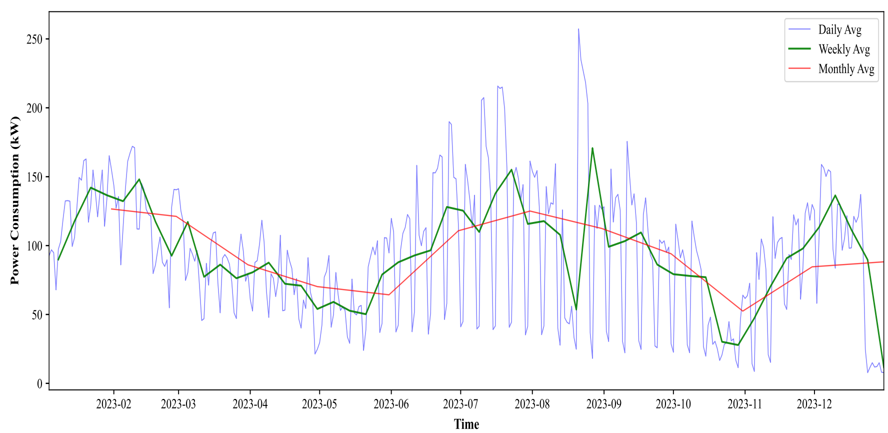


Figure 3.3: Daily, weekly, and monthly averages of active power consumption

3.2.2 Meteorological Dataset

To enhance the electrical dataset and to enrich the framework of the forecasting procedure, a multivariate meteorological dataset was collected from the available databases of the National Aeronautics and Space Administration (NASA) Power database (Chandler et al., 2013) and the Photovoltaic Geographical Information System (PVGIS) (Espino et al., 2024). These sources provide globally standardised and quality-checked environmental data for use in energy modelling applications. The meteorological variables for this research include:

- Outdoor temperature (°C)
- Humidity (g/kg)
- Wind speed (m/s)
- Beam irradiance (W/m^2)
- Diffuse irradiance (W/m^2)
- Sunshine duration (hours)
- Wind speed at 10 meters (m/s)

These parameters play critical roles in the prediction of the extent of energy consumption by a building, especially HVAC-related loads. The heating and cooling requirements are regulated by temperature, the heating gain within a building is regulated by irradiance, and the thermal exchange is regulated by wind. The meteorological dataset covers 2023 years, which allows for temporal alignment and for an increased feature set beyond the electrical dataset time period.

The electrical and meteorological data collectively have provided an extensive and realistic basis for the forecasting experiments and imputation experiments conducted in this study. Their real-world origin guarantees ecological validity, practical relevance, and the evaluation of the forecasting performance under conditions normally found in the environment of smart buildings.

3.3 Data Preprocessing

3.3.1 Outlier Detection and Removal

Outlier detection and removal are important first steps in the preparation of the electrical and meteorological datasets for modelling. The data of an operational smart building may contain irregular values because of sensor failure, electrical interference, network connection outage, or temporary equipment failure. These deviations, if not appropriately managed, may result in the alteration of statistical characteristics and also introduce substantial bias in the training processes of the models. Outliers in this research are identified through the Interquartile Range (IQR) measure, which is a powerful non-parametric technique that fits skewed or seasonally different time series (Fu et al., 2024). Observations below the lower bound value ($Q_1 - 1.5 \times IQR$) and above the upper bound value ($Q_3 + 1.5 \times IQR$) were flagged and removed/replaced by neighbouring temporal estimates. This operation ensures that the imputation and forecasting models are applied to consistent, reliable distributions, and the learning algorithms are not used to erroneously apply to the noise-based deviations.

3.3.2 Normalization

Normalisation is done to make sure that all the input variables contribute proportionately during the training process of the model. Because different parameters are present in the datasets, with various units and magnitudes, e.g., power (kW), temperature (C), irradiance (W/m^2), wind speed (m/s), etc., unscaled features can result in biased model convergence or unstable gradients in forecasting models' architectures (Passalis et al., 2020). The Min-Max normalisation technique is used to normalise the features between 0 and 1, keeping the original distribution intact, which helps relatively adaptive optimisation. This method of scaling is quite beneficial in the case of algorithms like LSTM, GRU, SVR, and neural networks, where gradient-based learning is sensitive to the scale of inputs. By making all features numerically consistent, normalisation improves the stability of the training, speeds up the convergence, and reduces the problems with domination by the high magnitude variables.

3.3.3 Temporal Feature Engineering

Temporal feature engineering is a way to enrich the data with time-dependent contextual features that are crucial to identify consumption patterns within smart buildings (Rakesh et al., 2025). The loads and patterns of demand can be substantially determined by such factors as occupancy schedules, operating hours, times of the year, and seasonal climatic cycles. In order to represent such dynamics, several temporal characteristics based on the index of the timestamp are obtained: hour of day, day of the week, indicator of weekend/weekday (weekend:1 / weekday:0), month, and season. These features allow the forecasting models to learn about cyclical patterns, such as daily peaks, weekly operational tendencies, seasonality of HVAC usage, etc. The temporal structure is explicitly encoded in the engineered features, and this approach therefore improves the generalisation capabilities of the models to predict more accurately over MTLF horizons out of simple historical trends, compared to when the models are not engineered.

3.3.4 Correlation-Based Feature Selection

Correlation-based feature selection is used to evaluate the relationship between electrical load variables and potential predictors (Verdonck et al., 2024). Pearson correlation coefficients are used to recognise features with strong positive or negative correlations to the dependent variable, and assist in removing redundant or lowly predictive predictors (J. Zhang et al., 2020). In this research, the Figure 3.4 shows the variables that have an absolute correlation value higher than 0.29 and lower than 0.88 are retained, which indicate some moderate and strong correlation relationships with explanatory power in forecasting tasks. This choice process helps in reducing dimension, reduces multicollinearity, and enhances the model interpretability while retaining important climatic and operational variables. By pre-processing the input space prior to model training, correlation-based feature selection helps to achieve more efficient computation, trading stability as well as forecasting power for the overall learning performance.

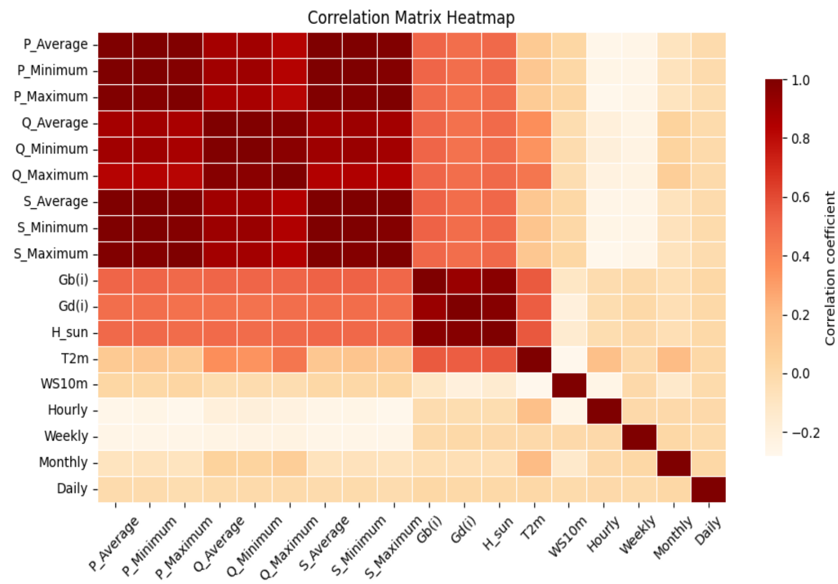


Figure 3.4: Pearson correlation among energy and meteorological features

3.4 Exploratory Data Analysis

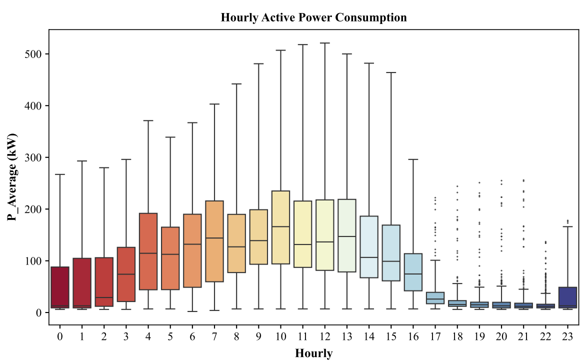
In order to develop a successful forecasting model, the temporal characteristics and the variation of data need to be analysed. This section discusses a multi-scale exploratory analysis performed with hourly, daily, weekly, and monthly data representations of the data. The visual insights allow obtaining an overall picture of load behaviour at various timescales, and it is helpful in the design of the sound forecasting models.

Four complementary perspectives on the energy usage patterns are shown in Figure 3.5. The hourly consumption profile Figure 3.5(a) shows large daily cycles, with peak loads occurring during business hours and much lower consumption during the night time. From 16:00, there is a sharp decrease in consumption, corresponding to lower occupancy and lower operational activity. This high-granularity view is especially relevant in the case of STLF tasks.

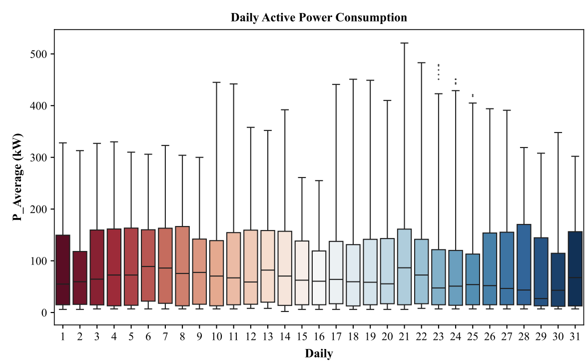
The daily load differences Figure 3.5(b) show the intra-week differences, with higher median values on weekdays than on the weekends, a typical feature of commercial buildings. Besides, the changes in the weekly dynamics may be related to external factors like weather, holidays, or non-operation of the building at full capacity. These factors lead to a non-stationarity of the data on short time scales.

In Figure 3.5(c), the comparison of weekdays and weekends clearly highlights the pattern of operation for the facility, validating the importance of the use of calendar-based features in forecasting models. Finally, the monthly resolution of Figure 3.5(d) reflects the greater seasonal trends. Although the total consumption throughout the year is relatively the same, there are some months where the consumption is higher due to the increased operation of the HVAC system during the summer or winter seasons.

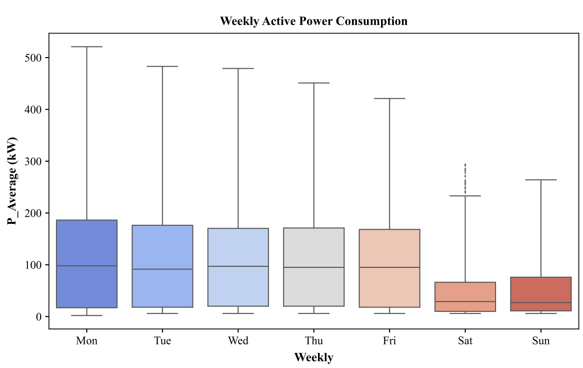
These four visualisations are useful to provide an overall picture of the temporal energy behaviour at various levels. Integrating such temporal heterogeneity into the input feature space can substantially improve both the capacity and prediction capabilities of both ML and ML models to predict energy.



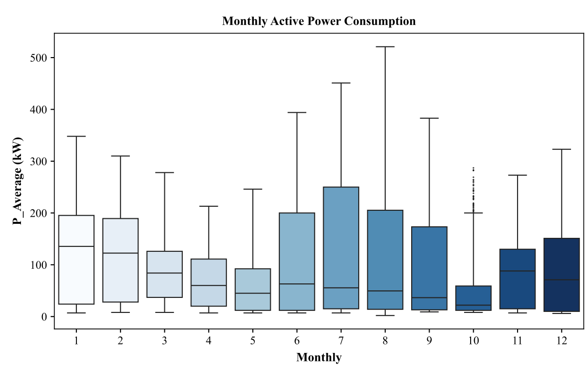
(a) Hourly consumption pattern



(b) Daily median consumption



(c) Weekday vs. weekend consumption



(d) Monthly load trends

Figure 3.5: Multi-resolution exploratory analysis of energy consumption: (a) hourly, (b) daily, (c) weekday/weekend comparison, and (d) monthly trends

3.5 Missing-Data Mechanisms

Reliable analysis of data requires a clear understanding of the mechanisms that are responsible for the missing values. Before coming up with strategies to handle insufficiencies from incomplete datasets, it is important to know the essential things that cause missing enter data. In the domain of statistical studies, the phenomenon of missing data is traditionally classified into three distinct categories (Osman et al., 2018): Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). These mechanisms are described below.

3.5.1 Missing Completely at Random (MCAR)

The probability of missingness is considered by the MCAR mechanism to be completely independent of both the observable and unseen variables. Under this assumption, the presence of missing values does not introduce bias, allowing unbiased estimation in principle. However, true MCAR conditions are rarely observed in the real world datasets (David et al., 2021). The MCAR mechanism can be expressed as:

$$f(Y \mid X, \theta) = f(Y \mid \theta) \quad \forall X, \theta \tag{3.1}$$

In Equation 3.1, X symbolizes a vector comprising observed values, Y signifies the indicators of missingness, θ represents an indeterminate parameter, and $f(\cdot)$ indicates the conditional probability distribution.

3.5.2 Missing at Random (MAR)

Under the MAR mechanism, missingness is associated only with the observed variables and not the unobserved components. A dataset that satisfies the MAR assumption may or may not produce biased estimates, depending on the modelling approach. Mathematically, MAR can be formulated as:

$$f(Y \mid X, \theta) = f(Y \mid X_{\text{obs}}, \theta) \quad \forall X_{\text{mis}}, \theta \tag{3.2}$$

In Equation 3.2, X_{obs} and X_{mis} represent the observed and missing components, respectively. The unknown parameter θ can be estimated by relating X_{obs} with other explanatory variables (Farewell et al., 2018).

3.5.3 Missing Not at Random (MNAR)

The MNAR mechanism occurs when missingness depends on unobserved variables, meaning the missing values are systematically related to the missing portion itself. As with MAR, MNAR datasets may or may not lead to biased parameter estimates. MNAR can be expressed in In Equation 3.3:

$$f(Y, X \mid \lambda, \theta) = f(Y \mid \lambda)\, f(Y \mid X, \theta) \tag{3.3}$$

where λ represents a parameter associated with the distribution of X estimated from the observed data, and θ characterizes the missingness pattern (Carreras et al., 2021).

Randomly missing data, particularly MAR, can distort patterns and reduce the predictive performance of models by introducing noise and bias (Lee et al., 2024). Therefore, effective imputation methods should be adopted in order to minimise the effects of the missing information and to achieve good predictive power.

3.6 Missing-Data Simulation

In this research, two distinct missing-data mechanisms are incorporated to replicate realistic disruptions that commonly occur in smart-building monitoring systems. The simulated missingness patterns are derived directly from the smart building electrical datasets for MTLF and imputation analysis. These patterns allow us to evaluate how forecasting accuracy declines in the presence of incomplete information and how different imputation strategies influence later prediction performance. Both missingness structures (linear block-type and random uniformly distributed) indicate the conditions found in actual operational deployments due to breakdown of equipment, instability of communications, and random network disturbances and their resulting data losses. The cumulative effect of all these mechanisms is a holistic basis of investigation into the resilience of ML, DL, and hybrid forecasting models in adverse data conditions.

3.6.1 Linear (Block-Type) Missing Data

Linear (block-type) missing-data mechanism is the structured data gaps through the permanent sensor outages, planned maintenance, or protracted communication failures. In our earlier study of the problem of missingness analysis using ML, block-type missingness was systematically introduced at severity levels from 6% to 30%, which showed that non-interrupted segments of missing values significantly affect temporal dependency that is crucial for MTLF

(Ahn et al., 2022). Figure 3.6 shows that block-type data gaps provide a mechanism of control over the strength of imputation techniques to retain continuity between daily, weekly, and seasonal cycles.

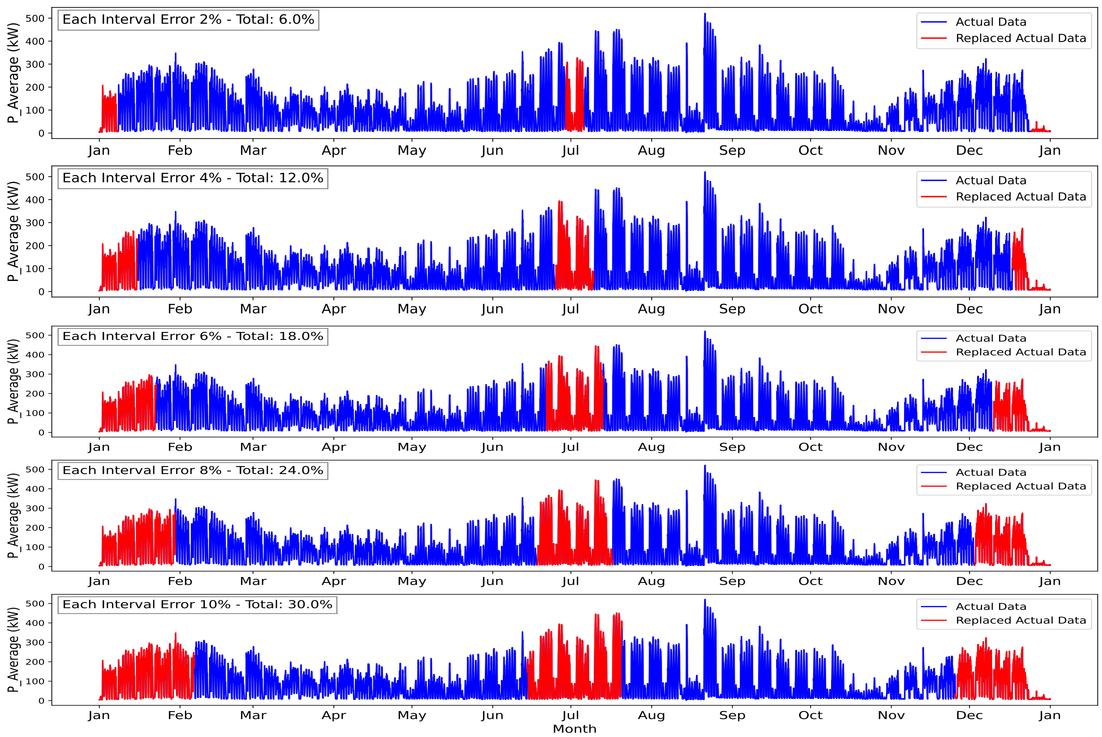


Figure 3.6: Impact of linear missing data intervals (6% to 30%) on active power consumption

3.6.2 Random Missing Data

The random missing data mechanism is a scenario of a missing data pattern in which individual observations are deleted without regard for time, throughout the time series. This type of missingness is typical of realistic operational conditions in smart building metering systems in which temporary disruptions, noisy network, or isolated sensor glitches result in irregular data loss at unpredictable time intervals. In contrast to the block-type data gaps, where missing values appear in contiguous segments, the random pattern is scattered throughout the entire dataset, resulting in small gaps appearing at different time locations with different frequencies.

In order to test the behaviour of the dataset under this kind of disruption, missing values were introduced at multiple predefined percentages between 5% and 40% (Joel et al., 2025). A probability of removal was assigned to each value in the series so that there was an unbiased and independent distribution of the missing points. This random missingness pattern is especially valuable when attempting to analyse the robustness of preprocessing and imputation methods because it is affected by irregular disturbance of short-term fluctuations and seasonal trends. The Figure 3.7 provides a good visual representation of how the continuity of the dataset is affected when the percentage of randomly missing values is increased, giving an insight into the complexity of the reconstruction of such incomplete time series data.

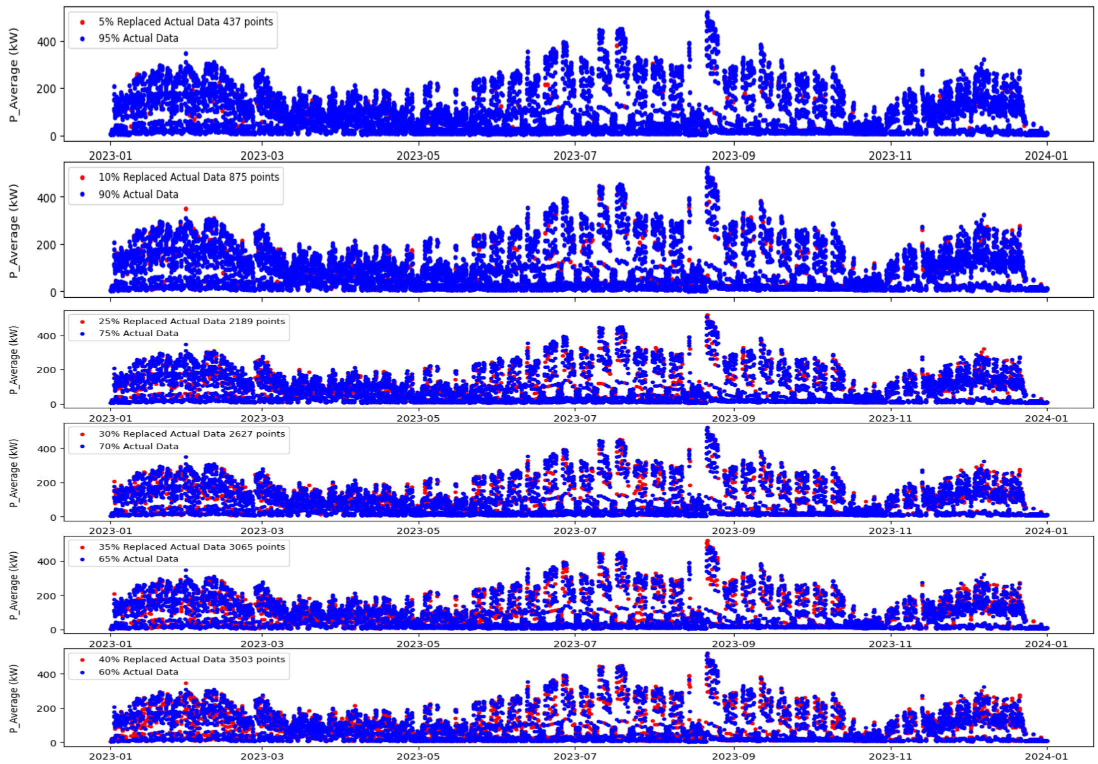


Figure 3.7: Impact of random missing data (6% to 40%) on active power consumption

3.7 Imputation Methodologies

Missing values are crucial in terms of making data-driven forecasting models reliable and robust. The imputation methodologies that were considered for this study belong to three broad families: statistical techniques, ML-based techniques, and hybrid approaches.

3.7.1 Statistical Imputation Techniques

Statistical techniques provide clear and computationally efficient strategies for the imputation of missing values. Mean and median substitution replace a missing entry x_i with the average or median of the observed samples (Alwateer et al., 2024), as shown in Eq. 3.4:

$$x_i = \begin{cases} x_i, & \text{if observed,} \\ \bar{x}, & \text{if missing (mean imputation),} \\ \tilde{x}, & \text{if missing (median imputation).} \end{cases} \tag{3.4}$$

For categorical variables, mode imputation is applied (Eq. 3.5):

$$x_i = \text{mode}(x). \tag{3.5}$$

Although easy to implement, these deterministic rules may distort data variance and weaken correlations.

3.7.1.1 Multiple Imputation by Chained Equations (MICE)

MICE is a multivariate statistical imputation method that iteratively models each variable with missing values conditional on the remaining variables (Buuren, 2011). At each iteration, missing values are estimated using:

$$x_{\text{miss}}^{(j)} = f_j\big(X^{(-j)}\big) + \varepsilon_j, \tag{3.6}$$

where $f_j(\cdot)$ denotes a variable-specific regression model, $X^{(-j)}$ represents the set of all other variables, and ε_j is a stochastic error term. By accounting for inter-variable dependencies, MICE preserves statistical relationships more effectively than simple imputation methods, at the cost of increased computational complexity.

3.7.2 ML-Based Imputation Techniques

Overall, these ML-based imputation techniques effectively capture nonlinear relationships and complex feature interactions present in real-world data. However, their performance is sensitive to hyperparameter selection and the availability of sufficient, representative training samples.

3.7.2.1 k-Nearest Neighbour (k-NN)

The kNN simplified formula for imputation, estimating the missing value using k as the closest samples (Emmanuel et al., 2021), as expressed in Eq. 3.7:

$$\hat{x}_i = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} x_j. \tag{3.7}$$

where $\hat{x}_i$ denotes the imputed value for sample i, k is the number of nearest Neighbours, $\mathcal{N}_k(i)$ represents the set of k samples closest to i, and x_j is the observed value from the j-th neighbor used for imputation.

3.7.2.2 Random Forest (RF)

Tree-based methods such as RF learn an ensemble of predictors (Tang et al., 2017); the imputed value is given by Eq. 3.8:

$$\hat{x}_i = \frac{1}{T} \sum_{t=1}^{T} f_t(X_i), \tag{3.8}$$

where $\hat{x}_i$ denotes the imputed value for sample i, T is the number of trees in the forest, $f_t(\cdot)$ represents the prediction function of the t-th tree, and X_i denotes the feature vector of sample i used as input to the model. This ensemble averaging mechanism enhances robustness by reducing variance and mitigating overfitting.

3.7.2.3 Support Vector Regression (SVR)

SVR imputes using the kernel-based representation in Eq. 3.9:

$$\hat{x}_i = \sum_{j=1}^{N} (\alpha_j - \alpha_j^*)\, K(x_j, x_i) + b, \tag{3.9}$$

where $\hat{x}_i$ denotes the imputed value for sample i, N is the number of training samples, α_j and α_j^* are the learned SVR dual coefficients, $K(x_j, x_i)$ is the kernel function measuring similarity between samples x_j and x_i, and b is the bias term (Smola et al., 2004). This approach enables SVR to model complex nonlinear relationships while maintaining strong generalisation performance.

3.7.3 Hybrid Imputation Techniques

Hybrid methods combine the strengths of multiple algorithms, for example, GB-MICE, RF-KNN, and Autoencoder-enhanced ML models. Although they are not associated with a single closed-form expression, they can be interpreted as composite models that leverage the complementary behaviour of the primary models under different missing-data mechanisms (MCAR, MAR, MNAR). These techniques facilitate a balance between the clarity of statistical analysis and the inherent flexibility of ML models.

3.8 Forecasting Methodologies

The forecasting models applied in this study are a combination of ML approaches, DL architectures, and hybrid ML-DL models. This section is to give a brief overview of the key approaches, without addressing each of the models separately.

3.8.1 ML Forecasting Models

ML algorithms such as DT, RF, kNN, SVR, GB, and XGBoost are based on nonlinear transformations that characterise the relationship between the input features and the associated target outcomes.

3.8.1.1 Decision Tree (DT)

Decision trees produce piecewise-constant predictions over regions R_m, as shown in Eq. 3.10:

$$\hat{y}_i = c_m \quad \text{if } x_i \in R_m. \tag{3.10}$$

where $\hat{y}_i$ denotes the predicted value for sample i, R_m represents the m-th region of the feature space defined by the decision tree splits, x_i is the feature vector of sample i, and c_m is the constant prediction assigned to all samples within region R_m. This structure allows DTs to capture nonlinear relationships through hierarchical partitioning of the input space (Klusowski et al., 2022).

3.8.1.2 Gradient Boosting (GB)

Ensemble methods such as GB update the model in a forward stage-wise manner (Zhu et al., 2022) Eq. 3.11:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \tag{3.11}$$

where $F_m(x)$ denotes the ensemble model at iteration m, $F_{m-1}(x)$ is the model from the previous iteration, $h_m(x)$ represents the m-th weak learner fitted to the residuals, and γ_m is the step size (learning rate) controlling the contribution of h_m. This iterative refinement process enables GB to progressively minimise prediction error.

3.8.2 DL Forecasting Models

DL architectures, including RNNs, LSTMs, and GRUs, demonstrate exceptional proficiency in modelling sequential data.

3.8.2.1 Recurrent Neural Networks (RNNs)

A standard RNN computes the hidden state and output according to to Eq. 3.12:

$$\begin{aligned} h_t &= \sigma(W_h h_{t-1} + W_x x_t + b_h), \\ \hat{y}_t &= W_y h_t + b_y, \end{aligned} \tag{3.12}$$

where h_t denotes the hidden state at time step t, h_{t-1} is the previous hidden state (Lipton, 2015), x_t is the input at time t, W_h, W_x, and W_y are learnable weight matrices, b_h and b_y are bias terms, $\sigma(\cdot)$ is a nonlinear activation function, and $\hat{y}_t$ is the model prediction at time t.

3.8.2.2 Long Short-Term Memory Networks (LSTMs)

LSTM networks alleviate vanishing-gradient problems through gating mechanisms. The input, forget, and output gates, as well as the cell and hidden states, are updated as shown in Eq. 3.13:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c), \\ h_t &= o_t \odot \tanh(c_t), \end{aligned} \tag{3.13}$$

where x_t is the input at time t, i_t, f_t, and o_t denote the input, forget, and output gates, respectively, c_t is the cell state, h_t is the hidden state, $W_{\{\cdot\}}$ and $U_{\{\cdot\}}$ are learnable weight matrices, $b_{\{\cdot\}}$ are bias terms, $\sigma(\cdot)$ is the sigmoid activation function, $\tanh(\cdot)$ is the hyperbolic tangent function, and $\odot$ denotes element-wise multiplication (Mena-Oreja et al., 2020).

3.8.2.3 Gated Recurrent Units (GRUs)

GRUs provide a lightweight alternative with fewer gates (Cho et al., 2014). The update and reset mechanisms are summarised in Eq. 3.14:

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}), \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}), \\ \tilde{h}_t &= \tanh(W_h x_t + U_h (r_t \odot h_{t-1})), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \end{aligned} \tag{3.14}$$

where x_t is the input at time t, z_t and r_t denote the update and reset gates, respectively, $\tilde{h}_t$ is the candidate hidden state, h_{t-1} and h_t are the previous and current hidden states,

$W_{\{\cdot\}}$ and $U_{\{\cdot\}}$ are learnable weight matrices, $\sigma(\cdot)$ is the sigmoid activation function, $\tanh(\cdot)$ is the hyperbolic tangent function, and $\odot$ denotes element-wise multiplication.

3.8.2.4 FireNet

FireNet is a DL model that was first created for image classification and, in this study, customised to apply to time-series data forecasting, building energy. The hierarchical convolutional feature extraction structure makes the model capable of capturing short/long-term temporal dependence among multi-source load signals. The model takes sequential input data that is fed through five stacked one-dimensional convolutional blocks, each consisting of max-pooling and batch normalisation operations.

The input to the network is a temporal matrix $X_{\text{input}} \in \mathbb{R}^{T\times C}$ (Eq. 3.15), where T denotes the number of time steps in the input window and C is the number of input features:

$$X_{\text{input}} \in \mathbb{R}^{T\times C}. \tag{3.15}$$

Each convolutional block learns temporal features at different resolutions. The first block (with 32 filters) captures short-term fluctuations such as hourly consumption spikes. The second block (64 filters) models daily cycles, while the third block (128 filters) extracts multi-day patterns. The fourth block (256 filters) focuses on weekly patterns, and the final block (512 filters) captures long-range dynamics driven by holidays, seasonal effects, or weather-induced variations (Dabrowski et al., 2020).

Table 3.1 summarises the FireNet architecture used in this study.

Table 3.1: FireNet architecture for energy forecasting.

Block	**Filters**	**Kernel Size**	**Pool Size**	**Output Shape**
1	32	5	2	$(T/2,\ 32)$
2	64	5	2	$(T/4,\ 64)$
3	128	3	2	$(T/8,\ 128)$
4	256	3	2	$(T/16,\ 256)$
5	512	3	–	$(T/16,\ 512)$

A convolutional block $l \in \{1,\ldots,5\}$ applies the transformation shown in Eq. 3.16:

$$F^{(l)} = \mathrm{ReLU}\big(W^{(l)} * I^{(l)} + b^{(l)}\big), \tag{3.16}$$

where $*$ denotes the 1D convolution operator, $I^{(l)}$ is the block input, and $W^{(l)} \in \mathbb{R}^{k\times f_{\text{in}}\times f_{\text{out}}}$

is the kernel tensor. The dimensionality of the kernel, denoted as k, is determined to be 5 in the initial two blocks, while it is configured to 3 in the subsequent blocks.

Following each convolution, max-pooling reduces the temporal resolution while retaining the dominant learned features. The pooling operation is defined in Eq. 3.17:

$$P^{(l)} = \mathrm{MaxPool}\big(F^{(l)},\, s = 2\big), \qquad T_{\text{out}} = \frac{T_{\text{in}}}{2}, \tag{3.17}$$

where s denotes the stride, and T_{out} is the reduced temporal length.

Batch normalisation is applied after pooling to stabilise and accelerate training. For a batch of activations x, batch normalization follows the transformation in Eq. 3.18:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \qquad B^{(l)} = \gamma^{(l)}\hat{x}^{(l)} + \beta^{(l)}, \tag{3.18}$$

where μ and σ^2 denote the batch mean and variance, ε ensures numerical stability, and γ and β are learnable parameters that scale and shift the normalized activations.

Through this hierarchy stack of convolution, pooling, and normalisation layers, FireNet is able to extract multi-scale temporal dependency, making it a powerful hierarchical energy forecasting architecture for short-term and intermediate-term forecasting tasks.

3.8.3 Hybrid ML-DL Forecasting Models

The hybrid models bring about the integration of DL feature generation and ML predictors so as to exploit the benefits of the two paradigms. We can take the example of LSTM or GRU layers, which can be utilised to obtain temporal representations, and then the temporal representations are fed to the XGBoost or the RF regressors, and the final prediction is made. This integrated design offers improved generalisation in complex, nonlinear, and seasonal forecasting tasks.

3.9 Hyperparameter Optimisation

The hyperparameters need to be optimised to enhance the stability and performance of ML and DL models. The architecture of a model and the properties of learning are defined by hyperparameters; the optimal choice of hyperparameters can significantly influence the performance in terms of generalisation and prediction accuracy. This research used two complementary strategies for tuning: Grid Search, searching systematically exhaustively across all parameters, and Optuna Bayesian optimisation to guide choices efficiently within high-dimensional parameter spaces.

3.9.1 Grid Search

Grid Search represents a systematic and comprehensive search methodology that assesses the efficacy of a model across the entirety of conceivable combinations of specified hyperparameter values (Bergstra et al., 2012). For a model with hyperparameter space $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, the optimal configuration is determined using the formulation shown in Eq. 3.19:

$$\theta^* = \arg\min_{\theta \in \Theta} \ \mathcal{L}_{\text{val}}(\theta), \tag{3.19}$$

where $\mathcal{L}_{\text{val}}$ denotes the validation loss under configuration θ. As Grid Search evaluates every possible parameter combination, its computational cost grows exponentially with the number of hyperparameters. For this reason, it was applied in this work primarily to baseline models with relatively small and well-defined search spaces.

3.9.2 Optuna Bayesian Optimisation

To efficiently tune models with large or continuous hyperparameter spaces, Optuna was employed (Hakyemez et al., 2024). Optuna uses a type of Bayesian optimisation with its Tree-structured Parzen Estimator (TPE) sampler that represents the distribution of both promising and non-promising hyperparameter regions. At each iteration t, the algorithm selects the next trial θ_t using the expected improvement (EI) criterion presented in Eq. 3.20:

$$\theta_t = \arg\max_{\theta} \ \mathrm{EI}(\theta), \tag{3.20}$$

The expected improvement function, formally defined in Eq. 3.21, quantifies how much a candidate hyperparameter configuration is expected to improve over the current best value:

$$\mathrm{EI}(\theta) = \mathbb{E}\big[\max(f^* - f(\theta), 0)\big], \tag{3.21}$$

where f^* represents the best objective value found so far. Optuna focuses its evaluations on the most promising regions of the hyperparameter space and avoids unnecessary exploration of low-performing configurations. Optuna is far more effective than comprehensive grid-based techniques related to this technique, particularly when handling continuous or high-dimensional hyperparameter distributions.

In the present investigation, Optuna was employed to optimise hyperparameters for an extensive array of predictive modelling techniques, encompassing Decision Trees, Random Forests, Support Vector Regression (SVR), XGBoost, FireNet, and LSTM architectures. The defined search spaces and corresponding optimal values selected by Optuna are summarised in Table 3.2.

Table 3.2: Hyperparameter tuning for different models using Optuna Bayesian optimisation

Model	**Parameter**	**Range**	**Optuna Value**
Decision Tree	max_depth	[10, 20, 30, 50, 100]	7
	min_samples_split	[2, 3, 5, 10]	10
	min_samples_leaf	[1, 5, 10]	8
	criterion	[mse, mae]	mse
	split_strategy	[best, random]	best
Random Forest	n_estimators	[50, 100, 150, 200, 300]	80
	max_depth	[10, 30, 50, 70, 100, None]	13
	min_samples_split	[2, 3, 5, 10]	7
	min_samples_leaf	[1, 2, 3, 4, 5]	5
	max_features	[auto, sqrt, log2]	sqrt
	bootstrap	[True, False]	True
XGBoost	n_estimators	[50, 300]	261
	max_depth	[3, 20]	4
	learning_rate	[0.01, 0.3]	0.051
	subsample	[0.5, 1.0]	0.949
	colsample_bytree	[0.5, 1.0]	0.813
	gamma	[0, 5]	3.704
	reg_alpha	[0, 5]	3.083
	reg_lambda	[0, 5]	3.442
SVR	C	[0.1, 100] (log)	34.95
	epsilon	[0.001, 1.0] (log)	0.244
	kernel	[rbf, linear]	rbf
	gamma	[scale, auto]	auto
FireNet	filters	[32, 64, 128]	64
	dropout_rate	[0.2, 0.5] (uniform)	0.2132
	learning_rate	[0.0005, 0.01] (log)	0.000507
	batch_size	[16, 32, 64]	64
LSTM	units	[32, 50, 64]	32
	dropout_rate	[0.1, 0.5] (uniform)	0.285
	learning_rate	[0.0005, 0.01] (log)	0.007215
	batch_size	[16, 32, 64]	16

In summary, Grid Search and Optuna achieved the efficient exploration of hyperparameter spaces of different complexity. While Grid Search gave good baseline configurations, Optuna ensured fast convergence towards optimal model configurations, especially for large-scale ML and DL forecasting architectures.

3.10 Model Training and Validation

Effective model development necessitates a serious training and validation approach in order to guarantee robustness, generality, and fairness in performance evaluation. This section explains the data splitting strategy and procedures of cross-validation with which this study is conducted.

3.10.1 Data Splitting Strategy

The dataset was split into train, validation, and test sets to avoid information leakage and provide unbiased evaluation. Let X denote the full dataset and T_{train}, T_{val}, and T_{test} represent the respective partitions. The splitting procedure is formally defined in Eq. 3.22, which ensures that the three subsets are mutually exclusive:

$$X = T_{\text{train}} \cup T_{\text{val}} \cup T_{\text{test}}, \qquad T_{\text{train}} \cap T_{\text{val}} = \varnothing, \quad T_{\text{val}} \cap T_{\text{test}} = \varnothing. \tag{3.22}$$

We used a time-based split, which is essential for time series forecasting. Data was split as training (70%), validation (15%), and test samples (15%). The training set was used to fit the model, and the validation set was used for hyperparameter tuning, preventing overfitting issues, as well as guiding model selection; Additionally, the test set served as an unseen dataset for final assessing performance.

3.10.2 Cross-Validation Method

For model selection and hyperparameter optimisation, time series cross-validation was employed. Traditional k-fold cross-validation is unsuitable for temporal datasets because random partitioning violates the chronological order of observations. To address this (Bergmeir et al., 2018), an expanding-window (rolling-origin) strategy was adopted. This ensures that training always comes before validation in time, and thus prevents information leakage.

Given an ordered dataset $\{x_1, x_2, \ldots, x_T\}$, the construction of training and validation folds follow the formulation shown in Eq. 3.23:

$$\begin{aligned}
&\text{Fold 1:} \quad \text{Train } [1 : t_1], \quad \text{Validate } [t_1 + 1 : t_2],\\
&\text{Fold 2:} \quad \text{Train } [1 : t_2], \quad \text{Validate } [t_2 + 1 : t_3],\\
&\quad\vdots\\
&\text{Fold } k: \quad \text{Train } [1 : t_{k-1}], \quad \text{Validate } [t_{k-1} + 1 : t_k],
\end{aligned} \tag{3.23}$$

where $t_1 < t_2 < \cdots < t_k \leq T$ denote predefined time indices used to split the ordered sequence. Each successive fold adds to the training horizon to allow the model to learn from more historical data with temporal integrity. This approach simulates real-world forecasting situations in which future values must be forecast on the basis of past information only.

In order to obtain a good estimate of the generalisation capability of the model, the validation loss is calculated for each fold and aggregated as shown in Eq. 3.24:

$$\mathcal{L}_{\text{cv}} = \frac{1}{k} \sum_{i=1}^{k} \mathcal{L}^{(i)}, \tag{3.24}$$

where $\mathcal{L}^{(i)}$ denotes the validation loss for the i-th fold. The k folds provide a robust measure of the model stability, help to reduce the risk of overfitting of the model, and ensure that the selected hyperparameters will generalise well to unseen temporal data.

3.10.3 Mean Squared Error (MSE)

The MSE gives the average of the squared difference of the predicted and observed values. As shown in Eq. 3.25, MSE penalises large deviations more heavily than small ones, making it particularly sensitive to significant forecasting errors (Shirzadi et al., 2021):

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2. \tag{3.25}$$

where N denotes the total number of observations, y_t represents the observed value at time step t, and $\hat{y}_t$ denotes the corresponding predicted value.

3.10.4 Root Mean Squared Error (RMSE)

The RMSE, defined in Eq. 3.26, is the square root of the MSE and therefore expressed in the same units as the original target variable (Almaghrebi et al., 2020). This enhances interpretability while retaining sensitivity to large errors:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}. \tag{3.26}$$

3.10.5 Mean Absolute Error (MAE)

The MAE provides a scale-dependent measure of average prediction error by computing the mean absolute deviation between actual and predicted values. Unlike MSE and RMSE, MAE penalises outliers less aggressively. Its formulation is presented in Eq. 3.27:

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|. \tag{3.27}$$

3.10.6 Mean Absolute Percentage Error (MAPE)

MAPE expresses prediction error as a percentage of the actual values, enabling scale-independent comparisons across different datasets. As defined in Eq. 3.28, MAPE may become unstable when the actual values approach zero (Amber et al., 2018), which must be considered when interpreting results:

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \tag{3.28}$$

3.10.7 Coefficient of Determination (R^2)

The R^2 score quantifies the proportion of variance in the observed data that is explained by the model (Chicco et al., 2021). As shown in Eq. 3.29, a value close to 1 indicates strong predictive power, while values below zero imply that the model performs worse than predicting the mean:

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2}, \tag{3.29}$$

where $\bar{y}$ denotes the mean of actual observations.

3.10.8 Computation Time Analysis

In addition to accuracy-based metrics, computation time was evaluated to assess the practical feasibility of deploying the models in real-world energy forecasting and control applications. Computation time measures the total wall-clock duration required for training or inference. The metric is given by in Eq. 3.30, where is the difference between the start and end timestamp of the execution process:

$$T_{\text{comp}} = t_{\text{end}} - t_{\text{start}}. \tag{3.30}$$

This metric brings up the trade-off between predictive accuracy and computational overhead. Models with high accuracy but too long computation time may not be suitable for real-time deployment, and models that are fast but have moderate accuracy may be better suited to applications that are limited by compute power.

Together, the aforementioned metrics defined in Eqs. 3.25–3.30 offer a comprehensive evaluation of the forecasting performance, ensuring a balanced evaluation of the error magnitude, scale dependency, explainability power, and computational efficiency across all ML, DL, and hybrid models used in this study.

3.11 Chapter Summary

This work proposes a comprehensive methodology for making robust forecasting models in conditions of incomplete and irregularly sampled energy data. The methodological pipeline is organised into integrated stages in which raw data processing starts from the first stage and ends with model evaluation.

The data obtained in terms of building energy was initially subjected to data preprocessing, which included data cleaning, normalisation, temporal alignment, and junk or inconsistent records. An exploratory data analysis was also undertaken in order to observe multi-resolution consumption patterns, focusing on hourly, daily, weekly, and monthly consumption, and the seasonality properties, as well as other sources of variations that might influence forecasting. This analysis informed the next step of treating missing data.

Missing-data behaviour was characterised according to three well-established mechanisms: MCAR, MAR, and MNAR. To mitigate the adverse effects of incomplete data, three approaches of imputation techniques were employed: statistical imputers (mean, median, and mode substitution), ML–based imputers (kNN, RF, SVR, XGBoost, and Autoencoders), and hybrid approaches that combine complementary model properties. These methods were applied to generate complete datasets for downstream forecasting tasks.

The forecasting framework comprised ML models (DT, RF, SVR, KNN, GB, XGBoost), DL architectures (RNN, LSTM, GRU), and hybrid ML–DL models that integrate temporal feature extraction with nonlinear regressors. Hyperparameter tuning was performed with the use of Grid Search in low-dimensional search spaces and Optuna Bayesian optimisation in multidimensional, continuous domains, in order to optimise the performance of these models.

The model training and validation were time-sensitive. The data was divided chronologically to create subsets of data applied for training, validation, and testing to ensure that no future information is introduced into the training process. In order to have a more rigorous assessment, an expanding-window cross-validation was employed, thus enabling the possibility of assessing across multiple temporal horizons.

Finally, model performance was quantified in terms of a set of evaluation metrics which reflect the magnitude of error (MSE, RMSE), relative performance (MAPE), explanatory power (R^2), and computational efficiency (computation time). A combination of these elements makes up a strict and replicable methodology of missing data analysis, imputation, and the construction of high-performance forecasting models of the building energy sector.

Chapter 4

Overview of Experimental Setup and Baseline Performance

Based on the multi-stage methodology described in Chapter 3, the empirical results obtained in this work provide a rigorous assessment of data imputation strategies and their cascading effects on MTLF in smart commercial buildings. Using hourly electrical load from a 30,000 m^2 multi-purpose building in Bergamo, Italy, from 1 Jan to 31 Dec 2023 (8,760 points), with meteorological covariates from NASA Prediction of Worldwide Energy Resources (POWER) database and Photovoltaic Geographical Information System (PVGIS), synthetic missing data scenarios: linear (block-type, aggregating 6%–30% in three time intervals) and random (uniformly, 5%–40%) were simulated to mimic common anomalies. Imputation paradigms included statistical baselines (e.g., Mean, Median, Mode Substitution, Last Observation Carried Forward (LOCF), MICE). ML ensembles (e.g., XGBoost, RF, SVR, kNN), and hybrid constructions (e.g., GB with MICE), followed by a DL forecasting pipeline: RNN, GRU in the case of linear irregularities (volatility, noise), and LSTM in the case of random irregularities.

Quantitative evaluation was performed using the standard measures including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) under chronological partition (70% training, 15% validation, 15% testing) as well as 5-fold time series cross-validation in order to protect against look-ahead bias. These outcomes confirm the robustness of the framework for addressing the data imperfections but also reveal complex dynamics between the typology of missingness, imputation fidelity, and predictive stability.

4.1 Comparative Analysis of Random Error Imputation Techniques

After determining the primary performance metrics (i.e., MSE, RMSE, and MSE), the results of a systematic comparative analysis between data imputation methods (statistical, ML, and hybrid methods) were performed to assess the effects of data imputation methods on electrical load forecasting performance in terms of accuracy and computational costs using the LSTM-based MTLF model. The obtained outcomes are summarised in tables and figures, each of which is dedicated to a particular imputation method.

In order to evaluate the accuracy and effectiveness of the imputation strategies, the performance evaluation metrics of the LSTM model for 24-hour-ahead load forecasting are reported in Table 4.1. Furthermore, this table also contains a benchmark case where all data is available so that a direct comparison of performance can be made. The results for the statistical imputation methods are described in Table 4.2 and Figure 4.1, which allows the visualisation and quantitative evaluation of the relative strengths and limitations of the methods.

Table 4.1: LSTM forecasting without missing data

Metric Parameter	**No Missing Data**
MSE	884.977
RMSE	29.93
MAPE (%)	41.145

The analysis of statistical imputation methods to handle missing data shows that forecasting error increases in proportion to the percentage of missing values, as shown by the increasing evaluation metrics and the average time needed to calculate them. However, the extent to which performance degradation varies from technique to technique. Simple methods, e.g., zero and mode imputation, have a lot of error amplification with increasing missingness, with zero imputation having the worst and most unstable MAPE in all cases. In contrast, whereas mode and median imputation require more processing time, they provide better accuracy under conditions of missing data. Among statistical methods, MICE represents a compromise between predictive accuracy and computational efficiency and can be considered a suitable choice for the handling of missing data in load forecasting applications.

Table 4.2: Performance of statistical imputation for LSTM forecasting under random missing data

Imputation Method	Time Avg. (s)	Metric Parameter	5%	10%	15%	20%	25%	30%	35%	40%
Zero	110.148	MSE	1811.47	1964.297	2527.194	2660.593	2990.181	3491.397	4631.182	4744.172
		RMSE	42.561	44.32	50.271	51.581	54.683	59.088	68.053	73.878
		MAPE (%)	4.71×10^8	6.89×10^8	9.74×10^8	1.09×10^9	1.58×10^9	2.01×10^9	2.74×10^9	3.12×10^9
Mean	115.843	MSE	1394.641	1433.627	1890.658	1991.166	2004.004	2168.529	2158.41	2394.124
		RMSE	37.345	37.863	43.482	44.622	44.766	46.567	46.459	48.93
		MAPE (%)	66.4	100.079	81.509	107.657	130.342	147.108	138.992	165.75
Mode	140.872	MSE	1569.351	1873.432	2364.658	2464.259	2683.373	3121.101	3148.997	3365.561
		RMSE	39.615	43.283	48.628	49.641	51.801	55.867	56.116	58.013
		MAPE (%)	88.913	129.123	125.92	188.14	188.304	197.165	173.295	246.58
Median	147.016	MSE	1285.373	1403.908	1754.768	1787.886	1897.027	2179.377	2320.485	2232.656
		RMSE	35.852	37.469	41.89	42.283	43.555	46.684	48.171	47.251
		MAPE (%)	64.944	70.161	77.546	91.736	82.422	98.113	83.282	85.52
MICE	138.633	MSE	1350.284	1466.694	2082.051	2063.719	2026.765	2174.838	2156.248	2183.75
		RMSE	36.746	38.297	45.629	45.428	45.02	46.635	46.435	46.731
		MAPE (%)	88.923	79.884	170.037	100.318	95.495	134.828	139.761	139.34

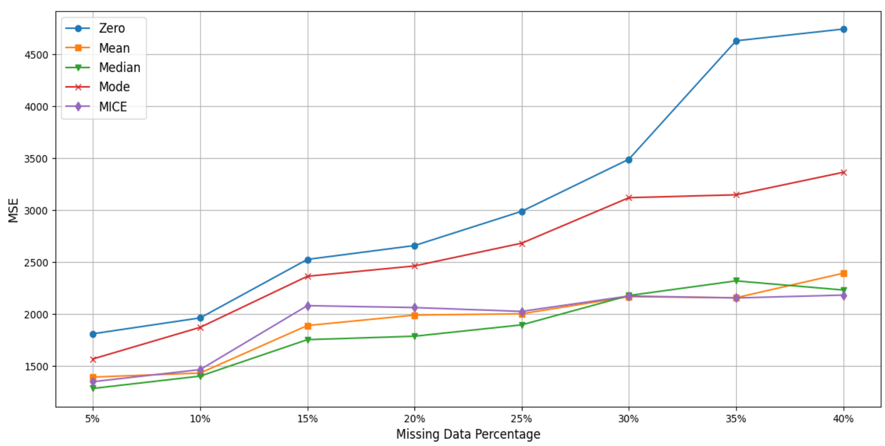


Figure 4.1: MSE of LSTM forecasting under statistical imputation across random missing data

Table 4.3 and Figure 4.2 show the results of ML imputation. All in all, these approaches provide a more accurate prediction of the electrical load at higher levels of missing data while requiring less computation time. Among them, SVR, RF, and autoencoders are found to be most consistent in their accuracy and efficiency performance across the varying proportions of missingness, while GB maintains low error values with a slightly increased execution time.

On the other hand, XGBoost is good at the low to moderate missing data rates (up to 30%) and has a sudden increase in error after that, making it less reliable for highly incomplete data. Similarly, kNN and Bayesian imputation are not good enough in case of high missingness and generate much higher error rates regardless of the metrics. The

ML imputation analysis identifies GB as a very stable algorithm, with good forecasting ability regardless of missing data volume, according to the various evaluation criteria. Its robustness at low and high levels of missingness and minimal degradation in performance make it a reliable option for use in real-world scenarios where missing data are inevitable.

Table 4.3: Performance of ML imputation for LSTM forecasting under random missing data

Imputation Method	**Time Avg. (s)**	**Metric Parameter**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**
SVR	130.161	MSE	903.844	1043.156	1024.665	1105.509	1129.354	1083.008	1049.756	1139.506
		RMSE	30.064	32.298	32.010	33.249	33.606	32.909	32.400	33.757
		MAPE (%)	75.231	65.455	109.563	67.108	113.264	71.852	69.967	110.717
DT	133.506	MSE	972.270	1187.604	1253.919	1219.310	1312.045	1379.520	1243.631	1546.253
		RMSE	31.181	34.462	35.411	34.919	36.222	37.142	35.265	39.322
		MAPE (%)	68.566	65.363	81.329	49.399	82.121	66.419	82.967	112.748
RF	134.410	MSE	931.825	1122.079	1032.019	1096.043	1053.758	1096.904	1081.073	1150.148
		RMSE	30.526	33.497	32.125	33.107	32.462	33.120	32.880	33.914
		MAPE (%)	46.221	85.120	55.141	81.409	71.610	70.817	86.742	81.058
KNN	185.245	MSE	924.117	1042.867	1058.853	1185.159	1227.086	1220.965	1059.614	1210.180
		RMSE	30.399	32.293	32.540	34.426	35.030	34.942	32.552	34.788
		MAPE (%)	60.965	64.197	58.366	92.701	100.489	101.623	44.841	48.697
XGBoost	131.341	MSE	972.851	1051.040	1099.579	1092.893	1122.374	1113.451	1152.888	1211.747
		RMSE	31.191	32.420	33.160	33.059	33.502	33.368	33.954	34.810
		MAPE (%)	84.326	38.351	90.332	68.458	89.818	59.932	74.115	91.717
Autoencoder	108.782	MSE	1043.464	1041.291	1087.503	1070.714	1048.222	1142.464	1022.913	1168.709
		RMSE	32.303	32.269	32.977	32.722	32.376	33.800	31.983	34.186
		MAPE (%)	80.363	66.958	48.543	56.856	74.703	84.887	48.942	90.364
Gradient Boost	140.399	MSE	939.607	1026.154	1023.307	1077.813	1065.240	1011.671	1037.875	1047.425
		RMSE	30.653	32.034	31.989	32.830	32.638	31.807	32.216	32.364
		MAPE (%)	79.739	61.302	56.225	74.629	78.847	60.596	81.131	66.924
Bayesian	210.217	MSE	980.872	1063.451	1196.004	1248.850	1415.374	1217.751	1352.565	1358.912
		RMSE	31.319	32.611	34.583	35.339	37.621	34.896	36.777	36.863
		MAPE (%)	61.901	72.142	86.168	72.114	98.834	64.922	57.664	64.607

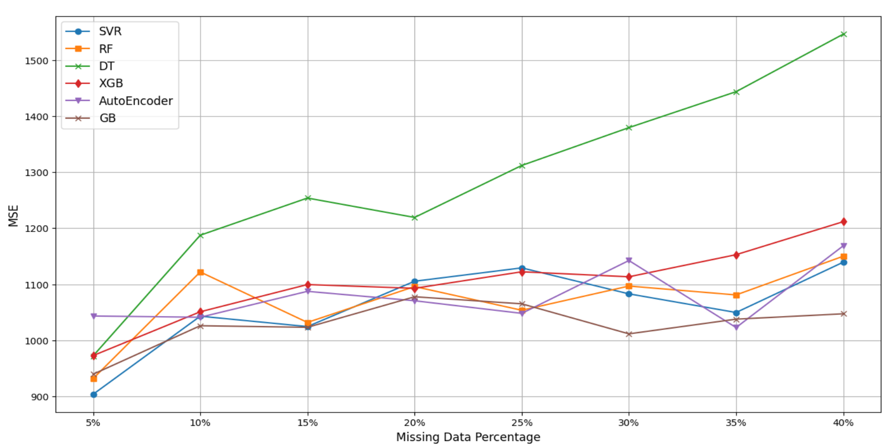


Figure 4.2: MSE of LSTM forecasting under ML imputation across random missing data

Finally, the performance of hybrid imputation is reported in Table 4.4 and Figure 4.3. The combination of GB and MICE becomes the most effective approach, providing the same accuracy under all percentages of missing data, with only moderate increments in MSE, RMSE, and MAPE. This hybrid model achieves an excellent balance between predictive accuracy and computational efficiency, with an average of 90.47 seconds on each execution.

In comparison, the combinations Random Forest with kNN and SVR with kNN show more performance variability, especially with increasing missingness, as there is a more significant increase in the error metrics and also large MAPE fluctuations. The MICE-kNN combination works well at the lower missing data values but shows a significant drop in accuracy at the higher data proportions. Moreover, the computational time in SVR-kNN is the highest among the hybrids. Overall, GB-MICE is the most reliable and efficient hybrid method while ensuring stable forecasting accuracy under various data incompleteness conditions and also maintaining the practical computational requirements.

Table 4.4: Performance of hybrid imputation for LSTM forecasting under random missing data

Imputation Method	**Time Avg. (s)**	**Metric Parameter**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**
GB & MICE	90.472	MSE	907.344	979.976	1018.830	1022.219	1032.026	1032.224	1023.011	1093.231
		RMSE	30.122	31.305	31.919	34.960	32.125	32.128	31.985	33.064
		MAPE (%)	63.191	54.777	58.814	89.471	64.361	58.845	72.072	83.600
RF & KNN	125.420	MSE	923.789	1115.258	1070.743	1116.988	1154.157	1094.315	1128.762	1187.418
		RMSE	30.394	33.395	32.722	33.421	33.973	33.080	32.074	34.459
		MAPE (%)	78.699	83.333	68.362	70.613	92.917	60.912	51.908	85.965
SVR & KNN	129.509	MSE	998.064	1098.811	1051.478	1132.401	1120.637	1191.336	1093.245	1145.111
		RMSE	31.592	33.148	32.427	33.651	33.476	35.935	33.064	33.839
		MAPE (%)	69.468	43.326	40.826	48.223	80.226	103.752	84.974	61.438
MICE & KNN	124.467	MSE	912.786	1067.978	1060.892	1146.699	1098.613	1111.362	1115.061	1223.348
		RMSE	30.212	32.680	32.571	33.863	33.145	33.337	33.393	34.976
		MAPE (%)	49.876	52.032	57.212	79.368	64.084	63.138	58.423	94.029

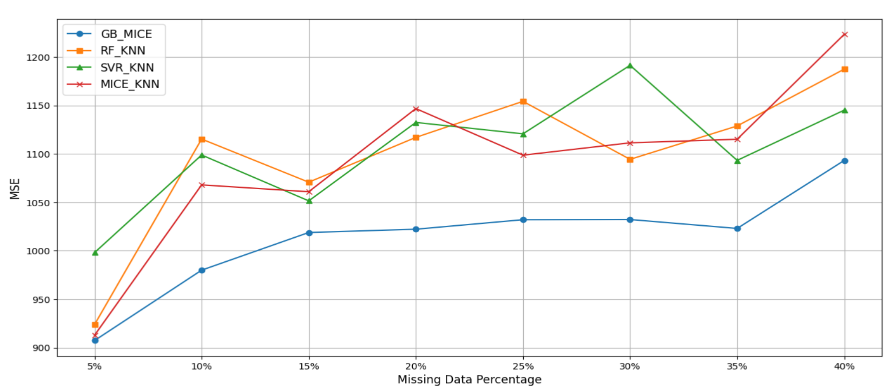


Figure 4.3: MSE of LSTM forecasting under hybrid imputation across random missing data

4.1.1 Random Imputation Conclusion

The results of this study highlight the significant role of the choice of the data imputation technique, depending on the percentage of missing data, in the success of robust MTLF. Basic statistical methods are not robust when missingness is higher than 30%, whereas ML-based methods have shown better strength and accuracy in such situations. Among the ML techniques, SVR, RF, and autoencoders are able to provide high predictive accuracy with lower computational overheads, while GB is an outstanding approach in terms of the stability it shows across all evaluated missing data levels. In the area of hybrid approaches, the combination of GB and MICE becomes the best solution with the highest accuracy, reliability, and computational efficiency in the entire range of missing data proportions (5–40%). When combined with a forecasting model using LSTM, the combination of ML and hybrid imputation techniques is the best available framework for maintaining MTLF performance in the presence of incomplete datasets.

Future work should focus on developing ML and hybrid imputation methods to enable broader applicability, including combining them with state-of-the-art forecasting architectures such as Transformers and Graph Neural Networks (GNNs). Evaluating imputation and forecasting validation on real-world and diverse datasets, such as those from energy communities, will help assess performance under complex and heterogeneous missing data patterns. Additionally, the development of an adaptive, data-driven framework for automated selection of the best imputation strategy based on dataset characteristics would make the framework much more robust and generalizable. Progress in such directions will help facilitate more reliable imputation pipelines, ultimately enhancing predictive modelling and decision making process across energy and its related industries.

4.2 Comparative Analysis of Linear Error Imputation Techniques

This section is a summary of the results of the comparative analysis of linear error imputation. Specifically, the forecasting performance of the RNN and GRU models from the original, complete dataset (i.e., without missing values and imputation) is shown in Table 4.5. The results clearly indicate that the GRU model is better than the RNN in terms of prediction accuracy by showing lower error metrics. In comparison, the consideration of the RNN model shows significantly faster computation, with only 32.8 seconds compared to the 68.1 seconds of the GRU, which is about half of the computation time. Performing as reference benchmarks, the values of execution in Table 4.5 allow the reference assessment of model degradation under the condition of data incompleteness. The detailed impact of the various imputation techniques and the increasing percentage of missing data values on the forecasting performance of RNN and GRU models is presented in Tables 4.6 and 4.7, respectively.

Table 4.5: RNN and GRU forecasting performance without missing data

Model	**MSE**	**RMSE**	**MAPE (%)**	$\boldsymbol{R^2}$	**Time (s)**
RNN	941	30.67	46.20	0.837	32.80
GRU	869	29.47	43.45	0.849	68.13

4.3 Linear Missing Data Patterns

Linear missing data (block-type: 6%–30%) was simulated in three annual segments to replicate outages. Four ML imputation methods (RF, SVR, KNN, XGBoost) were applied, followed by RNN and GRU forecasting (70/30 train/test split).

4.3.1 Imputation and Forecasting with RNN

Table 4.6 compares the performance of different types of imputation techniques combined with the RNN model. Among them, XGBoost shows the best accuracy under the minimum missing data (MSE = 895 at 6%), which shows good performance in the early stage. SVR is consistently performing with high R^2 values under any condition and is indicative of good predictive reliability. kNN provides acceptable forecasting accuracy in the case of missing data ranging between 6–18%, but it declines its performance in the case of missing data percentage beyond this range (e.g., MSE = 1259 at 30% missingness). In contrast, the RF imputation method offers balanced and stable results in a wide range of missing data proportions, moderate errors, and strong consistency. In summary, the RNN-XGBoost combination is best for scenarios with little missing data, while the RNN-RF combination seems to be more reliable and versatile for various and higher levels of data incompleteness.

In Fig. 4.4, the MSE trend as a function of the missing data percentage is given for all the employed imputation techniques, assuming the use of the RNN model for forecasting the active power demand.

Table 4.6: Performance of ML imputation for RNN forecasting under linear missing data

Imputation Technique	Computational Time (s)	Evaluation Metrics	6%	12%	18%	24%	30%
KNN	33.68	MSE	1009	1026	1021	1119	1259
		RMSE	31.76	32.03	31.95	33.45	35.49
		MAPE (%)	50.62	65.25	54.15	82.84	54.52
		R^2	0.825	0.820	0.820	0.802	0.775
RF	37.25	MSE	978	1006	1106	1143	1225
		RMSE	31.27	31.72	33.26	33.81	35.00
		MAPE (%)	77.70	48.18	69.60	55.51	86.47
		R^2	0.830	0.825	0.807	0.799	0.783
SVR	38.66	MSE	918	960.2	1039	1115	1302
		RMSE	30.30	30.99	32.23	33.39	36.09
		MAPE (%)	69.80	44.40	56.76	74.28	84.37
		R^2	0.841	0.833	0.819	0.804	0.768
XGBoost	38.41	MSE	895	1034	1122	1133	1237
		RMSE	29.92	32.16	33.50	33.66	35.18
		MAPE (%)	46.89	63.68	51.27	64.16	64.37
		R^2	0.845	0.820	0.804	0.801	0.780

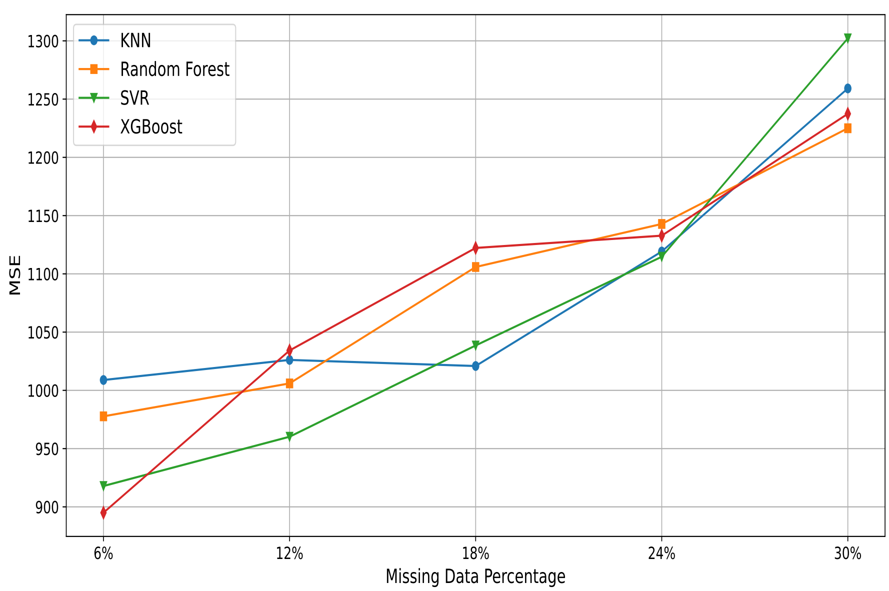


Figure 4.4: MSE of RNN forecasting under ML imputation across linear missing data

4.3.2 Imputation and Forecasting with GRU

In Table 4.7 shows the forecasting performance of the GRU model with respect to imputation techniques, where there is a little improvement over the RNN results, especially in MSE and R^2. XGBoost measures the lowest MSE for the lowest missing data level (MSE = 877 at 6% missingness), although this advantage vanishes as the missing data amount rises. SVR maintains high R^2 values but has varying MAPE for the set of missing data percentages. kNN has a stable performance at low missing levels, and progressive performance degradation when the rate of missing data increases. As can be seen in the case of RNN, the RF again shows better reliability with an excellent trade-off between the error measures and computational efficiency.

Table 4.7: Performance of ML imputation for GRU forecasting under linear missing data

Imputation Technique	Computational Time (s)	Evaluation Metrics	6%	12%	18%	24%	30%
KNN	81.84	MSE	874	969.0	1070	1097	1192
		RMSE	29.56	31.13	32.71	33.12	34.52
		MAPE (%)	50.89	76.01	64.69	45.75	66.21
		R^2	0.822	0.830	0.812	0.806	0.787
RF	83.67	MSE	895.2	969.8	1067	1096	1263
		RMSE	29.92	31.14	32.66	33.11	35.53
		MAPE (%)	45.26	59.32	54.71	51.00	80.35
		R^2	0.845	0.831	0.814	0.807	0.776
SVR	81.80	MSE	906.3	914.6	992.0	1179	1190
		RMSE	30.11	30.24	31.50	34.34	34.49
		MAPE (%)	85.84	50.19	57.27	44.25	60.09
		R^2	0.843	0.841	0.827	0.793	0.788
XGBoost	91.12	MSE	876.6	1041	1157	1143	1242
		RMSE	29.61	32.26	34.01	33.81	35.24
		MAPE (%)	59.10	51.64	89.20	64.66	59.70
		R^2	0.848	0.820	0.798	0.799	0.780

Similar to Figure 4.4, Figure 4.5 shows results on MSE versus percentage of missing data, where active power prediction is done by using a GRU model with a combination of imputation techniques discussed as KNN, RF, SVR, and XGBoost.

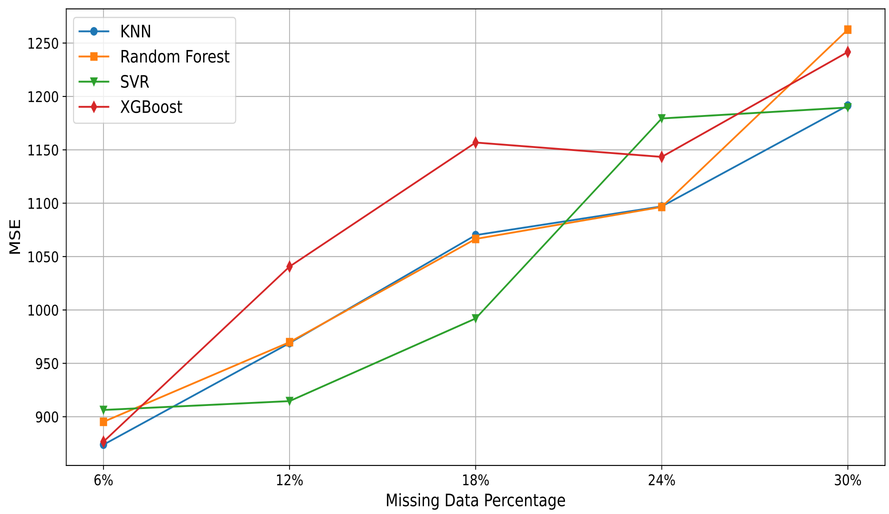


Figure 4.5: MSE of GRU forecasting under ML imputation across linear missing data

4.3.3 Linear Imputation Conclusion

The imputation approach has a considerable effect on the forecast accuracy, with a pronounced effect as the proportion of the missing data increases. XGBoost and SVR perform well under the conditions of low to moderate missingness, with superior MSE and R^2 values, but the MAPE values of SVR have fluctuations at various conditions. kNN has computational efficiency but does not have reliability with large missingness levels, especially when combined with RNN. In contrast, RF has consistently the most balanced and stable performance across a wide range of missing data percentage values (especially good compatibility when combined with the GRU model). The analysis further confirms that GRU outperforms RNN in accuracy across all the techniques in imputation, but with increased computational demands. Overall, the amount of missing data determines the optimal combination of forecasting and imputation models for obtaining acceptable prediction precision.

Further research in this area is recommended to aim at producing increasingly more evolved and accurate forecasting tools based on diverse input data, such as those collected from energy communities' profiles, weather data, socio-economic factors, or grid-level consumption patterns. This kind of extensive integration will improve the model's generalisability and robustness to different real-world conditions. Also, by experimenting with hybrid imputation models that integrate both statistical models and more sophisticated ML models, specifically, ensemble-based iterative imputation in non-stationary time series and adaptive variants of random forests in multi-modal gaps, it is possible to simultaneously handle several categories and classes of missingness. Such hybrid models have demonstrated effectiveness at enhancing the quality of restored data, especially in time series with different models.

Lastly, to have the case of uncertainty quantification and explainability enter into the realms of imputation and forecasting pipelines, both will be necessary to enable clear, credible, and actionable decision-making to be easy.

4.4 Chapter Summary

This chapter presented a comprehensive experimental evaluation of data imputation strategies and their impact on MTLF for smart commercial buildings. Using a real-world hourly electrical load dataset augmented with meteorological covariates, both random and linear missing data scenarios were systematically simulated to reflect practical data corruption patterns. A wide spectrum of imputation approaches, ranging from statistical baselines to ML and hybrid methods, was assessed in conjunction with DL forecasting models, including LSTM, GRU, and RNN architectures. Performance was rigorously evaluated using standard error metrics under chronological splitting and time-series cross-validation, ensuring robustness against look-ahead bias.

The results demonstrate that simple statistical imputation methods deteriorate rapidly as missingness increases, while ML-based and hybrid approaches exhibit superior resilience and stability. In particular, GB combined with MICE consistently delivered the most reliable performance across all levels of random missingness, whereas RF and XGBoost were more effective for structured linear gaps. Furthermore, GRU and LSTM models outperformed RNNs in forecasting accuracy, with GRU offering a favourable balance between predictive performance and computational efficiency. Overall, the findings confirm that the choice of imputation technique aligned with the missing data pattern is critical for maintaining forecasting reliability, and that hybrid imputation integrated with deep learning models provides a robust framework for real-world energy forecasting applications.

Chapter 5

Mid-Term Load Forecasting Techniques

5.1 Chapter Overview

The main goal in this chapter is to present and analyse the empirical results achieved from the experiments conducted on electricity consumption forecasting models on the basis of the methodology presented in Chapter 3 in detail. By synthesising the results of selected key studies, in this chapter, the performance of different forecasting techniques in handling the issue of data quality and improving the accuracy of MTLF for smart commercial buildings is evaluated. The emphasis is on quantifying model efficacy using standard metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2), and emphasising the role of these results in real applications for energy management, demand response, and sustainability. This is in line with the recent literature calling for strong MTLF in order to optimise resource allocation in urban energy systems, where accurate predictions can lead to a reduction of the operation costs of 10–20% and support the integration of renewables (Misiurek et al., 2025b; Rodrigues et al., 2023b).

5.2 Data Description and Interpretation

The data sets to be used in this research were taken from smart commercial buildings with intelligent metering infrastructure. The main data set includes hourly electrical load observations made during 2023 and thus consists of 8760 observations and 11 variables. These signals were provided by intelligent circuit breakers, communication systems, and cloud-based energy logs. Datasets include active, reactive, and apparent power metrics at an hourly resolution in terms of minimum, maximum, and averages at an hourly granularity.

The combination of these parameters can be used to gather complete information on the varying demand patterns in smart buildings. Peak values of active and apparent power are key parameters of the level of activity, which are especially important during the working day in the conditions of commercial activity.

Sample records describing the power profile are presented in Table 5.1 below. The data display high hour-by-hour variations in electrical loads, which highlight the impact of vibrant occupancy behaviours and equipment timetable on the total energy need of buildings. Notably, these variations point to the existence of intra-hour peaks, which can be attributed to intermittent HVAC operations and simultaneous activation of high-power appliances, emphasising the requirement for high-resolution (for capturing such volatility) forecasting models in order to effectively manage energy.

Table 5.1: Sample data of linear power consumption parameters for the year 2023

Date	**Time**	$\boldsymbol{P}_{\text{Avg}}$ (kW)	$\boldsymbol{P}_{\text{Min}}$ (kW)	$\boldsymbol{P}_{\text{Max}}$ (kW)	$\boldsymbol{Q}_{\text{Avg}}$ (kvar)	$\boldsymbol{Q}_{\text{Min}}$ (kvar)	$\boldsymbol{Q}_{\text{Max}}$ (kvar)	$\boldsymbol{S}_{\text{Avg}}$ (kVA)	$\boldsymbol{S}_{\text{Min}}$ (kVA)	$\boldsymbol{S}_{\text{Max}}$ (kVA)
10/01/2023	00:00:00	152	123	220	11	6	20	155	125	223
10/01/2023	01:00:00	222	202	242	21	18	23	225	206	246
10/01/2023	02:00:00	230	210	246	22	19	23	233	213	249
10/01/2023	03:00:00	232	196	258	22	17	25	235	199	261
10/01/2023	04:00:00	217	190	255	20	16	25	220	192	258

Environmental variables were included to account for load variation that depends on the weather. The National Aeronautics and Space Administration (NASA) POWER database includes wind speed (m/s), temperature (°C), and humidity (g/kg). Since HVAC loads are very sensitive to ambient conditions, these characteristics play an important role in mid-term prediction, as shown in the sample records in Table 5.2.

Table 5.2: Sample data of meteorological parameters from NASA POWER

Date	**Time**	**Temperature** (°C)	**Humidity** (g/kg)	**Wind Speed** (m/s)
10/01/2023	00:00:00	3.51	3.36	3.46
10/01/2023	01:00:00	3.16	3.30	3.90
10/01/2023	02:00:00	2.73	3.17	4.26
10/01/2023	03:00:00	2.55	3.05	4.41
10/01/2023	04:00:00	2.43	2.99	4.46

A second dataset of the year 2023 includes additional meteorological variables from the Photovoltaic Geographical Information System (PVGIS) as illustrated in Table 5.3, specifically beam and diffuse irradiance, sunshine duration, and wind speed at an elevation of

10 meters (at higher elevation). These climate signals bring richer representation of the renewable influence and solar-related heat gains, of particular relevance in MTLF models.

Table 5.3: Sample data of meteorological parameters from PVGIS

Date	Time	Beam Irradiance (Gb(i)) ($\mathbf{W/m^2}$)	Diffuse Irradiance (Gd(i)) ($\mathbf{W/m^2}$)	Sunshine Duration (h)	Temperature (°C)	Wind Speed @10m (m/s)
1/01/2023	08:00:00	56.005	43.800	8.39	−1.440	1.585
1/01/2023	09:00:00	103.139	75.933	14.83	1.100	0.997
1/01/2023	10:00:00	142.206	96.400	19.25	3.866	0.483
1/01/2023	11:00:00	186.007	93.733	21.20	4.933	0.593
1/01/2023	12:00:00	165.273	97.200	20.49	5.143	0.745

These other weather variables are especially important because, as irradiance increases, thermal gain through the building's interior is increased, often increasing cooling loads during summer seasons.

5.3 Data Preprocessing and Feature Engineering

Data preprocessing played an important role in stabilising the model's behaviour. After removing outliers and interpolating missing values, the 2023 dataset provided a clean input space, allowing clearer comparison between ML and DL models. Feature engineering attributes such as temperature, season, and calendar indicators contributed to the improved performance of models that rely on contextual patterns, which is reflected in the forecasting accuracy reported in Chapter 3, Section 3.3.

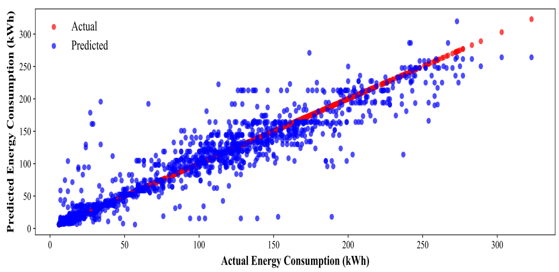
5.4 Evaluation of Forecasting Models

To achieve a solid and comparative analysis of the forecasting models, different ML, DL, and hybrid models were tested with standard error-based and accuracy-based metrics. The ability of particular performance metrics to quantify both absolute and relative deviations from actual loads is determined by their selection. The first study was conducted using ML models such as DT, RF,SVR, kNN, and XGBoost, and the second study used DL models such as LSTM, FireNet, and hybrid FireNet-XGBoost models. This diverse set of models, from two studies, offers a comprehensive and complementary evaluation of forecasting capabilities in smart building environments.

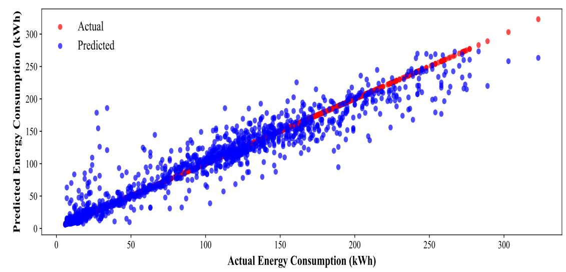
The evaluation metrics MAPE, MSE, and RMSE were used as the main performance criteria in order to measure the deviation between forecasting demand and actual demand for ML models. MAPE provides an implicit explanation for the relative error in load magnitude, making simple comparisons possible. On the other hand, MSE and RMSE are more sensitive to large errors, so they can quickly identify models that are prone to making severe mistakes under extreme load times. Collectively, these measures offer a complete measure of error along with the risk of deviation.

5.4.1 ML Forecasting Models

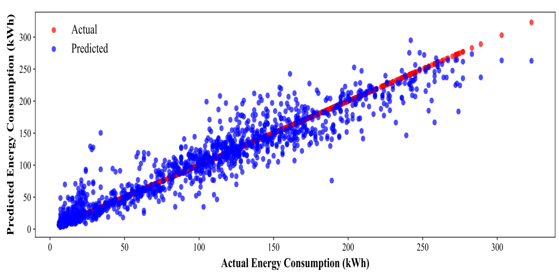
The ML models selected for evaluation include DT, RF, XGBoost, and SVR. Figure 5.1 compares actual power consumption with predicted values generated by these models.



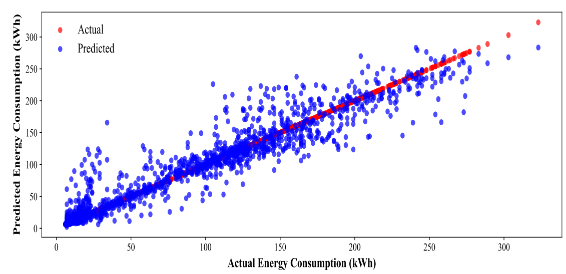
(**a**) DT: actual vs. predicted



(**b**) RF: actual vs. predicted



(**c**) XGBoost: actual vs. predicted



(**d**) SVR: actual vs. predicted

Figure 5.1: Forecasting performance of ML models (**a**) DT (**b**) RF (**c**) XGBoost (**d**) SVR

As summarised in Table 5.4, XGBoost demonstrated the highest predictive accuracy and computational efficiency, achieving the lowest MSE (i.e., 518.5), RMSE (i.e., 22.77), and a higher R^2 score (i.e., 0.9010), with a training time of only 0.051 s. RF also performed competitively, offering a strong balance between accuracy and runtime, with slightly higher values of MSE, RMSE, and a lower R^2 score (i.e., 0.8992), and training time of 1.101 s. The DT model exhibited the fastest training time (i.e., 0.0162 s), though it exhibited slightly lower accuracy, as reflected by the R^2 score (i.e., 0.8514). In contrast, SVR showed the weakest performance, recording the highest MSE, RMSE, and the lowest R^2 score (i.e., 0.7460), with a significantly higher runtime of 41.13 s. These findings confirm that XGBoost and RF offer the most accurate and computationally efficient solutions among the evaluated ML models.

Table 5.4: Evaluation of ML forecasting models using standard performance metrics

Model	**Training Time (s)**	**MSE**	**RMSE**	**R^2 Score**	**MAPE (%)**
DT	0.01620	778.2	27.90	0.8514	35.68
RF	1.101	527.7	22.97	0.8992	33.63
XGBoost	0.05100	518.5	22.77	0.9010	37.17
SVR	41.13	2111	45.95	0.7460	44.79

Figure 5.2 shows the comparison of actual and predicted energy consumption for the last 2 months in the year 2023. The two methods, XGBoost and RF, closely match the real data, being able to fit peaks and trends as well as fluctuations. The performance of the DT is fit, with the largest deviations from actual values. On the other hand, SVR always underestimates energy consumption and does not account for sharp variations. These visual results are consistent with the numerical measures reported above, and further demonstrate that XGBoost and RF have better forecasting accuracy.

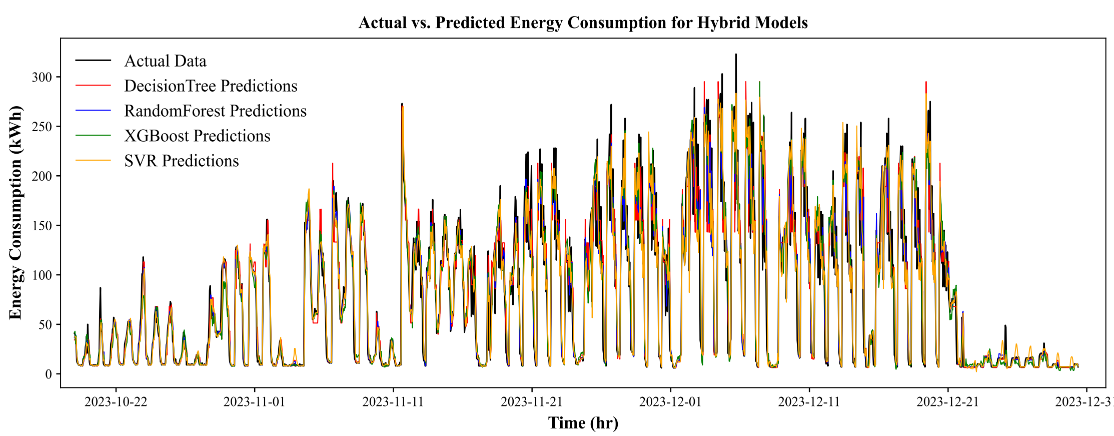
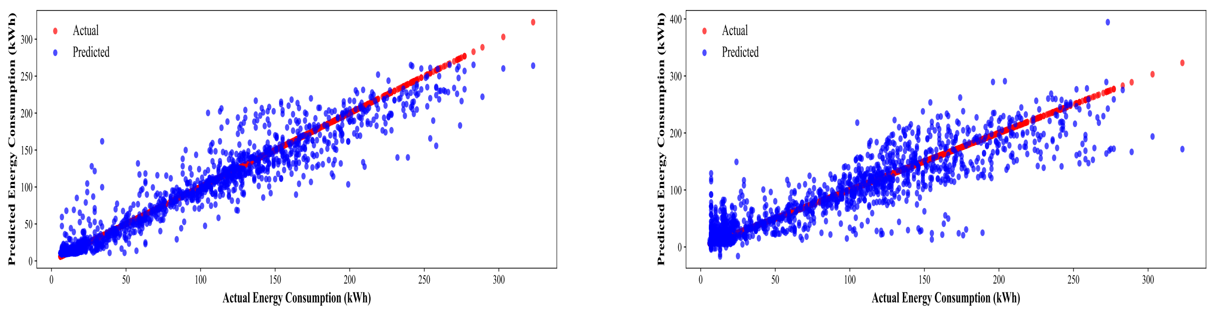


Figure 5.2: Actual vs. predicted energy consumption for ML models

5.4.2 Deep Learning Forecasting Models

Besides classic ML models, the expanded research tested sophisticated DL models, such as LSTM and FireNet, and other advanced models. The research focuses on two sequence models, LSTM and FireNet, and their hybridisation with tree ensembles to leverage complementary strengths in temporal pattern learning and non-linear feature interaction. Consistent with expectations for hourly, MTLF, the standalone DL models improved the representation of diurnal/weekly cycles relative to simpler baselines; however, the best accuracy emerged from hybrid stacking, where a DL extractor provides temporal features and an ML learner (e.g., XGBoost) maps those features to load values with strong generalisation. Figure 5.3 presents the forecasting performance of both considered DL models.



(**a**) FireNet: actual vs. predicted (**b**) LSTM: actual vs. predicted

Figure 5.3: Forecasting performance using DL models (**a**) FireNet (**b**) LSTM

The comparative results are listed in Table 5.5. FireNet showed better predictive performance than that of LSTM, in particular for the learning of non-linear and time-dependent energy usage patterns. FireNet, in particular, had a smaller difference in MSE (i.e., 923.5 against 1280), as well as less RMSE (i.e., 30.39 against 35.78) and smaller MAPE (i.e., 39.90% against 96.21%). Furthermore, FireNet obtained a higher R^2 (i.e., 0.8237 against 0.7558). Considering the computational time, the FireNet is faster, requiring only half of the time needed by the LSTM. Overall, FireNet performed significantly better than LSTM across all metrics and provided improvements in accuracy and computational efficiency.

Table 5.5: Evaluation of DL forecasting models using standard performance metrics

Model	**Training Time (s)**	**MSE**	**RMSE**	R^2 **Score**	**MAPE (%)**
FireNet	22.61	923.5	30.39	0.8237	39.90
LSTM	46.97	1280	35.78	0.7558	96.21

In Figure 5.4, the actual energy consumption is compared against the forecasts delivered by the selected DL models. As expected, FireNet's forecasts closely align with actual values, especially during peak periods, while LSTM predictions are noticeably overestimated in the case of peak demand periods.

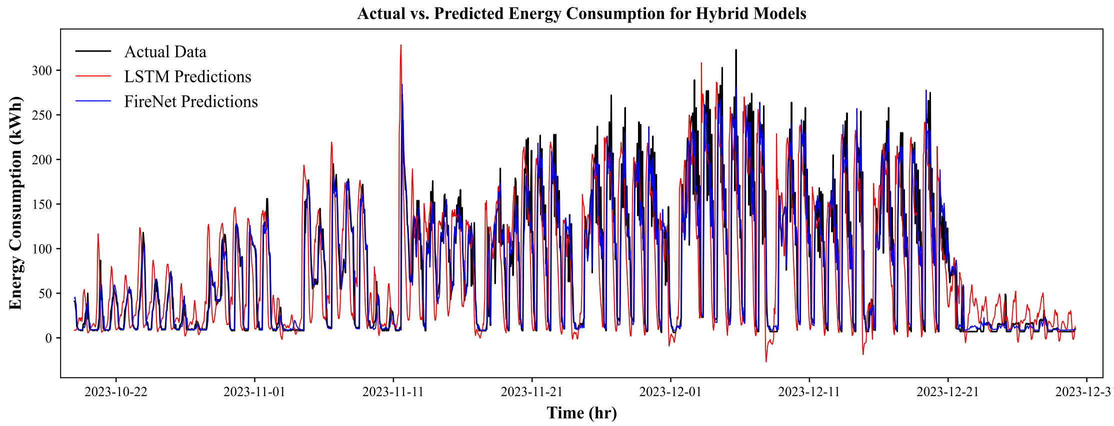
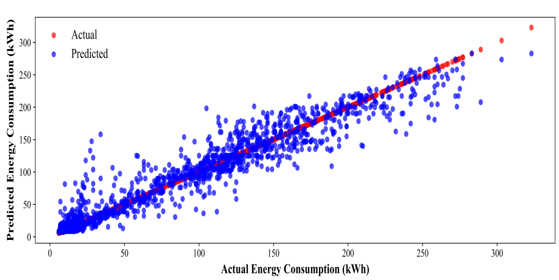


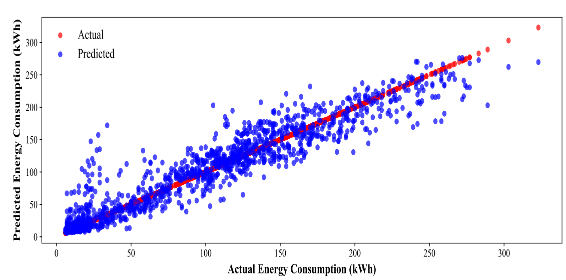
Figure 5.4: Actual vs. predicted energy consumption for DL models

5.4.3 Hybrid Model Performance

Based on the previous results, XGBoost and RF are identified as the better-performing ML models, while FireNet ensures superior forecasting performance in terms of DL models. Based on these results, two hybrid models have been developed by integrating the predictive ability of ensemble-based ML models with the FireNet sequential learning capabilities (i.e., XGBoost and FireNet).



(**a**) FireNet–XGBoost: actual vs. predicted



(**b**) FireNet–RF: actual vs. predicted

Figure 5.5: Forecasting performance of hybrid (ML-DL) models (**a**) FireNet–XGBoost (**b**) FireNet–RF

In Figure 5.5, the energy consumption forecasts provided by both hybrid models (i.e., FireNet–XGBoost and FireNet–RF) are shown, while the corresponding key performance parameters are reported in Table 5.6. The FireNet–XGBoost hybrid model demonstrated superior performance due to lower MSE (i.e., 350.2 against 418.4), reduced RMSE (i.e., 18.71 against 20.45), smaller MAPE (i.e., 27.00% against 29.26%), along with higher R^2 (i.e., 0.9334 against 0.9205). On the other hand, the FireNet–RF hybrid model required less computation time (i.e., 21.11 s against 33.69 s). Overall, the FireNet–XGBoost model outperformed the FireNet–RF model across all evaluated metrics, indicating enhanced precision, except for the computational effectiveness.

Table 5.6: Evaluation of hybrid forecasting models using standard performance metrics

Model	Training Time (s)	MSE	RMSE	R^2	MAPE (%)
FireNet-XGBoost	33.69	350.2	18.71	0.9334	27.00
FireNet-RF	21.11	418.4	20.45	0.9205	29.26

Figure 5.6 illustrates that the FireNet–XGBoost model closely follows actual energy consumption patterns, particularly during peak demand periods, whereas the FireNet–RF model tends to slightly overestimate energy usage across similar intervals.

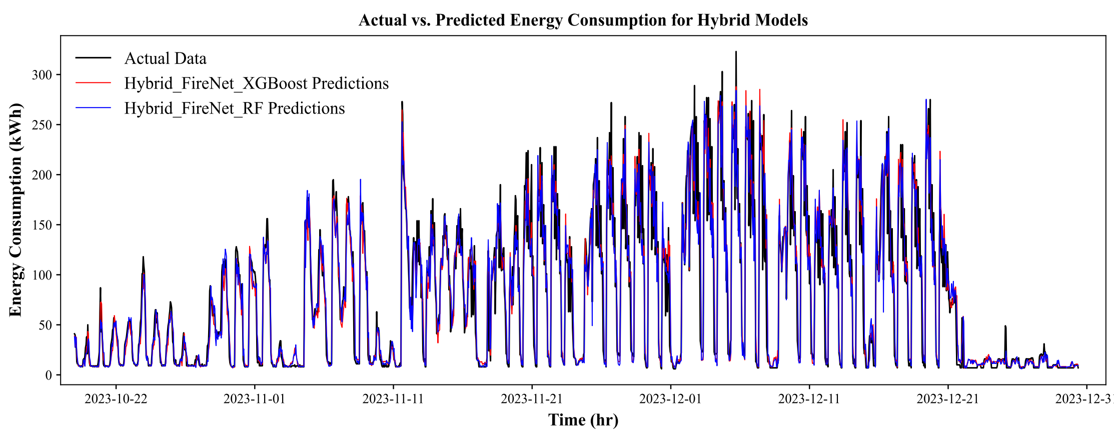


Figure 5.6: Actual vs. predicted energy consumption for hybrid models.

5.5 Chapter Summary

This chapter presented a comprehensive evaluation of MTLF techniques applied to smart commercial buildings, using high-resolution electrical and meteorological datasets. A wide range of ML, DL, and hybrid models was systematically assessed to quantify their predictive accuracy, robustness, and computational efficiency. The analysis highlighted the importance of enriched feature sets, including meteorological and apparent power variables, in improving forecasting performance. Among standalone ML models, ensemble-based methods, particularly XGBoost and RF, demonstrated superior accuracy and stability, achieving low MSE and high R^2 values while maintaining reasonable computational costs. In contrast, simpler models such as DT and SVR exhibited either reduced accuracy or increased runtime, limiting their applicability for operational forecasting.

DL models further improved temporal pattern representation, with FireNet outperforming LSTM in both accuracy and computational efficiency, particularly in capturing nonlinear and time-dependent load dynamics. The strongest overall performance, however, was achieved through hybrid architectures that combined the temporal feature extraction capability of DL models with the generalisation strength of ensemble ML learners. In particular, the FireNet–XGBoost hybrid consistently delivered the lowest error metrics and highest explanatory power, effectively capturing peak demand and load variability. These results confirm that hybrid ML–DL frameworks provide a powerful and scalable solution for MTLF in smart buildings, offering tangible benefits for energy management, demand response planning, and the integration of renewable energy resources.

Chapter 6

Implementation of the Energy Forecasting in a Multi-Platform Environment

The entire forecasting process will help users move raw data to meaningful insights in a seamless process using an interactive and user-friendly interface. The process starts when users upload their past electrical power consumption details into the system. When uploading a file, the default system does the verification of the format, checks the authenticity of the contents, and presents a preview to make sure that the information they provide is correct before going ahead.

The MTLF application is developed with the assistance of the Streamlit web application engine, using which Python scripts may be transformed into interactive and web-based applications. Streamlit was selected as it has a user-friendly interface, and a person with no technical skills can interact with complex forecasting models without having to write any code or comprehend the underlying dynamics.

Users can easily open the app with a single command and upload data, preprocess it with provided preprocessing options, train models, and view the results in real-time. It is an approach that creates advanced forecasting and analytics access to anyone, including energy planners, without the need to use data science skills. Streamlit has a responsive design that makes all actions automatically update the screen, be it previewing the data, visualising predictions, or comparing performance. The software incorporates familiarity with strong Python packages, allowing it to mix tools, such as Prophet, LSTM, and Random Forests, in a single, user-friendly setting.

6.1 Overview of the Electrical Power Forecasting Platform

This visualization tool outlines a summary of key metrics, time-oriented graphical illustrations, and complete reports about data accuracy (including the counting of absent data points along with their distribution trends). When the information is loaded successfully, one can access it and navigate through it using an analytics dashboard. These enable users to easily interpret the design and pattern of their dataset in spotting trends, in their seasonal changes, or in their inconsistency, which might require consideration.

Once users have reviewed the dataset, they move to the preprocessing phase. In this case, they are able to set the missing value idiom, identify and eliminate anomalies with provided statistical functions, and provide optional feature engineering settings to automatically create useful time-based variables, e.g., hour-of-day or day-of-week predictor. Then, the correlation analysis module gives a detailed insight into the relationship between various features and each other, as well as power consumption, in general. This step can be used to get better users decide on the most influential factors to include in their predictive models.

The user proceeds to the model training phase after the data has been cleaned and prepared. The system offers multiple modelling techniques to support various user requirements and ability levels. To apply various types of traditional statistical models, such as Prophet and SARIMAX, ML-based models, such as DT and RF, or DL based networks, such as LSTM and GRU networks that can learn complex time-varying patterns, can be used by the user. To be more automated, it is possible to choose H2O AutoML, which will test, tune, and select the best-performing model with the minimum number of manual parameters.

These models are then trained on 70% of the data, testing and evaluation on 30% data to assess their performance. The RMSE, MAPE, and R^2 are calculated by the platform to evaluate, for predicting future consumption using different models.

Once the modelling stage is completed, the system extends its functionality beyond basic forecasting. It has an energy emissions calculator that differentiates the estimated energy consumption from the measurable carbon emissions. The findings are categorised in terms of seasons and presented in day-by-day trend charts, and this provides the user with a clear understanding of how their performance varied across different seasons.

In addition to having exact power consumption forecasts at the end of the workflow, the users are also able to obtain valuable data regarding their carbon footprint. These factors are combined using the platform, data exploration, preprocessing, modelling, and environmental analysis into one user-friendly system that allows organisations to make smarter, data-driven, and sustainable energy decisions using available data.

6.2 System Architecture

6.2.1 Development Environment and Libraries

The system is built upon a clearly defined programming language version, virtual environment, and a suite of software libraries that together standardise the workflow for the entire forecasting system. In defining a consistent environment, the system provides that data preparation, model training, assessment, and deployment are repeatable and uniform across multiple machines and users. To prevent conflicts between packages, dependencies are isolated within a virtual environment, ensuring a stable environment. The selected libraries offer specialised functionality to manipulate data, visualise statistical models, ML, DL, and interactive application development with Streamlit-Core Frameworks.

6.2.1.1 Core Frameworks

- **Streamlit (v1.34.0)**: Streamlit powers the entire forecasting workflow through a web interface, allowing users to upload data, adjust preprocessing settings, start training, and view results directly in the browser, no coding required. The interactive tools, visual representations, and performance metrics are updated immediately, consequently delivering a thorough understanding of each stage in the operation.

- **Pandas (v2.2.2)** and **NumPy (v1.26.4)**: Pandas handles the management of tabular time series data, including timestamp parsing, gap filling, table reshaping, and merging of external features such as weather data. NumPy provides the optimised numerical layer beneath, accelerating operations like scaling, array slicing, and feature engineering.

- **Openpyxl (v3.1.2)**: With the Openpyxl library, it is easy to transfer predictive analytics, daily data compilations, and assessment metrics into Excel files. Reports can include multiple sheets such as raw forecasts, daily aggregates, and metric summaries with consistent formatting throughout. These results are easily shared with the business stakeholders who can easily work with spreadsheets to review or analyse further.

6.2.1.2 Visualization Tools

- **Matplotlib (v3.9.0)** and **Seaborn (v0.13.2)**: Matplotlib plotting is based on the underlying plotting functionality, and Seaborn builds these underlying tools with high-level visual effects, colour schemes, and faster statistical charting applications. Together, they lead to the publication and quality visualisation of time series, boxplots, to investigate the outliers, heatmaps, to carry out correlation analysis, and distribution charts.

- **Plotly (v5.24.0)**: Plotly is the one that drives the interactive charts within the Streamlit interface. The app provides the ability to zoom in, hover to see detailed tooltips, filter data ranges, and export images of high quality, all within the app, to allow operations teams to analyse scenarios and insights immediately. Forecast overlays, residual diagnostics, and seven-day projections are also interactive and exploratory.

6.2.1.3 Modelling Libraries

This system is constructed based on a collection of fundamental toolkits and AutoML frameworks that allow statistical, ML, and DL predictions. Collectively, they ensure streamlining of the model development, optimisation, and deployment across the varied forecasting approaches.

- **Scikit-learn (v1.5.0)** and **XGBoost (v2.0.3)**: Scikit-learn provides one of the most integrated tools that facilitate ML workflows, which includes splitting data, feature scaling, and evaluating the outcomes in terms of RMSE, MAE, and R^2. It also offers a tree-based base model and easy-to-use utilities to construct and maintain pipelines. XGBoost builds on this premise and introduces gradient-boosted decision trees, which help to model complex, non-linear interactions and provide strong performance. These libraries together assist in the entire ML process, including preprocessing, model training, and performance assessment.

- **Statsmodels (v0.14.2)** and **Prophet (v1.1.5)**: Statsmodels has the strength to provide the statistical modelling functions, such as SARIMAX, ARIMA, and exogenous analysis of autoregressive, seasonal, and exogenous relationships. PROPHET is also a tool that complements this method since it breaks down time series into trend, seasonality, and event-based effects, and it needs little manual tuning. Together, they compose the statistical forecasting layer, which includes results that are easy to understand and interpret, quick experimentation, and solid statistical models to analyse trends and seasonality.

- **TensorFlow (v2.15.0)**: TensorFlow is the implementation of the computational engine of DL, and it provides GPU acceleration, flexibility of the computational graph, as well as customisation of training callbacks. The implementation of sophisticated sequence models, LSTM and GRU, can be easily implemented through their Keras API, which has configurable layers, dropout, and monitoring features.

6.2.1.4 AutoML Frameworks

- **H2O (v3.42.0.2)**, **FLAML (v2.1.1)**, **TPOT (v0.11.1)**, and **Auto-sklearn (v0.15.0)**: The search for optimal models (hyperparameters and combinations of features) is automated by these AutoML frameworks and does not require the intensive manual tuning of the search. They search many pipelines over specified time constraints, compare the performance on validation measures, and tend to improve results through the collection of top-performing models into an ensemble to achieve much better performance.

6.2.2 DATASET

The dataset .csv file contains 8,759 hourly records spanning a complete year from April 1, 2022, to March 31, 2023, providing comprehensive coverage of seasonal consumption patterns across all four seasons. The dataset of linear parameters is summarised in Table 6.1.

Table 6.1: Feature description of temporal and electrical power measurements

Category	Feature Name	Description
Temporal Features	Date	Date of the observation
	Time	Time of the observation
Active Power Measurements	P_Average	Average active power
	P_Minimum	Minimum active power
	P_Maximum	Maximum active power
Reactive Power Measurements	Q_Average	Average reactive power
	Q_Minimum	Minimum reactive power
	Q_Maximum	Maximum reactive power
Apparent Power Measurements	S_Average	Average apparent power
	S_Minimum	Minimum apparent power
	S_Maximum	Maximum apparent power

$P\_Average$, $P\_Minimum$, and $P\_Maximum$ denote the active power consumption values that reflect the natural daily cycle, ranging from low power usage during nighttime to significantly higher demand during daytime hours. Measurements of reactive power are made by $Q\_Average$, $Q\_Minimum$ and $Q\_Maximum$, and $S\_Average$, $S\_Minimum$ and $S\_Maximum$ recorded apparent power values.

Table 6.2 summarises the non-linear meteorological features that complement the electrical load data. These variables, such as ambient *Temperature*, *Humidity*, *Wind_ Speed*, and several solar irradiance measures, etc., reflect the effect of weather and solar conditions on building energy requirements. By including these parameters in the forecasting models, seasonal effects, variations driven by HVAC systems, and solar gain effects that are essential for MTLF can be learn more effectively in smart commercial buildings.

Table 6.2: Feature description of non-linear meteorological parameters

Category	Feature Name	Description
Meteorological Parameters	Temperature	Ambient air temperature (°C)
	Humidity	Atmospheric humidity (g/kg)
	Wind_Speed	Wind speed at measurement height (m/s)
Solar Irradiance Parameters	Gb(i)	Beam (direct) solar irradiance (W/m^2)
	Gd(i)	Diffuse solar irradiance (W/m^2)
	H_sun	Sunshine duration (hours)
	WS10m	Wind speed at 10 m elevation (m/s)

The electrical power consumption data in 2023 is of robust quality with just one missing value in the power columns, showing 0.01% of the entire records, thus such data is quite appropriate to use in time series forecasting.

The hourly granularity also allows the models to capture intra-day regularities such as morning starting peaks, midday plateaus, evening surges, and the lows of demand at night, which are typical of the power usage in the real world.

6.2.3 Data Preprocessing

At the preprocessing stage, the CSV data is imported through a robust data loading module that identifies all types of missing entries (such as Not Available (NA), Not Applicable (N/A), or blank). It does not change zero values, which may be an actual low-demand sensitivity of power consumption. In Figure 6.1, the platform also verifies whether or not it is an empty file and also verifies that it is in the right format.

6.2.4 Outlier Detection

After the dataset is uploaded, the application allow user to choose the Interquartile Range (IQR) multiplier of outlier removal. The default value for the IQR threshold is 1.50, as shown in Figure 6.2, which balances between the removal of outliers and retaining the data set integrity in most cases. When the multiplier is changed to 3.0, the outlier filter becomes

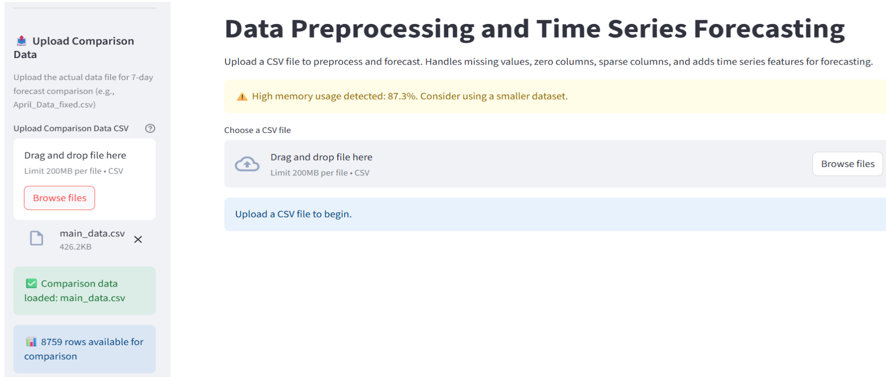


Figure 6.1: Application Interface: User will upload dataset file

less strict: $[Q1 - 3(IQR), Q3 + 3(IQR)]$ gets larger, and more extreme observations could stay in as part of the data. Conversely, when the multiplier is reduced to 0.5 sharply narrows the detection window (which results in extensive removal of valid data points) as bounds are more strict and many data points are flagged as outliers and removed due to the tighter bounds.

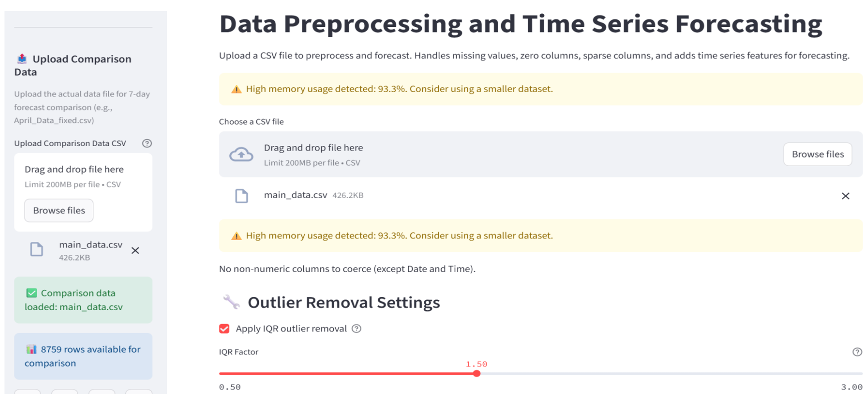


Figure 6.2: Dynamic outlier removal configuration

Using the Interquartile Range method, outliers are identified by calculating:

- $Q1$ (25th percentile) and $Q3$ (75th percentile)
- $IQR = Q3 - Q1$
- Lower bound: $Q1 - 1.5\,IQR$
- Upper bound: $Q3 + 1.5\,IQR$

The interquartile range (IQR) plot of $P\_Average$ is to show the distribution of power consumption values. The blue box shows the interquartile range that covers 50% of the observations. The red dashed lines, on the other hand, represent upper and lower outlier boundaries that were computed according to the user-provided value of the IQR multiplier.

Blue markers, data plotted above the upper threshold value, indicate detected outliers, where power consumption exceeds the permitted bounds.

Figures 6.3 and 6.4 illustrate the minimum and maximum IQR-based outlier detection cases.

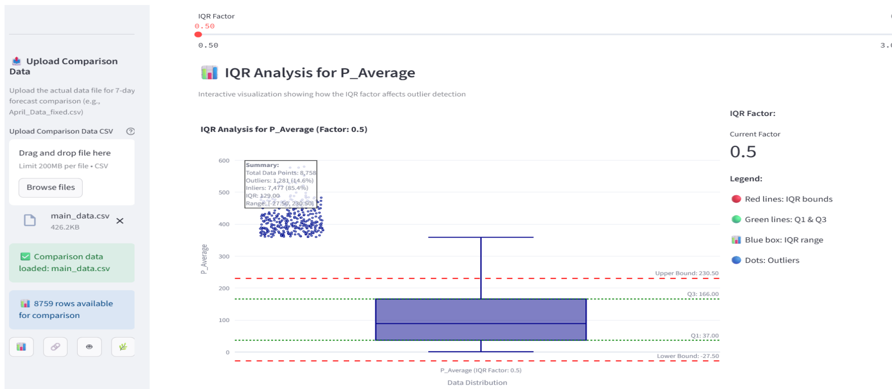


Figure 6.3: IQR plot with minimum IQR factor

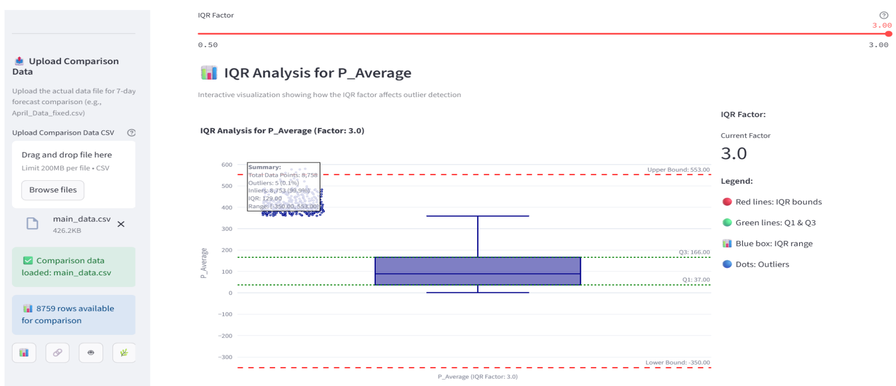


Figure 6.4: IQR plot with maximum IQR factor

This application is capable of handling missing values by applying mean imputation, where each numerical feature with missing entries is filled using its corresponding column average. This retains the dataset's overall balance, especially considering that the percentage of missing data is less than 1%. On large files, it works in blocks, then authenticates and fills in any missing data with zeros, records all the imputation information to be transparent. The Figure 6.5 shows the number of missing data for each variable, which can be selected and reviewed directly within the interface.

The system recognises the target column, *P_Average* (mean active power), and generates the corresponding columns for model training. This maintains consistency of the target variable throughout all the steps of modelling. Once the preprocessing is complete, a preview

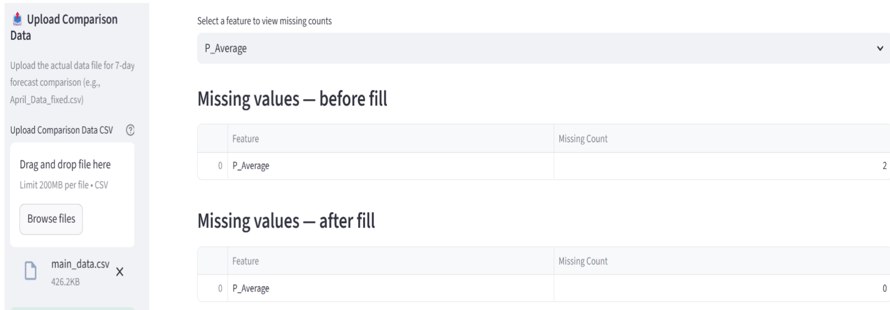
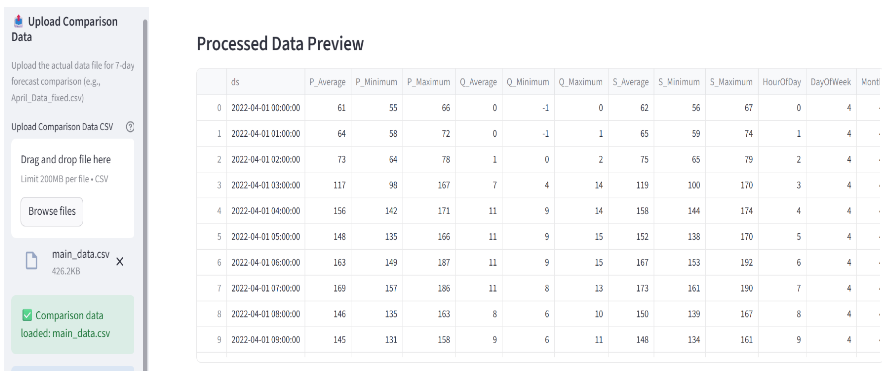


Figure 6.5: Result of mean interpolation in data preprocessing

of the processed data of the Excel sheet is shared on the Streamlit interface, as illustrated in Figure 6.6.



	ds	P_Average	P_Minimum	P_Maximum	Q_Average	Q_Minimum	Q_Maximum	S_Average	S_Minimum	S_Maximum	HourOfDay	DayOfWeek	Montl
0	2022-04-01 00:00:00	61	55	66	0	-1	0	62	56	67	0	4	
1	2022-04-01 01:00:00	64	58	72	0	-1	1	65	59	74	1	4	
2	2022-04-01 02:00:00	73	64	78	1	0	2	75	65	79	2	4	
3	2022-04-01 03:00:00	117	98	167	7	4	14	119	100	170	3	4	
4	2022-04-01 04:00:00	156	142	171	11	9	14	158	144	174	4	4	
5	2022-04-01 05:00:00	148	135	166	11	9	15	152	138	170	5	4	
6	2022-04-01 06:00:00	163	149	187	11	9	15	167	153	192	6	4	
7	2022-04-01 07:00:00	169	157	186	11	8	13	173	161	190	7	4	
8	2022-04-01 08:00:00	146	135	163	8	6	10	150	139	167	8	4	
9	2022-04-01 09:00:00	145	131	158	9	6	11	148	134	161	9	4	

Figure 6.6: Overview of the output of the processed data

6.2.5 Correlation Analysis

In Figure 6.7, the correlation analysis module explores the investigation of the correlation in a systematic way, the connectivity of power consumption with relevant variables of impact by means of visual tools and statistical measures, which are interactive. It visualises heatmaps with colour-coded heatmaps strength (between +1 and -1) of correlation and bar charts of ranking the features in order of influence on the target variable. The level of correlation has been divided into levels, which include very strong, strong, moderate, and weak, to be easily interpreted. In addition to its numerical results, the characteristics of significant features are also brought out by this analysis influence consumption to inform model development, energy planning, and capacity decisions. As an example, time-of-use pricing can be informed by strong hourly correlations, and seasonal trends can support infrastructure planning, turning data insights into actionable strategies for smarter energy management.

The correlation matrix visually represents how all the input features in the model re-

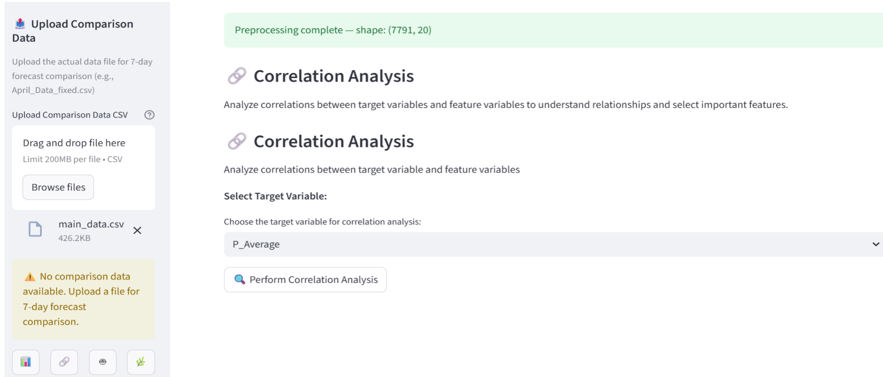
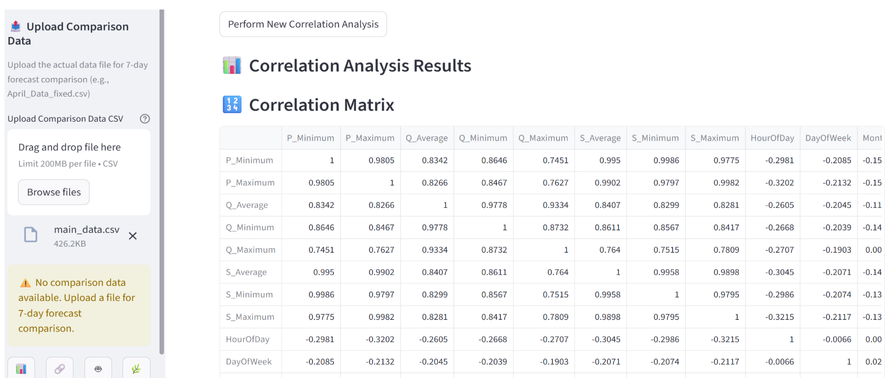


Figure 6.7: Correlation analysis tab

late to one another. It assists in establishing the variables that increase or decrease in the same direction or those that increase or decrease to a low degree. This matrix is created in Figure 6.8, to be explicitly useful in demonstrating the strength and direction of relationships between each feature, enabling one to have a hint at how various factors affect the consumption of power and how they interrelate in the data.



	P_Minimum	P_Maximum	Q_Average	Q_Minimum	Q_Maximum	S_Average	S_Minimum	S_Maximum	HourOfDay	DayOfWeek	Mont
P_Minimum	1	0.9805	0.8342	0.8646	0.7451	0.995	0.9986	0.9775	-0.2981	-0.2085	-0.15
P_Maximum	0.9805	1	0.8266	0.8467	0.7627	0.9902	0.9797	0.9982	-0.3202	-0.2132	-0.15
Q_Average	0.8342	0.8266	1	0.9778	0.9334	0.8407	0.8299	0.8281	-0.2605	-0.2045	-0.11
Q_Minimum	0.8646	0.8467	0.9778	1	0.8732	0.8611	0.8567	0.8417	-0.2668	-0.2039	-0.14
Q_Maximum	0.7451	0.7627	0.9334	0.8732	1	0.764	0.7515	0.7809	-0.2707	-0.1903	0.00
S_Average	0.995	0.9902	0.8407	0.8611	0.764	1	0.9958	0.9898	-0.3045	-0.2071	-0.14
S_Minimum	0.9986	0.9797	0.8299	0.8567	0.7515	0.9958	1	0.9795	-0.2986	-0.2074	-0.13
S_Maximum	0.9775	0.9982	0.8281	0.8417	0.7809	0.9898	0.9795	1	-0.3215	-0.2117	-0.13
HourOfDay	-0.2981	-0.3202	-0.2605	-0.2668	-0.2707	-0.3045	-0.2986	-0.3215	1	-0.0066	0.00
DayOfWeek	-0.2085	-0.2132	-0.2045	-0.2039	-0.1903	-0.2071	-0.2074	-0.2117	-0.0066	1	0.02

Figure 6.8: Correlation matrix of dataset

6.2.6 Feature Correlation Analysis

The correlation analysis in Figure 6.9 shows that power measurement features are the most influential predictors of energy consumption. Apparent and active power variables ($S\_Average$, $S\_Minimum$, $S\_Maximum$, $P\_Minimum$, $P\_Maximum$) show very strong positive correlations ranging from 0.9875 to 0.9988, while reactive power features ($Q\_Average$, $Q\_Minimum$, $Q\_Maximum$) exhibit strong correlations between 0.75 and 0.86. Other time features like hour, day, and month are much less correlated (below $|0.4|$), indicating that their patterns can be better captured by non-linear models such as neural networks.

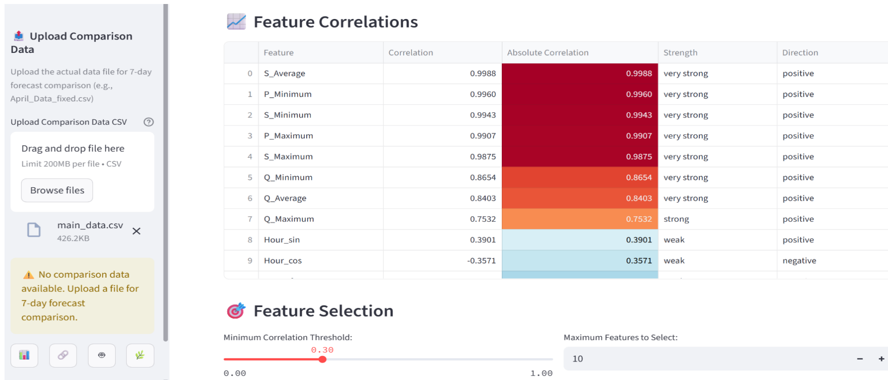


Figure 6.9: Feature correlations of the dataset (Heatmap view)

6.3 Model Training and Forecasting Process

A pre-cleaned set of 8,370 hourly power data for one year is used to begin the training of the model. Data has been divided on the basis of time-series; 80% (approximately 6696 hours) of training and 20 percent (approximately 1674 hours) of testing to provide adequate time-based validation. The final number of data points changes according to the IQR threshold specified for outlier detection because different ranges remove various amounts of data.

All the models are trained on the same data using tuned hyperparameters, sequence lengths of 24–168 hours, 50–100 epochs using early stopping, the batch size of 32, and using Adam as an optimiser (learning rate 0.001).

- **Training Monitoring:** The model's training can be viewed in real-time as the model learns. The interface shows the visualisation of the current epoch, completion percentage, and loss values, enabling more frequent monitoring of the learning behaviour and early identification of problems without interrupting training.

- **Model Saving:** The application automatically saves a checkpoint every time the validation loss decreases to a new minimum. This makes sure that the most effective iteration of the model is reliably kept intact, which removes the requirement for manual file handling.

- **Forecast Generation:** Once training is complete, the model generates predictions on the test set using a recursive (one-hour-at-a-time) forecasting strategy. In this approach, the most recent training split hours of sequence length are fed to the model to generate a single-step prediction. This predicted value is then added to the rolling input sequence and is again used in the subsequent prediction. This is repeated on a per-hour basis in the test set and then on the 7-day future horizon so that realistic, time-understandable forecasts are obtained, which maintain the temporal correlations that exist in real-world load profiles.

- **Performance Evaluation:** The system measures accuracy by computing four important measures, namely RMSE, MAE, MAPE, and R^2. When combined, these measures give a comprehensive assessment of the results of the forecasting. They indicate how the model captures the behaviour of the underlying data and the size of the deviations between the expected and actual values, so the performance of the model can be analysed and viewed in a number of complementary ways.
- **Model Selection:** The most successful model is selected in accordance with the assessment of parameter metrics.
- **Final Forecasting:** A seven-day (168-hour) hourly basis forecasting is then generated by the model chosen. This forward-looking outlook facilitates planning towards scheduling, procurement, and resource or emission management.

Three kinds of forecasting models are involved in the training phase:

1. **Statistical Models (Prophet, SARIMAX):** The models assume that a time series may be split into three salient components, which include trend, seasonality, and noise. The individual parts are modelled on the interpretable parameters that reveal the general composition of the data. The Prophet model is highly applicable in the context of multiple seasonality, holiday effects with minimal tuning, whereas the SARIMA is the extension of ARIMA that incorporates autoregressive and moving average terms in addition to exogenous terms (external), combining them can create stable baselines and give unambiguous information on prevailing temporal trends.
2. **ML Models (Decision Trees, Random Forests):** Tree-based models break down the data into easy-to-understand decision rules and include both nonlinear relationships and interactions between values like weather, calendar effects, and operational factors. RF improves this method by training a large number of trees using bootstrapped data sets and averaging their estimates to reduce variance and increase stability. The models indicate that quick training ability is capable of managing different kinds of data and is significant in detecting the primary characteristics that result in the outputs.
3. **Deep Learning Models (LSTM, GRU, Transformer):** Sequence-driven neural networks are effective at finding patterns from past data, using earlier findings to forecast upcoming behaviours. LSTM and GRU architectures retain memory of recent

trends through a gated mechanism that controls information flow, while Transformer models use attention layers to focus on the most relevant time steps without relying on recurrence. These approaches scale efficiently to long historical datasets and can uncover subtle, evolving dependencies that simpler models might overlook.

The drop-down menu, as shown in Figure 6.10, provides users with the ability to select the specific forecasting model to be trained.

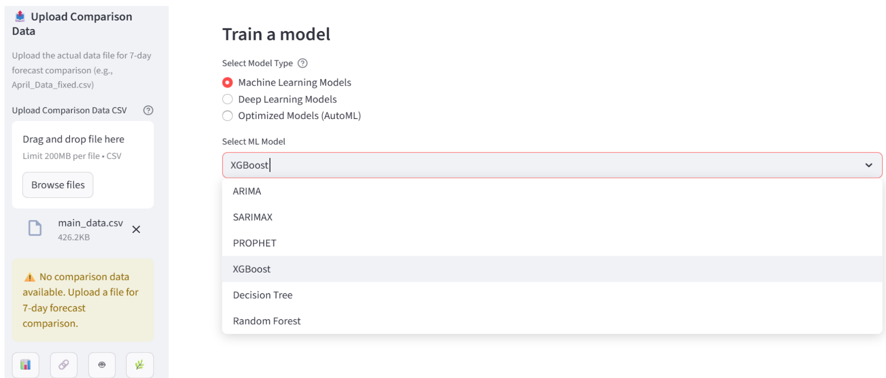


Figure 6.10: User-based model selection on the forecasting interface

As shown in Figure 6.11, features can be included or excluded, allowing users to evaluate how each variable influences model training.

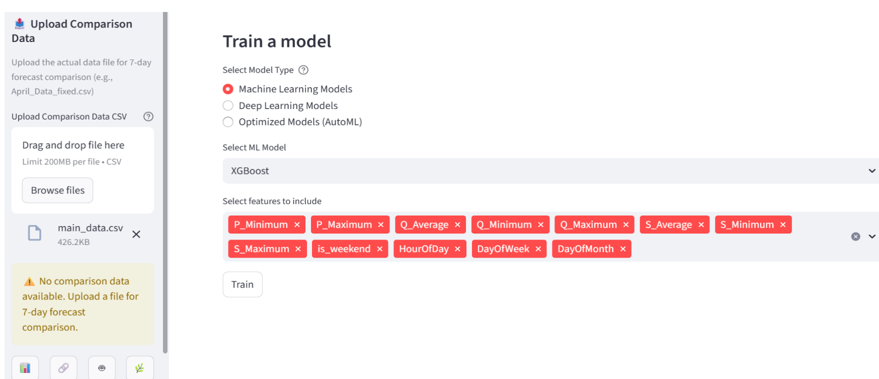


Figure 6.11: Selection of features for modelling

6.4 Energy Forecasting Output and Visualisation

After model training is completed, the system displays the test-set forecasting metrics, including RMSE, R^2, MAE, and MAPE. As illustrated in Figure 6.12, the Forecast Visualisation panel provides an interactive comparison of the training and test segments of the dataset. In this phase, the final seven days of observations are shown to represent the test forecast performance.

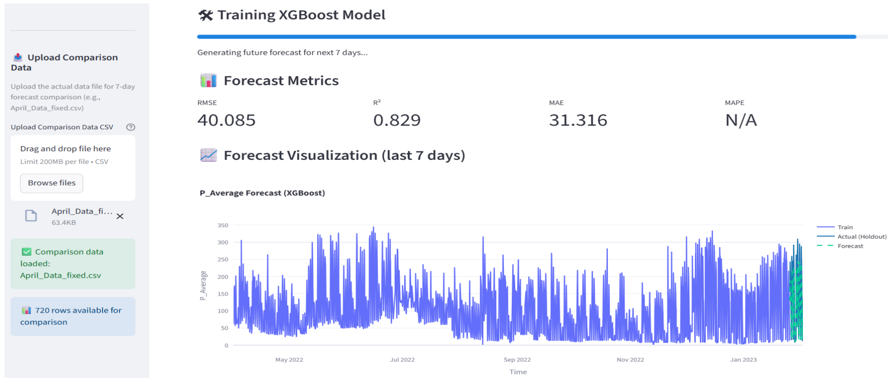


Figure 6.12: Result of XGBoost forecasting

Residual analysis, shown in Figure 6.13, is performed after the forecast results are generated. It involves computing the difference between the actual and predicted values to evaluate the distribution and magnitude of the model's errors. The resulting residual plots help us find any potential bias and whether the model might have, or if there is any systematic pattern in our forecasts. This evaluation also delivers a comprehensive perspective on how the model may misinterpret both under and over various time intervals. It's important to identify these patterns to enhance the model's robustness and ensure that forecasting performance remains stable under varying operating conditions.

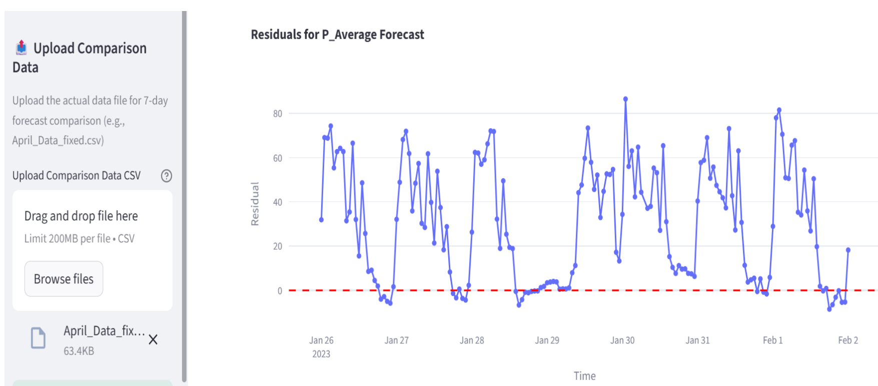


Figure 6.13: Residual analysis of XGBoost model

The comparative visualisation of the actual and forecast values is shown in Figure 6.14. This allows a clear understanding of the numerical values being represented and how closely the predictions follow the real load patterns. It also assists in identifying any systematic deviations across different hours of the evaluation period.

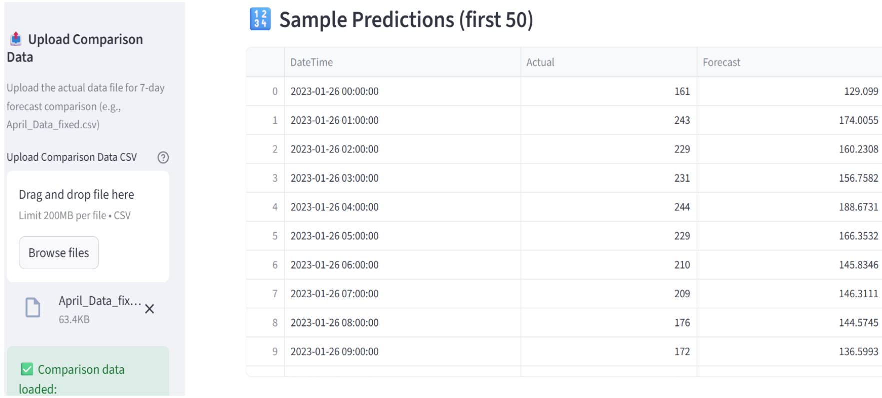


Figure 6.14: Overview of forecasting

The seven-day forecast is presented through an interactive chart that visualises the predicted power consumption as a continuous 168-hour time series beyond the final point of the training dataset. As shown in Figure 6.15, the advance forecast values are first displayed in a structured table, where the corresponding original readings are retrieved from the comparison CSV file uploaded by the user. This tabular comparison will be used to confirm that there is a similarity between actual and forecasted values at each hourly interval. It also gives a good understanding of the basis upon which multiplication of short-term deviations and examining the extent to which the model can be extrapolated beyond the training horizon can be evaluated.

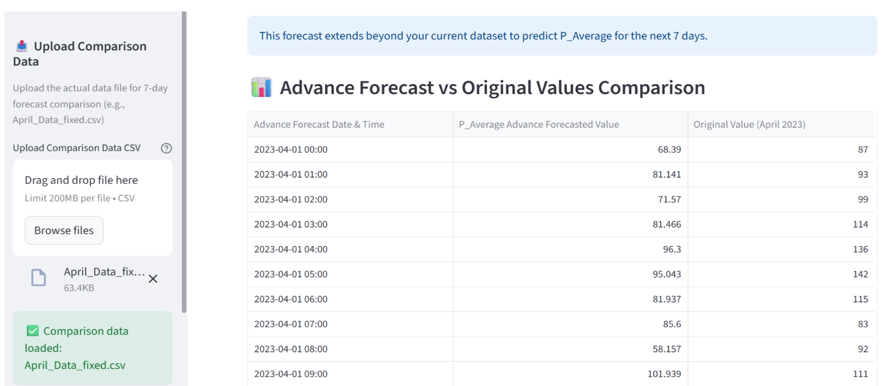


Figure 6.15: Advance forecasting vs original values

The advance forecasting versus actual values curve shown in Figure 6.16 shows the difference between the forecast and the actual power consumption across the seven-day comparison window. The two curves indicate both the times when they are closely matched and those times when they are deviated greatly, especially during peak demand times. This visual comparison can be used to determine the generalisation of the model outside the training horizon and the temporal behaviour of the model when faced with unknown data.

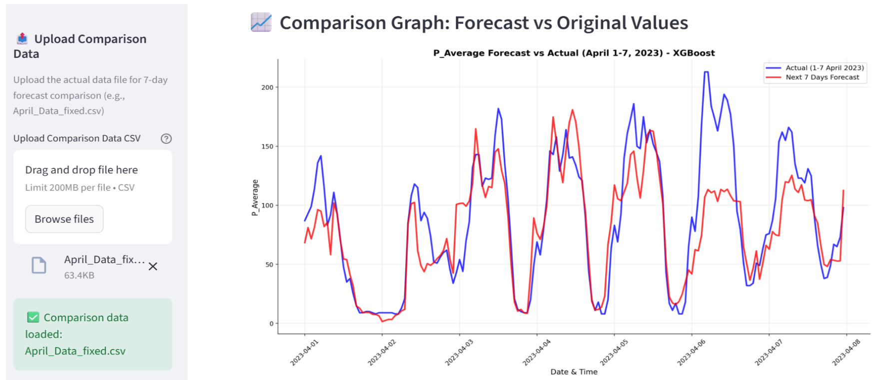


Figure 6.16: Comparison plot of advance forecasting vs actual values

The results can be easily exported in multiple formats: as a CSV file (which facilitates rapid spreadsheet review), having been converted into Excel format for more detailed daily summaries; as a JSON document integrated with other systems and application programming interfaces (APIs). These built-in flexibilities ensure that the findings are extremely useful for any kind of practical decision or strategy making. They may also be exploited in operational planning, resource management, and capacity forecasting.

6.4.1 Implementation of AutoML for Energy Forecasting

The H2O AutoML will be configured in this project to automatically train and compare up to ten models with three of the most effective algorithm families, which are Gradient Boosting Machines (GBM), XGBoost, and Distributed Random Forest (DRF). In order to determine which model is the best, it also uses the RMSE to compare each model's output. In case the H2O cluster cannot start for any reason, the system is switched to a standalone XGBoost model without failures in order to make sure that the system keeps the forecasting process going, as shown in Figure 6.17.

H2O AutoML provided excellent results in terms of forecasting and a high level of precision, also indicating the strength of AutoML in conducting an automated search through a variety of algorithms and parameter settings to discover the most performing model.

The comparison, in general, shows that H2O AutoML is superior compared to traditional models and, therefore, a more efficient option to predict events accurately and plan the strategic design of the energy.

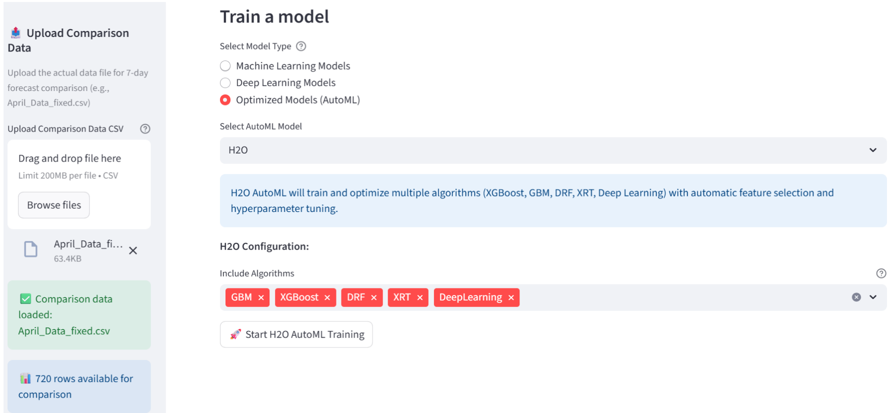


Figure 6.17: AutoML model test forecasting results

6.5 CO_2 Emission and Cost Estimation Modules

The energy emissions module converts power-usage data into meaningful carbon-impact insights. It sums all target variable values, such as ($P\_Average$), to compute the total energy consumption (in kWh) and multiplies this quantity by an emission factor (default 0.233 kg CO_2e/kWh) to estimate overall emissions. The data is then categorised by season in the application, and daily outputs are determined when and where the carbon output is greatest. Besides this, the module also draws attention to the seasonal pattern of emissions, and the user can see the time when the carbon output is high and what causes it. The formula used for energy emission calculation is:

$$\text{CO}_2 \text{ Emissions (kg)} = \text{Energy Consumption (kWh)} \times \text{Emission Factor (kg CO}_2\text{e/kWh)}$$

Where:

- **Energy Consumption (kWh)** = Sum of all P_Average values in the dataset
- **Emission Factor** = 0.233 kg CO_2e/kWh (default value, adjustable by user)

For daily emissions, the formula is applied after resampling (IPCC, 2006):

$$\text{Daily CO}_2 \text{ Emissions (kg)} = \text{Daily Energy Sum (kWh)} \times \text{Emission Factor (kg CO}_2\text{e/kWh)}$$

For seasonal emissions:

$$\text{Seasonal CO}_2 \text{ Emissions (kg)} = \text{Seasonal Energy Total (kWh)} \times \text{Emission Factor (kg CO}_2\text{e/kWh)}$$

The emission factor of 0.233 represents the carbon intensity of electricity generation, indicating that each kilowatt-hour of consumed electricity corresponds to the release of 0.233 kilograms of CO_2 equivalent into the atmosphere. This parameter, illustrated in Figure 6.18, serves as the basis for estimating carbon emissions from the forecast energy consumption.

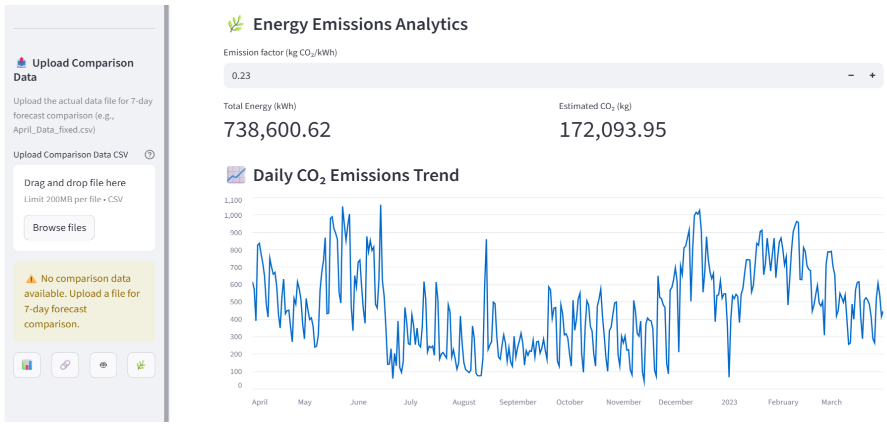
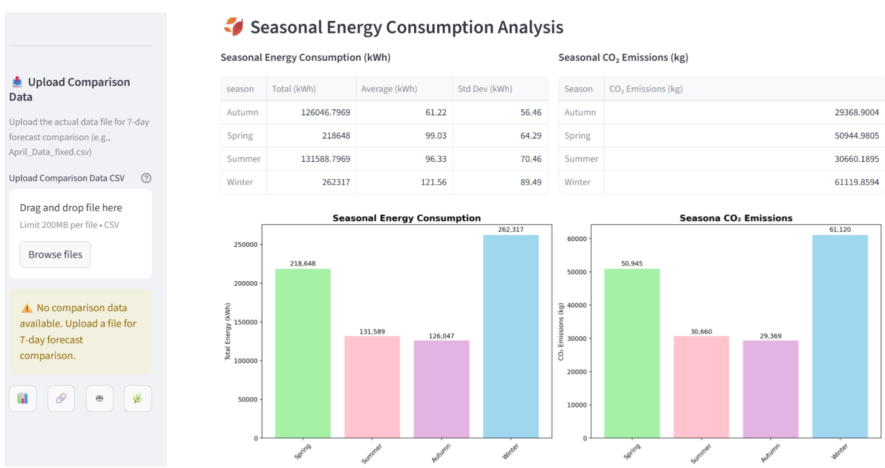


Figure 6.18: Energy emission trend

The seasonal dashboard provides a proper understanding of the seasonal energy consumption and the carbon emissions. It emphasises that the periods when there is an increased energy demand and an increase in emissions are the colder months, with the milder seasons demonstrating lower usage as a result of less heating or cooling requirement. The seasonal variations can be easily observed in Figure 6.19, using bar charts, and this will assist organisations to determine the peak used energy and areas where efficiency efforts are most beneficial. This understanding can be helpful in more effective demand management, renewable integration, and carbon reduction programs planning during outburst periods.



season	Total (kWh)	Average (kWh)	Std Dev (kWh)
Autumn	126046.7969	61.22	56.46
Spring	218648	99.03	64.29
Summer	131588.7969	96.33	70.46
Winter	262317	121.56	89.49

Season	CO₂ Emissions (kg)
Autumn	29368.9004
Spring	50944.9805
Summer	30660.1895
Winter	61119.8594

Figure 6.19: Seasonal energy consumption dashboard

6.5.1 Discussion on Forecasting Errors and Practical Implications

Although several studies report lower forecasting errors for ML and DL based MTLF models, these results are often obtained under idealised conditions, such as highly controlled datasets, simplified load profiles, or shorter forecasting horizons. In contrast, the present study evaluates forecasting performance under realistic operational conditions, using complete but highly variable real-world commercial building data, strict separation between training and unseen test sets, and a seven-day forecasting horizon at hourly resolution. Consequently, higher error values (not lower than 27% MAPE) are observed. While such errors may limit direct real-time control actions, they remain valuable for energy planning, demand trend analysis, and CO_2 emission estimation within Energy Management Systems (EMS). Importantly, these results highlight the trade-off between forecasting accuracy and robustness in real-world deployments.

6.6 Chapter Summary

This platform demonstrates that it does not need a complex approach to advanced energy forecasting. A combination of powerful algorithms and a clean, user-friendly interface using the Streamlit platform has ensured that a program that used to be a technical, code-heavy process has become a refined and intuitive experience that anyone can use. One is now able to enter their power consumption data, view it in interactive charts, clean and prepare it, run many forecasting models, and compare their output without writing a single line of code.

The system integrates a large variety of forecasting techniques, such as statistical predictive models (e.g., Prophet and ARIMA) and ML technologies (e.g., DT and RF), and automated tools, such as H2O AutoML. Among the evaluated methods, AutoML demonstrated an excellent approach in terms of overall forecasting performance and predictive accuracy relative to the actual consumption patterns. These findings suggest that AutoML systems can greatly simplify the model selection and tuning process and still perform at the level of a forecasting solution designed by experts.

The difference with this project is that it is capable of integrating the technical complexity and reality. Each of the data uploads, as well as pre-processing, is via the interactive web dashboard, which reacts to user input in real-time, and the entire model training and analysis of the emissions is run. The availability of a carbon emissions calculator offers an extra facet of insight since the user is able to attribute the utilisation of energy to its effects on the environment. This transforms the tool into an easy-to-predict tool into a heavyweight decision-making tool that supports efficiency and sustainability.

In practice, the profits are far-flung. The predictions allow utility operators to organise the manner in which they will produce the power, not to mention the reduction of energy wastage. The facility managers can forecast the peak of demand and calculate the loads in a balanced manner. The sustainability departments can understand the maximum periods of emissions and come up with smarter control solutions. The system can be formed to any desired outcome, thus it can be helpful with the operations, as well as with environmental conservation.

In one of the futures, the future versions will include the incorporation of real-time weather information, holiday effects, as well as more forecasting to make even more precise forecasting. With these enhancements, the platform could be turned into an energy intelligence system where it is not only expected to make predictions related to consumption but also to help people manage energy smarter, greener, and in a more sustainable way in the future.

Chapter 7

Conclusion and Future Work

7.1 Summary of Contributions

This dissertation provides a complete framework to deal with missing data and construct accurate prediction models for building energy consumption. The work combines methodological, experimental, and system-level innovations into a unified pipeline, which is capable of addressing real data challenges.

A notable contribution is observed in the thorough examination of the dynamics of absent data as affected by simulated mechanisms of MCAR, MAR, and MNAR, including both stochastic and block structured configurations. This analysis provided clear insight into how missingness disrupts temporal structure and impacts forecasting models. Building on this, the study evaluated a diverse range of imputation techniques, statistical, ML, DL, and hybrid approaches demonstrating that advanced and hybrid methods offer more reliable reconstruction of energy consumption sequences.

This research also investigated a wide variety of forecasting methods from classical ML and DL models to hybrid ML–DL models. An important development is the transformation of the FireNet architecture into a 1D convolutional network for time series prediction, allowing for multi-scale representations to be learned for both short- and long-term time windows. The combination of these forecasting techniques and automatic hyperparameter tuning (using Grid Search and Optuna) allowed for an accurate comparison between models.

In addition to theoretical contributions, a complete forecast application platform was designed. This framework combines preprocessing, missing data treatment, imputation, prediction, optimisation, and visualisation within an integrated system. It enables experimental consistency and demonstrates the feasibility of deploying the proposed workflow to real energy management applications.

7.2 Major Findings

The results obtained from the extensive experimentation reveal several important insights. First, the behaviour and structure of missing data strongly influence forecasting accuracy. MAR patterns and block-type gaps create the most severe degradation, confirming that understanding missingness is essential for reliable forecasting. The imputation study showed that hybrid and advanced imputers consistently outperform simple statistical techniques, particularly when the missing rate is high.

The forecasting experiments indicated that the DL models (i.e., LSTM, GRU, and FireNet) are good at capturing temporal dependencies. On the other hand, ML models like XGBoost and Random Forest achieve a comparable performance at much lower computational effort. FireNet performed with notable performance in learning multi-scale temporal patterns, obtaining high accuracy not only on the complete data but also on the imputed data. Hybrid forecasting models took advantage of both the strengths of temporal encoders and nonlinear regressors and achieved better overall predictive performance.

The assessment further emphasised the importance of imputation quality on predictive accuracy, also in terms of models trained over well-imputed data always performing significantly better. Furthermore, the use of Optuna Bayesian optimisation significantly reduced tuning time and resulted in better hyperparameters than computational search techniques.

7.3 Limitations of the Study

Although the results of this thesis are promising, there are several limitations that need to be addressed. The study was based on a single building and may not generalise to patterns experienced by buildings of different types in different climates. Moreover, the application of DL models, and specifically CNNs and RNNs, requires considerable computational power for training, which makes it possible to use them only on high-performance devices.

Another constraint is problems regarding the assumptions on missing-data mechanisms. Although analysis of MCAR, MAR, and MNAR simulations provides useful information, actual missingness can have more complex causal configurations. In addition, the forecasting models built only on fixed-length sliding windows, and employing more sophisticated methods for temporal signal processing, such as adaptive windowing or attention-based selection, are expected to be beneficial. Finally, while a few hybrid models are shown to be highly accurate in practice, their intensive computational requirement may limit their applicability in real-time.

7.4 Future Directions

The research in this thesis suggests a number of valuable directions for future work. One possible direction is to design distributed learning algorithms such that models are learned collaboratively from data of multiple buildings, while raw data privacy can be maintained and model generalisation ability improved.

Another direction involves edge AI, where lightweight, compressed, or quantised versions of the models, especially FireNet and LSTM, could be deployed on embedded controllers and IoT devices for real-time forecasting. Transformer-based architectures and attention mechanisms may further enhance the modelling of long-range temporal dependencies. A further interesting field is self-supervised learning, which enables models to learn deeper temporal representations before fine-tuning on predicting tasks by utilising huge amounts of unlabeled energy data.

Further research may also focus on causal modelling of missing data to better understand the origins and propagation of missingness in building-management systems. Finally, cross-building transfer learning could reduce the need for large training datasets by adapting models from one building to others with minimal retraining effort.

7.5 Final Remarks

In summary, this dissertation offers important methodological, algorithmic, and practical contributions to the state of the art in building energy forecasting. This work establishes that with the use of missing data analysis, imputation, ML, and DL techniques, as well as hybrid forecasting models developed in an integrated framework supported by a working forecasting application system, it is possible to accomplish credible and accurate predictions even under difficult data scenarios. The findings obtained contribute to the further development of intelligent data-driven Energy Management Systems (EMSs) and promote the overall idea of energy-efficient, sustainable, and autonomous smart buildings.

Bibliography

[1] Okanlawon, O. E. *A comparative analysis of time series method for energy demand forecasting in Nigeria*. Bachelor's thesis. 2023. URL: https://www.academia.edu/113149119/Okanlawon_final_project.

[2] Misiurek, K., Olkuski, T., and Zyśk, J. "Review of methods and models for forecasting electricity consumption". In: *Energies* 18.15 (2025), p. 4032. DOI: 10.3390/en18154032.

[3] Baek, M. and Seo, Y. "Hybrid forecasting of university electricity demand using time series and deep learning". In: *Energy and Buildings* (2025). Advance online publication. DOI: 10.1016/j.enbuild.2025.116400.

[4] Tzortzis, A. M., Pelekis, S., Spiliotis, E., Karakolis, E., Mouzakitis, S., Psarras, J., and Askounis, D. "Transfer learning for day-ahead load forecasting: A case study on European national electricity demand time series". In: *Mathematics* 12.1 (2024), p. 19. DOI: 10.3390/math12010019.

[5] Ali, S., Bogarra, S., Riaz, M. N., Phyo, P. P., Flynn, D., and Taha, A. "From time-series to hybrid models: Advancements in short-term load forecasting embracing smart grid paradigm". In: *Applied Sciences* 14.11 (2024), p. 4442. DOI: 10.3390/app14114442.

[6] Nti, I. K., Teimeh, M., Nyarko-Boateng, O., and Adekoya, A. F. "Electricity load forecasting: a systematic review". In: *Journal of Electrical Systems and Information Technology* 7 (2020), p. 13. DOI: 10.1186/s43067-020-00021-8.

[7] Rodrigues, F., Cardeira, C., Calado, J. M. F., and Melicio, R. "Short-term load forecasting of electricity demand for the residential sector based on modelling techniques: A systematic review". In: *Energies* 16.10 (2023), p. 4098. DOI: 10.3390/en16104098.

[8] Mystakidis, A., Koukaras, P., Tsalikidis, N., Ioannidis, D., and Tjortjis, C. "Energy forecasting: A comprehensive review of techniques and technologies". In: *Energies* 17.7 (2024), p. 1662. DOI: 10.3390/en17071662.

[9] Ugbehe, P. O., Diemuodeke, O. E., and Aikhuele, D. O. "Electricity demand forecasting methodologies and applications: A review". In: *Sustainable Energy Research* 12 (2025), p. 19. DOI: 10.1186/s40807-025-00149-z.

[10] Mamun, A. A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., and Hossain, E. "A comprehensive review of the load forecasting techniques using single and hybrid predictive models". In: *IEEE Access* 8 (2020), pp. 134911–134939. DOI: 10.1109/ACCESS.2020.3010702.

[11] Almazrouee, M. A. et al. "A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives and applications". In: *Energy Conversion and Management: X* 22 (2025), p. 100922. DOI: 10.1016/j.ecmx.2025.100922.

[12] Kolkova, A. and Kljucnikov, A. "Demand forecasting: AI-based, statistical and hybrid models vs practice-based models – the case of SMEs and large enterprises". In: *Economics and Sociology* 15.4 (2022), pp. 39–62. DOI: 10.14254/2071-789X.2022/15-4/2.

[13] Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y., and Imran, M. "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons". In: *Sustainability* 12.15 (2020), p. 5931. DOI: 10.3390/su12155931.

[14] Madin, V., Salykova, O., Ivanova, I., Bizhanova, O., and Aldasheva, D. "Enhancing electricity consumption forecasting in the Republic of Kazakhstan using machine learning". In: *Journal of Applied Engineering and Technological Science* 6.2 (2025), pp. 1166–1196. DOI: 10.37385/jaets.v6i2.7425.

[15] Makris, I., Moschos, N., Radoglou-Grammatikis, P., Andriopoulos, N., Tzanis, N., Malamaki, K.-N., Fotopoulou, M., Kaousias, K., Lampsidis Tompros, G., and Sarigiannidis, P. "Exploring load forecasting: Bridging statistical methods to deep learning techniques in real-world environment". In: *Proceedings of the IEEE Global Communications Conference*. IEEE, 2024, pp. 1–6. DOI: 10.1109/DCOSS-IoT61029.2024.00078.

[16] Karageorgou, E. "Forecasting Spain's electricity load: A comparative analysis of classical time series, neural networks, and deep learning models". Master's thesis. MA thesis. University of Macedonia, Thessaloniki, Greece, 2023. URL: https://dspace.lib.uom.gr/bitstream/2159/30101/4/KarageorgouEvangeliaMsc2023.pdf.

[17] Rai, S. and De, M. "Analysis of classical and machine learning based short-term and mid-term load forecasting for smart grid". In: *International Journal of Sustainable Energy* 40.8 (2021), pp. 783–802. DOI: 10.1080/14786451.2021.1873339.

[18] Jahan, I. S., Snasel, V., and Misak, S. "Intelligent systems for power load forecasting: A study review". In: *Energies* 13.22 (2020), p. 6105. DOI: 10.3390/en13226105.

[19] Oreshkin, B. N., Dudek, G., Pełka, P., and Turkina, E. "N-BEATS neural network for mid-term electricity load forecasting". In: *Applied Energy* 293 (2021), p. 116918. DOI: 10.1016/j.apenergy.2021.116918.

[20] Zimmermann, M. and Ziel, F. "Efficient mid-term forecasting of hourly electricity load using generalized additive models". In: *Applied Energy* 388 (2025), p. 125444. DOI: 10.1016/j.apenergy.2025.125444.

[21] Zhang, L., Wu, D., and Luo, X. "An Error Correction Mid-term Electricity Load Forecasting Model Based on Seasonal Decomposition". In: *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2023. DOI: 10.1109/SMC53992.2023.10394531.

[22] Cai, Y., Chen, Q., Wang, Y., Li, W., Ren, J., and Qu, Y. "Forecasting method of electric vehicle charging load based on virtual prediction parameter estimation strategy". In: *Archives of Electrical Engineering* 73.2 (2024), pp. 355–372. DOI: 10.24425/aee.2024.149921.

[23] Razghandi, M., Zhou, H., Erol-Kantarci, M., and Turgut, D. "Smart home energy management: Sequence-to-sequence load forecasting and Q-learning". In: *Proceedings of the IEEE Global Communications Conference*. IEEE, 2021, pp. 1–6. DOI: 10.48550/arXiv.2109.12440.

[24] Widmark, N. "Short-term electricity consumption forecasting using deep learning and external variables". Master's thesis. MA thesis. Luleå University of Technology, Luleå, Sweden, 2022. URL: https://www.diva-portal.org/smash/get/diva2:1668934/FULLTEXT01.pdf.

[25] Lotfi, H., Vafadoost, P., Rokhsati, H., and Parsadust, H. "An optimized machine learning approach for short-term electricity demand forecasting considering environmental and historical influencing factors". In: *Discover Applied Sciences* 7 (2025), p. 1015. DOI: 10.1007/s42452-025-07681-z.

[26] Muqtadir, A., Li, B., Ying, Z., Songsong, C., and Kazmi, S. N. "Nowcasting the next hour of residential load using boosting ensemble machines". In: *Scientific Reports* 15 (2025), p. 7157. DOI: 10.1038/s41598-025-91767-6.

[27] Zairi, S. and Freihat, M. "Electric load forecasting using machine learning for peak demand management in smart grids". In: *Engineering, Technology & Applied Science Research* 15.3 (2025), pp. 23335–23346. DOI: 10.48084/etasr.10687.

[28] Percuku, A., Minkovska, D., and Hinov, N. "Enhancing electricity load forecasting with machine learning and deep learning". In: *Technologies* 13 (2025), p. 59. DOI: 10.3390/technologies13020059.

[29] Choubey, M., Mamun, A. A., and Khuntia, S. R. "Robust electric load forecasting through ensemble learning: A stacking approach with empirical mode decomposition and transfer learning". In: *Computers & Electrical Engineering* 118 (2025), p. 109454. DOI: 10.1016/j.compeleceng.2025.110511.

[30] Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., and Tzovaras, D. "A multivariate ensemble learning method for medium-term energy consumption forecasting". In: *Neural Computing and Applications* 35 (2023), pp. 1–25. DOI: 10.1007/s00521-023-08777-6.

[31] Wang, A., Yu, Q., Wang, J., Yu, X., Wang, Z., and Hu, Z. "Electric load forecasting based on deep ensemble learning". In: *Applied Sciences* 13.17 (2023), p. 9706. DOI: 10.3390/app13179706.

[32] Rubow, A. "Electricity demand forecasting with machine learning: A case study on the implications of data availability and economic impact for German municipal utilities". Master's thesis. MA thesis. Karlsruhe Institute of Technology, 2025. URL: https://kth.diva-portal.org/smash/get/diva2:1993933/FULLTEXT01.pdf.

[33] Satyanarayana, S. V. and Madhavi, P. "Machine learning for load forecasting in power systems". In: *E3S Web of Conferences*. Vol. 391. 2023, p. 01002. DOI: 10.1051/e3sconf/202345301008.

[34] Ibrahim, B., Rabelo, L., Gutierrez-Franco, E., and Clavijo-Buritica, N. "Machine learning for short-term load forecasting in smart grids". In: *Energies* 15.20 (2022), p. 8079. DOI: 10.3390/en15218079.

[35] Jain, A. and Gupta, S. C. "Evaluation of electrical load demand forecasting using various machine learning algorithms". In: *Frontiers in Energy Research* 12 (2024), p. 1334663. DOI: 10.3389/fenrg.2024.1408119.

[36] Bae, H.-J., Park, J.-S., Choi, J.-H., and Kwon, H.-Y. "Learning model combined with data clustering and dimensionality reduction for short-term electricity load forecasting". In: *Scientific Reports* 15.1 (2025), p. 3575. DOI: 10.1038/s41598-025-86982-0.

[37] Tavakoli, N., Zare, M., and Nasiri, S. "Clustering time series data through autoencoder-based deep learning models". In: *arXiv preprint* (2020). arXiv: 2004.07296. URL: https://arxiv.org/abs/2004.07296.

[38] Moradzadeh, A., Khashei, M., and Mohammadi-Ivatloo, B. "Short-term electricity demand forecasting via variational autoencoders and batch training-based bidirectional long short-term memory". In: *Sustainable Energy Technologies and Assessments* 52 (2022), p. 102209. DOI: 10.1016/j.seta.2022.102209.

[39] Li, S., Wang, P., and Goel, L. "A novel wavelet-based ensemble method for short-term load forecasting with hybrid optimization and feature selection". In: *IEEE Transactions on Smart Grid* 14.1 (2023), pp. 123–134. DOI: 10.1109/TPWRS.2015.2438322.

[40] Varanasi, J. and Tripathi, M. "Short-term PV power generation forecasting using a hybrid WPD-PCA-LSTM network". In: *Renewable Energy* 141 (2019), pp. 465–477. DOI: 10.3390/su17125277.

[41] Dong, Q., Huang, R., Cui, C., Towey, D., Zhou, L., Tian, J., and Wang, J. "Short-term electricity-load forecasting by deep learning: A comprehensive survey". In: *arXiv preprint* (2025). arXiv: 2408.16202v2.

[42] Yang, Z., Li, J., Liu, C., and Wang, H. "Forecasting very short-term power load with hybrid interpretable deep models". In: *Systems Science & Control Engineering* 13.1 (2025), p. 2486136. DOI: 10.1080/21642583.2025.2486136.

[43] Wen, X., Liao, J., Niu, Q., Shen, N., and Bao, Y. "Deep learning-driven hybrid model for short-term load forecasting and smart grid information management". In: *Scientific Reports* 14 (2024), p. 13720. DOI: 10.1038/s41598-024-63262-x.

[44] Zhong, B. "Deep learning integration optimization of electric energy load forecasting and market price based on the ANN–LSTM–transformer method". In: *Frontiers in Energy Research* 11 (2023), p. 1292204. DOI: 10.3389/fenrg.2023.1292204.

[45] Zare, M. S., Nikoo, M. R., Chen, M., and Gandomi, A. H. "Capturing complex electricity load patterns: A hybrid deep learning approach with proposed external-convolution attention". In: *Energy Strategy Reviews* 52 (2025), p. 101638. DOI: 10.1016/j.esr.2025.101638.

[46] Abbas, M., Che, Y., Maqsood, S., Yousaf, M. Z., Abdullah, M., Khan, W., Khalid, S., Bajaj, M., and Shabaz, M. "Kolmogorov–Arnold network optimized with self-adaptive differential evolution for short-term load forecasting". In: *Scientific Reports* 15 (2025), p. 05918. DOI: 10.1038/s41598-025-05918-w.

[47] Demir, Emrah and Gunal, Serkan. “Short-term electricity consumption forecasting with deep learning”. In: *Journal of Supercomputing* 81.10 (2025), p. 1108. DOI: 10.1007/s11227-025-07564-5.

[48] Ashtar, D., Mohammadi Ziabari, S. S., and Alsahag, A. M. M. “Hybrid forecasting for sustainable electricity demand in The Netherlands using SARIMAX, SARIMAX-LSTM, and sequence-to-sequence deep learning models”. In: *Sustainability* 17.16 (2025), p. 7192. DOI: 10.3390/su17167192.

[49] Hussain, A., Giangrande, P., Franchini, G., Fenili, L., and Messi, S. “Analyzing the effect of error estimation on random missing data patterns in mid-term electrical forecasting”. In: *Electronics* 14.7 (2025), p. 1383. DOI: 10.3390/electronics14071383.

[50] Su, Z., Zhang, J., Yang, Z., and Ma, L. “A hybrid monthly electricity demand forecasting model combining an Hodrick-Prescott filter, recurrent neural networks, and autoregressive integrated moving average”. In: *Energy and AI* 22 (2025), p. 100600. DOI: 10.1016/j.egyai.2025.100600.

[51] Sinha, A., Tayal, R., Vyas, A., Pandey, P., and Vyas, O. P. “Forecasting electricity load with hybrid scalable model based on stacked non linear residual approach”. In: *Frontiers in Energy Research* 9 (2021), p. 720406. DOI: 10.3389/fenrg.2021.720406.

[52] Elakkiya, E., Antony Raj, S., Arunkumar, Balakrishnan, Bhavyasri, Sanisetty, and Revanth Balaji, Bandaru. “Stacked hybrid model for load forecasting: Integrating transformers, ANN, and fuzzy logic”. In: *Scientific Reports* 15 (2025), p. 19688. DOI: 10.1038/s41598-025-04210-1.

[53] Khwaja, A. S., Anpalagan, A., Naeem, M., and Venkatesh, B. “Joint bagged-boosted artificial neural networks: Using ensemble machine learning to improve short-term electricity load forecasting”. In: *Electric Power Systems Research* 179 (2020), p. 106080. DOI: 10.1016/j.epsr.2019.106080.

[54] Srivastava, A. K., Singh, D., and Pandey, A. S. “Short term load forecasting using regression trees: Random Forest, Bagging and M5P”. In: *International Journal of Advanced Trends in Computer Science and Engineering* 9.2 (2020), pp. 1898–1902. DOI: 10.30534/ijatcse/2020/152922020.

[55] Qinghe, Z., Wen, X., Boyan, H., Jong, W., and Junlong, F. “Optimised extreme gradient boosting model for short term electric load demand forecasting of regional grid system”. In: *Scientific Reports* 12 (2022), p. 19282. DOI: 10.1038/s41598-022-22024-3.

[56] Semmelmann, L., Henni, S., and Weinhardt, C. "Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data". In: *Energy Informatics* 5.Suppl 1 (2022), p. 24. DOI: 10.1186/s42162-022-00212-9.

[57] Wang, C., Bäck, T., Hoos, H. H., Baratchi, M., Limmer, S., and Olhofer, M. "Automated machine learning for short-term electric load forecasting". In: *Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2019, pp. 313–320. DOI: 10.1109/SSCI44817.2019.9002839.

[58] Praca, I., Pinto, T., and Vale, Z. "Energy consumption forecasting using ensemble learning algorithms". In: *Advances in Intelligent Systems and Computing*. Springer, 2020, pp. 1–9. DOI: 10.1007/978-3-030-23946-6_1.

[59] Nespoli, L. and Medici, V. "Global forecasting models for residential load flexibility and grid optimization". In: *Frontiers in Energy Research* 13 (2025), p. 1547617. DOI: 10.3389/fenrg.2025.1547617.

[60] Hale, E., Stoll, B., and Mai, T. *Capturing the impact of storage and other flexible technologies on electric system planning [NREL/TP-6A20-65726]*. Tech. rep. National Renewable Energy Laboratory, 2016. URL: https://www.nrel.gov/docs/fy16osti/65726.pdf.

[61] Gumerman, E., Chan, P., Lesieutre, B., Marnay, C., and Wang, J. *Modeling interregional transmission congestion in the National Energy Modeling System [LBNL-59076]*. Tech. rep. Lawrence Berkeley National Laboratory, 2006. URL: https://eta-publications.lbl.gov/sites/default/files/report-lbnl-59076.pdf.

[62] Pierpont, B., Nelson, D., Goggins, A., and Posner, D. *Flexibility: The path to low-carbon, low-cost electricity grids [CPI Report]*. Tech. rep. Climate Policy Initiative, 2017. URL: https://www.climatepolicyinitiative.org/wp-content/uploads/2017/04/CPI-Flexibility-the-path-to-low-carbon-low-cost-grids-April-2017.pdf.

[63] Kiguchi, Y. "Residential demand response using electricity smart meter data". Doctoral dissertation. PhD thesis. University of Cambridge, 2021. URL: https://core.ac.uk/download/484016990.pdf.

[64] Dogoulis, P. "Mid-term load forecasting from smart meter data". Master's thesis. MA thesis. Aristotle University of Thessaloniki, 2022. URL: https://ikee.lib.auth.gr/record/343438/files/GRI-2022-37404.pdf.

[65] Javedian, S. "Application of machine learning and deep learning methods for load prediction in institutional buildings". Master's thesis. MA thesis. Concordia University, 2020. URL: https://www.concordia.ca/content/dam/research/chairs/smart-cities/docs/thesis-sahar-javedian.pdf.

[66] Lin, X., Zamora, R., Baguley, C. A., and Srivastava, A. K. "A hybrid short-term load forecasting approach for individual residential customer". In: *IEEE Transactions on Power Delivery* (2022). DOI: 10.1109/TPWRD.2022.3178822.

[67] Ruiz-Abellón, M. C., Fernández-Jiménez, L. A., Guillamón, A., Falces, A., García-Garre, A., and Gabaldón, A. "Integration of demand response and short-term forecasting for the management of prosumers' demand and generation". In: *Energies* 13.1 (2019), p. 11. DOI: 10.3390/en13010011.

[68] Gupta, T., Bhatia, V., Singh, R., and Sharma, S. "An ensemble-based enhanced short and medium term load forecasting model using machine learning techniques". In: *PLoS ONE* 20.7 (2025), e0307890. DOI: 10.1371/journal.pone.0307890.

[69] Hussain, A., Franchini, G., Akram, M., Ehtsham, M., Hashim, M., Fenili, L., Messi, S., and Giangrande, P. "Hybrid ML/DL Approach to Optimize Mid-Term Electrical Load Forecasting for Smart Buildings". In: *Applied Sciences* 15.18 (2025), p. 10066. DOI: 10.3390/app151810066.

[70] Chandler, W. S., Hoell, J. M., Westberg, D., Zhang, T., and Stackhouse, P. W. *NASA Prediction of Worldwide Energy Resource High Resolution Meteorology Data for Sustainable Building Design*. 2013. URL: https://ntrs.nasa.gov/api/citations/20130013357/downloads/20130013357.pdf.

[71] Espino, D., Dietrich, K., González-Ávila, I., and Pathak, S. "A geographic information system-based procedure to assess the coupled implementation of photovoltaic panels and green roofs for improved energy efficiency". In: *Elsevier BV*. 2024, pp. 453–474. DOI: 10.1016/b978-0-443-13613-9.00007-6.

[72] Fu, J., Mandolfo, M., Wang, J., Li, Y., Zhang, D., and Noci, G. "Day-Ahead Tariff Prediction Method for Power Trading Market Based on IQR-RANSAC". In: 2024, pp. 223–227. DOI: 10.1109/acfpe63443.2024.10801075.

[73] Passalis, N. and Tefas, A. "Global Adaptive Input Normalization for Short-Term Electric Load Forecasting". In: *IEEE Symposium Series on Computational Intelligence*. 2020, pp. 1–8. DOI: 10.1109/SSCI47803.2020.9308553.

[74] Rakesh, S. and Pavankumar, T. "A meta-learning framework with temporal feature integration for electricity load forecasting". In: *Energy Informatics* 8.1 (2025). DOI: 10.1186/s42162-025-00572-y.

[75] Verdonck, T., Baesens, B., Óskarsdóttir, M., and Broucke, S. van den. "Special issue on feature engineering editorial". In: *Machine Learning* 113.7 (July 2024), pp. 3917–3928. DOI: 10.1007/s10994-021-06042-2.

[76] Zhang, J., Xu, Z., and Wei, Z. "Absolute logarithmic calibration for correlation coefficient with multiplicative distortion". In: *Communications in Statistics - Simulation and Computation* 52.2 (2020), pp. 482–505. DOI: 10.1080/03610918.2020.1859541.

[77] Osman, M. S., Abu-Mahfouz, A. M., and Page, P. R. "A Survey on Data Imputation Techniques: Water Distribution System as a Use Case". In: *IEEE Access* 6 (2018), pp. 63279–63291. DOI: 10.1109/ACCESS.2018.2876098.

[78] David, S., Azariya, S., Mohanraj, V., Emilyn, J. J., and Jothi, G. "A Comparison of Missing Data Handling Techniques". In: *ICTACT Journal on Soft Computing* 11 (2021), pp. 2433–2437. DOI: 10.21917/ijsc.2021.0347.

[79] Farewell, D., Daniel, R., and Seaman, S. *Missing at Random: A Stochastic Process Perspective*. 2018. arXiv: 1801.06739. URL: https://arxiv.org/abs/1801.06739.

[80] Carreras, G., Miccinesi, G., Wilcock, A., Preston, N., Nieboer, D., Deliens, L., Groenvold, M., Lunder, U., Heide, A. van der, Baccini, M., et al. "Missing Not at Random in End of Life Care Studies: Multiple Imputation and Sensitivity Analysis on Data from the ACTION Study". In: *BMC Medical Research Methodology* 21 (2021), p. 13. DOI: 10.1186/s12874-020-01220-4.

[81] Lee, K., Lim, H., Hwang, J., and Lee, D. "Evaluating Missing Data Handling Methods for Developing Building Energy Benchmarking Models". In: *Energy* 308 (2024), p. 132979. DOI: 10.1016/j.energy.2024.132979.

[82] Ahn, Hyun, Sun, Kyunghee, and Kim, Kwanghoon P. "Comparison of Missing Data Imputation Methods in Time Series Forecasting". In: *Computers, Materials & Continua* 70.1 (2022), pp. 767–779. DOI: 10.32604/cmc.2022.019369.

[83] Joel, L. O., Doorsamy, W., and Paul, B. S. "A comparative study of imputation techniques for missing values in healthcare diagnostic datasets". In: *International Journal of Data Science and Analytics* 20.7 (2025), pp. 6357–6373. DOI: 10.1007/s41060-025-00825-9.

[84] Alwateer, M., Atlam, E.-S., Abd El-Raouf, M. M., Ghoneim, O. A., and Gad, I. "Missing Data Imputation: A Comprehensive Review". In: *Journal of Computer and Communications* 12.11 (2024), pp. 53–75. DOI: 10.4236/jcc.2024.1211004.

[85] Buuren, Stef van. "Multiple Imputation of Multivariate Data by Chained Equations". In: *Journal of Statistical Software* 45.3 (2011), pp. 1–67. DOI: 10.18637/jss.v045.i03.

[86] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O. "A survey on missing data in machine learning". In: *Journal of Big Data* 8.1 (2021), p. 140. DOI: 10.1186/s40537-021-00516-9.

[87] Tang, F. and Ishwaran, H. "Random Forest missing data algorithms". In: *Statistical Analysis and Data Mining* 10.6 (2017), pp. 363–377. DOI: 10.1002/sam.11348.

[88] Smola, Alex J. and Schölkopf, Bernhard. "A Tutorial on Support Vector Regression". In: *Statistics and Computing* 14.3 (2004), pp. 199–222. DOI: 10.1023/B:STCO.0000035301.49549.88.

[89] Klusowski, Jason M. and Tian, Peter M. "Large Scale Prediction with Decision Trees". In: *Journal of the American Statistical Association* 119.545 (2022), pp. 525–537. DOI: 10.1080/01621459.2022.2126782.

[90] Zhu, J., Dong, H., Zheng, W., Li, S., Huang, Y., and Xi, L. "Review and prospect of data-driven techniques for load forecasting in integrated energy systems". In: *Applied Energy* 321 (2022), p. 119269. DOI: 10.1016/j.apenergy.2022.119269.

[91] Lipton, Z. "A Critical Review of Recurrent Neural Networks for Sequence Learning". In: (2015). URL: https://doi.org/10.48550/arXiv.1506.00019.

[92] Mena-Oreja, J. and Gozalvez, J. "A Comprehensive Evaluation of Deep Learning-Based Techniques for Traffic Prediction". In: *IEEE Access* 8 (2020), pp. 91188–91212. DOI: 10.1109/ACCESS.2020.2994415.

[93] Cho, K., Merriënboer, B. van, Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. "Learning phrase representations using RNN encoder–decoder for statistical machine translation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1724–1734. URL: https://arxiv.org/abs/1406.1078.

[94] Dabrowski, J. J., Zhang, Y., and Rahman, A. "ForecastNet: A Time-Variant Deep Feed-Forward Neural Network Architecture for Multi-step-Ahead Time-Series Forecasting". In: *Neural Information Processing. ICONIP 2020. Lecture Notes in Computer Science*. Vol. 12534. 2020. DOI: 10.1007/978-3-030-63836-8_48.

[95] Bergstra, J. and Bengio, Y. "Random Search for Hyper-Parameter Optimization". In: *Journal of Machine Learning Research* 13 (2012), pp. 281–305. URL: http://jmlr.org/papers/v13/bergstra12a.html.

[96] Hakyemez, T. C. and Adar, O. "Testing the Efficacy of Hyperparameter Optimization Algorithms in Short-Term Load Forecasting". In: *arXiv* (2024). eprint: 2410.15047. URL: http://arxiv.org/abs/2410.15047.

[97] Bergmeir, C., Hyndman, R. J., and Benítez, J. M. "On the use of cross-validation for time series predictor evaluation". In: *Information Sciences* 191 (2018), pp. 192–213. DOI: https://doi.org/10.1016/j.ins.2011.12.028.

[98] Shirzadi, N., Nizami, A., Khazen, M., and Nik-Bakht, M. "Medium-Term Regional Electricity Load Forecasting through Machine Learning and Deep Learning". In: *Designs* 5 (2021), p. 27. URL: https://www.mdpi.com/2411-9660/5/2/27.

[99] Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K., and Alahmad, M. "Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods". In: *Energies* 13 (2020), p. 4231. URL: https://doi.org/10.3390/en13164231.

[100] Amber, K. P., Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., and Khan, M. S. "Intelligent Techniques for Forecasting Electricity Consumption of Buildings". In: *Energy* 157 (2018). URL: https://doi.org/10.1016/j.energy.2018.05.155.

[101] Chicco, D., Warrens, M., and Jurman, G. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation". In: *PeerJ Computer Science* 7 (2021), e623. DOI: 10.7717/peerj-cs.623.

[102] Misiurek, K., Olkuski, T., and Zyśk, J. "Review of methods and models for forecasting electricity consumption". In: *Energies* 18.15 (2025), p. 4032. DOI: 10.3390/en18154032.

[103] Rodrigues, F., Cardeira, C., Calado, J. M. F., and Melicio, R. "Short-term load forecasting of electricity demand for the residential sector based on modelling techniques: A systematic review". In: *Energies* 16.10 (2023), p. 4098. DOI: 10.3390/en16104098.

[104] IPCC. *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2: Energy, Chapter 2*. Provides standardized methodologies for calculating greenhouse gas emissions from energy use, including fuel combustion and electricity consumption. Geneva, Switzerland: Intergovernmental Panel on Climate Change, 2006. URL: https://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html.

Borsa di dottorato cofinanziata con risorse dell'Unione europea-*NextGeneration EU*
Piano Nazionale di Ripresa e Resilienza Missione 4, componente 2 "Dalla Ricerca all'Impresa" - Investimento 3.3 "Introduzione di dottorati innovativi che rispondono ai fabbisogni di innovazione delle imprese e promuovono l'assunzione dei ricercatori dalle imprese" – CUP F83C22001100004

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | UNIVERSITÀ DEGLI STUDI DI BERGAMO