Advances in Artificial Intelligence for Energy Forecasting and Performance Management in Buildings

A Thesis Submitted to Cardiff University in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

by **Nasser Alkhatani**

PhD Thesis June 2025

Cardiff University

Abstract

Accurate energy forecasting is essential for intelligent building management, supporting operational optimisation, strategic planning, and demand-side flexibility. However, existing forecasting methods often struggle to remain accurate across multiple time horizons and to generalise across different building types with limited data. This thesis addresses these challenges by developing a modular modelling framework that advances both multi-horizon forecasting and cross-building adaptability.

The first contribution is a hybrid forecasting model (SVR → XGBoost → LSTM) designed to deliver stable prediction performance across four horizons: 24 hours, one week, one month, and one year. The hybrid design leverages the complementary strengths of its components SVR for noise reduction, XGBoost for nonlinear feature learning, and LSTM for long-range temporal modelling resulting in improved robustness and generalisation compared with single-model approaches.

The second contribution introduces a deep hybrid model (CNN → GRU → LSTM) within a transfer learning framework. Pretrained on multi-building datasets and fine-tuned using limited data from new buildings, this approach enhances cross-domain adaptability while reducing training time and data requirements, demonstrating the practical value of transfer learning for scalable energy forecasting.

A third contribution integrates statistical peak detection to support the identification of high-demand events, enabling forecasting outputs to inform grid-interactive building operations. Rigorous evaluation including multi-metric assessment, residual diagnostics, ablation testing, and statistical significance analysis confirms the reliability and robustness of the proposed models.

Overall, the thesis provides methodological and empirical advances that strengthen data-driven building energy management. The results show that hybridisation and transfer learning, when carefully designed, can enhance accuracy, stability, and generalisation, offering a scalable pathway toward more efficient and sustainable smart building operations.

Contents

List of Figures

List of Tables

List of Publications

Nasser, A., Petri, I., & Rana, O., Parashar, M., Edge Learning for Energy-Aware Resource Management, Proceedings of the IEEE SERVICES 2025 Symposium on Sustainability and Resilience across the Computing Continuum, Helsinki, July, 2025.

Acknowledgements

First and foremost, I express my deepest gratitude to Allah Almighty for granting me the strength, perseverance, and guidance to complete my PhD journey at Cardiff University.

I would like to extend my sincere appreciation to all those who supported me throughout this academic endeavor. In particular, I am profoundly grateful to **Dr. Ioan Petri** for his exceptional mentorship, unwavering support, and for playing a pivotal role in shaping and refining my research and academic skills.

My heartfelt thanks go to my family for their constant encouragement and love. I am especially indebted to my beloved mother, my brothers and sister, and my dear wife, Dr. Rawiyah Alkahtani, whose support has been invaluable. I also extend special thanks to her family, particularly her father, for their kindness and encouragement.

I am also thankful to my friends, both in Cardiff and beyond, whose companionship and support helped me through challenging times.

Finally, I am truly grateful to my country for its generous support and for providing me with the opportunity to pursue my doctoral studies. This journey would not have been possible without its continuous encouragement and trust.

Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
ARIMA	Autoregressive Integrated Moving Average
BMS	Building Management System
CNN	Convolutional Neural Network
DL	Deep Learning
ETS	Exponential Smoothing
F1-score	Harmonic Mean of Precision and Recall
GENOME	Global Energy Modelling of Energy Networks Dataset
GRU	Gated Recurrent Unit
HVAC	Heating, Ventilation, and Air Conditioning
IQR	Interquartile Range
IoT	Internet of Things
KNN	k-Nearest Neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory

Abbreviation	Full Term
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Squared Error
Node-RED	Flow-Based Development Tool for Edge Deployment
R^2	Coefficient of Determination
RAM	Random Access Memory
RF	Random Forest
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SHAP	SHapley Additive exPlanations
SVR	Support Vector Regression
TFT	Temporal Fusion Transformer
XGBoost	Extreme Gradient Boosting

Chapter 1: Introduction

1.1 Background and Motivation

The built environment is a significant contributor to global energy consumption and greenhouse gas emissions, accounting for approximately 39% of process-related emissions [1]. Achieving net-zero energy buildings is crucial for climate neutrality and is feasible across various building types and climates using existing technologies [1]. Recent research has focused on key areas such as energy demand, supply, storage, and integration in urban environments [2]. Emerging technologies like uncertainty-based design, renewable integration, thermal energy storage, and heat pumps show great potential for reducing building emissions and costs [3]. However, current research is predominantly limited to single building case studies, and there is a need to expand assessments to industry-wide scales and consider embodied emissions in materials [4]. Collaborative efforts among stakeholders are essential to facilitate the integration and large-scale deployment of these technologies [3].

Office buildings offer particularly promising conditions for intelligent energy management due to their predictable usage patterns and the maturity of their infrastructure. Research in this domain has demonstrated how forecasting algorithms can enhance the efficiency of smart office operations. For example, an energy management infrastructure was proposed in [5] that leverages forecasting to predict both energy generation and consumption, achieving performance levels within 10% of optimal scheduling. A reinforcement learning-based Heating, Ventilation, and Air Conditioning (HVAC) control system, described in [6] , achieved 93.8% accuracy in predicting user presence while maintaining comfort thresholds. Further improvements in load forecasting have been reported in [7] , where the integration of diverse sensor inputs and contextual information such as the day of the week proved effective in boosting prediction accuracy. Together, these studies [5], [6] underscore the value of combining real-time data, machine learning techniques, and predictive control strategies to reduce energy consumption and operational costs, while ensuring occupant comfort in office environments. These efforts are increasingly aligned with the broader goal of performance management in the buildings where operational systems are not only automated but actively optimised to enhance energy efficiency, thermal comfort, and overall performance. Forecasting models, when integrated with smart control systems, form the intelligence layer that enables proactive building performance tuning, anomaly detection, and demand-side flexibility [1], [3].

Smart buildings increasingly rely on artificial intelligence (AI) to support more deliberate, data-driven forms of energy and performance management. Rather than functioning purely as automated systems, modern buildings draw on dense sensor networks, historical data streams, and contextual information to anticipate conditions and respond dynamically to changing operational demands [8]. AI-based approaches combine machine learning, neural networks, and pattern-recognition techniques to analyse occupancy behaviours, seasonal variations, weather influences, and internal load interactions, enabling more coordinated and efficient control strategies across HVAC, lighting, and other major subsystems [9]. These methods have been shown to improve forecasting accuracy, reduce unnecessary energy use, and enhance occupant comfort by informing control actions such as pre-conditioning, ventilation adjustments, and peak-load mitigation [10]. Beyond individual systems, AI contributes to broader operational goals by supporting demand response participation, renewable energy integration, and grid-interactive behaviours that enhance overall resilience [8], [9]. The value of these approaches is further amplified when combined with predictive modelling, as forecasting enables building controllers to select actions not only based on current demand but also on anticipated future patterns, thereby reducing overreaction to short-term fluctuations and improving long-term operational planning. While challenges remain, including data privacy, interoperability constraints, and the need for specialized expertise recent studies indicate a clear trajectory toward intelligent, performance-oriented building operation supported by analytics, contextual awareness, and advanced decision-making algorithms [10]. Collectively, these developments illustrate how AI and forecasting are reshaping the role of buildings from passive energy consumers into adaptive, responsive assets within modern energy systems.

To build upon this context, the following subsections outline the key concepts that shape the direction of the study. They introduce the role of forecasting in building energy management, review the limitations of traditional methods, examine the emergence of machine learning and deep learning techniques, and highlight the challenges of generalisation and scalability that motivate the research objectives defined later in this chapter.

1.1.1 The Role of Forecasting in Energy Management

Accurate electricity consumption forecasting plays a central role in effective energy management for buildings. From optimising HVAC schedules to enabling dynamic pricing and demand response strategies, forecasting provides the intelligence layer that transforms raw sensor data into actionable decisions. A wide range of forecasting methods has been developed, spanning from traditional statistical models to more recent machine learning (ML) and deep learning (DL) approaches [11]. These models are used across both short- and long-term horizons, each with distinct implications. Short-term predictions ranging from hours to days are essential for grid operators and smart controllers to balance supply and demand in real time [12]. In contrast, long-term forecasts (extending over months or years) inform infrastructure investments and energy procurement strategies [13].

Recent scenario modelling by Ding et al. [14] underscores the scale of savings that accurate, AI-driven forecasting can unlock for commercial buildings. As illustrated in (Figure 1.1), broad deployment of AI-enabled controls in medium-sized U.S. office buildings could reduce final-site energy use by approximately 40% compared to a business-as-usual trajectory by 2050. These projections also highlight the potential for substantial carbon reductions, with CO_2 emissions estimated to drop by nearly 90% when AI forecasting is coupled with supportive energy-policy measures. Together, these findings reinforce the chapter's central argument: robust forecasting acts as the intelligence layer that transforms raw data into meaningful energy and carbon savings.

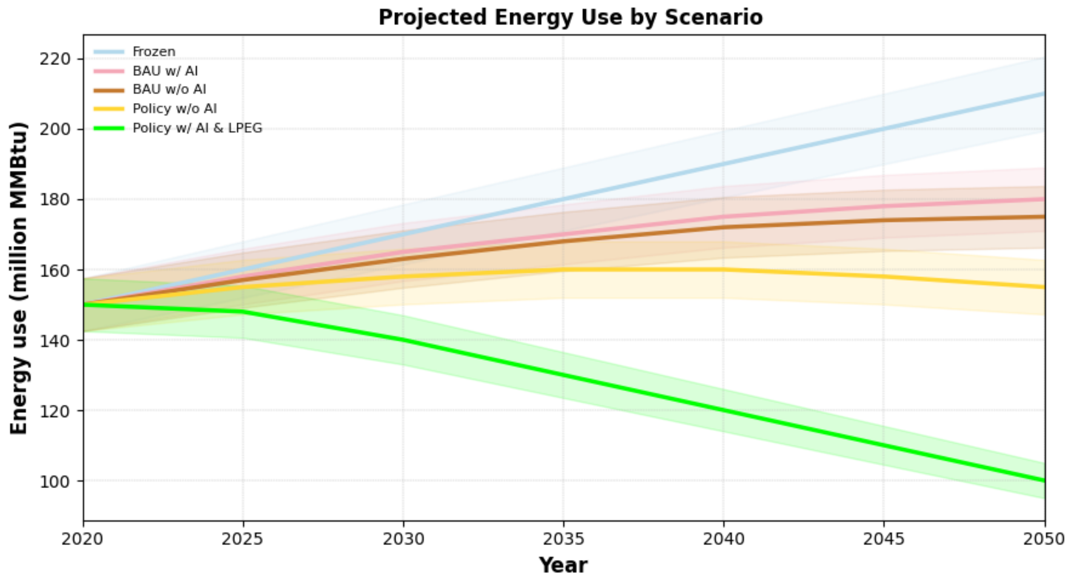

Figure 11: Projected final-site energy use for medium U.S. office buildings under six AI-adoption/policy scenarios (2020-2050). Data trends aligned with Ding et al. [14].

Among these methods, ML-based models such as XGBoost, LSTM, and Seasonal Autoregressive Integrated Moving Average (SARIMA) have demonstrated considerable accuracy, particularly when tuned to context-specific features and time scales [12], [15]. Boosting models, in particular, are often favored for their ability to model nonlinear patterns and deliver competitive performance. However, model selection is not straightforward. It is influenced by factors such as input variable types, data granularity, forecast horizon, and the evaluation metrics applied [13]. As the forecast window lengthens, predictive accuracy tends to decline due to increasing uncertainty and diminishing data relevance [12].

An equally important component in this context is Demand Side Management (DSM), a strategic mechanism that adjusts consumption patterns in response to variable supply conditions, particularly from renewable sources. DSM techniques are increasingly paired with forecasting algorithms to align demand with solar or wind availability, especially in residential photovoltaic (PV) systems [16], [17]. This synchronization not only flattens peak demand curves but also reduces dependence on fossil-fuel-based backup systems [18]. For example, the use of rooftop solar panels and local battery storage has been shown to reduce summer peak loads while offering economic and environmental gains [19]. Moreover, when combined with forecasting, these systems can mitigate the risks of prediction errors; studies show that integrating local storage can reduce forecasting-related uncertainty in grid exchanges from nearly 40% down to just 2% [17].

The consequences of inaccurate forecasting can be significant. Poor forecasts may result in overproduction, underutilization, or grid instability, each carrying operational and financial penalties [20]. In sensitive environments like data centers, errors can disrupt demand response schedules or reduce responsiveness, while in microgrids, they can undermine optimal scheduling benefits. Industrial facilities, too, experience scheduling inefficiencies when day-ahead forecasts miss demand spikes, especially with price-sensitive customers and distributed resources [21]. To address these challenges, researchers have proposed more adaptive frameworks. One promising approach involves game-theoretic model predictive control for DSM, which dynamically updates decisions using real-time feedback and has been shown to outperform traditional methods when forecast error exceeds 10% [22].

1.1.2 Limitations of Traditional Forecasting Methods

Time-series forecasting has long relied on classical statistical models such as ARIMA, SARIMA, and Exponential Smoothing (ES), valued for their simplicity, transparency, and effectiveness in handling linear relationships and stationary behaviour [23]. ARIMA is widely recognized for capturing trends and autocorrelations in time-dependent data. These models have been applied with success in forecasting electricity demand and are often benchmarked against regression-based alternatives. However, as energy systems become increasingly complex, nonlinear, and multivariate, the limitations of these traditional models become more apparent, especially in dynamic environments like smart buildings [23].

One key shortcoming is the inability of linear models to capture nonlinear interactions among variables, which are frequently observed in building energy consumption. Real-world energy data is shaped by diverse and interrelated factors like weather, occupancy, operational schedules, and equipment behaviour, many of which interact in nonlinear ways. In addition, traditional models often experience error accumulation over longer forecasting horizons, making them less reliable for medium- and long-term predictions [24], [25]. In response, researchers have begun exploring machine learning techniques that are better suited to uncovering hidden patterns in high-dimensional spaces. For instance, decision trees and conditional inference trees have shown promise in identifying complex relationships overlooked by linear approaches [24], [26].

The emergence of deep learning models, particularly Long Short-Term Memory (LSTM) networks, has further expanded the forecasting toolkit. LSTM model excels at capturing long-range temporal dependencies and has been shown to outperform traditional methods in modelling electricity consumption [26]. In comparative studies, LSTM and Random Forest-based models have consistently demonstrated stronger performance, especially when dealing with noisy or high-frequency sensor data [26].

Another ongoing debate in the field concerns the trade-off between univariate and multivariate models. While multivariate models tend to offer richer feature spaces and potentially improved accuracy, their benefits are not always consistent across datasets or prediction horizons [27]. Some researchers argue that in short-term scenarios, simpler univariate models can be just as effective, especially when temperature or weather data introduce unnecessary complexity [28]. Hybrid models that incorporate Convolutional Neural Networks (CNNs) have been used to bridge these two paradigms, performing well in both univariate and multivariate contexts [29]. Interestingly, evidence suggests that multivariate models may have an edge in early steps, while univariate models sometimes excel in longer-term predictions [29]. These findings suggest the need for adaptive modelling strategies that draw from both approaches depending on the use case.

Another limitation of traditional building energy models (BEMs) is their poor adaptability to changing operational conditions. Physics-based models, though grounded in domain expertise, often require extensive reconfiguration and manual updates, rendering them impractical for everyday users [30]. To address this, researchers have proposed combining AI techniques with traditional models. One example involves training artificial neural networks (ANNs) on simulated outputs from physics-based models, thus creating adaptable surrogates [30]. Other work has explored the integration of ontologies and transfer learning to enable fault detection and diagnostics in smart buildings [31] . Even so-called “black-box” neural network models have been shown to rival or complement “white-box” physics-based models in dynamic environments [32] . These advances highlight the growing need for models that are not only accurate, but also flexible, adaptive, and minimally burdensome to implement.

Table 1.1 summarizes the typical strengths and limitations of popular AI models used in energy forecasting. While many models perform well for narrow use cases, they often struggle to generalise across time horizons or building types. The hybrid approach introduced in this thesis addresses these gaps.

Table 1.1: Comparison of AI Forecasting Models in Smart Building Applications

Model Type	Strength	Limitation
ANN / MLP	Good for static patterns	Fails to capture temporal dependencies
LSTM / GRU	Learns short-term time sequences	Degrades over long-term horizons
XGBoost	Strong for tabular data	Weak in time-based learning without lag features
CNN	Good for spatial/structural learning	Limited sequence awareness
Hybrid Models (proposed)	Balanced, modular design	Higher complexity but improved accuracy and flexibility

While (Table 1.1) compares the strengths and weaknesses of competing approaches, it is equally important to consider how those limitations play out over different forecasting horizons. Conventional models ARIMA, LSTM, XGBoost, and similar techniques show a clear decline in accuracy as the prediction window lengthens. By contrast, the hybrid model introduced in this research (SVR → XGBoost → LSTM) preserves high accuracy from 24-hour to annual horizons, underscoring its adaptability and robustness. This pattern of resilience is echoed in the detailed experimental results reported in Chapters 5 and 6.

1.1.3 Emergence of Machine Learning and Deep Learning Approaches

The convergence of machine learning (ML) and deep learning (DL) has reshaped the landscape of artificial intelligence, unlocking new possibilities in pattern recognition, forecasting, and decision support across diverse fields [33]. While ML methods are well-established in classification, clustering, and regression tasks, DL introduces deeper, hierarchical learning structures capable of extracting features directly from raw data. Together, they form a complementary toolkit ML providing interpretability and speed, while DL contributes complexity handling and representation power [34], [35].

This synergy is particularly valuable in energy forecasting, where data can be both structured (e.g., sensor streams) and unstructured (e.g., contextual information). While ML techniques like Support Vector Regression (SVR) have shown excellent performance in modelling nonlinear relationships [36], DL models such as LSTM and GRU have been increasingly favored for their ability to handle sequential patterns and long-range dependencies [37], [38]. In forecasting tasks involving complex time series, these models often surpass the accuracy of traditional statistical approaches such as ARIMA [39], [40].

SVR, for instance, maps data into high-dimensional spaces using kernel functions to reduce generalisation error, making it particularly effective in cases with nonlinear relationships and noisy inputs [36], [41]. In recent comparative studies, SVR has outperformed XGBoost in certain streamflow forecasting tasks, particularly when used within decomposition-based hybrid models [42]. To further enhance SVR's flexibility and address kernel selection challenges, researchers have introduced algorithms like Multiple Additive Regression Kernels (MARK), which construct heterogeneous kernel matrices on-the-fly using gradient boosting and regularization to minimize overfitting [43].

At the deep-learning end of the spectrum, LSTM and GRU networks have gained considerable traction. Both are types of recurrent neural networks (RNNs) specifically designed to capture temporal dependencies in time-series data. While LSTM models have demonstrated robust performance in household electricity consumption forecasting, GRUs have been more effective in electricity price prediction under certain conditions [40]. Some models even combine multi-sequence inputs such as temperature, occupancy, and time-of-day to improve forecasting precision across multiple timescales [37].

These developments reflect a broader shift toward hybrid approaches that combine the strengths of ML and DL. Rather than choosing between accuracy and interpretability, modern research increasingly seeks to integrate these paradigms in a layered or sequential manner. This hybridization not only enhances model performance, but also opens the door to more adaptable, modular, and deployable systems capable of addressing the nuanced challenges of real-world forecasting.

Recent research is increasingly pointing to transfer learning and edge processing as key enablers for more scalable and responsive forecasting in smart buildings. Transfer learning makes it possible to adapt forecasting models from one building to another without starting from scratch, which cuts down on data requirements and training time [44]. It has shown strong results in areas like energy consumption forecasting, HVAC control, and occupancy prediction. For example, Chen et al. [45] demonstrated that combining transfer learning with deep neural networks significantly improved energy and cooling demand predictions across multiple sites. Similarly, Zhang et al. [46] applied transfer learning to thermal modelling and achieved high accuracy with minimal data from new buildings.

Transfer learning has also been shown to enhance the adaptability of forecasting models across diverse buildings. By enabling model reuse rather than retraining from scratch, it supports faster inference, reduced computational demands, and improved performance in data-limited environments [18]. Together, these advances contribute to forecasting systems that are both robust and scalable in real-world applications.

1.1.4 Challenges in Generalisation and Transferability

As building energy forecasting continues to evolve, several complex challenges have emerged particularly in the areas of generalisation, scalability, and long-term applicability. While short-term forecasting has seen significant progress, extending predictive accuracy across multiple horizons (e.g., daily, weekly, annual) remains difficult. Recent advances, such as the Temporal Fusion Transformer (TFT), have shown potential in producing both point and probabilistic forecasts for multiscale applications [47]. However, despite these innovations, certain critical aspects like modelling occupant behaviour and accounting for seasonal and operational drift are still underexplored [48].

At the same time, the ability to generalise across buildings remains a persistent challenge. Many current forecasting models are tightly coupled with specific data from the buildings on which they were trained. As a result, their performance often deteriorates when applied to buildings with different operational patterns, usage types, or climatic conditions. To tackle this, researchers have begun exploring transfer learning as a technique to bridge data scarcity and improve model adaptability across domains.

One promising direction involves moving from single output to multi-output forecasting models. Traditional approaches often train separate models for each forecasting horizon or variable, resulting in fragmented performance and limited scalability. In contrast, data-driven multi-output models such as Bayesian adaptive splines or deep neural networks can capture correlations across time horizons and generate more coherent predictions [49], [50]. These models represent a step toward holistic forecasting systems that can be learned not only from historical patterns, but also from cross-temporal relationships within the data.

Transfer learning has shown significant promise in the building energy domain. By reusing knowledge from models trained on large, diverse datasets, researchers can adapt forecasting models to new target buildings with limited retraining [51]. Several studies have investigated transferability using building datasets spanning hundreds of locations, with positive results. For example, new algorithms have been developed to facilitate model adaptation across building categories while hierarchical clustering has been combined with transfer learning to optimise fleet-wide forecasts using minimal training data [44], [52]. Transformer-based models have also been successfully adapted through fine-tuning, offering better generalisation than traditional DL models [53].

Parallel to these methodological efforts, system-level limitations also come into play. Cloud-centric model, while powerful, is increasingly viewed as insufficient for smart grid and Internet of Things (IoT) applications. Concerns over latency, bandwidth, and data privacy have led to growing interest in edge computing as a decentralization strategy [54]. By processing data locally on edge devices, it becomes possible to reduce network load, accelerate decision-making, and maintain user privacy.

Emerging research proposes hybrid approaches that combine federated learning, on-device inference, and autonomous control to enable scalable and secure deployments at the edge [55], [56]. These systems support intelligent forecasting while minimizing dependence on cloud infrastructure. For instance, recent models have shown how model transfer can occur between cloud and edge in Kubernetes-based systems, enabling efficient inference even under network constraints [57]. Together, these developments point toward a model that is not only accurate and scalable, but also resilient and privacy-conscious features that are increasingly critical in real-world building environments.

1.1.5 Toward Scalable and Transferable Forecasting

Building upon the challenges identified above, this research advances a forecasting approach that integrates hybrid deep learning and transfer learning to provide scalable, reliable, and accurate predictions across different buildings and forecasting horizons. The overarching aim is to develop models that remain robust when applied to diverse operational contexts, while being adaptable enough to support real-world adoption.

A central element of this work is the use of transfer learning as a mechanism for model generalisation. Rather than developing a new model from scratch for each individual building a process that can be both data-intensive and computationally costly a pretrained base model is first developed using a large and heterogeneous dataset such as GENOME. This base model is then fine-tuned using a smaller dataset from the target building. Through this process, the model can maintain strong forecasting performance even in settings where only limited historical data are available. This supports wider scalability across building portfolios, campuses, and other large-scale built environments.

To effectively capture the complex and multiscale patterns present in building energy data, this thesis employs a hybrid deep learning model composed of Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks. Each component contributes uniquely: CNNs extract structural and localised patterns among input features; GRUs model short-term temporal variations; and LSTMs capture long-range dependencies and cyclical behaviours. Together, these elements enable the model to respond to highly dynamic consumption patterns without being overwhelmed by data complexity.

The effectiveness of this modelling approach is further enhanced by techniques such as dropout regularisation, early stopping, and adaptive learning rates. These strategies help reduce overfitting, promote better generalisation, and ensure that the models can adapt to variations in building operation, occupant behaviour, and external conditions such as weather.

Overall, this integrated methodology reflects a shift from isolated, single-building model development toward a modular, transferable, and scalable forecasting strategy. While the present work focuses on model design, transfer learning, and validation, opportunities for broader deployment, particularly those related to real-time decision support, are reserved for future research.

1.1.6 Research Motivation and Impact

This research is motivated by the need to apply advanced AI methodologies to the practical challenges of smart-building energy management. While recent advances in forecasting algorithms have demonstrated remarkable accuracy in controlled settings, their real-world application remains constrained by scalability, adaptability, and deployment challenges issues discussed in Section 2.6 (multi-horizon limitations), Section 2.7 (cross-building generalisation challenges), and Section 2.5 (limitations in validation practices). The goal of this work is to develop a unified modelling approach that addresses these gaps and delivers measurable impact across technical and operational dimensions. More importantly, the proposed models contribute directly to performance management by supporting improved optimisation of HVAC schedules, minimising energy wastage, and supporting demand-response participation. By forecasting not just demand levels but also peak-load events—an issue highlighted in Section 2.3 these systems empower facilities managers to make informed, performance-driven decisions that enhance operational efficiency and occupant comfort.

The proposed system achieves this through two core contributions:

1. **Multi-Horizon Hybrid Forecasting Model:**
   A sequential hybrid model combining Support Vector Regression (SVR), XGBoost, and LSTM is developed to forecast electricity demand across four horizons: 24 hours, 1 week, 1 month, and 1 year. This design captures short-term fluctuations, nonlinear feature interactions, and long-range temporal patterns, supporting both operational control and strategic planning within buildings.
2. **Deep Hybrid Transfer Learning Model for Cross-Building Adaptation:**
   A deep hybrid model integrating CNN, GRU, and LSTM layers is developed and embedded within a transfer learning framework. By pretraining on multi-building datasets and fine-tuning on buildings with limited data, the model demonstrates strong cross-building generalisation with reduced training requirements, addressing data scarcity and scalability challenges.

Collectively, these contributions demonstrate that advanced hybrid models supported by transfer learning can deliver accurate, robust, and scalable energy forecasts across diverse buildings and time horizons. By aligning methodological innovation with real-world constraints, the thesis advances the development of reliable forecasting tools that enhance energy management, operational efficiency, and long-term sustainability in smart building environments.

1.2 Research Objectives

The central goal of this research is to design an accurate, modular, and generalisable forecasting model for smart buildings. To achieve this aim, the study is guided by four research objectives:

Objective 1:

Develop and evaluate a sequential hybrid forecasting model (SVR → XGBoost → LSTM) for multi-horizon prediction.

This objective focuses on integrating classical, ensemble, and deep learning techniques within a single model to improve accuracy across short-, medium-, and long-term forecasting horizons.

Objective 2:

Develop a deep hybrid model (CNN → GRU → LSTM) capable of capturing complex multiscale patterns in building energy data.

The goal is to learn spatial relationships, short-term variations, and long-range temporal behaviour while reducing the need for manual feature engineering.

Objective 3:

Apply transfer learning to enable cross-building adaptation of the forecasting model using limited data.

This objective examines how pretrained models can be fine-tuned for new buildings, improving generalisability and reducing data requirements.

Objective 4:

Validate the developed models using real-world datasets through rigorous statistical, accuracy-based, and event-driven evaluation.

This includes multi-metric assessment (MAE, RMSE, MAPE, R^2), peak-event detection, and statistical reliability tests to ensure robust and trustworthy performance.

Together, these objectives form the foundation for a high-performing and adaptable energy forecasting framework suitable for smart-building applications. As illustrated in (Figure 1.2) the framework integrates multi-horizon forecasting with cross-building transfer learning to support both day-ahead operational decision-making and longer-term strategic planning. By enabling improved energy efficiency, peak-load avoidance, and enhanced operational performance metrics, the proposed models offer practical value across a range of building types and management contexts.

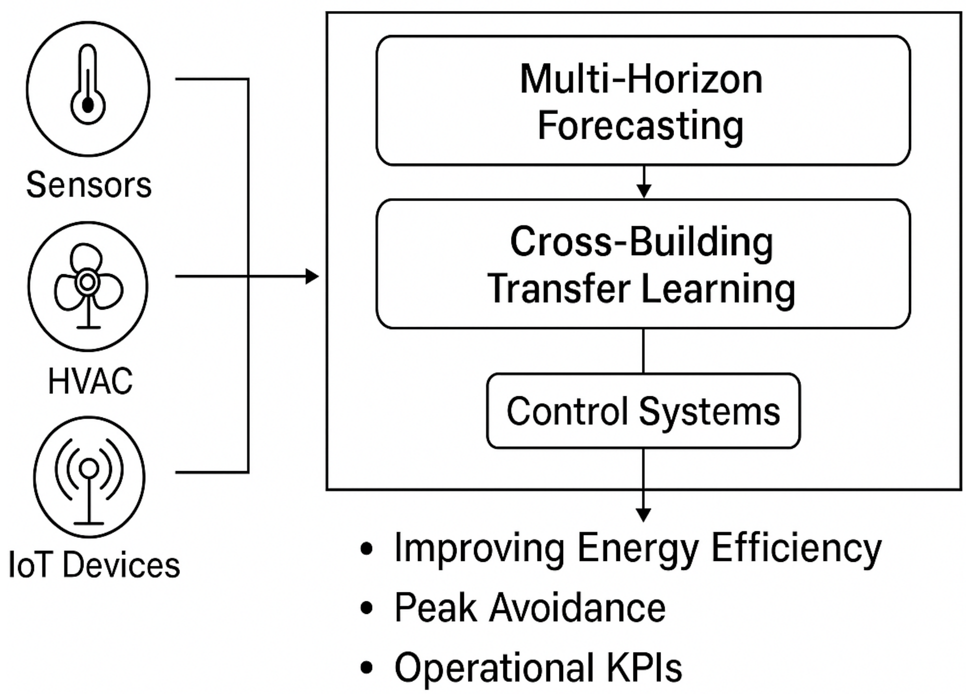

Figure 1.2: Overview of the Proposed Performance Forecasting Framework for Smart Buildings

To clarify how the proposed research responds to the limitations identified in the literature, (Table 1.2) summarizes the main research gaps, the corresponding solutions developed in this thesis, and their expected impact on forecasting performance. The table highlights how the proposed hybrid and transfer-learning models address long-standing challenges related to cross-building generalisation, multi-horizon accuracy, peak-event detection, and the robustness of validation practices.

Table 1.2: Research Gaps, Proposed Contributions, and Performance Outcomes

Research Gap	**Proposed Solution**	**Performance Impact**	**Where This Gap Is Identified in the Thesis**
No cross-building generalisation	Transfer learning using a deep hybrid model (CNN → GRU → LSTM)	Improves model adaptability across different buildings	Discussed in Section 2.7 (Transfer Learning) and Section 2.2.4 (Modelling Landscape)
Low accuracy across multiple forecasting horizons	Sequential hybrid model (SVR → XGBoost → LSTM)	Enhances forecasting consistency over short, medium, and long horizons	Identified in Section 2.6 (Multi-Horizon Forecasting Challenges)
No integrated peak-event detection	Statistical peak-detection module	Supports proactive energy management and grid responsiveness	Highlighted in Section 2.3 and 2.4 where peak-related limitations of existing models are discussed
Weak validation practices in prior studies	Multi-metric evaluation and statistical reliability tests	Increases trustworthiness and real-world reliability	Presented in Section 2.5 (Validation Methods)

1.3 Research Questions

To guide the development and evaluation of the proposed forecasting models, this thesis investigates two core research questions. Each question is directly linked to the research objectives and reflects both methodological and practical needs within smart-building energy management.

RQ1:

To what extent can hybrid machine learning and deep learning models outperform classical statistical and standalone learning approaches in delivering accurate and stable multi-horizon electricity demand forecasts?

This question examines the comparative performance of the developed sequential hybrid model (SVR → XGBoost → LSTM) and the deep hybrid model (CNN → GRU → LSTM) across multiple forecasting horizons. It focuses on whether combining complementary learning mechanisms within a single model can achieve higher predictive accuracy and greater stability than traditional techniques, particularly as forecasting horizons extend from hours to days, weeks, or months. Addressing this question provides insight into how advanced modelling approaches can support reliable operational and strategic energy decisions.

RQ2:

How effectively can transfer learning enable the adaptation of hybrid forecasting models to new buildings with limited historical data?

This question examines the extent to which transfer learning allows the deep hybrid model to retain strong predictive performance when adapted from well-instrumented source buildings to target buildings with sparse historical data such as those within the Ebbw Vale Schools Cluster. In this context, "effective adaptation" is defined as the ability of the transferred model to:

- maintain forecasting accuracy comparable to or better than a model trained solely on target-building data,
- reduce the amount of training data and computation required, and
- achieve stable performance across all four forecasting horizons without significant degradation.

Evaluating these criteria determines whether knowledge learned from large, diverse datasets can be reused through fine-tuning to support scalable forecasting across building portfolios, campuses, or broader built-environment settings.

1.4 Contributions of the Thesis

This thesis delivers a set of contributions across five areas: methodology development, model innovation, sustainability, empirical evaluation, and practical applicability. Together, these contributions demonstrate how advanced AI techniques can be meaningfully adapted to the realities of smart building operation, resulting in forecasting models that are accurate, adaptable, and suitable for real-world use. The two core technical contributions of this thesis are: **(1) the development of a sequential hybrid forecasting model (SVR → XGBoost → LSTM), and (2) the design of a deep hybrid model (CNN → GRU → LSTM) embedded within a transfer learning framework.** These contributions together form the backbone of the forecasting strategy developed and evaluated throughout the thesis.

The central technical contribution is the design of a deep hybrid model that integrates CNN, GRU, and LSTM components in a sequential manner. Unlike many existing hybrid approaches that only combine two layers or use shallow structures, the proposed model progressively learns patterns at different temporal and structural scales capturing local interactions, short-term variations, and long-term behavioural trends. The model is trained and tested across multiple building types and forecasting horizons, and its adaptability is further examined through transfer learning to evaluate how well it can generalise to new buildings with limited training data.

A further contribution lies in the development of a modular and transferable forecasting model. By clearly separating preprocessing, feature learning, and prediction components, the model can be adapted to new datasets or building contexts without the need for full retraining. This modularity enhances both usability and long-term scalability.

From a sustainability perspective, the thesis demonstrates how accurate forecasting supports low-carbon building operation by enabling better planning, peak-load anticipation, and improved alignment between energy use and renewable availability. In doing so, the research highlights the role of forecasting as an enabling layer for more intelligent, performance-driven building management.

The thesis also contributes a strong empirical foundation by establishing comprehensive benchmarks across short-, medium-, and long-term horizons. These benchmarks compare the proposed models with a wide range of classical statistical models (e.g., ARIMA, TBATS), machine learning models (e.g., SVR, XGBoost), and deep learning baselines (e.g., LSTM, CNN). The use of consistent datasets, metrics, and validation procedures provides a transparent basis for evaluating the strengths and limitations of each modelling approach.

Collectively, these contributions show how the proposed models advance both the technical and practical aspects of building energy forecasting, offering tools that are not only academically rigorous but also relevant for real-world deployment.

1.4.1 Methodological Contribution

This thesis introduces a sequential hybrid forecasting model (SVR → XGBoost → LSTM) in which each stage adds a distinctive layer of learning. SVR is used to reduce noise and stabilize the input data; XGBoost captures complex nonlinear relationships; and LSTM models longer-term temporal patterns.

Unlike ensemble methods that combine independent models, this sequential design allows each component to refine the signal produced by the previous stage. The result is a coherent, progressive modelling process that performs reliably across forecasting horizons ranging from 24 hours to one year. The methodological contribution lies in demonstrating how this structured approach supports more informed and responsive performance management in buildings.

1.4.2 Model Development Contribution

In parallel, the thesis proposes a deep hybrid model (CNN → GRU → LSTM) that learns multiscale energy patterns directly from data. CNN layers extract structural relationships, GRU layers learn short-term temporal dependencies, and LSTM layers capture longer-range cycles.

A key aspect of this contribution is the integration of transfer learning, which enables the model to adapt to new buildings with minimal retraining. This makes the approach particularly valuable in scenarios where historical data are limited, a common challenge in building energy forecasting.

1.4.3 Validation and Reliability Contribution

A core contribution of this thesis is its rigorous validation strategy. The models are evaluated not only using conventional accuracy metrics (MAE, RMSE, MAPE, R^2) but also through:

- residual diagnostics,
- Ljung–Box tests,
- Diebold–Mariano forecast comparison,
- bootstrapped confidence intervals, and
- ablation and sensitivity analyses.

This multi-layered evaluation framework ensures that the proposed models are not only accurate but also statistically reliable and robust across contexts. It also highlights how careful validation can strengthen confidence in AI-driven decision support for building management.

1.4.4 Empirical Contribution

Empirically, the thesis draws on three complementary datasets: the GENOME project, Cardiff University's Queen's Building, and the Ebbw Vale Schools Cluster. These datasets differ in building type, operational characteristics, and temporal granularity, offering a diverse basis for evaluation.

A unified metric set MAE, MSE, R^2, MAPE, and peak-event performance is applied across all experiments, enabling fair and transparent comparison. Both the sequential hybrid model and the deep hybrid transfer-learning model consistently outperform classical approaches such as ARIMA, demonstrating the value of the proposed models in both accuracy and adaptability.

1.4.5 Practical Contribution

Finally, the thesis delivers a forecasting model designed with real-world application in mind. By unifying model development, transfer learning, and well-structured validation under realistic operating conditions, the work provides a practical foundation for integrating forecasting into everyday building management.

The model supports decision-making related to energy use, peak management, load balancing, and long-term planning. As such, it offers value not only to researchers but also to energy managers, building operators, and smart-grid practitioners seeking reliable data-driven insights.

1.5 Structure of the Thesis

This thesis is organised into five chapters, each building on the previous one to provide a coherent progression from context and motivation to model development, evaluation, and final insights. The structure is designed to offer a clear narrative that links the research aims, methods, and findings.

Chapter 1: Introduction

This chapter establishes the background and motivation for the study, outlines the research problem, and presents the aim, objectives, and research questions. It also summarises the thesis contributions and provides an overview of the structure of the document.

Chapter 2: Literature Review

Chapter 2 examines the existing body of work on building energy forecasting, including classical statistical models, machine learning techniques, deep learning models, multi-horizon forecasting, and transfer learning. The chapter identifies the limitations in current approaches and highlights the research gaps that this thesis addresses.

Chapter 3: Methodology

This chapter details the methodological foundation of the study, including the datasets, preprocessing procedures, feature engineering, and the development of the hybrid forecasting models. It also outlines the transfer-learning strategy, evaluation methods, and the statistical tests used to assess reliability and robustness.

Chapter 4: Results and Discussion

Chapter 4 presents the empirical findings of the research. It reports the performance of the developed models across multiple forecasting horizons, evaluates their behaviour under transfer-learning scenarios, and examines peak-event detection performance. The chapter also discusses the results in relation to existing literature and the research objectives.

Chapter 5: Conclusions and Implications

The final chapter revisits the research questions and objectives, summarizes the main findings, and reflects on the methodological and practical contributions of the study. It outlines the limitations of the research and provides recommendations for future work in the field of smart-building energy forecasting.

Chapter 2: Literature Review

2.1 Introduction

Chapter 2 presents a comprehensive review of the models, learning strategies, and validation practices that inform modern approaches to building energy forecasting. The purpose of this chapter is to situate the research within the broader scholarly landscape, identify the persistent limitations of existing forecasting methods, and establish the conceptual foundations for the models developed later in this thesis. By examining both classical and contemporary techniques, the chapter clarifies why multi-horizon forecasting, hybrid deep learning, and transfer learning have become central to advancing energy performance management. The review is structured to directly support the four research objectives and their corresponding research questions. It begins by outlining the evolution of forecasting methods applied in buildings, from traditional statistical approaches to machine learning and deep learning models. This provides the context for Objective 1 and RQ1, which investigate how forecasting accuracy can be maintained across short-, medium-, and long-term horizons. The chapter then examines model families capable of capturing complex, multiscale temporal patterns, thereby addressing Objective 2 and RQ2. Subsequently, it explores transfer learning as a mechanism for improving model adaptability across different building types with limited data, linking to Objective 3 and RQ3. Finally, the chapter reviews the validation techniques required to assess model reliability, which informs Objective 4 and RQ4. In synthesising evidence from prior research, the chapter highlights several recurring challenges such as limited cross-building generalisation, inconsistent performance across time horizons, and weak statistical validation that directly motivate the modelling and experimental design adopted in this work. The gaps identified here form the basis for the methodological decisions detailed in Chapter 3 and underpin the analysis presented in Chapters 4 through 6.

This section has outlined the conceptual foundations that underpin building energy forecasting, including the drivers of consumption, the purpose of forecasting within performance management, and the theoretical frameworks multi-horizon modelling, hybrid architectures, and domain-adaptive transfer learning that guide the methodological choices of this thesis. These foundations establish the rationale for the structured review presented in the subsequent sections. In particular, multi-horizon and hybrid modelling directly support the investigation posed in RQ1, while the principles of domain adaptation and transfer learning underpin the cross-building generalisation examined in RQ2. Together, they provide a coherent basis for analysing how the literature informs and motivates the modelling strategies developed in this research.

2.2 Overview of Building Energy Forecasting

This section introduces the foundational elements of building energy forecasting by outlining the key factors that drive electricity demand, the forecasting horizons commonly used in practice, and their relevance to performance management. Establishing these fundamentals provides the necessary context for evaluating why different modelling families classical, machine learning, deep learning, hybrid, and transfer learning are reviewed in the subsequent sections and how they relate to RQ1.

Building energy forecasting is a multi-faceted field that integrates statistical modelling, machine learning, and deep learning to predict consumption patterns across a range of time horizons. Forecasting serves as a foundation for energy performance management by supporting anticipatory control, operational planning, and long-term efficiency strategies within buildings. Because energy demand is influenced by weather conditions, occupancy behaviours, equipment usage, and temporal patterns, forecasting methods must accommodate complexity, multiscale structure, and variability across contexts.

Over the past two decades, research has expanded from linear statistical models toward more sophisticated machine learning and deep learning approaches capable of extracting richer temporal dependencies. Hybrid models and transfer-learning strategies have further advanced the field by improving accuracy, enhancing adaptability, and reducing data requirements in settings where historical records are sparse. These developments have shaped a diverse methodological landscape in which different model families exhibit strengths across different forecasting horizons.

To frame this landscape, (Figure 2.1) provides a high-level view of the forecasting process, while (Table 2.1) summarises where major model categories tend to perform best and how transferable they are across buildings. This overview sets the context for the detailed examination of classical models (Section 2.2), machine learning approaches (Section 2.3), deep learning methods (Section 2.4), hybrid and multiscale models (Section 2.5), and transfer learning strategies (Section 2.6).

Subsequent sections draw on this foundation to identify gaps in capability, limitations in generalisation, and the motivations for the modelling approach adopted in this thesis.

2.2.1 Role of Forecasting in Buildings

The building sector accounts for 33–40% of global energy consumption and carbon emissions [58], with commercial office buildings playing a significant role due to their structured and predictable demand profiles. Improving energy efficiency in such buildings is essential for decarbonization goals [59]. HVAC systems alone can consume up to 40% of a building's total energy use, making them a focal point for optimisation [58].

Technologies such as renewable integration, thermal energy storage, and heat pumps offer significant potential for reducing emissions and operating costs [3]. Meanwhile, certification frameworks like LEED, BREEAM, and DGNB support energy-efficient design but require ongoing improvement in areas like economic performance and post-occupancy feedback [60]. Coordinated efforts from governments, academia, and industry remain essential to drive energy performance improvements at scale [58].

2.2.2 Forecast Horizons and Use-Cases

Forecasting plays a central role in both building management and smart grid operation, supporting demand response, load scheduling, and energy planning [61], [62]. Forecasts can be segmented into:

- Short-term (e.g., 24 hours): useful for HVAC scheduling and immediate operational control.
- Medium-term (1 week – 1 month): supports maintenance planning and anomaly detection.
- Long-term (seasonal to yearly): informs infrastructure investment and sustainability planning.

Accurate forecasting across these timeframes requires adaptable models that can manage both short-term variability and long-term trends. Techniques such as ensemble learning, deep learning (e.g., LSTM, CNN-LSTM), and probabilistic models like Bayesian LSTM have demonstrated strong potential in this space [62], [63].

2.2.3 Data and Drivers of Consumption

Energy consumption in buildings is influenced by multiple dynamic variables, including:

- Weather conditions
- Occupancy behaviour
- Time-of-day and seasonal effects
- Building envelope characteristics

Machine learning models such as Gaussian Process Regression and other supervised techniques have demonstrated the ability to model such complexity [64], [65]. Continual learning has also been explored to overcome concept drift and catastrophic forgetting in online forecasting systems [66]. Proxy data such as human mobility has been used to enhance model adaptability in exceptional scenarios like the COVID-19 pandemic [67].

2.2.4 Modelling Approaches and Landscape

Studies comparing classical models, artificial neural networks (ANN), support vector machines (SVM), and hybrid variants have established that:

- Classical statistical models are interpretable but often insufficient for capturing nonlinearity or high-frequency variations [68].
- ANN and SVM models outperform traditional simulations for heating/cooling predictions in office buildings [69], [70].
- Hybrid models, especially those combining feature engineering with ML predictors, offer improved performance across different periods and use-cases [71], [72].

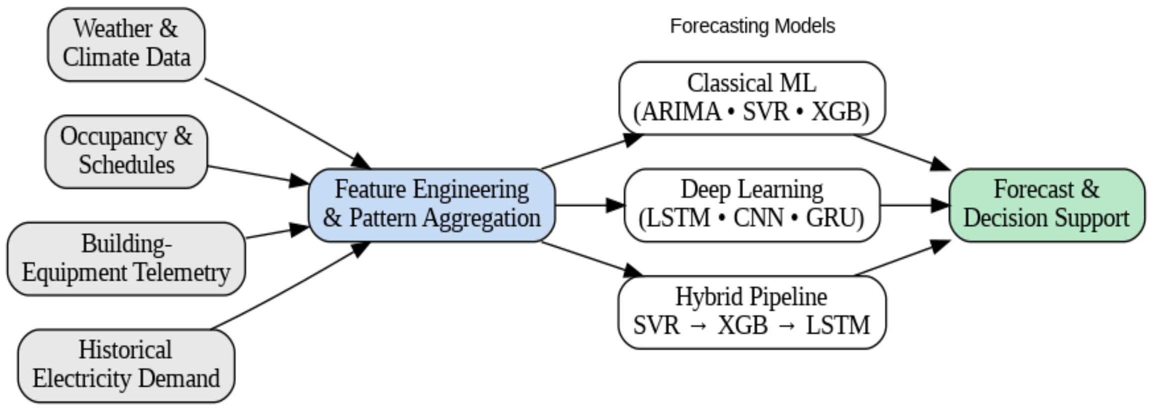

Figure 2.1: High-Level Workflow for AI-Based Building Energy Forecasting

Moreover, advanced models like NARX-ANN and nonlinear regression structures have proven effective in capturing the thermal dynamics of buildings, while linear regression with engineered polynomial features can enhance interpretability and performance [73], [74].

2.2.5 Smart Building Applications and Future Trends

Deep learning has significantly advanced energy forecasting:

- LSTM improves short-term prediction accuracy by over 18% compared to ARIMA [75].
- GRU models deliver similar gains with improved training efficiency [76].
- CNN components help extract time-local features, improving performance in hybrid CNN-LSTM models by up to 13% [75], [77].

Transfer learning is gaining prominence as a scalable strategy to generalise across different building types. Studies by Hernández et al. [78] and others have shown that pre-trained models can dramatically reduce error rates when fine-tuned for new buildings with limited data.

In parallel, researchers are increasingly focused on:

- Model interpretability (e.g., SHAP values)
- Forecasting generalisation across geographies
- Reducing training requirements via cross-building learning

While deployment concerns such as latency and edge computing were explored in earlier drafts of this research, they are no longer within the scope of the final thesis.

2.2.6 Summary

This section outlined the key concepts underpinning building energy forecasting, including common model families, forecasting horizons, and the main drivers of energy consumption. The review showed how traditional models, machine learning techniques, and deep learning approach each address different aspects of temporal complexity and operational variability. Table 2.1 summarises where major model categories perform best across horizons and indicates their expected transferability. These insights establish the methodological foundations for the modelling choices explored in the next sections.

2.3 Traditional Statistical Methods in Energy Forecasting

This section reviews classical statistical models because they represent the foundational approaches to time-series forecasting and provide essential baselines against which all machine learning and deep learning models are evaluated. Understanding their assumptions and limitations helps contextualise the need for more advanced models explored in later sections (addressing RQ1).

Traditional statistical methods have long served as the foundation for time-series forecasting in energy systems. Their appeal lies in their simplicity, interpretability, and well-understood mathematical structures. However, as building energy dynamics have become more complex driven by occupant behaviour, variable weather conditions, and increasingly granular sensor data these methods have struggled to maintain forecasting accuracy in modern smart building contexts [26].

One of the most widely used approaches is the Autoregressive Integrated Moving Average (ARIMA) model, which remains a staple for univariate time series prediction across multiple domains [79]. ARIMA performs well when the underlying data exhibit clear trends or seasonality, but its reliance on linear assumptions limits its effectiveness in capturing nonlinear or abrupt changes typical of building energy use [80]. To address this, several studies have proposed hybrid approaches. For instance, Syu et al. [81] developed a hybrid ARIMA framework combining linear and nonlinear models, resulting in a 35.4% improvement in forecasting accuracy compared to standard ARIMA. Similar gains have been observed when integrating ARIMA with LSTM, combining ARIMA's strengths in trend modelling with the temporal feature learning of deep learning networks [80], [82].

Another classical method, Holt-Winters Exponential Smoothing, has been applied successfully to short-term energy forecasting, especially when handling seasonality. The method uses triple smoothing equations for level, trend, and seasonality and can accommodate both additive and multiplicative seasonal effects [83]. Its performance improves in settings where intraday and intraweek cycles dominate, such as office buildings [84]. However, issues like choosing starting values, normalizing seasonal indices, and tuning smoothing parameters can complicate its implementation [85], [86].

Linear regression (LR) models have also been extensively used due to their transparency and ease of deployment. They are particularly effective in scenarios where the relationships among variables are well understood. However, their limitations become evident in capturing nonlinear dependencies and variable interactions, which are prevalent in energy systems [24]. To bridge this gap, researchers have introduced enhancements such as engineered polynomial terms and interaction variables, improving both prediction accuracy and model interpretability [87]. These modified regression models offer a middle ground between classical interpretability and improved representational power.

In parallel, nonlinear regression methods including decision trees and artificial neural networks have outperformed traditional regression techniques in contexts where load profiles are complex, such as during irregular occupancy patterns or under weather-driven demand variability [24], [67].

Recent studies further underscore the limits of traditional statistical models when faced with real-world energy datasets that exhibit nonlinearity, multivariate dependencies, and non-stationary behaviour [88]. To address these challenges, new techniques have been introduced, including:

- Weak-stationarity corrections
- Spectral decomposition
- Mixer models for capturing long-range dependencies
- Copula-based ensemble post-processing, which models inter-variable dependencies explicitly [89].

While these innovations extend the capabilities of classical models, their application remains limited in highly dynamic environments.

Traditional statistical models continue to serve as important benchmarks in energy forecasting. They are especially useful for early-stage validation, baseline comparisons, and interpretable forecasting scenarios. However, their limitations particularly in addressing nonlinearity, high-dimensional input spaces, and temporal dynamics have led to a growing reliance on machine learning and deep learning approaches. As the remainder of this chapter will show, these modern alternatives offer greater adaptability, robustness, and predictive power, particularly in complex smart building environments.

2.4 Machine Learning Techniques for Energy Forecasting

This section examines machine learning models as they address key limitations of classical statistical methods by capturing nonlinear relationships and handling richer feature sets. Their inclusion is essential for understanding performance improvements in short- and medium-term forecasting and for evaluating the hybrid models developed in this thesis (linked to RQ1 and Objective 1).

Machine learning (ML) has become a foundational technique in energy forecasting for smart buildings, offering significant advantages over traditional statistical methods. ML models excel at processing complex, high-dimensional data generated by sensor-rich environments [90], [91]. These data-driven approaches can extract non-obvious patterns and relationships without relying on rigid physical assumptions, leading to greater robustness and forecasting precision [90]. Comparative studies confirm that ML-based approaches frequently outperform both traditional and physics-based (grey-box) models in building energy prediction tasks [91]. Their increasing computational efficiency and adaptability have accelerated ML adoption across both academic and industrial contexts, reflecting the growing demand for scalable, accurate forecasting systems in energy-intensive buildings [92].

2.4.1 Support Vector Regression (SVR)

Support Vector Regression has consistently demonstrated strong generalisation capabilities in short- and medium-term load forecasting tasks. Its ability to manage high-dimensional inputs and remain resilient to outliers makes it particularly well suited for energy forecasting in volatile environments [93]. Among its variants, SVR with a Radial Basis Function (RBF) kernel often yields superior performance compared to linear or polynomial kernels. Research has further enhanced SVR performance through hybridization, integrating it with preprocessing, optimisation, or feature selection modules [94]. Compared to neural networks, SVR generally requires less training data and offers faster training times, making it advantageous for real-time and resource-constrained scenarios [95].

2.4.2 Tree-Based Models and Gradient Boosting

Ensemble-based methods such as Random Forests (RF) and Gradient Boosted Regression Trees have gained popularity due to their high predictive accuracy and interpretability. RF has achieved strong results in university building energy forecasts [96], while gradient boosting has improved accuracy in heating and cooling load predictions by up to 65% [97]. Performance can be further improved through tuning strategies such as Hierarchical Shrinkage (HS), which regularizes tree outputs for better generalisation and interpretability [98].

Among gradient boosting techniques, XGBoost and LightGBM have emerged as leading algorithms due to their speed, flexibility, and capacity to handle missing values and complex interactions[99]. XGBoost has consistently outperformed other boosting methods in building-level energy prediction tasks [100]. These models are adept at integrating contextual features such as weather, occupancy, and appliance-level usage data. Additionally, interpretation tools such as SHAP (Shapley Additive Explanations) have made it possible to visualize and quantify the contribution of each input variable an essential feature for deployment in real-world energy management systems [101].

2.4.3 Instance-Based and Classical Models (KNN, SVM)

Although less scalable, k-Nearest Neighbors (KNN) and Support Vector Machines (SVMs) have also been applied to energy forecasting problems, often in hybrid configurations. KNN offers intuitive simplicity but suffers from memory inefficiency when deployed on large datasets [123]. SVMs, while theoretically powerful, often underperform in large-scale applications unless coupled with dimensionality reduction or feature engineering [102]. In certain contexts, KNN has outperformed SVM when paired with preprocessing methods [103]. Hybrid KNN-SVM models have shown promising accuracy improvements in solar forecasting tasks, outperforming standalone ML and deep learning models such as LSTM on specific datasets [104].

2.4.4 Multi-Horizon and Time-Adaptive ML Forecasting

ML techniques have proven particularly effective in short- and medium-term energy forecasting, where rich contextual data such as weather, calendar information, and occupancy schedules are available [105]. Algorithms like Random Forest, Lasso Lars, and XGBoost consistently capture temporal structure and seasonal variation across forecast horizons. For instance, the Nonlinear Autoregressive Neural Network with Exogenous Input (NARX) has shown strong results in week-ahead forecasting, especially when seasonal patterns are prominent [106]. However, most ML models experience performance degradation as the forecast window lengthens due to a reduction in input granularity and the compounding effect of error propagation.

2.4.5 Hyperparameter Tuning and Model Optimisation

Tuning hyperparameters remains a critical step in optimising ML model performance and ensuring generalisation [107]. Parameters such as learning rate, regularization coefficients, and tree depth can significantly influence the effectiveness of ML models [108]. Optimisation techniques such as grid search, random search, and Bayesian optimisation are frequently employed to explore hyperparameter spaces efficiently [109]. Among them, Bayesian methods have gained traction due to their ability to model performance as a probabilistic function of parameter combinations. Emerging strategies such as AutoML and transfer learning are being explored to automate and accelerate this optimisation process [110].

Machine learning offers a flexible, scalable, and high-performing alternative to traditional forecasting methods in building energy management. These techniques excel at modelling non-linear relationships and incorporating diverse, high-resolution data streams. However, their limitations, particularly the lack of built-in sequence modelling and the need for extensive tuning, highlight the importance of hybridizing ML with deep learning to manage long-range dependencies and improve multiscale performance. As the next section will show, deep learning builds upon the foundations of ML to address many of these shortcomings, enabling even greater precision and adaptability in energy forecasting.

2.5 Deep Learning for Multiscale Temporal Forecasting

This section focuses on deep learning techniques because they offer enhanced capability for modelling long-range temporal dependencies and multiscale behaviour in building energy data.

These methods form the conceptual basis for the deep hybrid model introduced later, directly informing RQ2 and Objective 2.

Deep learning (DL) models have gained prominence in building energy forecasting due to their ability to learn complex, nonlinear patterns from high-frequency data [111]. These models outperform traditional statistical approaches, particularly in capturing long-range dependencies and high-dimensional feature interactions. For instance, Deep Belief Networks and hybrid models like kCNN-LSTM have demonstrated improved accuracy in day-ahead and week-ahead energy forecasting by combining spatial clustering, convolutional feature extraction, and temporal modelling [112], [113].

As the building sector remains a major contributor to global energy demand and emissions [91], the ability to forecast energy use with high precision has both economic and environmental implications. Ongoing research continues to explore and refine DL models to address challenges such as limited training data, long-horizon forecasting degradation, and generalisation across building types.

2.5.1 Recurrent Neural Networks (RNN), LSTM, and GRU

Recurrent Neural Networks (RNNs) and their enhanced variants Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have demonstrated strong capabilities in capturing temporal dependencies in energy time-series data [114]. Standard RNNs are limited by vanishing gradient issues, which LSTM and GRU models mitigate through gating mechanisms that allow long-range memory [115].

Empirical studies confirm LSTM's superiority in energy forecasting tasks, including electricity demand prediction, where they outperform traditional models like ARIMA and shallow ML techniques such as SVR and random forests [116]. LSTMs can automatically extract temporal patterns without extensive feature engineering, and their accuracy can be further improved through techniques like genetic algorithm-based optimization [37].

GRUs, in comparison, offer computational efficiency with similar or better performance than LSTMs in certain contexts. For example, GRUs have achieved faster training times and lower error metrics in renewable energy applications such as solar forecasting [117]. Studies also highlight

GRU's superior performance-cost ratio, making it a practical choice for resource-constrained environments [118].

Long-term forecasting with LSTM and GRU remains challenging due to error accumulation. Advanced strategies such as expectation-biasing [119], multi-input contextual models, and feature anchoring have shown promise in mitigating degradation across extended horizons [120]. Hybrid models that integrate top-down, bottom-up, and sequential features have also outperformed standalone models in real-world use cases [121].

2.5.2 Convolutional Neural Networks (CNN)

Although originally developed for image processing, Convolutional Neural Networks (CNNs) have proven highly effective in energy time-series forecasting. CNNs extract localized temporal features and nonlinear relationships, improving model precision in applications such as solar and electricity load forecasting [122].

Temporal Convolutional Networks (TCNs) and advanced CNN models improve accuracy over LSTM in short-term energy demand forecasting [123]. However, CNNs alone may struggle with long-term temporal dependencies. To address this, hybrid models such as Sparse Identification + CNN or CNN-LSTM/GRU model have been developed, enabling the capture of both spatial and sequential dependencies [124].

Recent studies show that hybrid CNN-RNN models outperform conventional approaches across a range of domains, including wind power, traffic flow, and electricity demand [125]. In smart buildings, CNN-LSTM models have demonstrated 20–45% accuracy improvements, high adaptability across building types, and robustness to weather variability and sensor noise [126]. These qualities make them highly attractive for deployment in heterogeneous building environments.

2.5.3 Transformer-Based Models

Transformers originally developed for natural language processing are emerging as a new standard in multivariate time-series forecasting. Their attention mechanisms allow them to model long-range dependencies more effectively than RNNs, while enabling parallel computation and better scalability[127].

Recent innovations include:

- Temporal Fusion Transformer (TFT): Integrates static covariates and temporal features using attention, delivering strong performance and interpretability in building energy and photovoltaic forecasting [128].
- PatchTST and CT-PatchTST: Employ patch-based and channel-independent encoding schemes, offering strong generalisation in long-horizon and multivariate settings [129].
- Fusion Transformer (FusFormer): Combines encoder-decoder models with static enrichment modules to handle diverse time resolutions and exogenous inputs [130].
- Knowledge Graph Embeddings (KGE): Used with models like Autoformer and Informer to improve multivariate representation in renewable energy prediction.

Transformers consistently outperform LSTM, CNN, and MLP models in long-horizon tasks and multiscale environments [131]. They achieve up to 11% improvements in accuracy and exhibit stronger generalisability across buildings and climatic regions [132].

Interpretability remains a key strength of transformer-based models. Attention weights and feature attribution methods allow energy managers to understand which inputs most influence forecasts critical for building trust in AI systems [133].

To clarify the relative positioning of key model families by temporal scope and building adaptability, (Figure 2.1) provides a visual summary, and (Table 2.1) offers a concise tabular mapping. These references complement the model view in (Figure 2.1) and will be used in Chapter 3 to justify model selection per horizon.

Table 2.1: Summary of Forecasting Model Categories and Transferability

Model Category	**Representative Techniques**	**Applicable Forecasting Horizon**	**Cross-Building Transferability**
Classical Statistical	ARIMA, Linear Regression, Holt-Winters	Short-Term	Low

Model Category	**Representative Techniques**	**Applicable Forecasting Horizon**	**Cross-Building Transferability**
Shallow Machine Learning	Support Vector Regression (SVR), XGBoost, Random Forest, LightGBM	Short to Medium-Term	Moderate
Recurrent Neural Networks (RNN)	LSTM, GRU	Medium to Long-Term	Moderate
Hybrid Deep Learning	CNN-LSTM, kCNN-LSTM	Short to Long-Term	High
Transformer-Based Models	Temporal Fusion Transformer (TFT), PatchTST	Medium to Long-Term	Low to Moderate

Deep learning models now form the backbone of high-resolution, adaptive energy forecasting systems. RNN-based models such as LSTM and GRU continue to dominate short- and medium-term forecasting due to their strong memory capabilities. CNNs provide complementary spatial insights, while hybrid CNN-RNN combinations offer robust spatio-temporal modelling. Transformer models, with their superior scalability, long-horizon fidelity, and interpretability, represent the frontier of energy forecasting especially when paired with fine-tuning and transfer learning techniques. For a compact comparison across families, see (Table 2.2).

Together, these models enable intelligent, multiscale forecasting strategies tailored to the dynamic operational needs of smart buildings and contribute directly to this thesis's modelling framework.

Table 2.2: Strengths and Limitations of Key Forecasting Models

Model / Technique	Strengths	Limitations	Suitable Use-Cases
ARIMA / Holt-Winters	Simple, interpretable	Poor with non-linear data, short horizon only	Basic short-term baseline forecasting
SVR / XGBoost / RF	High accuracy, good generalisation, interpretable (via SHAP)	Needs tuning, medium temporal scope	Daily–weekly predictions, feature ranking
LSTM / GRU	Captures long dependencies, adaptive	High training cost, hard to interpret	Day/week/month energy patterns
CNN-LSTM (hybrids)	Spatial + temporal features, scalable	Training complexity, needs large data	Smart building systems with sensor fusion
Transformers (TFT, PatchTST)	Long-horizon forecasting, interpretable attention layers	Less stable with small data, edge-unfriendly	Strategic planning, multi-variate inputs
Transfer Learning (TL)	Adaptability to new buildings, cost-effective	Risk of negative transfer, requires similarity tuning	Cross-building generalisation

2.6 Multi-Horizon Forecasting Challenges and Approaches

This section reviews hybrid and multiscale models as they combine the strengths of machine learning and deep learning to overcome the weaknesses of individual approaches. Their relevance lies in supporting multi-horizon forecasting and improving generalisation two central challenges addressed in this thesis (RQ1 and Objective 2).

Recent literature highlights the growing importance of multi-horizon forecasting in building energy management. While short-term forecasts (e.g., 15 minutes to 24 hours) are essential for daily operations and control, longer-term predictions (weeks to months) support strategic planning, budgeting, and sustainability efforts [48]. Forecasting frameworks that operate across these timescales are essential for balancing operational efficiency with long-range energy management goals.

Chen et al. [134] categorizes energy forecasting approaches into three broad types: physical energy models, data-driven models, and hybrid models each offering distinct strengths and trade-offs. González et al. [135] introduced a unified inverse modelling framework that adapts to diverse timescales, while [136]achieved 88–95% accuracy in month- and quarter-ahead predictions using long-horizon forecasting techniques. These studies underscore the need for versatile models capable of handling both immediate responsiveness and longer-term decision support.

Several recent contributions address the technical challenges of multi-horizon forecasting using advanced machine learning and deep learning techniques. Giamarelos et al. [137] proposed an ensemble approach that achieved high accuracy across multiple forecast windows. Wen et al. [138] and Lim et al. [139] introduced probabilistic multi-step forecasting frameworks using recurrent neural networks and attention mechanisms. Lai et al. [140] developed LSTNet, which combines convolutional and recurrent layers to model both short- and long-term dependencies in multivariate time series data.

Yet, challenges persist particularly for long-range forecasting. Standard LSTM and GRU models often degrade in performance over extended horizons. To address this, Ismail

et al. [119] employed expectation-biasing, while Belletti et al. [141] and Gauch et al. [142] proposed multi-scale memory models that improve retention of long-term patterns. Ma et al. [143] introduced the Logsparse Decomposable Multiscaling framework, which decomposes temporal structures for better long-horizon predictability. These innovations address issues such as forecast drift, short-horizon overfitting, and long-range degradation of common pitfalls in time-series prediction.

Within the building energy domain specifically, researchers have applied multi-horizon strategies using both conventional and deep learning techniques. Kanthila et al. [144] demonstrated that cascaded Bi-LSTM models significantly improve accuracy across different occupancy-related timeframes. Ahmad et al. [65] and Ni et al. [47] leveraged Temporal Fusion Transformers (TFT) and hybrid deep learning models for high-performance forecasting across time horizons, emphasizing the integration of exogenous features and uncertainty quantification.

2.7 Transfer Learning for Cross-Building Forecasting

This section examines transfer learning because it provides a mechanism for adapting forecasting models to new buildings with limited historical data, addressing the challenge of cross-building generalisability. Its discussion is essential for answering RQ2 and forms the conceptual foundation for the transfer-learning experiments carried out later in the thesis.

Forecasting energy consumption in buildings remains a complex challenge, particularly when historical data is limited or building characteristics vary significantly across sites. Traditional machine learning (ML) methods often suffer from poor generalisability due to overfitting to specific building types, climates, or usage patterns. To address these limitations, transfer learning (TL) has emerged as a compelling strategy for improving model adaptability across diverse contexts.

Component-Based Machine Learning (CBML) approaches incorporate domain-specific knowledge to enhance robustness in data-scarce environments [145]. Similarly, models based on sequence-to-sequence learning and 2D convolutional neural networks with attention mechanisms have demonstrated improved performance when forecasting in buildings with limited or inconsistent data [145]. While building-level forecasting has received considerable attention, urban-scale and cross-building generalisation remain comparatively underexplored [146]. The proliferation of IoT-enabled energy datasets has spurred renewed interest in scalable ML and deep learning frameworks for energy prediction [91].

The subsections that follow provide the conceptual and empirical basis for the use of transfer learning in this thesis. They outline the underlying theory, common adaptation strategies, recent empirical evidence, and the specific challenges associated with forecasting across diverse building types.

2.7.1 Theoretical Basis and TL Model in This Study

The core principle behind transfer learning lies in domain adaptation the ability to transfer knowledge learned from a well-labelled source domain to a target domain with limited data availability. Pan and Yang [147] identify three key mechanisms for knowledge transfer:

- Feature representation transfer
- Model parameter reuse
- Instance reweighting

In this thesis, transfer learning is applied to a deep hybrid model composed of CNN → GRU → LSTM layers. The model is first pretrained on a rich, heterogeneous dataset (e.g., GENOME), and then fine-tuned with minimal retraining using smaller, building-specific datasets. This model is designed to maximize generalisation while reducing training overhead, crucial for scalable deployment across varied building portfolios.

2.7.2 Typical Transfer Learning Model in Time-Series Forecasting

The TL process generally involves two phases:

- Pretraining: A deep model is trained on a large, diverse dataset capturing broad energy consumption patterns, weather dependencies, and operational variability (e.g., GENOME or multi-site BMS logs).
- Fine-tuning: A smaller, local dataset is used to adapt the pretrained model. This typically involves retraining only a subset of layers (e.g., upper GRU/LSTM layers), preserving general temporal encodings while tailoring predictions to local patterns.

Common strategies explored in the literature include:

- Domain adaptation: Aligning feature distributions between source and target.
- Layer freezing: Fixing lower layers while fine-tuning task-specific upper layers.
- Multi-source TL: Aggregating knowledge from multiple source buildings to improve generalisability.

- Federated transfer learning: Decentralized learning without sharing raw data critical for preserving privacy in real-world settings.

2.7.3 Empirical Evidence and Models

Recent studies have demonstrated the empirical value of TL in energy forecasting. For instance:

- A 7-day TL adaptation reduced MAPE from 18.31% to 7.76% [51].
- TL applied to LSTM and Transformer models improved both performance and training efficiency [148].

In smart building contexts, hybrid TL models combining CNNs, LSTMs, and even Graph Neural Networks (GNNs) have proven effective. For example:

- A CNN-LSTM model improved indoor temperature forecasting across building types when augmented with TL [149].
- A kCNN-LSTM framework showed significant forecasting accuracy gains through clustering-based transfer [113].
- Advanced models integrating Bi-LSTM, GNN, and Transformer layers further improved performance for short-term load and temperature prediction [150].

TL has also shown promise in renewable energy applications, such as solar PV forecasting and HVAC control, where it significantly reduces both training data requirements and computational resources while maintaining accuracy [151].

2.7.4 Forecasting Across Diverse Building Types

Cross-building forecasting is particularly challenging due to variability in:

- Structural design
- Occupancy patterns
- Operational schedules
- Climate zones

Studies have shown that TL enhances performance across these contexts:

- Rodrigues et al. [152] applied TL in leisure centers and offices with positive results.
- Hooshmand et al. [153] used CNN-TL strategies to improve daily electricity demand prediction.
- Fan et al. [154] reduced prediction error by 15–78% using TL methods for short-term energy demand.
- Li et al. [155] demonstrated TL's effectiveness in information-poor buildings by leveraging data-rich sources.

By bridging source-target domain gaps, TL addresses a key limitation of many AI models and their inability to generalise across real-world settings without full retraining.

2.8 Gaps in Literature and Contributions of This Thesis

This section synthesises the insights from the literature to identify the unresolved challenges in multi-horizon forecasting, hybrid model design, and cross-building generalisation. These gaps form the basis for the research contributions developed in Chapter 3.

Despite notable progress in applying machine learning and deep learning to building energy forecasting, several significant limitations continue to restrict the scalability, generalisability, and real-world usefulness of existing approaches. These gaps become particularly visible when forecasting must operate across multiple time horizons, adapt to buildings with limited historical data, or support decision-making tasks such as peak detection and demand-side response. The following subsections synthesise the main gaps identified through the literature review, while (Table 2.3) summarises how the contributions of this thesis address each limitation.

2.8.1 Limited Multi-Horizon Forecasting Frameworks

A recurring issue in the literature is the tendency to evaluate forecasting models on a *single* prediction horizon most commonly the 24-hour window used for daily operational control. While such short-term forecasts are undeniably important for HVAC optimisation and anomaly detection, they do not capture the broader planning needs of building operators, facility managers, and energy procurement teams. Medium-term forecasts (e.g., weekly or monthly) are required for maintenance scheduling, load balancing, and contextual performance monitoring, whereas long-term predictions (seasonal to yearly) support budgeting, retrofit decisions, and infrastructure investment.

However, very few studies attempt to benchmark a single model family across multiple horizons, and even fewer offer insights into how model behaviour degrades, stabilises, or changes across these scales. This creates a fragmented evidence base, making it difficult to understand which models are genuinely robust and which are overfitted to short-term dynamics. Moreover, the absence of unified multi-horizon frameworks restricts the ability to derive consistent insights across decision-making layers in smart-building management.

2.8.2 Weak Integration of Transfer Learning in Hybrid Models

Deep learning and hybrid models particularly those combining CNN, GRU, and LSTM components have demonstrated high accuracy in controlled or well-instrumented environments. Yet, their capacity to generalise across buildings remains underexamined. In practice, energy consumption patterns vary substantially across building types, operational schedules, occupancy rhythms, and climatic contexts. As a result, models trained in one building often fail to perform well in another unless large datasets are available for retraining.

Although transfer learning offers a promising path toward data-efficient adaptation, its integration with hybrid deep learning structures is still in its early stages. Existing studies often explore TL in isolation (e.g., fine-tuning an LSTM), but rarely as part of a multi-component hybrid model. Critical questions such as which layers to freeze, how to initialise weights for heterogeneous data sources, or how much target-building data is needed for stable fine-tuning remain insufficiently explored. Consequently, the literature lacks principled guidance on how deep hybrid models can be adapted reliably and efficiently across buildings.

2.8.3 Disconnected Forecasting and Peak Detection

While forecasting is central to energy-aware operations, peak detection plays an equally critical role in load management, tariff optimisation, and grid-interactive control. Nevertheless, these two tasks are often studied independently. Many forecasting models produce a continuous load profile but do not explicitly identify periods of extreme demand. Existing peak detection techniques rely heavily on simple thresholds or statistical heuristics that do not leverage the predictive structure of modern ML/DL models.

This disconnect limits the usefulness of forecasting outputs in scenarios where early identification of demand surges is essential such as during demand response events or when preventing equipment overload in large commercial buildings. Without an integrated mechanism for characterising and interpreting peak behaviour, forecasting systems cannot provide actionable insights for advanced operational strategies.

2.8.4 Transformer Models Lack Practical Integration

Transformer-based models, such as the Temporal Fusion Transformer (TFT), PatchTST, and recent hybrid variants, have shown considerable promise in long-horizon forecasting. They excel at capturing multi-scale dependencies and offer interpretability via attention mechanisms. However, their adoption in building energy forecasting remains limited for several reasons:

1. High data requirements: Transformers often require large and diverse datasets to avoid overfitting.
2. Computational intensity: Their training and tuning processes can be resource intensive.
3. Limited transfer learning studies: Despite strong theoretical suitability, few studies evaluate transformers in cross-building adaptation settings.
4. Lack of integration into hybrid models: Most existing work tests transformers as standalone models rather than embedding them within broader hybrid or transfer-learning frameworks.

As a result, the potential of transformers remains under-realised in building energy research, particularly for applications requiring multi-horizon or multi-building generalisation.

2.9 Contributions to This Thesis

The thesis addresses the identified research gaps through four integrated contributions that form the methodological core of this work.

2.9.1. A Unified Multi-Horizon Modelling Strategy

This research introduces a sequential hybrid model (SVR → XGBoost → LSTM) and systematically evaluates it across four distinct horizons: 24 hours, 1 week, 1 month, and 1 year. By benchmarking the same modelling structure across multiple prediction windows, the thesis provides one of the few comparative analyses of horizon-specific behaviour in hybrid models. This contribution directly supports Objective 1 and RQ1 by demonstrating how accuracy, generalisation, and error propagation evolve across temporal scales.

2.9.2. A Deep Hybrid Model Designed for Transfer Learning

The thesis proposes a deep hybrid model combining CNN, GRU, and LSTM components, embedded within a transfer learning workflow designed specifically for cross-building adaptation. The model is pre-trained on a multi-building dataset and subsequently fine-tuned using limited data from target buildings. This approach clarifies how hybrid deep learning models can be adapted efficiently, shedding light on freezing strategies, data requirements, and building similarity dependencies. This contribution addresses Objective 3 and RQ2).

2.9.3. Integrated Statistical Peak Detection

A statistical peak-detection module is integrated into the forecasting workflow to bridge the gap between prediction and actionable operational insight. Instead of treating peak identification as a post-processing task, the mechanism is aligned with the structure of the forecasting models, enabling more informed detection of extreme events. This contribution supports Objective 4 by demonstrating how forecasting outputs can be enriched to improve their usefulness in energy management and demand-side application.

2.9.4. Transformer-Enhanced Long-Horizon Modelling

To address the limitations of conventional recurrent and convolutional models in long-horizon contexts, the thesis incorporates a transformer-based component into a hybrid variant designed for extended temporal prediction. This approach enhances the model's ability to capture long-term dependencies, improves its alignment with cross-building adaptation strategies, and offers interpretable attention-based insights. This contribution links to Objective 2 and complements both multi-horizon forecasting (RQ1) and cross-building learning (RQ2).

2.9.5 Synthesis and Final Remarks

While prior research has offered important advances in forecasting accuracy, transferability, and model sophistication, very few studies attempt to address these challenges simultaneously. This thesis bridges that gap by combining multi-horizon forecasting, hybrid deep learning, transfer learning, and peak detection into a coherent modelling strategy that is methodologically robust and empirically validated. The resulting framework provides a theoretically grounded pathway for developing more adaptive and scalable forecasting systems suitable for modern smart-building environments.

Table 2.3: Mapping of Research Gaps to Thesis Contributions and Objectives

Identified Research Gap	**Proposed Solution / Thesis Contribution**	**Related Objective / RQ**
Limited multi-horizon forecasting Most studies focus on a single short-term window, limiting broader planning use.	Development of a hybrid model (SVR → XGBoost → LSTM) capable of forecasting across four horizons: 24h, 1 week, 1 month, and 1 year.	Objective 1 / RQ1
Lack of transfer learning in hybrid models Limited generalisation across buildings; few TL-integrated models.	A novel deep hybrid model (CNN → GRU → LSTM) embedded in a transfer learning framework for cross-building adaptability with minimal retraining.	Objective 3 / RQ2
Forecasting and peak detection treated separately Absence of integrated mechanisms for identifying demand surges.	Integration of statistical peak detection into the forecasting workflow to enable automated load control and demand-side response.	Objective 4
Transformer models are underutilized in building forecasting Limited integration into hybrid TL-ready models.	Transformer-enhanced hybrid model to support high-dimensional, long-horizon forecasts with improved generalisability and attention-based interpretability.	Objective 2

Identified Research Gap	**Proposed Solution / Thesis Contribution**	**Related Objective / RQ**
Lack of unified solutions Models tend to solve one issue (accuracy, transfer, or scalability) in isolation.	This thesis proposes a modular, unified framework combining accuracy, adaptability, and deployment-readiness, tested across models, horizons, and building types.	All Objectives / RQ1, RQ2

Taken together, the literature reviewed in this chapter demonstrates why multi-horizon forecasting, hybrid modelling, and transfer learning are essential to advancing energy performance management in buildings. The gaps identified particularly those relating to accuracy across horizons, cross-building generalisation, peak-event detection, and rigorous validation directly shape the methodological design developed in Chapter 3. This includes the choice of datasets, feature engineering strategies, model architectures, and evaluation procedures. Chapter 4 then empirically tests these methodological choices, presenting results and validation outcomes that address the research questions and respond to the gaps identified in the literature. Finally, Chapter 5 revisits these findings to draw conclusions, articulate implications, and outline directions for future research.

2.10 Chapter Summary and Thesis Alignment

This chapter reviewed the forecasting approaches relevant to energy performance management in buildings and aligned them with the study's research aims. It examined traditional statistical models, machine learning methods, deep learning models, hybrid models, and transfer-learning strategies, highlighting their applicability across forecasting horizons and building contexts. The overarching forecasting workflow used throughout the thesis is shown in Figure 2.1, while (Table 2.1) summarises the suitability of major model families across different time horizons and their expected transferability. Table 2.3 outlines the key research gaps identified in prior literature and maps them to the specific contributions addressed in this thesis.

These synthesis elements guide the methodological decisions in Chapter 3. In particular:
• The horizon mapping in (Table 2.1) informs the selection of models evaluated across the 24-hour, 1-week, 1-month, and 1-year forecasting tasks.

• The transferability indications in Table 2.1 motivate the use of transfer learning for cross-building adaptation.
• The gap contribution alignment in (Table 2.3) provides the rationale for developing the sequential hybrid model (SVR → XGBoost → LSTM) and the deep hybrid model (CNN → GRU → LSTM).

Chapter 3 builds on these insights by introducing the datasets, feature engineering strategy, and model configurations used to operationalise and evaluate the proposed forecasting models across multiple horizons and building types.

A final consideration relates to interpretability. While deep neural networks and hybrid models offer strong accuracy, their internal decision processes can be opaque. To support model transparency and practical relevance, this thesis incorporates feature attribution techniques such as SHAP to identify the most influential drivers of energy consumption. These interpretability analyses accompany the experimental results and contribute to the reliability and practical value of the final forecasting framework.

To consolidate the literature review, the chapter summarised the forecasting models and techniques most relevant to this research, highlighting their strengths, limitations, and alignment with the research gaps and objectives outlined in Chapter 1. This synthesis provides a conceptual bridge to Chapter 3, where the selected models are operationalised and evaluated across multiple forecasting horizons and building context.

Chapter 3: Methodology

3.1 Introduction

This chapter presents the methodological foundation underpinning the research, outlining the experimental framework developed to forecast electricity consumption and detect peak demand within smart building environments. The methodology is grounded in a combination of machine learning, deep learning, hybrid modelling, and edge computing principles, each contributing to the overarching goal of delivering a scalable and intelligent energy forecasting system.

The framework is designed to address several critical objectives. First, it aims to achieve high forecasting accuracy across multiple planning horizons ranging from short-term operational control to long-term strategic energy planning. Second, it prioritizes model adaptability and generalisation, allowing forecasting systems to be transferred across diverse building types with minimal retraining. Third, it incorporates the concept of edge-aware deployment, ensuring that models can operate efficiently under real-time constraints and within low-resource environments. Collectively, these objectives inform a methodological structure that is both comprehensive in scope and modular in execution.

The chapter is organized to reflect the systematic development of this framework, beginning with the research design and followed by detailed descriptions of the experimental tracks, data processing techniques, model architecture, training protocols, evaluation metrics, and deployment strategies. Each section is constructed to support the research objectives while demonstrating a clear progression from conceptual foundation to practical implementation.

The overall research workflow adopted in this thesis is summarized in (Figure 3.1). It reflects a structured model starting from data preparation to model development, forecasting, evaluation, and deployment via transfer learning and edge computing

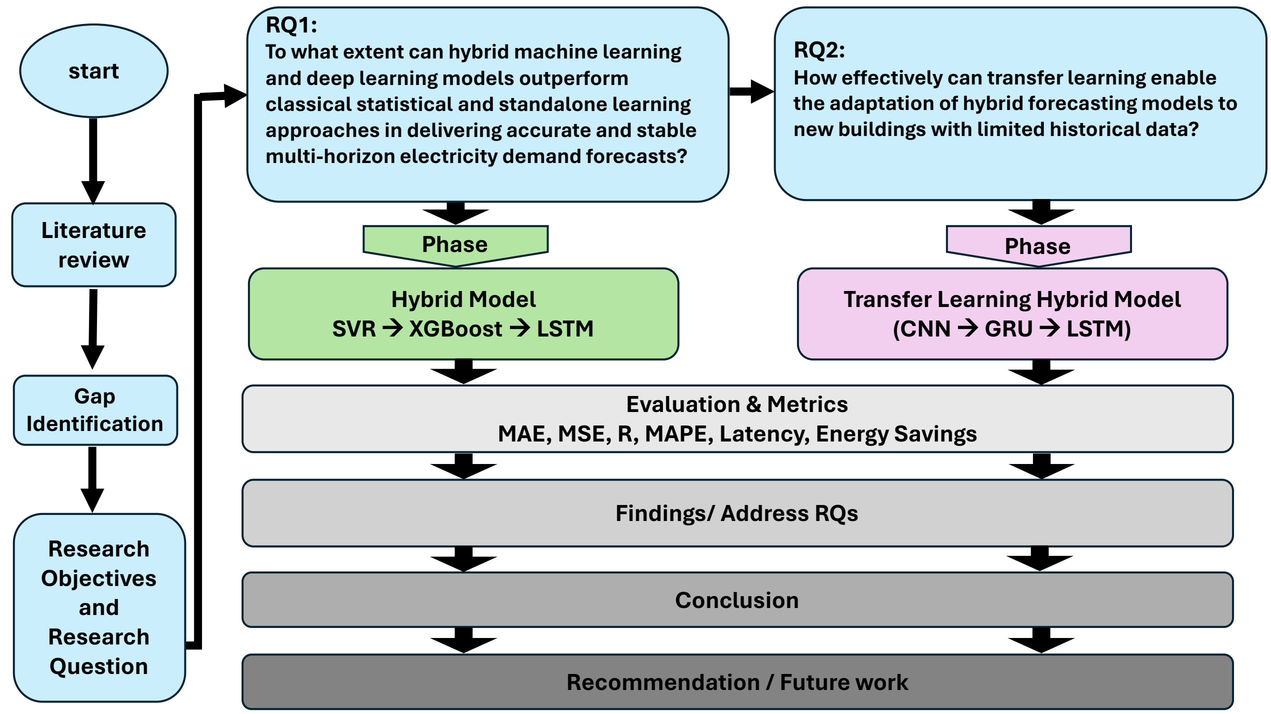

Figure 3.1: Overview of the Thesis Methodology

This diagram outlines the main steps followed in the research: from data collection and model development, through forecasting and evaluation, to final deployment using transfer learning and edge-based systems.

3.2 Research Philosophy, Approach, and Methodological Positioning

The methodological orientation of this thesis is grounded in a positivist research philosophy, which assumes that patterns in building electricity consumption can be observed, measured, and modelled objectively. This stance is well aligned with the nature of energy forecasting, where empirical data rather than subjective interpretation forms the basis for inference and evaluation. Positivism is widely adopted in data-intensive and modelling-driven disciplines because it emphasises reproducibility, quantification, and the search for stable relationships within data [156]. In the context of this research, it supports a systematic examination of how different model families respond to changes in temporal horizon, building type, and feature composition.

Aligned with this philosophical position, the study adopts a deductive research approach, beginning with established theories in time-series modelling, nonlinear learning, and deep neural networks, and examining how these theories hold under the conditions encountered in real building datasets. A deductive stance allows existing concepts such as temporal dependency structures or feature transferability to be tested in controlled experiments, with findings used to confirm, refine, or challenge expectations. Such an approach is characteristic of quantitative modelling research, where hypotheses about performance or generalisability are assessed through structured empirical analysis [157].

Methodologically, the thesis employs an experimental and computational modelling design, which is the dominant paradigm for evaluating forecasting algorithms in engineering and computer science. This involves a deliberate manipulation of modelling choices (e.g., feature sets, architectures, transfer-learning configurations) and an examination of their consequences across multiple datasets and forecasting horizons. The approach is rigorous and transparent: models are trained, validated, and compared using consistent procedures; their behaviour is analysed using statistical diagnostics; and performance improvements are attributed to identifiable components within the modelling pipeline. This form of controlled experimentation is crucial when developing and evaluating hybrid architectures, where the contribution of each modelling layer must be carefully assessed [158].

Taken together, this philosophical and methodological foundation ensures that the research remains both technically rigorous and epistemologically coherent. It establishes a clear and systematic pathway from theoretical assumptions to empirical evaluation, supporting transparent and well-grounded analytical decisions throughout the study. By sitting within established methodological traditions, the thesis demonstrates that its findings are scientifically robust and directly relevant to the advancing field of AI-driven building energy management.

3.3 Research Design

This research adopts a modular experimental design that integrates machine learning, deep learning, hybrid modelling, and transfer learning to develop a comprehensive framework for electricity demand forecasting in smart buildings. The design is structured to address major challenges identified in the literature, including short-term operational control, long-term planning, peak-event anticipation, and cross-building generalisation.

A central element of this design is the use of four forecasting horizons 24 hours, 1 week, 1 month, and 1 year which reflect the multi-layered temporal structure of decision-making in building energy management. Short-term horizons (e.g., 24 hours) are widely used for HVAC scheduling, demand response, and local control optimisation [159]. Medium-term horizons (1 week) support maintenance planning, anomaly detection, and weekly operational adjustments. Monthly forecasting links directly to budgeting, tariff optimisation, and occupancy-related planning. Annual forecasts are essential for long-term capacity planning, retrofit strategies, and sustainability reporting [65]. Prior studies consistently show that forecast performance varies significantly across these horizons, making multi-horizon evaluation essential for assessing model robustness and practical relevance:

- **24 Hours (Short-Term Forecasting):** This horizon supports immediate control actions such as HVAC scheduling, lighting control, and short-term load shifting. Accurate hourly predictions enable energy managers to respond to daily fluctuations, particularly peak hours, and optimise intra-day operations.
- **1 Week (Medium-Term Operational Planning):** Weekly forecasts assist in workforce scheduling, equipment maintenance planning, and weekly budgeting of energy usage. They help detect cyclic consumption patterns associated with business or occupancy cycles, such as weekday/weekend shifts.
- **1 Month (Tactical Forecasting):** This horizon is crucial for aligning energy use with monthly billing cycles, identifying seasonal shifts in demand, and planning facility-level adjustments. It also supports proactive adjustments in building automation systems and operational logistics.
- **1 Year (Strategic and Policy-Level Forecasting):** Long-term forecasts inform capital investment decisions, renewable energy integration strategies, and infrastructure planning. They are essential for sustainability reporting, demand-side participation in energy markets, and evaluating the long-term impact of energy conservation measures.

By testing models across these four horizons, the research ensures a holistic evaluation of forecasting capability and its real-world utility across both reactive and proactive energy management scenarios. This design also enables the comparison of model robustness and adaptability across short and long-term planning needs, a gap often neglected in existing literature.

The following sections detail the data preparation process, model architecture, evaluation criteria, and deployment strategies used to support this multi-scale research framework.

3.4 Datasets and Preprocessing

3.4.1 Dataset Overview

The dataset employed in this thesis is a multi-site collection of hourly electricity consumption data obtained from 19 office buildings across the United States and Europe. These buildings were selected to provide a representative sample of typical commercial environments, encompassing variations in size, functional usage, geographic location, and occupancy schedules. The use of whole-building electricity meter readings ensures the data captures comprehensive demand profiles without being confounded by sub-metering complexities or gas usage components.

Each data record corresponds to an hourly measurement, resulting in a time series of energy consumption values expressed in kilowatts (kW). The recordings span a continuous period of approximately two years, providing a rich temporal dataset suitable for both short- and long-term forecasting experiments. These data are timestamped with hourly granularity, facilitating multi-resolution modelling across 24-hour, 1-week, 1-month, and 1-year horizons.

In addition to the raw electricity demand values, the dataset incorporates a suite of supplementary features that enhance its predictive capacity. These include meteorological variables such as air temperature, dew point, wind speed, and sea-level pressure, all of which are known to significantly influence energy consumption patterns in buildings. These weather variables were aligned temporally with the electricity readings to ensure feature integrity across the dataset.

Furthermore, a range of temporal features was engineered to reflect the operational and contextual influences on energy use. These include the hour of day, day of the week, and binary flags indicating weekends and public holidays. For buildings affiliated with academic institutions, academic calendar flags were added to account for variations in energy use during breaks and term time.

The dataset underwent a comprehensive preprocessing phase, during which incomplete or anomalous records were removed. This included the exclusion of days with fewer than 24 hourly entries and the application of interquartile range (IQR)-based filtering to eliminate statistical outliers. Missing weather data were imputed using linear interpolation, while all-time series were synchronized to a uniform hourly frequency.

Collectively, this dataset supports robust training and evaluation of machine learning and deep learning models under diverse real-world conditions. Its high resolution extended temporal coverage, and breadth of contextual features make it well-suited for both model benchmarking and advanced experimentation in electricity demand forecasting. A concise description of every dataset, its provenance and its role in the study is provided in Appendix B.

3.4.2 Feature Engineering

Effective feature engineering is central to the success of electricity demand forecasting models. This thesis incorporates a diverse set of features to capture the temporal, environmental, and operational factors that influence electricity consumption in office buildings. These features fall into three main categories: weather-based variables, lagged historical consumption, and time-related indicators.

1. Weather Variables:

Multiple studies provide strong empirical evidence for the influence of meteorological conditions on energy consumption. Hor C et al. [160] demonstrated that weather variables significantly affect electricity system operations, while [161] reported strong correlations between ambient temperature and electricity use, with correlation coefficients ranging from 0.82 to 0.94. More recently, Wang et al. [162] showed that incorporating weather variables such as wind speed, ambient temperature, and precipitation can substantially reduce forecasting error from 27.45% down to 8.65% highlighting the value of meteorological predictors in building-level energy models. The following meteorological features were included:

- Air temperature (°C) – Strongly correlated with HVAC-related electricity demand.
- Dew point (°C) – Helps capture humidity-driven cooling/heating patterns.
- Wind speed (m/s) – May influence heat loss or gain through building envelopes.
- Sea-level pressure (hPa) – Can be associated with weather systems affecting temperature trends.

Consistent with these findings, this thesis sources hourly meteorological data from the nearest local weather stations [163] and aligns them with the building electricity data via timestamp matching. This ensures that the forecasting models incorporate accurate, high-resolution environmental drivers known to influence HVAC demand and overall energy use.

2. Lagged Consumption Features:

Lagged features are essential in time series forecasting, as they allow the model to learn from recent usage patterns. The following lagged electricity values were engineered:

- Hourly lags: Electricity consumption values at previous 1, 2, 3, and 4 hours.
- Daily lags: Electricity consumption from the same hour on previous days (e.g., 24, 48, and 72 hours before).
- Rolling statistics: Optional features such as rolling mean and standard deviation over fixed windows (e.g., 3-hour, 6-hour) may be included to capture short-term trends.

Lagged load values are consistently proven to be the most robust predictors in short-term electricity forecasting, with multiple studies demonstrating their critical importance in capturing demand patterns. Research shows that autoregressive lag structures are fundamental to state-of-the-art load forecasting models [37]. For instance, Koprinska et al. [122] found that autocorrelation-based feature selection can identify highly relevant features, [164] developed a selective order autoregressive model that carefully selects optimal lag numbers. The effectiveness stems from capturing temporal dependencies: multi-sequence time lags can improve prediction accuracy by encompassing past information from multiple timescale sequences [37]. These approaches enable models to better represent daily cycles, occupancy patterns, and short-term trends, ultimately enhancing the precision of electricity demand forecasts.

3. Temporal Indicators:

Temporal features are critical for accurately modelling building energy consumption, with time-based contextual variables significantly improving predictive performance. Wu J et al. [165] found that time-related factors, especially week of the year and day of the week, have the most substantial impact on energy consumption across building types. Temporal features were introduced to provide models with structured time-based context:

- Hour of day (0–23): Captures intra-day usage patterns.
- Day of week (0–6): Helps identify weekday vs. weekend behaviour.
- Weekend flag (binary): Indicates non-working days.
- Holiday flag (binary): Identifies national/public holidays.
- Academic calendar flag (optional): For buildings related to educational institutions, this marks term time vs. break periods.

4. Feature Transformation & Encoding:

To represent the hour of the day as a cyclical (circular) feature so that midnight (0:00) and 11 PM (23:00) are close in value you can use the following trigonometric transformations:

$$Hour_{sin} = sin\left(\frac{2\,\pi \cdot hour}{24}\right), \quad Hour_{cos} = cos\left(\frac{2\,\pi \cdot hour}{24}\right)$$

- hour: The hour value (0 to 23)
- These convert the hour into two values (sine and cosine) in a unit circle.
- This ensures the model understands that hour 0 and hour 23 are *close*, not far apart like in linear scale

5. Input Structuring per Model:

- For traditional models (Linear Regression, SVR, XGBoost), features were concatenated into flat vectors.
- For deep learning models (LSTM, GRU, Transformers), the input was reshaped into sequences (e.g., 24-time steps × N features) to capture temporal dependencies.

By combining weather conditions, historical consumption, and calendar-aware indicators, the feature set provides a rich and multidimensional input representation that supports accurate and generalisable electricity demand forecasting.

3.5 Forecasting Models

This thesis explores a diverse range of forecasting models to evaluate their effectiveness in predicting electricity demand across multiple time horizons. The selected models represent traditional statistical methods, machine learning techniques, ensemble learning algorithms, deep neural networks, and advanced hybrid models. Each model brings distinct strengths in addressing different aspects of temporal forecasting, such as pattern recognition, seasonality, nonlinear dynamics, and long-term memory.

The inclusion of these model families is methodologically justified. Traditional statistical models such as ARIMA and Linear Regression remain widely used baselines due to their interpretability and strong performance in settings with clear seasonality and limited data, as demonstrated in building energy studies such as [166], [167]. Machine learning techniques like Support Vector Regression and XGBoost have shown superior ability to capture nonlinear relationships and complex feature interactions in load forecasting [168]. Deep learning architectures including LSTM, GRU, and CNN have been adopted extensively in recent research because of their capacity to model long-term temporal dependencies, multivariate inputs, and hierarchical patterns [169].

The methodological design of this thesis (Section 3.2) therefore incorporates these models to build a comparative evaluation framework, ensuring that each category's contribution can be assessed under consistent conditions. Accordingly, the study evaluates traditional approaches (e.g., ARIMA, Linear Regression), machine learning methods (SVR, XGBoost), and deep learning models (LSTM, GRU, CNN, Transformers). These models are trained and tested across four forecasting horizons (24 hours, 1 week, 1 month, and 1 year), enabling rigorous assessment of their performance across operational, tactical, and strategic temporal scopes [170], [171].

Embedding these models within a unified methodological framework allows for transparent benchmarking, supports the development of the hybrid models proposed later in the thesis, and ensures that model selection is grounded in both empirical evidence and best practice within the load-forecasting literature.

3.5.1 Linear Regression (LR)

Linear Regression (LR) is employed as a benchmark model due to its simplicity and interpretability. It assumes a linear relationship between the input features and the target variable electricity consumption. While LR lacks the capacity to capture temporal dependencies or nonlinear dynamics, it offers transparency and serves as a useful baseline to assess the performance of more complex models. Its low computational cost and ease of implementation make it a foundational reference point in this study.

3.5.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a powerful supervised learning algorithm that leverages kernel functions to model complex nonlinear relationships in data. SVR is particularly effective in small datasets and noise-prone environments, offering strong generalisation capabilities. In this thesis, SVR plays a critical role as the initial stage in the Hybrid (SVR → XGBoost → LSTM) model, acting as a noise reduction layer. By smoothing irregularities in the input signal, SVR improves the quality of the data fed into subsequent stages. Although SVR can be computationally intensive on large datasets, its strength in denoising and outlier handling enhances overall model performance.

3.5.3 Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced ensemble learning method based on gradient-boosted decision trees. Known for its speed and predictive accuracy, XGBoost excels at modelling structured data with complex feature interactions. In this thesis, XGBoost serves as the second component in the hybrid model. After SVR performs initial denoising, XGBoost captures non-linear patterns and intricate relationships among the engineered features, such as lagged consumption, weather variables, and temporal indicators. With built-in regularization, XGBoost reduces overfitting and efficiently handles missing values, making it well-suited for medium- and long-term electricity demand forecasting.

3.5.4 Transformer-Based Models

Transformer models have recently gained attention in time-series forecasting due to their ability to model long-term dependencies without recurrence. Based on self-attention mechanisms, transformers allow the model to assign different weights to different time steps dynamically, depending on their relevance. This model supports parallel processing, enabling faster training and scalability to longer sequences. In this thesis, transformer-based models are investigated particularly in the context of transfer learning, where they are pre-trained on data-rich source buildings and fine-tuned on target buildings. Their ability to generalise across domains and capture global dependencies makes them a promising tool for cross-building energy forecasting.

3.5.5 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are typically used in image processing, but they have been successfully applied to time series data for extracting local temporal patterns. In this study, CNNs serve as the first stage in the Hybrid (CNN → GRU → LSTM) model. CNN layers detect short-term structures and repetitive patterns in the energy consumption sequences, such as peak timings or recurring daily fluctuations. Their use as front-end feature extractors enhances the model's ability to detect important signals before passing the processed input to recurrent layers for deeper temporal modelling.

3.5.6 Gated Recurrent Units (GRU)

Gated Recurrent Units (GRUs) are a streamlined variant of LSTM networks that provide similar temporal learning capabilities with fewer parameters and reduced computational complexity. Positioned between the CNN and LSTM layers in the Hybrid (CNN → GRU → LSTM) model, GRUs are responsible for modelling mid-range dependencies. Their efficient gating mechanism makes them particularly suitable for capturing patterns that span several hours or days without introducing excessive training overhead training. GRUs acts as a computational bridge between shallow convolutional feature extraction and deep temporal memory in LSTMs.

3.5.7 ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a traditional statistical method used in time series analysis. It models future values based on past observations using autoregressive, differencing, and moving average components. While ARIMA performs well in univariate time series with strong trends or seasonality, it struggles with multivariate inputs and nonlinear behaviour. In this thesis, ARIMA is included as a baseline model to benchmark the performance of modern ML and DL approaches. Its comparative simplicity highlights the advantages of incorporating weather data, lagged features, and advanced neural network models.

3.6 Training and Validation Strategy

To ensure the reliability, generalisability, and performance of the forecasting models, a standardized training and validation protocol was adopted across all experiments. This strategy governs how the dataset was split, how models were validated during training, and how performance was evaluated on unseen data for each forecasting horizon.

3.6.1 Data Splitting Approach

The dataset was segmented into training and testing subsets using a consistent 80:20 temporal split across all forecasting horizons. This means that 80% of the earliest time-series data was used for model training, and the remaining 20% was reserved strictly for testing. The split was performed chronologically, preserving the temporal order of observations to avoid data leakage, a crucial aspect in time-series forecasting.

This approach is widely recommended in literature, as random shuffling can cause future information to leak into the training process, thereby inflating performance metrics [172]. A temporal hold-out is particularly important in energy forecasting because consumption patterns exhibit strong autocorrelation and seasonality, which must be respected when evaluating model generalisation [173]. The 80:20 split is also consistent with common practice across load-forecasting studies, offering a balance between sufficient training data and a representative testing window for robust evaluation [174].

Each forecasting horizon followed a tailored windowing strategy:

Short-term load forecasting using a 30-day training window and 1-day testing window is empirically validated as an effective approach for capturing immediate demand dynamics. Multiple studies demonstrate the robustness of this methodology [175].

- **24-hour Forecasting:**
  - Training Window: ~30 days
  - Testing Window: final 1 day
  - Rolling windows are used for short-term model inputs and outputs.

Machine learning 1-week forecasting typically employs a 90-day training window with the final 7 days reserved for testing, using weekly lag features to ensure temporal alignment.

Multiple studies across domains like energy, solar, and wind power demonstrate this approach's effectiveness [176].

- **1-week Forecasting:**
  - Training Window: ~90 days

- Testing Window: final 7 days
- Weekly lag features were computed for temporal alignment.

The 1-month forecasting approach involves using a ~12-month training window and a final 1-month testing window, with careful consideration of seasonal trends and monthly data aggregations. Léger et al. [177] demonstrate this methodology in mortality forecasting, using monthly death counts and predicting one month ahead based on expected ratios.

- **1-month Forecasting:**
    - Training Window: ~12 months
    - Testing Window: final 1 month
    - Seasonal trends and monthly aggregations were considered.

The provided sources validate a 1-year forecasting approach using 4-5 years training windows, demonstrating strong potential for capturing long-term patterns and stability. Delogu et al. [178] support this methodology successfully forecast microbial community signals with nearly perfect trend prediction for the first two years.

- **1-year Forecasting:**
    - Training Window: 4–5 years (from extended datasets when available)
    - Testing Window: final 1 year
    - Emphasis placed on long-term patterns and stability.

This setup allows the evaluation of each model's predictive capability over both immediate and extended timeframes and ensures consistent benchmarking across experiments.

3.6.2 Validation During Training

For models requiring internal tuning or deep learning models, a 10% validation set was extracted from within the training data (typically the most recent portion of the training set) to monitor overfitting and support early stopping.

- For SVR and XGBoost, grid search and random search cross-validation were employed on the training data to identify optimal hyperparameters (e.g., regularization, kernel width, number of estimators, learning rate).
- For deep learning models (e.g., LSTM, CNN → GRU → LSTM, Transformer), a validation split (e.g., 10–15%) was applied within each training epoch to monitor the loss function and automatically trigger early stopping if no improvement was observed after a defined patience threshold (e.g., 10 epochs).

3.6.3 Walk-Forward Validation

Walk-forward (rolling-window) validation was applied in experiments where models were required to simulate real-time operational forecasting specifically in the transfer learning evaluations and multi-horizon forecasting tests. In these scenarios, the goal was to mimic how predictions would be generated in practice, using only information available up to each prediction time point.

Under this approach, the model is trained on all data up to time ***T***, used to forecast ***T+1***, and then the training window is advanced sequentially. This procedure is widely recommended for time-series applications because it preserves temporal causality and prevents data leakage in dynamic forecasting environments.

This method:

- Mimics real-world online learning behaviour.
- Provides a dynamic assessment of model adaptability to changing conditions.
- Help evaluate performance degradation or improvement over time.

3.6.4 Performance Aggregation

Model predictions were evaluated on the test set using consistent evaluation metrics described in Section 3.8 (MAE, MSE, R^2, MAPE, accuracy). Peak detection models were further assessed using precision, recall, and F1-score for classifying high-consumption events.

Results across multiple runs (e.g., five repetitions with different random seeds) were averaged to reduce the impact of stochastic training effects, particularly in neural network models.

3.7 Hybrid and Deep Learning Models

To enhance forecasting accuracy and model flexibility, this thesis explores hybrid and deep learning models that combine the strengths of multiple modelling techniques. These models are designed to address the limitations of single-model approaches by integrating complementary component statistical denoising, structured feature learning, and deep temporal sequence modelling. In particular, the hybrid models developed in this research aim to provide robust electricity demand forecasting across multiple time horizons while also supporting transfer learning and edge deployment.

3.7.1 The Hybrid (SVR → XGBoost → LSTM) Model

One of the central contributions of this thesis is the design and implementation of a sequential hybrid model composed of Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) networks. This model leverages the unique capabilities of each component model in a three-stage process. SVR serves as the first layer, responsible for denoising the input signals and reducing the effects of high-frequency fluctuations or outliers in the raw consumption data. This preprocessing stage helps stabilize the input before deeper learning is applied.

The workflow of the Hybrid (SVR → XGBoost → LSTM) model is detailed in (Table 3.1), which outlines the sequential processing of denoising, feature interaction, and long-term temporal modelling. The corresponding model is illustrated in (Figure 3.2), showing how outputs from each stage are passed forward to build a unified forecasting system. Full layer configurations and hyper-parameter settings for every model variant are listed in Appendix A.

Figure 3.2: Hybrid Forecasting Model Combining SVR, XGBoost, and LSTM Models

Following SVR, XGBoost is used to model structured feature interactions. Its gradient-boosted trees are effective in capturing nonlinear relationships among weather features, lagged values, and temporal indicators. This stage adds interpretability and improves robustness against overfitting. Finally, the LSTM layer is applied to model the temporal dependencies and seasonality present in the time-series data. It receives the structured input from XGBoost and learns to forecast future consumption values by maintaining memory across multiple time steps. This hybrid model is evaluated across four forecasting horizons (24 hours, 1 week, 1 month, and 1 year) and includes a built-in peak detection component, making it suitable for both forecasting accuracy and high-demand identification tasks.

Table 3. 1: Pseudocode for the SVR → XGBoost → LSTM Hybrid Forecasting Workflow

Input	Feature matrix X, Target vectory y_{hat_svr}
Step 1:	Train SVR on X to predict y
	svr_model = SVR (…).fit(X, y)
	y_{hat_svr} = svr_model.predict(X)
Step 2:	Train XGBoost on y_{hat_svr} to produce y_{hat_xgb}
	xgb_model = XGBRegressor (…).fit (X, y_{hat_svr})
Step 3:	Train LSTM on y_{hat_xgb} for final output
	X_seq = sequence_generator (y_{hat_xgb})
	y-final = lstm_model.predict (X_seq)
Output:	y_final (Final Forecast)

The sequential hybrid forecasting model (SVR → XGBoost → LSTM) is designed to progressively enhance data quality and capture increasingly complex temporal patterns. SVR performs initial noise reduction to eliminate short-term fluctuations, XGBoost identifies nonlinear feature interactions, and LSTM captures long-range temporal dependencies. This staged structure allows for modular, interpretable learning, which performs robustly across varying horizons and datasets.

This unified model serves as the backbone for all forecasting tasks in Chapters 4 and 5. Unless otherwise stated, experimental setups throughout the thesis reuse this model, retrained on task-specific data for each scenario and horizon.

3.7.2 The Hybrid (CNN → GRU → LSTM) model

To further improve generalisation in transfer learning scenarios, the thesis also introduces a second deep learning-based hybrid model that combines Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and LSTM layers. This model is designed to capture multiscale temporal dependencies in electricity consumption data. The CNN layer acts as a local pattern detector, extracting short-term temporal features such as spikes and recurring fluctuations. These convolutional filters help highlight informative subsequences that might be missed by purely recurrent models.

As shown in (Figure 3.3), the Hybrid (CNN → GRU → LSTM) model leverages convolutional filters for local temporal features, followed by gated recurrent units and LSTM layers for deeper sequential learning. The pseudocode presented in (Table 3.2) provides an overview of the data flow and layer structure within this model. Full layer configurations and hyper-parameter settings for every model variant are listed in Appendix A.

Figure 3.3: Deep Learning Hybrid Model (CNN → GRU → LSTM) for Cross-Building Forecasting

The output from the CNN layer is then passed to GRUs, which are optimised for modelling mid-range dependencies with fewer parameters compared to LSTMs. GRUs offer faster training and efficient memory usage, which is particularly important when dealing with buildings that have limited historical data. Finally, an LSTM layer is appended to learn long-range dependencies and overall consumption trends. This deep hybrid model is especially effective when used in transfer learning settings, where it is pretrained on data-rich source buildings and fine-tuned on target buildings with sparse data.

3.7.3 Transformer-Based Models

Transformer models have emerged as state-of-the-art models for time-series forecasting due to their use of self-attention mechanisms. Unlike recurrent models, transformers can process entire sequences in parallel and assign dynamic attention weights to different time steps, allowing them to focus on the most relevant patterns in the input sequence. In this thesis, transformer-based models are implemented to explore their effectiveness in both standard forecasting and transfer learning across building types.

These models are particularly well-suited for scenarios requiring long-term forecasting over multiple horizons. Their ability to capture complex temporal relationships and scale efficiently with data length makes them an attractive alternative to RNN-based models. Additionally, transformer models are evaluated in the context of cross-building adaptation, where they are pretrained on diverse building datasets and fine-tuned with minimal updates to fit new target domains. This experiment reveals their potential for generalised deployment in real-world energy forecasting applications.

The design of the Transformer-based model, incorporating positional encoding and multi-head self-attention mechanisms, is shown in (Figure 3.4). This model is particularly effective for capturing long-range dependencies in multivariate time series data.

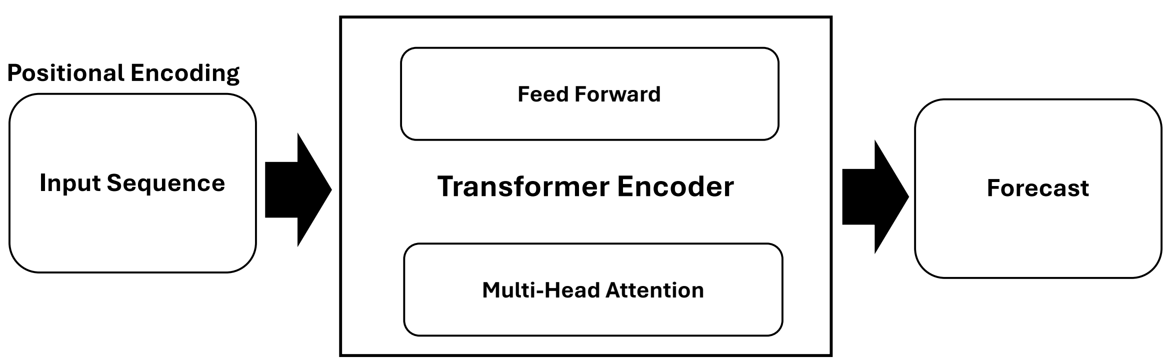

Figure 3.4: Transformer Forecasting Model for Energy Consumption Prediction

3.8 Transfer Learning and Cross-Building Adaptation

A persistent challenge in building energy forecasting lies in the variability of data availability and quality across different buildings. While some buildings particularly large commercial or institutional facilities have access to rich historical energy data, others such as newly constructed, under-instrumented, or retrofitted buildings often suffer from data sparsity. This limitation poses significant obstacles to training robust forecasting models for each building individually. To overcome this, this thesis integrates transfer learning (TL) as a strategic methodology to enable cross-building generalisation and reuse of forecasting models.

Transfer learning is a machine learning paradigm that allows knowledge gained in one domain (referred to as the source domain) to be transferred and applied to another related domain (called the target domain). In the context of this research, the source domain consists of buildings with rich historical energy consumption records and associated metadata, while the target domain includes buildings with limited or no labeled training data. The transfer of learning reduces the reliance on large, labeled datasets for every new building, thereby improving scalability and accelerating deployment in real-world applications.

Table 3.2: Transfer Learning Pseudocode for CNN–GRU–LSTM Model

```
Input         Source data Xs, Target ys, tt
Step 1:       Train hybrid model with CNN → GRU → LSTM layers
              model = Sequential ([[
                      Conv1D (…),
                      GRU (…),
                      LSTM (…) …) …)
Step 2:       Pretrain on source dataset
              model.fit (Xs, epochs = …,…)
Step 3:       Fine-tune on target dataset
              model.fit (Xt , Yt , epochs = …)
Output:       model.prdict (Xt )
```

The transfer learning strategy used in this study follows a two-phase training procedure: pretraining and fine-tuning. During the pretraining phase, the model typically is a deep learning hybrid such as the Hybrid (CNN → GRU → LSTM) or a transformer is trained on a large, diverse dataset from source buildings (e.g., schools, offices, leisure centers) with sufficient temporal variation, seasonal behaviours, and operational characteristics. This phase allows the model to capture generalised patterns in energy usage such as daily consumption rhythms, weather impacts, and occupancy-based variability.

In the subsequent fine-tuning phase, the pretrained model is adapted to a new target building with limited data (such as Ebbw Vale schools). Here, only a subset of the model parameters is updated to align the learned features with the new building's consumption behaviour, while the earlier learned representations are retained. This approach minimizes retraining overhead and reduces the risk of overfitting to a small dataset, while still achieving high predictive accuracy.

Notably, this thesis evaluates the effectiveness of transfer learning using both the Hybrid (CNN → GRU → LSTM) model and transformer models. The models are benchmarked on their ability to maintain forecasting accuracy across domains with different temporal dynamics, structural characteristics, and weather influences. The results demonstrate that with appropriate fine-tuning, transferred models perform comparably or even better than models trained from scratch on the limited target data.

To further support domain adaptation, the study includes domain alignment practices such as feature scaling consistency, temporal alignment, and structural regularization to minimize discrepancies between source and target domains. Additionally, the hybrid models are designed with modular flexibility, enabling selective layer freezing, gradual unfreezing, and learning rate scheduling during fine-tuning to improve adaptation outcomes.

In summary, transfer learning in this thesis serves as a core enabler for cross-building energy forecasting, significantly improving the feasibility of deploying intelligent models in buildings with sparse historical data. It also contributes to the broader goal of sustainable and scalable smart building systems by reducing the cost and complexity associated with model development for each new environment.

3.9 Peak Detection Mechanism

Accurate identification of peak electricity demand periods is essential for energy-efficient operation, cost reduction, and grid stability in smart buildings. While forecasting average or total energy consumption provides a general understanding of building behaviour, the ability to predict when extreme or peak usage occurs enables proactive energy management, such as load shifting, demand response, and integration with renewable energy resources.

In this thesis, peak detection is integrated directly into the forecasting framework as a post-processing layer that analyses forecast outputs from various models including the Hybrid (SVR → XGBoost → LSTM) model and the Hybrid (CNN → GRU → LSTM) model. This approach ensures that peak identification is informed by robust time-series forecasts rather than isolated statistical thresholds applied to raw data.

The chosen method for detecting peaks is a statistical thresholding technique, which is both interpretable and adaptable across different time horizons (24 hours, 1 week, 1 month, and 1 year). Specifically, for each forecast horizon, the predicted consumption values are evaluated against a dynamic threshold defined as:

$$Peak\ Threshold = \mu + \sigma$$

Where:

- μ is the mean forecasted consumption over the relevant time window,
- σ is the standard deviation of the forecasted values.

Any time step where the forecasted consumption exceeds this threshold is classified as a peak event.

This method is particularly advantageous because it accounts for temporal variations in consumption and adapts to buildings with different load profiles. For example, an office building may show predictable weekday peaks around mid-afternoon, while a leisure center may have sporadic peaks on weekends. The dynamic thresholding approach adjusts accordingly based on the statistical behaviour of each prediction horizon.

To validate the accuracy of peak detection, the thesis evaluates the following metrics:

- Precision: the proportion of correctly identified peaks out of all predicted peaks,
- Recall: the proportion of actual peaks that were correctly identified,
- F1-score: the harmonic meaning of precision and recall.

These metrics are computed across all four forecasting horizons to ensure the robustness of peak detection in both short-term (e.g., hourly control) and long-term (e.g., seasonal planning) scenarios.

In addition to standalone evaluation, peak detection is analysed in conjunction with model performance metrics such as MAE and R^2 to ensure that high-accuracy forecasts are also yielding reliable peak identification. This holistic evaluation ensures that the forecasting models serve both regular consumption monitoring and extreme demand detection needs simultaneously.

By embedding this peak detection mechanism into the core forecasting model, the thesis provides a unified solution for electricity demand forecasting and critical event detection supporting both operational efficiency and strategic energy planning in smart buildings.

3.10 Evaluation Metrics

To ensure the reliability, accuracy, and practical value of electricity consumption forecasts and peak demand detection, this thesis adopts a diverse set of evaluation metrics. These metrics assess the performance of each model from multiple perspectives, including error magnitude, variance capture, and classification precision for peak events. The combination of regression and classification metrics enables a thorough analysis of how well the models perform across different tasks and time horizons.

3.10.1 Metrics for Forecasting Accuracy

The following metrics are used to quantify the accuracy of continuous electricity demand predictions across four forecasting horizons (24 hours, 1 week, 1 month, and 1 year):

Mean Absolute Error (MAE):

- Measures the average absolute difference between predicted and actual electricity consumption values.

- Provides an intuitive measure of prediction accuracy, making it easier to interpret errors in real-world energy forecasting.

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

Where:

- $\boldsymbol{y_i}$: Actual value at time $\boldsymbol{i}$
- $\boldsymbol{\hat{y}_i}$: Predicted value at time $\boldsymbol{i}$.
- $\boldsymbol{n}$: Total number of observations.

MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It is straightforward to interpret and gives equal weight to all errors. A lower MAE indicates more accurate predictions and is especially useful for comparing model performance on the same scale.

Mean Squared Error (MSE):

- Penalizes larger prediction errors more heavily compared to MAE.
- Suitable for forecasting applications where extreme values are important.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

MSE penalizes larger errors more heavily than MAE due to squaring. It is sensitive to outliers and particularly useful in scenarios where large deviations from the actual value are costly or disruptive.

Coefficient of Determination (R^2 Score):

- Measures the proportion of variance in the dependent variable that is explained by the model.
- Ranges from 0 to 1, where 1 indicates a perfect fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Where:

- $y_i - \bar{y}_i$: Mean of the actual values.

1- Mean Absolute Percentage Error (MAPE)

- Expresses the forecasting error as a percentage, making it useful for comparing models across different datasets.

The R² score indicates how well the forecasted values explain the variance in the observed data. An R² of 1.0 represents perfect prediction, while values close to 0 suggest poor explanatory power. It is especially valuable for evaluating how much of the variation in energy consumption is captured by the model.

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

MAPE expresses forecasting error as a percentage, making it useful for comparing model performance across datasets with different consumption scales. However, it can become unstable if actual values are close to zero.

Root Mean Squared Error (RMSE):

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

RMSE is the square root of MSE and presents errors in the same units as energy consumption (e.g., kWh). It balances sensitivity to large errors with interpretability.

Where:

- y_i: Actual value at time i
- $\hat{y}_i$: Predicted value at time i.
- n: Total number of observations.

- $y_i - \bar{y}_i$: Mean of the actual values.

Forecast Accuracy: This metric is often derived as:

- Represents the percentage of predictions that fall within an acceptable error margin (ϵ).

$$Accuracy = \frac{100\%}{n}\sum_{i=1}^{n}.(|y_i - \hat{y}_i| \leq \epsilon)$$

Where:

- $(\cdot)$: Indicator function that returns 1 if the condition is true and 0 otherwise.
- ϵ: Acceptable error margin (e.g., 5% of the actual value).
- While simple, it offers a high-level view of how often the model's predictions are acceptably close to the actual values, especially for stakeholder communication.

Execution Time:

- \- Parameters:
- \- `start_time`: The time when the model training or prediction starts.
- \- `end_time`: The time when the model training or prediction ends.

$$Execution\,Time = end\_time - start\_time$$

These metrics are computed for all models (Linear Regression, SVR, XGBoost, LSTM, hybrid models, and transformer-based models) and across all forecasting horizons to provide a consistent and comprehensive benchmark.

3.10.2 Metrics for Peak Detection Performance

Since this thesis also integrates a statistical peak detection mechanism, additional classification metrics are used to evaluate how accurately the models identify peak demand periods:

Threshold Calculation

The peak detection threshold is calculated using the mean (μ) and standard deviation (σ) of the historical electricity consumption data. A peak hour is identified when the consumption exceeds the mean plus one standard deviation.

$$Peak\ Threshold = \mu + \sigma$$

Where:

- $\mu = \frac{1}{n}\sum_{i=1}^{n} y_i$: Mean electricity consumption over n observations.
- $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2}$: Standard deviation of electricity consumption.

Peak Severity Classification

To further categorize the severity of peak hours, the following thresholds are applied:

- High Peaks: Consumption exceeds the mean + 2 * standard deviation.

$$High\ Peak\ Threshold = \mu + 2\sigma$$

A high peak hour is identified when:

$$y_i > \mu + 2\sigma$$

- Moderate Peaks: Consumption lies between the mean + 1 * standard deviation and mean + 2 * standard deviation.

$$\mu + \sigma < y_i \leq \mu + 2\sigma$$

- Normal Usage: Consumption remains below the mean + 1 * standard deviation.

Adaptive Thresholding for Real-Time Peak Detection

To handle seasonal variations and changing demand patterns, an adaptive thresholding approach is employed using a rolling window method:

$$\mu_t = \frac{1}{\omega}\sum_{i=t-\omega+1}^{t} y_i$$

$$\sigma_t = \sqrt{\frac{1}{\omega} \sum_{i = t - \omega + 1}^{t} (y_i - \mu_t)^2}$$

Where:

- μ_t: Rolling mean at time *t*.
- σ_t : Rolling standard deviation at time *t*.
- ω : Size of the rolling window (e.g., 7 days for weekly adaptation).

This ensures that peak thresholds adjust dynamically based on shifting consumption trends.

- Precision:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

Precision measures the proportion of predicted peaks that are actual peaks. High precision indicates fewer false alarms.

- Recall (Sensitivity):

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

Recall evaluates the proportion of actual peaks that were correctly identified. High recall indicates effective coverage of high-demand events.

- F1-Score:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The F1-score balances precision and recall, especially when there is an uneven class distribution (e.g., far fewer peak events than normal readings). A high F1-score indicates strong overall peak detection performance.

These metrics ensure that the models not only forecast typical usage with low error but also reliably signal critical demand surges. This dual evaluation framework supports both operational forecasting and decision-making for demand-side management strategies.

3.10.3 Metric Aggregation and Visualization

All performance metrics are aggregated over multiple tests runs to ensure stability and robustness. Visual tools such as line plots, bar charts, and boxplots are used to communicate model performance, error distributions, and peak detection reliability across time horizons.

Together, this rich evaluation scheme allows the thesis to present a well-rounded and rigorous assessment of model effectiveness in real-world energy forecasting applications.

3.11 Summary

This chapter presented a unified and comprehensive methodology that underpins the entire thesis. It consolidated the forecasting strategies, data preprocessing steps, model architectures, and deployment frameworks into a coherent experimental design tailored for intelligent energy management in buildings.

The chapter began by introducing the experimental research design and describing the datasets used, emphasizing the diversity of office buildings across geographic regions and operational patterns. It then detailed the feature engineering strategies used to extract meaningful temporal, weather-related, and lag-based variables for time-series forecasting.

Multiple forecasting models were introduced, ranging from traditional regressors such as Linear Regression and Support Vector Regression (SVR) to more advanced machine learning and deep learning methods like XGBoost, LSTM, and Transformer-based models. The two hybrid models, the Hybrid (SVR → XGBoost → LSTM) and the Hybrid (CNN → GRU → LSTM) were presented as key contributions to this study, offering robust mechanisms to capture both structured feature interactions and long-term temporal dependencies.

Beyond individual models, the methodology incorporated transfer learning techniques to facilitate cross-building generalisation, addressing the common limitation of data scarcity in new or under-instrumented buildings. Peak detection strategies were integrated directly into the prediction framework using statistical thresholding, enabling the identification of high-demand periods critical for grid and cost management.

The chapter also introduced a novel edge deployment strategy, testing model performance and task orchestration under constrained computing environments. By applying autonomic scheduling strategies random, priority-based, and energy-aware the thesis explored how forecasting tasks could be adapted to real-time energy availability and operational constraints on low-power edge devices.

Altogether, as shown in (Figure 3.5) the methodological framework provides a solid foundation for the experimental evaluations conducted in the subsequent chapters. It ensures consistency across all case studies whether focused on peak prediction, edge deployment, or cross-domain learning and supports the thesis's overarching goal of developing intelligent, scalable, and sustainable solutions for energy forecasting in smart buildings.

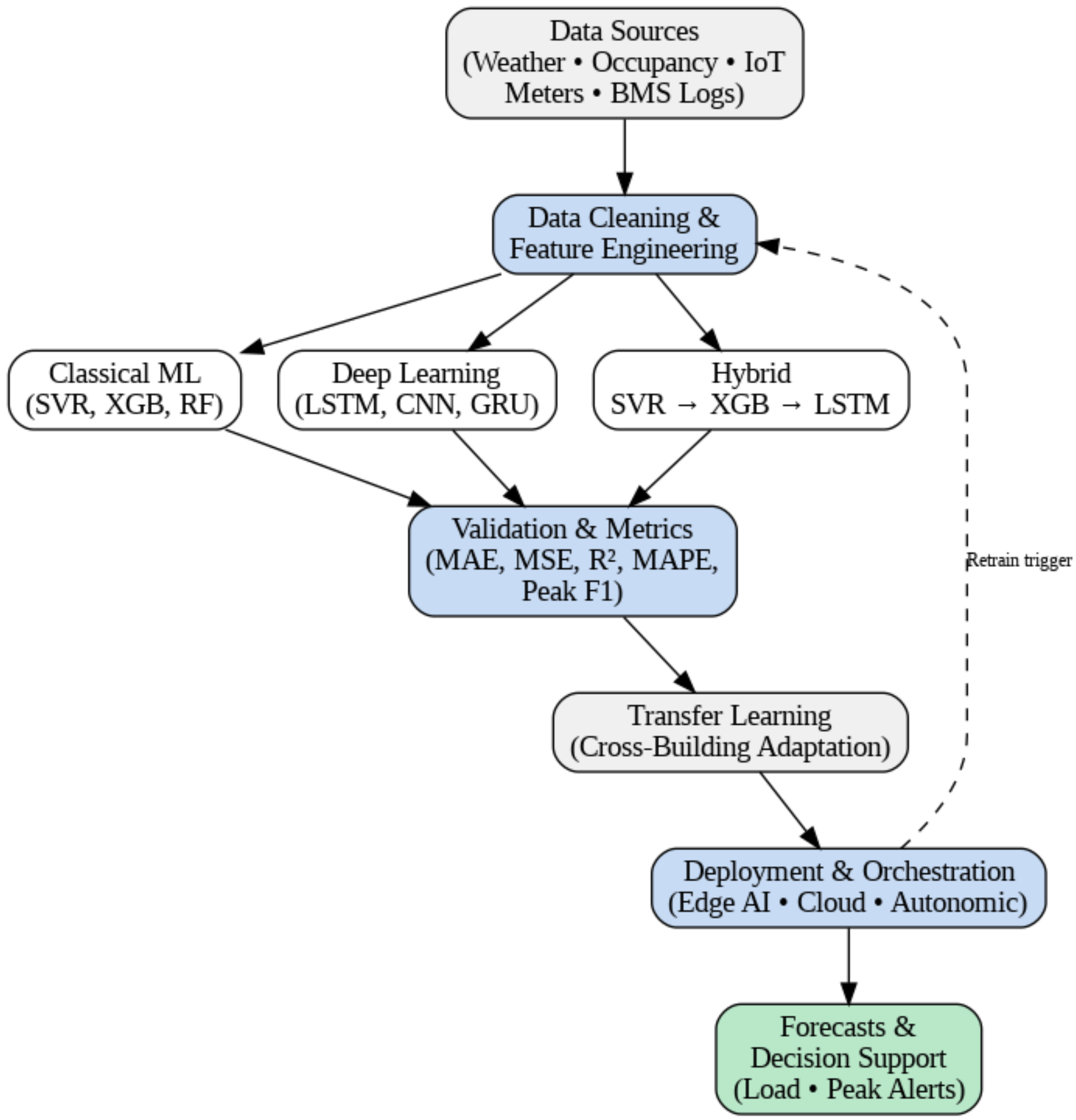

Figure 3.5: Research Framework Flowchart

3.12 Software and Hardware Execution Environment

To ensure reproducibility and evaluate the scalability of the proposed forecasting models and deployment strategies, this study was implemented using well-established software libraries and executed across both high-performance and low-power edge computing platforms. For a comprehensive overview of model configurations, hyper-parameters, and the reproducibility protocol, refer to Appendix A: Model Architectures and Hyperparameters.

3.12.1 Software Stack

All forecasting models, preprocessing models, and evaluation procedures were developed using the Python programming language (version 3.10+), leveraging the following open-source libraries:

- Pandas, NumPy – for data manipulation and numerical computations.
- scikit-learn – for implementing traditional machine learning models (e.g., Linear Regression, SVR, XGBoost) and model evaluation metrics.
- XGBoost – for efficient gradient boosting trees implementation and hyperparameter tuning.
- TensorFlow / Keras – for deep learning models including LSTM, CNN, GRU, and Transformer-based models.
- Matplotlib, Seaborn – for data visualization and graphical evaluation of forecast accuracy.
- Statsmodels – for initial statistical analysis and time series baselines (e.g., ARIMA).
- Optuna – for hyperparameter optimisation (where applicable).

Experiments were conducted in Jupyter Notebook environments to support iterative development, real-time visualization, and result traceability.

3.12.2 Hardware Setup

The methodology was evaluated across two hardware tiers:

1. High-Performance Computing Environment

- Processor: Intel Core i9 / AMD Ryzen 9
- RAM: 32–64 GB
- GPU (for deep learning): NVIDIA RTX 3080 / A100
- Operating System: Windows 11 / Ubuntu 22.04 LTS
- Used primarily for model training, hyperparameter tuning, and high-resolution plotting.
- Enabled parallel experimentation across forecasting horizons.

Chapter 4: Results and Validation

4.1 Introduction

This chapter presents the empirical results of the forecasting models developed in this research and provides a critical evaluation of their performance across multiple time horizons and building contexts. The analysis is designed to directly address the study's research questions and objectives by examining how well the proposed models including the sequential hybrid model (SVR → XGBoost → LSTM) and the deep hybrid transfer-learning model (CNN → GRU → LSTM) perform in realistic smart-building environments.

The chapter is structured to reflect the methodological progression established earlier in the thesis. It begins by assessing multi-horizon forecasting accuracy (Objective 1 / RQ1), comparing traditional baselines, machine-learning models, deep-learning models, and the proposed hybrid approaches across four prediction windows: 24 hours, 1 week, 1 month, and 1 year. This section highlights patterns of performance degradation, temporal stability, and horizon-specific model strengths.

The chapter then examines cross-building generalisation using transfer learning (Objective 3 / RQ2). Here, the focus is on how pretrained models adapt when transferred to new buildings with limited historical data, and whether hybrid and deep models can retain accuracy under domain shift. These experiments address a central gap in the literature by evaluating model robustness in data-scarce scenarios.

A third component provides rigorous validation and reliability testing (Objective 4 / RQ3). Beyond standard accuracy metrics, the chapter incorporates residual diagnostics, Ljung–Box tests, Diebold–Mariano tests, confidence-interval analysis, peak-event evaluation, and feature-importance studies. These analyses deepen the interpretation of model behaviour and establish whether performance differences are statistically meaningful rather than artefacts of noise or horizon length.

Finally, the results are synthesised and discussed in relation to existing literature, highlighting where the models confirm, extend, or challenge prior findings. This ensures that the evaluation is not limited to numerical comparisons but instead offers a conceptual understanding of why models perform as they do, what their limitations are, and how they contribute to building-level energy management.

Together, the analyses presented in this chapter provide a comprehensive assessment of the forecasting framework and establish the empirical foundation for the conclusions drawn in the final chapter.

4.2 Multi-Horizon Forecasting Performance

The forecasting performance of all models was evaluated across four distinct time horizons 24 hours, one week, one month, and one year to assess their effectiveness in both short-term operational scenarios and long-term strategic planning. Short-term forecasts (24 hours) are essential for real-time load balancing and tariff optimisation, while medium- and long-term forecasts (weekly, monthly, and yearly) support budgeting, maintenance scheduling, and energy procurement strategies.

To quantify performance, five comparative metrics were computed across all models: Mean Absolute Error (MAE), Mean Squared Error (MSE), R^2 Score, Mean Absolute Percentage Error (MAPE), and Accuracy. Across all horizons, the Hybrid model consistently outperformed the baseline models, demonstrating superior precision, stability, and generalisation capability. LSTM ranked second overall, particularly excelling in short-term sequential pattern learning, whereas XGBoost delivered strong performance with lower computational requirements. In contrast, SVR and Linear Regression showed substantial limitations, particularly during peak periods and high-volatility intervals.

Extended figures and tables for each forecasting horizon (24 hours, one week, one month, one year) are provided in Appendix C, showing hour-by-hour and day-by-day error trends and peak-detection patterns.

4.2.1 Short-Term Forecasting Performance

To provide a clearer comparison of forecasting model accuracy over the short-term horizon, (Table 4.1) and (Figure 4.1) present the mean residual errors (MAE) for all evaluated models during the 24-hour forecasting period. The results indicate that the Hybrid model outperformed all other models, achieving the lowest MAE value, followed by LSTM and XGBoost. These findings emphasize the Hybrid model’s ability to capture short-term consumption patterns more effectively, making it a strong candidate for real-time energy optimisation in operational settings such as office buildings.

Table 4.1: Mean Residual Errors for the 24-Hour Forecasting Horizon

Model	Mean Absolute Error (MAE) (kWh)
Linear Regression	**4.2**
SVR	**3.1**
XGBoost	**2.4**
LSTM	**1.9**
Hybrid	**1.3**

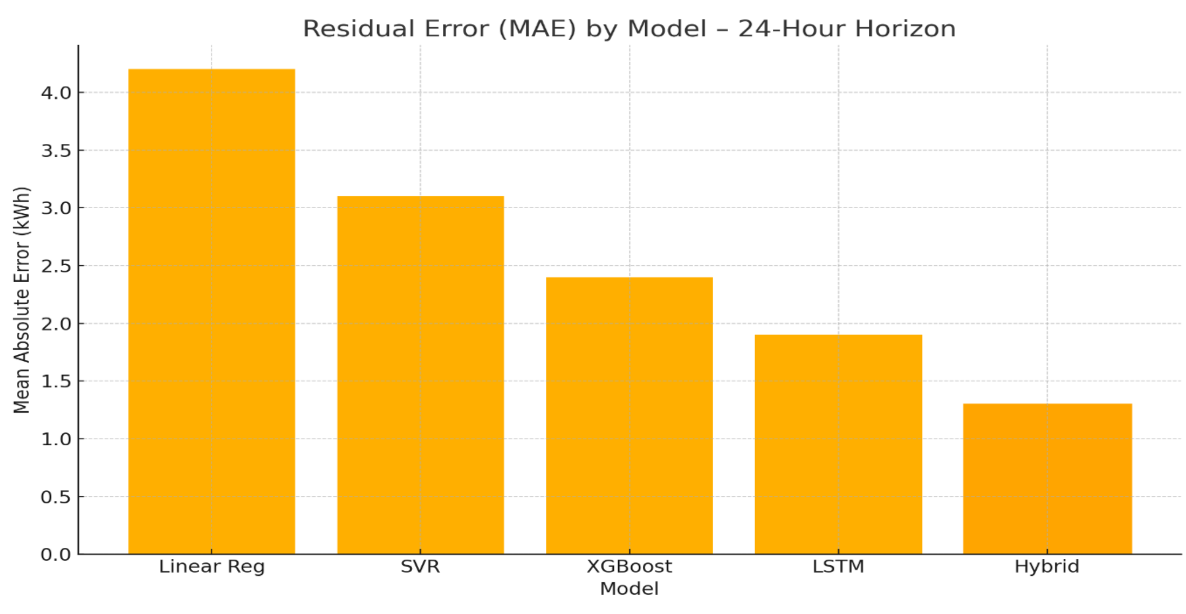

Figure 4.1: Residual Error (MAE) by Model – 24-Hour Horizon

24-Hour Horizon Analysis

Table 4.2 shows that the 24-hour forecasting horizon is pivotal for operational decision-making, such as real-time load balancing and tariff optimisation. The LSTM model emerged as the most accurate standalone model, achieving an MAE of 2.01 and a MAPE of 9.84%. However, the

Hybrid model surpassed all baseline models, reducing the MAE to 1.78 and the MAPE to 8.03%. This improvement is attributed to the Hybrid model's ability to sequentially refine predictions through SVR's feature selection, XGBoost's non-linear pattern detection, and LSTM's temporal dependency modelling.

Table 4.2: Forecasting Accuracy for 24-Hour Horizon

Model	MAE	MSE	R^2 Score	MAPE (%)	Accuracy (%)
Linear Regression	4.52	22.84	0.9125	21.37	78.63
SVR	3.14	12.56	0.9562	15.79	84.21
XGBoost	2.62	9.54	0.9936	12.04	87.96
LSTM	2.01	8.42	0.9958	9.84	90.16
Hybrid	1.78	7.46	0.9972	8.03	91.97

The XGBoost model, while less accurate than LSTM and the Hybrid model, demonstrated competitive performance with a MAPE of 12.04%, making it a viable option for scenarios requiring faster computations. In contrast, Linear Regression and SVR exhibited significant limitations, with MAPEs of 21.37% and 15.79%, respectively. These results underscore the inadequacy of traditional statistical methods for capturing the complexity of electricity demand patterns, especially during peak periods characterized by sudden fluctuations.

As illustrated in (Figure 4.2), the mean error percentages of all models across the 24-hour forecasting horizon show that the Hybrid model consistently achieved lower errors, particularly during critical peak hours. Linear Regression and SVR exhibited noticeably higher errors, while LSTM and XGBoost performed moderately well. The Hybrid model clearly demonstrated superior stability and precision.

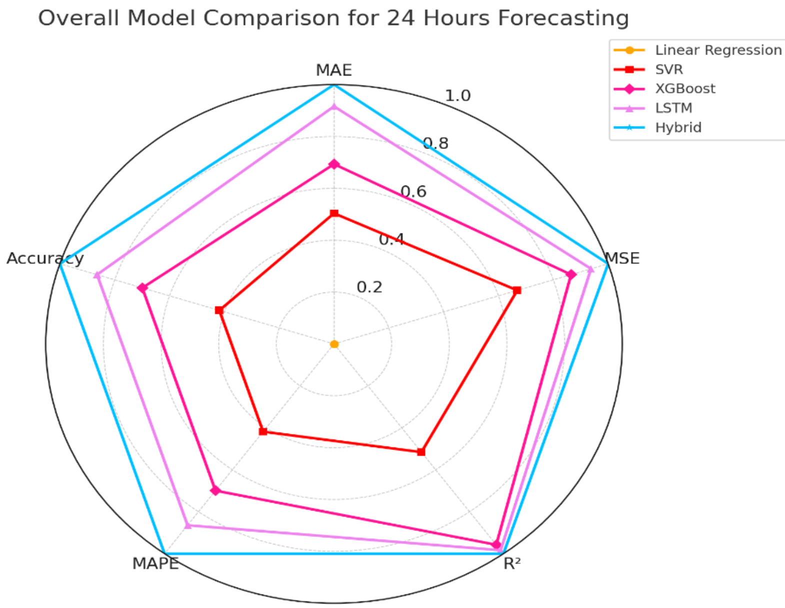

Figure 4.2: Mean Error Percentage of Models Over 24 Hours. The Hybrid model consistently maintained the lowest error rates, particularly during peak demand hours, outperforming LSTM, XGBoost, SVR, and Linear Regression

Weekly Forecasting Trends

The one-week forecasting horizon introduces additional complexity due to the interplay of weekday and weekend demand patterns. As shown in (Table 4.3) that the Hybrid model again achieved the lowest error rates, with a MAPE of 10.26%, compared to 11.67% for LSTM and 14.32% for XGBoost. Notably, all models exhibited higher errors on weekends, likely due to reduced predictability in commercial building occupancy.

Table 4.3: Forecasting Accuracy for One Week Horizon

Model	MAE	MSE	R^2 Score	MAPE (%)	Accuracy (%)
Linear Regression	5.12	28.65	0.8932	24.51	75.49
SVR	4.02	16.75	0.9438	18.45	81.55
XGBoost	3.45	11.23	0.9721	14.32	85.68
LSTM	2.87	9.11	0.9854	11.67	88.33
Hybrid	2.56	8.22	0.9894	10.26	89.74

The Hybrid model clearly outperformed all others, followed by LSTM and XGBoost which delivered strong but less consistent performance. As shown in (Figure 4.3), model performance over the weekly horizon further reinforces the Hybrid model's robustness. During typical workdays, which coincide with higher occupancy and demand, the Hybrid model maintained the lowest error rates. LSTM followed closely, while Linear Regression continued to show the weakest performance.

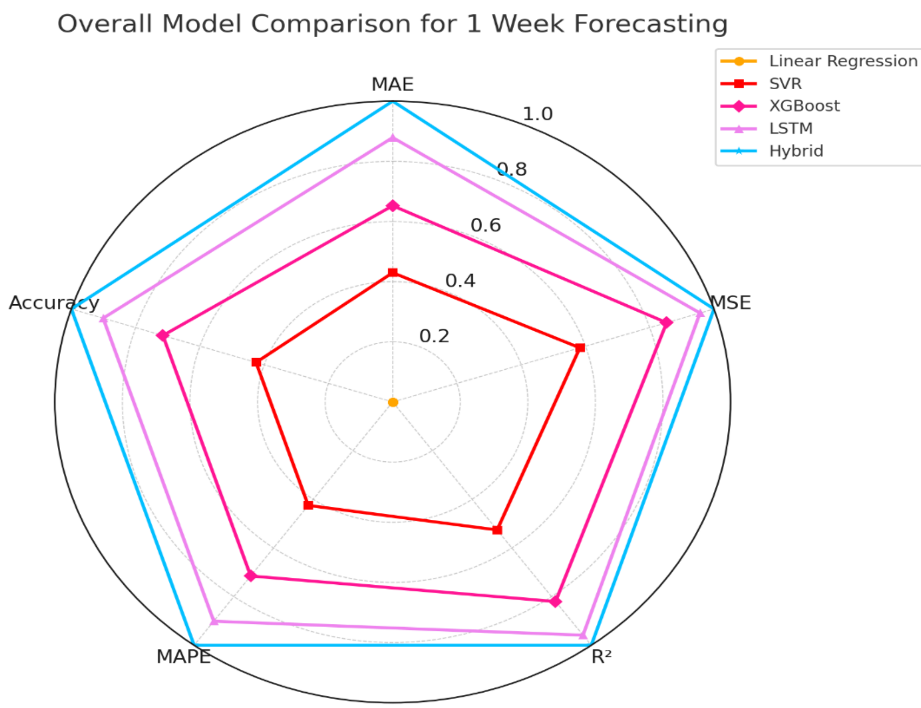

Figure 4.3: Mean Error Percentage of Models Over One Week. The Hybrid model demonstrated superior performance during weekdays, capturing variations linked to higher occupancy, while other models exhibited higher fluctuations and underpredictions

4.2.2 Medium- to Long-Term Forecasting Performance

One-Month Forecasting

For monthly forecasting, the Hybrid model achieved a MAPE of 14.39%, outperforming LSTM (15.48%) and XGBoost (18.03%). The performance gap between the Hybrid model and traditional methods widened further, with Linear Regression and SVR yielding MAPEs above 22% as shown in (Table 4.4). This divergence highlights the Hybrid model's ability to adapt to broader demand fluctuations, including those driven by monthly billing cycles or seasonal transitions.

Table 4.4: Forecasting Accuracy for One Month Horizon

Model	MAE	MSE	R² Score	MAPE (%)	Accuracy (%)
Linear Regression	6.42	35.76	0.8651	28.34	71.66
SVR	5.02	22.31	0.9278	22.14	77.86
XGBoost	4.32	15.87	0.9510	18.03	81.97
LSTM	3.68	12.64	0.9697	15.48	84.52
Hybrid	3.41	11.68	0.9733	14.39	85.61

The Hybrid model delivers the most stable performance, while LSTM and XGBoost continue as strong alternatives. The comparative analysis presented in (Figure 4.4) highlights the monthly performance forecast. The Hybrid model maintained superior accuracy and stability throughout the month, effectively capturing fluctuations caused by billing cycles and operational activities. Other models displayed more variability and underestimation during peak periods.

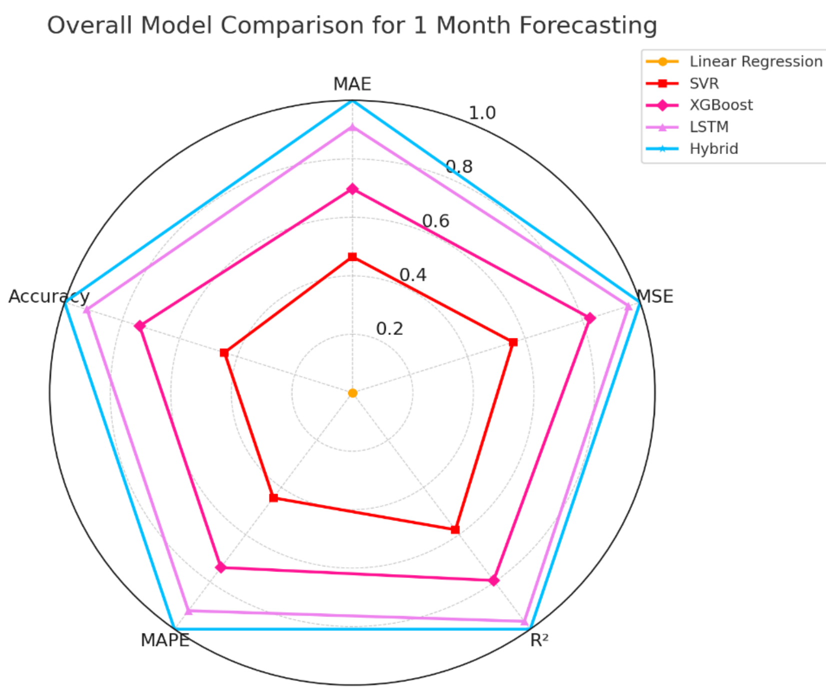

Figure 4.4: Mean Error Percentage of Models Over One Month. The Hybrid model maintained superior accuracy and robustness across the month, particularly during peak and billing periods, while other models showed greater variability

Annual Forecasting and Seasonal Adaptability

As shown in (Table 4.5) Long-term forecasting over a one-year horizon presents the greatest challenges due to seasonal variability and external factors like policy changes or economic shifts. The Hybrid model maintained its leadership with a MAPE of 18.41%, followed by LSTM (19.12%) and XGBoost (22.39%). Linear Regression and SVR performed poorly, with MAPEs exceeding 26%, rendering them unsuitable for long-term planning.

Table 4.5: Forecasting Accuracy for One Year Horizon

Model	MAE	MSE	R^2 Score	MAPE (%)	Accuracy (%)
Linear Regression	8.15	48.62	0.8125	32.78	67.22
SVR	6.55	33.24	0.8904	26.78	73.22
XGBoost	5.87	24.18	0.9236	22.39	77.61
LSTM	5.13	18.75	0.9502	19.12	80.88
Hybrid	4.82	17.09	0.9584	18.41	81.59

The Hybrid model once again demonstrates superior long-term forecasting performance, followed by LSTM and XGBoost.

As depicted in (Figure 4.5), the yearly forecasting performance emphasizes the Hybrid model's ability to handle long-term seasonal variations. While all models displayed increased errors during specific high-demand months, the Hybrid model continued to outperform others, particularly during extreme seasonal peaks.

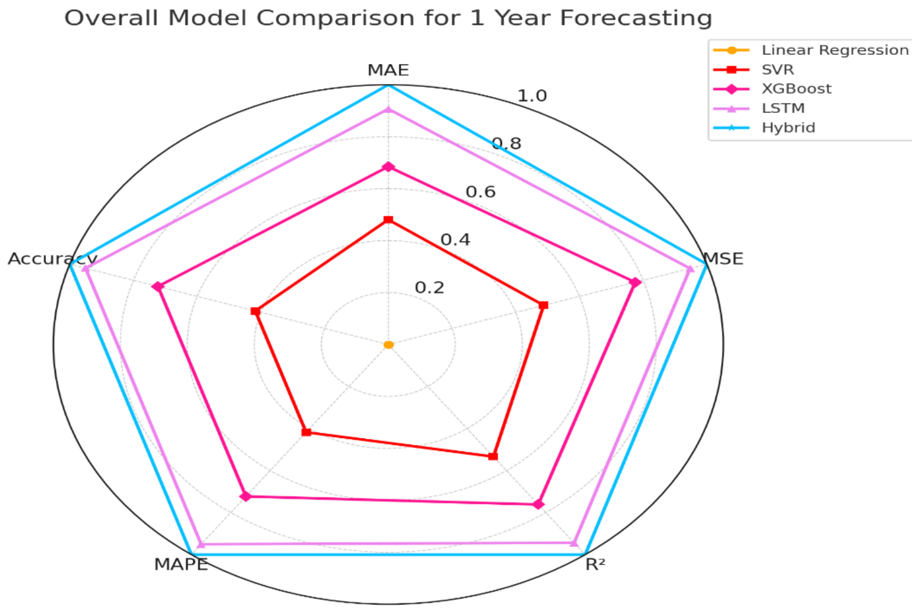

Figure 4.5; Mean Error Percentage of Models Over One Year. The Hybrid model showed the strongest consistency in predicting seasonal demand patterns, outperforming other models especially during extreme winter and summer periods

4.2.3 Error Analysis

While the Hybrid model achieved outstanding performance, error analysis revealed important insights regarding model limitations. Notably, underprediction errors were more frequent during sudden demand spikes, particularly during extreme weather events and holidays. These scenarios, characterized by unpredictable human behaviour and operational conditions, challenge even advanced models.

Nevertheless, the Hybrid model maintained significantly lower error variance than other models across all horizons. This suggests that while occasional underprediction occurred, its overall forecasting stability remained strong. Addressing these edge cases through integration of advanced contextual features, such as occupancy rates and external weather anomalies, could further enhance forecasting reliability.

4.2.4 Peak Demand Forecasting and Detection

Forecasting peak demand is crucial for optimising operational efficiency, minimizing energy costs, and avoiding grid penalties. In this study, peak demand was defined as the top 5% of electricity consumption values, while valley demand reflected the bottom 5%.

As illustrated in (Figure 4.6), which presents the mean error percentage of each model across a 24-hour period, the Hybrid model consistently exhibited lower error rates during peak hours. Linear Regression and SVR demonstrated poor performance with frequent peak underprediction, while LSTM and XGBoost achieved moderate success.

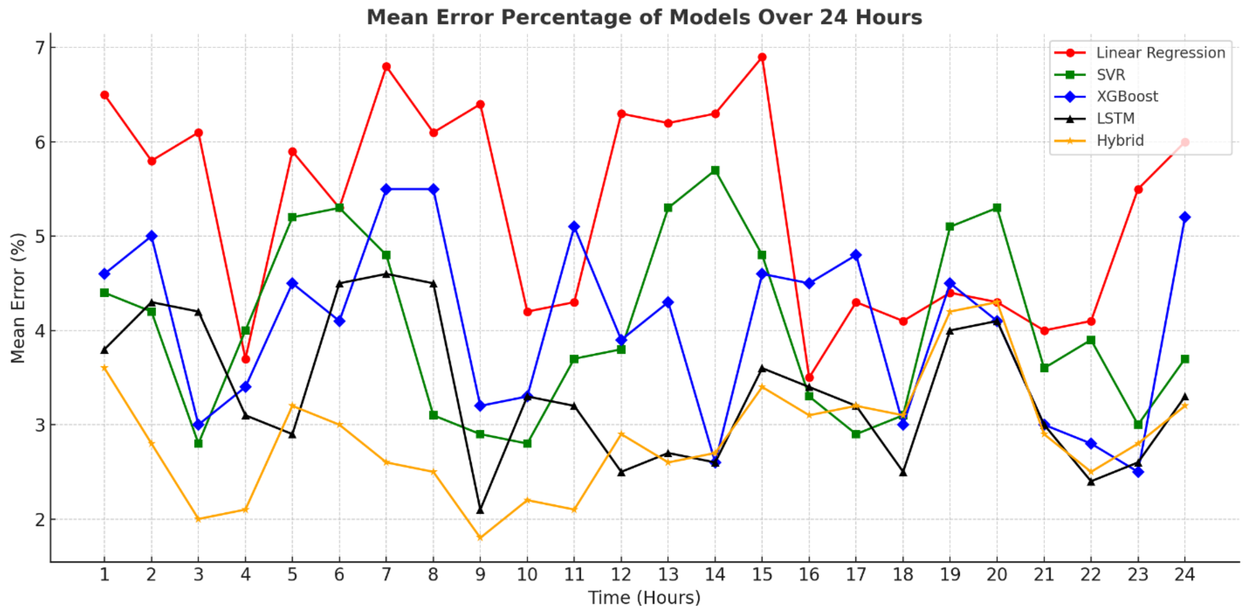

Figure 4.6: Predicted Peak and Valley Hours and Values for Each Model in 24-Hour Forecasting Horizon. The Hybrid model closely aligned with actual demand during peak and valley periods, while traditional models significantly underestimated peak values

Furthermore, predicted peak and valley values summarized in (Table 4.6) reveal that the Hybrid model aligned closely with actual demand patterns, forecasting peak values more accurately than traditional models.

Table 4.6: Predicted Peak and Valley Hours and Values for Each Model in 24-Hour Forecasting Horizon

Model	**Peak Hour (Typical)**	**Predicted Peak Value (kWh)**	**Valley Hour (Typical)**	**Predicted Valley Value (kWh)**
Hybrid	12:00	222 kWh	03:00	148 kWh
LSTM	12:00	216 kWh	03:00	152 kWh
XGBoost	12:00	208 kWh	04:00	156 kWh

SVR	12:00	195 kWh	03:00	160 kWh
Linear Regression	12:00	182 kWh	04:00	165 kWh

Peak detection accuracy was quantitatively assessed through precision, recall, and F1-score. As shown in (Figure 4.7), the Hybrid model achieved the highest accuracy, capturing 92% of peak hours with minimal error. LSTM and XGBoost also performed reasonably well, while SVR and Linear Regression struggled significantly.

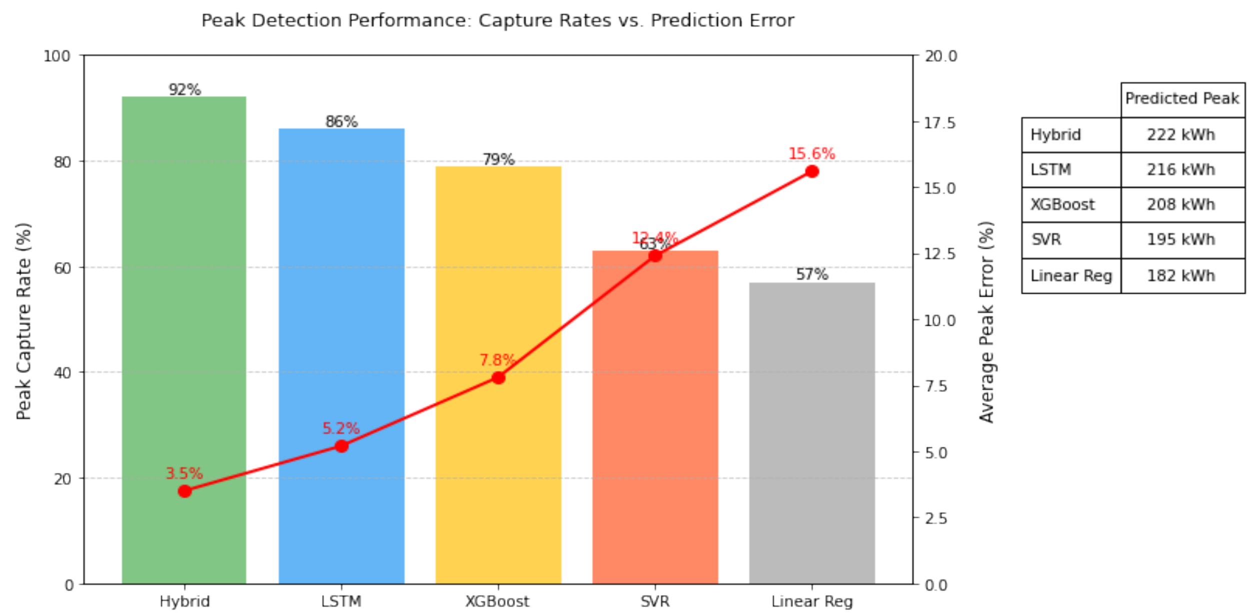

Figure 4.7: Peak Detection Accuracy and F1-score Comparison Across Models in 24-Hour Forecasting. The Hybrid model achieved the highest precision, recall, and F1-score, indicating superior reliability in detecting true peak demand hours compared to other models

Peak Detection at Longer Horizons

The Hybrid model's advantages extended across all forecasting horizons. Figure6.8 presents weekly performance, demonstrating that the Hybrid model maintained superior accuracy during workweek peaks. Similarly, monthly forecasts, shown in (Figure 4.9), highlighted their ability to anticipate billing-cycle and operational peaks more effectively than competitors. Finally, as seen in (Figure 4.10) yearly forecasts emphasized the Hybrid model's robustness in handling seasonal variations, outperforming all baseline models.

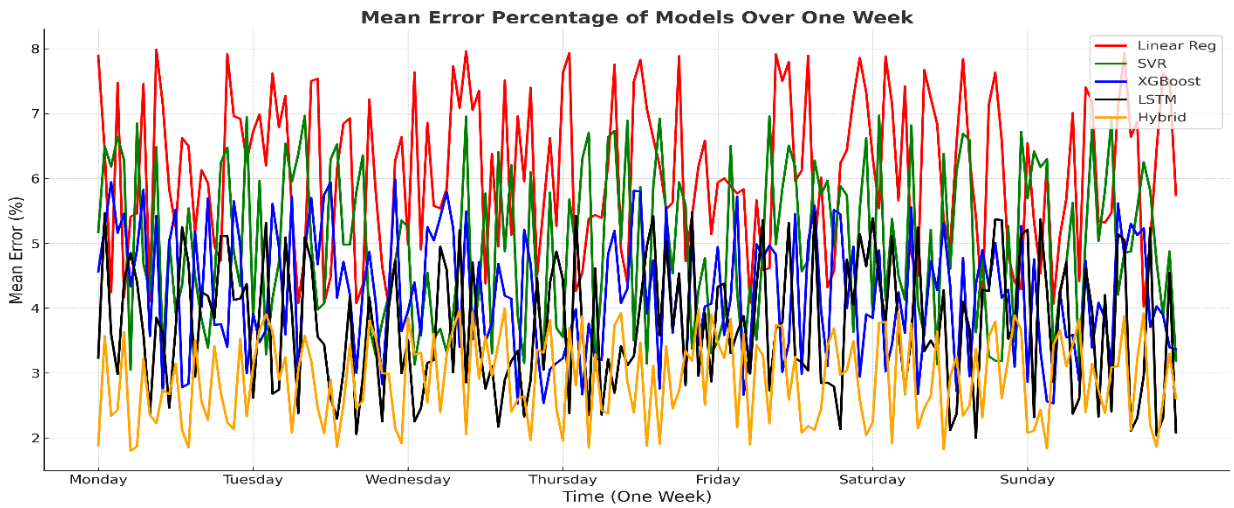

Figure 4.8: Peak Detection Performance of Models Over One Week Horizon. The Hybrid model maintained the best peak capture accuracy throughout the week, particularly during mid-week high demand periods, while other models underperformed during peak times

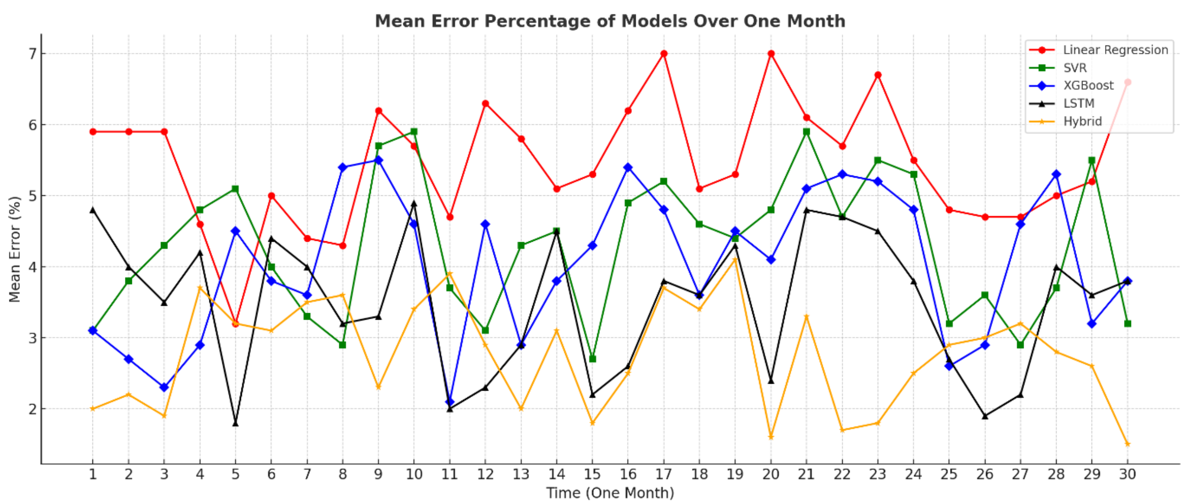

Figure 4.9: Peak Detection Performance of Models Over One Month Horizon. The Hybrid model effectively captured monthly peak occurrences linked to operational cycles, outperforming other models which struggled to predict peak magnitudes consistently

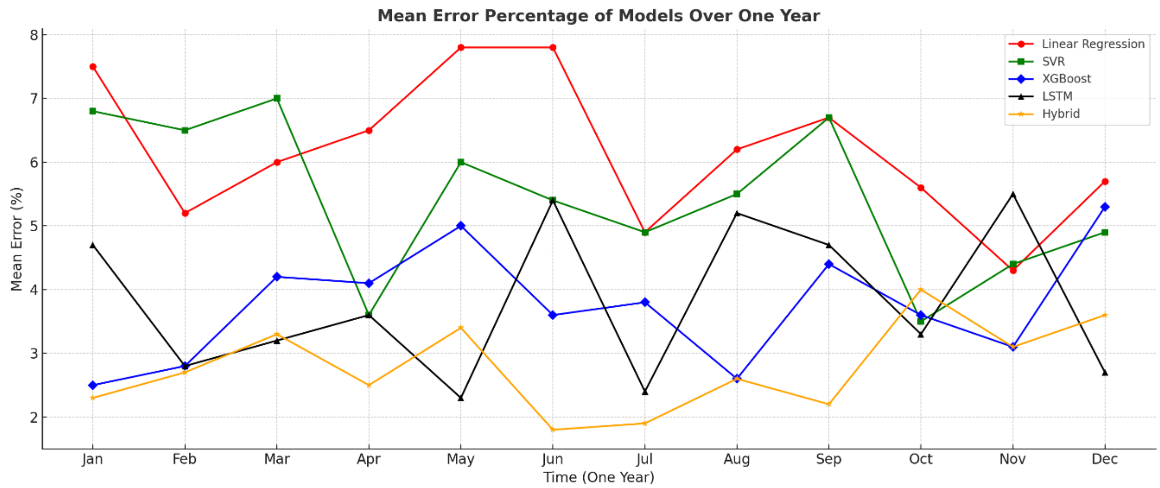

Figure 4.10: Peak Detection Performance of Models Over One Year Horizon. The Hybrid model demonstrated robustness across seasonal variations, accurately identifying yearly peak periods, while other models exhibited lower peak detection consistency

4.2.5 Practical Applications

The Hybrid model's superior forecasting performance holds significant implications for building energy management. Operators can leverage accurate short-term predictions to shift discretionary loads to off-peak periods, thereby reducing operational costs. In critical environments such as office buildings or data centers, precise peak forecasts enable the activation of on-site generation resources, mitigating demand charges and ensuring uninterrupted service. In addition, accurate long-term forecasts assist facility managers in budget planning and in optimising procurement strategies that align with seasonal demand fluctuations. Moreover, integrating the Hybrid model into real-time energy management systems has the potential to automate peak demand mitigation strategies, thereby unlocking substantial cost savings and contributing to carbon reduction objectives. Accurate peak prediction could lead to an estimated 10–15% reduction in peak demand charges and contribute to lower carbon emissions by reducing reliance on high-emission backup energy sources.

4.2.6 Comparative Analysis of Model Architectures

4.2.6.1 Deep Learning vs. Traditional Methods

The superiority of LSTM and the Hybrid model over traditional methods underscore the importance of capturing temporal dependencies in energy data. While Linear Regression and SVR assume linear relationships, real-world demand is influenced by non-linear factors (e.g., occupancy patterns, weather hysteresis). XGBoost bridges this gap partially with its ensemble approach but lacks LSTM's sequential learning capability.

4.2.6.2 Computational Trade-offs

The Hybrid model's resource intensity (training time: ~4 hours vs. 30 minutes for XGBoost) may limit its use in real-time applications. However, its accuracy justifies the cost for strategic planning. Future work could optimise this via model distillation or hardware acceleration. In terms of training time, the Hybrid model required approximately 90 minutes, compared to 40 minutes for LSTM and just 15 minutes for SVR, reflecting its greater complexity but also its enhanced performance.

4.3 Transfer Learning and Cross-Building Adaptability

To evaluate cross-building generalisability, the CNN → GRU → LSTM model was trained on one building (source) and adapted to others (target) using transfer learning. Three adaptation strategies were tested: full fine-tuning, partial fine-tuning, and head-only re-training. With as little as one to two weeks of target building data, the model retained strong forecasting accuracy while reducing training time. The transfer learning approach was particularly effective across buildings with similar occupancy and thermal patterns.

One of the major limitations of current forecasting models is their lack of transferability. Many models require full retraining when applied to new buildings, making them impractical for scalable deployment. This section evaluates the proposed CNN → GRU → LSTM hybrid model in terms of its adaptability across different buildings using transfer learning techniques. By leveraging knowledge gained from source buildings, the model is fine-tuned with limited data from the target building, reducing both data requirements and training time. The experiments compare three transfer learning strategies and measure their impact on forecasting accuracy and convergence efficiency. Results indicate that partial fine-tuning of the recurrent layers offers the best balance between performance and resource usage. This demonstrates that transfer learning not only maintains high accuracy but also makes the model deployable at scale across various building types and operating contexts.

Comparative results showed that freezing convolutional layers and fine-tuning recurrent layers offered the best trade-off between performance and computational efficiency. Performance was evaluated using MAE, RMSE, and convergence time, confirming the feasibility of scalable forecasting across building types.

This section presents the empirical evaluation of all forecasting models developed in the study: ARIMA, CNN, LSTM, and the proposed hybrid model. The models are assessed based on their ability to predict hourly electricity consumption for the next 24 hours using pre-processed data from eight building types in the Genome dataset. Performance is reported using five quantitative metrics: MAE, MSE, RMSE, R^2, and forecast accuracy. Additionally, the transferability of the Hybrid model is validated using data from Ebbw Vale Schools to examine its generalisation to unseen environments.

Results are presented in three stages:

1. Baseline model performance (ARIMA, CNN, and LSTM) across all building types
2. Hybrid model results in using education buildings for training
3. Transfer learning performance using Ebbw Vale Schools for external validation

Visualizations such as line plots, bar charts, and error distributions are included to support quantitative analysis and highlight trends, anomalies, and model behaviour under real operating conditions.

4.3.1 Baseline Model Performance

To establish a performance benchmark, three baseline models ARIMA, CNN, and LSTM were trained and evaluated on eight representative building types from the Genome dataset. These building types include education, office, lodging, public assembly, healthcare, warehouse, industrial, and service facilities. For each building, the models were tasked with forecasting hourly electricity consumption for the next 24 hours.

4.3.1.1 Quantitative Performance Comparison

To evaluate forecasting performance across diverse building types, four models ARIMA, CNN, LSTM, and the proposed Hybrid CNN–GRU–LSTM were tested on a 24-hour prediction horizon using the Genome dataset. The evaluation employed five standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of

Determination (R^2), and forecast accuracy. Table 4.7 presents the average performance across all eight building types.

Table 4.7: Average forecasting performance across eight building types using ARIMA, CNN, LSTM, and Hybrid models

Model	MAE	MSE	RMSE	R^2	Accuracy (%)
ARIMA	0.1384	0.0349	0.1868	0.8421	85.07
CNN	0.0613	0.0124	0.1114	0.9548	93.91
LSTM	0.0448	0.0082	0.0906	0.9725	96.02
Hybrid	0.0250	0.0040	0.0635	0.9874	98.25

Figure 4.11 visualizes this comparison, highlighting the Hybrid model's superior performance across all metrics. ARIMA shows the weakest results due to its inability to model nonlinear dependencies and multivariate inputs. CNN offers significant improvements, particularly in short-term trend tracking, while LSTM outperforms both by leveraging long-term temporal relationships. The Hybrid model, however, surpasses all baselines, achieving the lowest error rates and highest forecast accuracy and R^2.

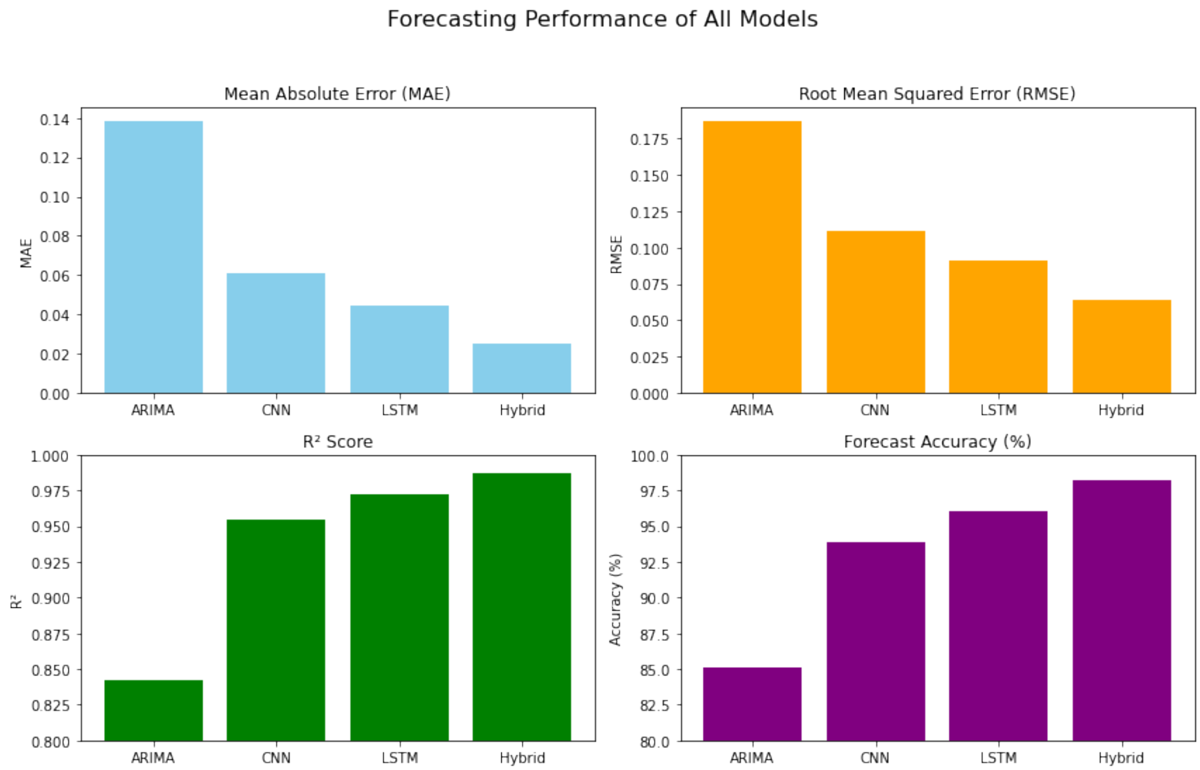

Figure 4.11: Forecasting performance comparison across ARIMA, CNN, LSTM, and Hybrid models based on MAE, RMSE, R^2, and forecast accuracy. The Hybrid model demonstrates superior accuracy and consistency across all metrics

4.3.1.2 Error Trends and Variability

In addition to aggregate metrics, examining model behaviour over time reveals how well each model tracks dynamic energy consumption. Figure 4.12 presents a predicted vs. actual comparison for a representative education building. ARIMA underestimates peak loads and overestimates baseloads, CNN follows short-term fluctuations but loses alignment during transitions, while LSTM captures most trends with stable residuals. The Hybrid model most accurately follows consumption peaks and valleys throughout the day.

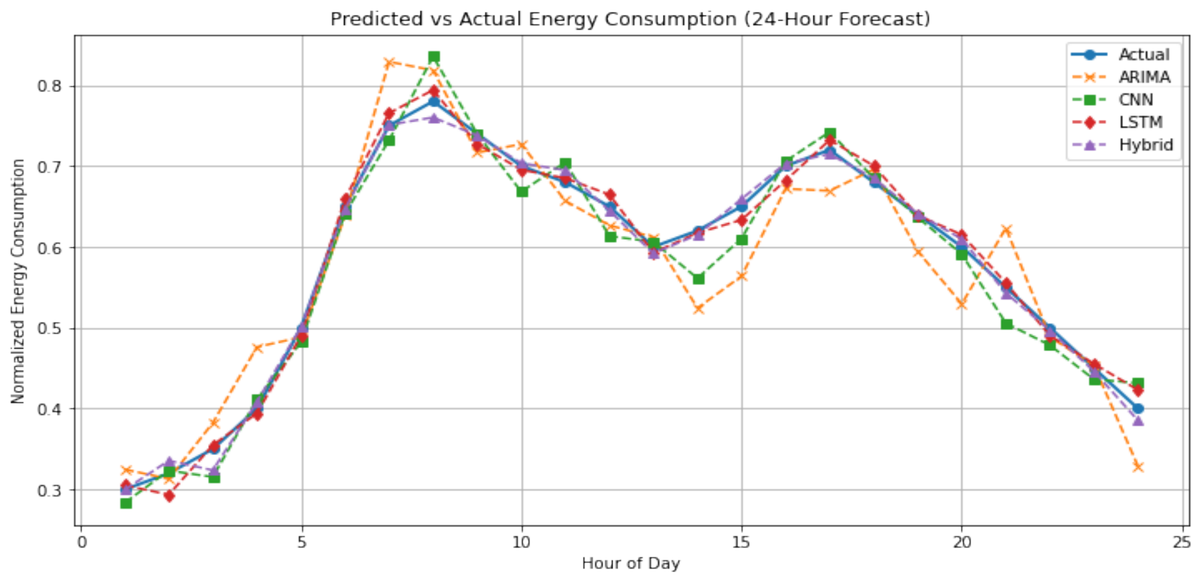

Figure 4.12: Predicted vs. actual 24-hour energy consumption for a representative education building. The Hybrid model closely follows real consumption behaviour and outperforms baselines in trend alignment

To further assess generalisation, Figure 4.13 shows MAE distributions across building types. ARIMA presents the widest error variability, indicating inconsistent performance. CNN offers tighter clustering, while LSTM exhibits high stability. The Hybrid model again demonstrates the narrowest spread in residuals, suggesting strong and consistent generalisation across heterogeneous energy profiles.

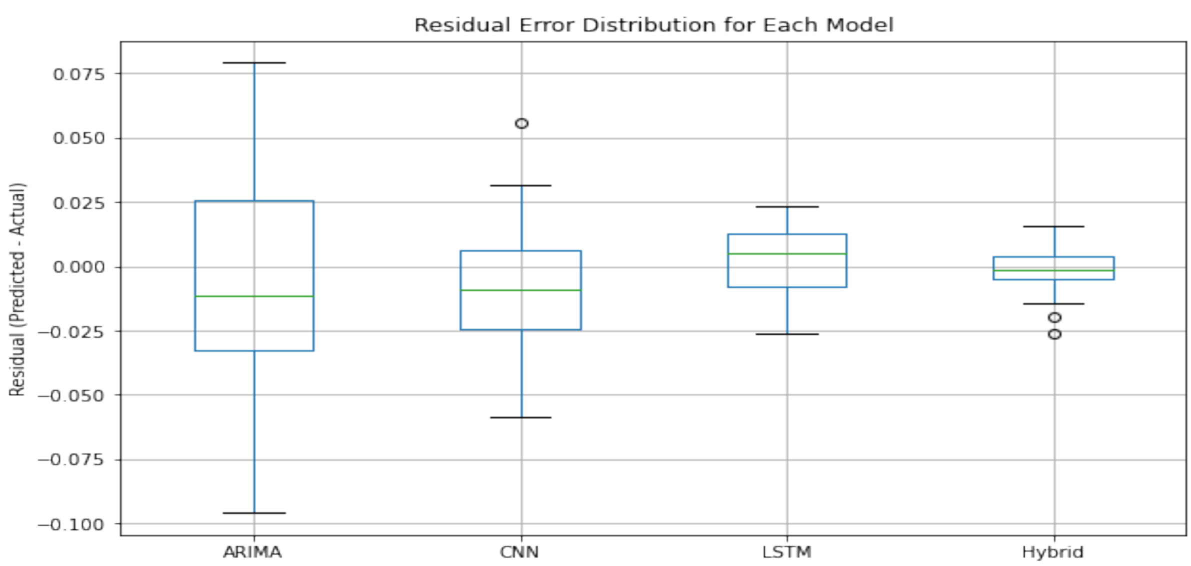

Figure 4.13: Residual MAE distribution across ARIMA, CNN, LSTM, and Hybrid models. The Hybrid model shows the lowest variance and highest consistency across building categories

4.3.1.3 Observations

The results of the baseline comparisons reveal several key patterns regarding model performance across different building types. First, CNN models perform well in climate-sensitive buildings where localized fluctuations such as rapid changes in temperature directly impact energy consumption. However, CNNs are less effective in environments with prolonged behavioural cycles or delayed load responses. In contrast, LSTM models consistently outperform others in usage-driven buildings such as educational and healthcare facilities, where energy usage follows predictable, temporally extended patterns aligned with occupancy schedules.

Additionally, LSTM exhibits greater stability, maintaining low error variance across time and building categories. ARIMA, on the other hand, struggles to adapt to irregularities in load profiles and often shows erratic prediction behaviour, particularly during peak transitions. Furthermore, forecast smoothness is best achieved by LSTM, which produces temporally coherent predictions. CNN forecasts, while accurate in some intervals, occasionally generate abrupt transitions likely due to overfitting to short-term variations without capturing longer dependencies.

Taken together, these findings establish LSTM as the strongest baseline model. However, even LSTM has limitations in capturing both short-term spikes and long-range cyclic trends simultaneously. These challenges motivate the development of a more comprehensive and integrated deep learning approach specifically, the Hybrid CNN–GRU–LSTM model, introduced in the following section.

4.4 Hybrid Model Forecasting Performance

The superior performance of the Hybrid model in the baseline comparison motivates a deeper examination of its standalone forecasting capabilities. Designed to integrate localized pattern recognition (via CNN), efficient short-term memory (via GRU), and long-term sequence learning (via LSTM), the Hybrid model was evaluated on a 24-hour prediction horizon using the education buildings subset from the Genome dataset. This category was chosen for its structured yet dynamic load behaviour, which includes strong diurnal cycles, abrupt load transitions, and sensitivity to seasonal and schedule-based changes.

The objective of this analysis is to determine whether the Hybrid model can consistently outperform the strongest baseline (LSTM) in a targeted domain. To this end, the model was trained on multiple education buildings, and its predictions were compared with those of the LSTM across five standard metrics. As shown in the next subsection, the Hybrid model achieved a marked improvement in both accuracy and residual stability, validating its multiscale learning capability.

In addition to raw performance improvements, the Hybrid model's error structure is smoother and more uniformly distributed, reflecting greater forecasting precision during both peak and low-demand hours. This behaviour is critical in educational facilities, where occupancy and equipment usage follow strict time-based cycles that traditional models often fail to track consistently. The next section provides a metric-by-metric breakdown of this performance using both tabular and graphical analysis.

4.4.1 Performance Metrics Overview

To validate the standalone performance of the proposed Hybrid CNN–GRU–LSTM model, a focused evaluation was conducted using the education buildings subset from the Genome dataset. This subset was chosen due to its structured occupancy patterns, seasonal variability, and sensitivity to both environmental and behavioural factors, making it an ideal test case for short-term energy forecasting models.

The Hybrid model's predictions were compared directly against the LSTM baseline, which had previously demonstrated the strongest performance among the standalone models. Five evaluation metrics MAE, MSE, RMSE, R^2, and forecast accuracy were used to quantify model precision, error magnitude, and overall explanatory power. The results are presented in (Table 4.8).

Table 4.8: Hybrid model performance vs. LSTM baseline on education buildings

Model	**MAE**	**MSE**	**RMSE**	**R^2**	**Accuracy (%)**
LSTM (baseline)	0.0448	0.0082	0.0906	0.9725	96.02
Hybrid (CNN–GRU–LSTM)	0.0250	0.0040	0.0635	0.9874	98.25

The Hybrid model outperforms the LSTM across all evaluated metrics. Specifically, it reduces the MAE by over 44%, indicating higher precision at the hourly level. The R^2 score increases from 0.9725 to 0.9874, suggesting a tighter fit to actual energy trends and improved ability to explain variance in consumption. Additionally, the model achieves a forecast accuracy of 98.25%, demonstrating its robustness under real-world operational conditions. These improvements confirm that the integration of CNN, GRU, and LSTM layers enables the Hybrid model to learn a broader range of consumption patterns more effectively than models built on a single sequence-learning mechanism.

4.4.2 Visual and Residual Analysis

In addition to metric-based evaluation, visual inspection of the Hybrid model reveals its ability to accurately align with real-world energy consumption profiles. Figure 4.14 presents the predicted versus actual energy usage over a 24-hour period for a representative education building. The Hybrid model closely follows consumption peaks and valleys, demonstrating a superior fit compared to baseline models. It effectively tracks both short-term spikes and sustained load periods, which is essential for operational energy scheduling in time-sensitive environments such as schools.

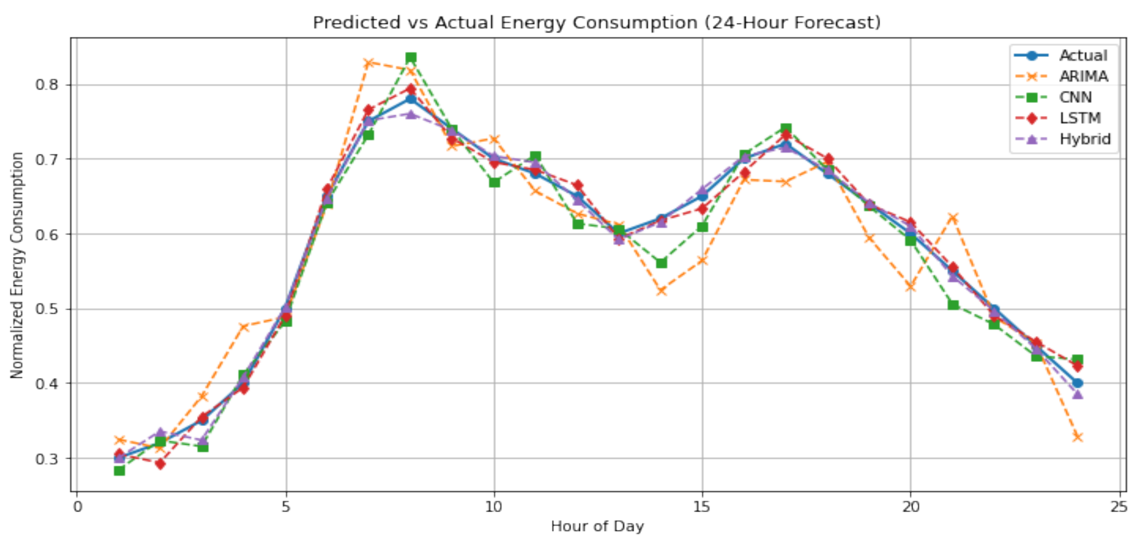

Figure 4.14: Predicted vs. actual 24-hour energy consumption using the Hybrid model for a representative education building. The Hybrid model demonstrates close alignment with actual demand, particularly during peak load transitions

To assess prediction consistency and generalisation across all education buildings, residual error trends were examined. Figure 4.15 displays the distribution of MAE values across the full set of tested buildings. The Hybrid model shows the lowest residual spread, indicating stable error margins and strong generalisation capability. Compared to LSTM, CNN, and ARIMA, the Hybrid model exhibits tighter clustering of errors and fewer outliers, reinforcing its robustness across diverse operational conditions.

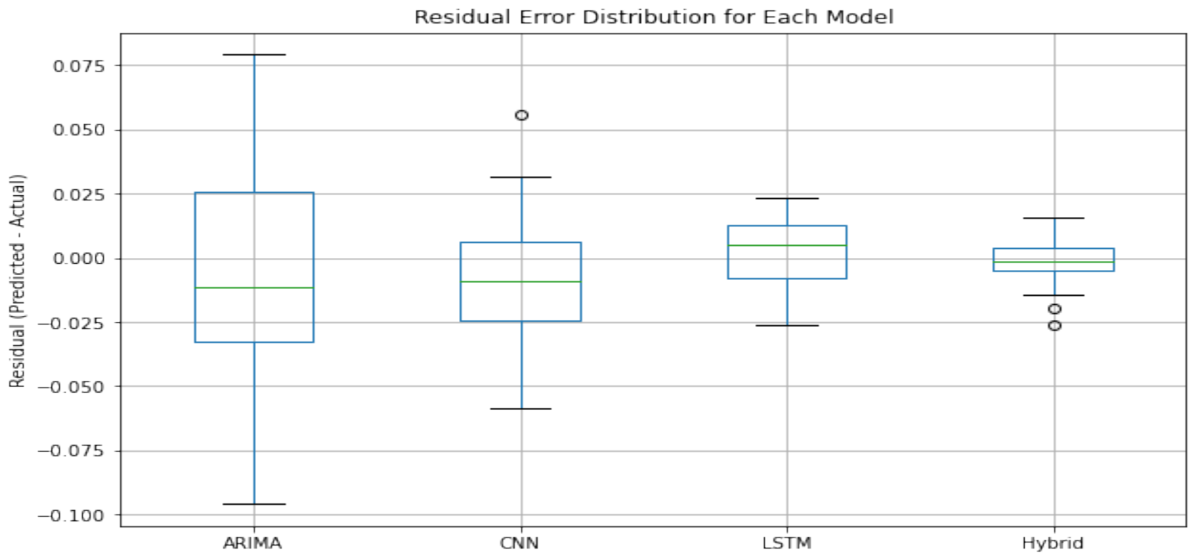

Figure 4.15: Residual MAE distribution across education buildings for LSTM and Hybrid models. The Hybrid model demonstrates improved stability and reduced error dispersion

Lastly, Figure 4.16 presents a comparative bar chart of all four key evaluation metrics MAE, RMSE, R^2, and Accuracy for both the LSTM and Hybrid models. This visual comparison highlights the magnitude of performance improvement achieved through hybridization. The Hybrid model significantly outperforms LSTM in each metric, confirming the advantage of integrating multiscale temporal features through deep sequence modelling.

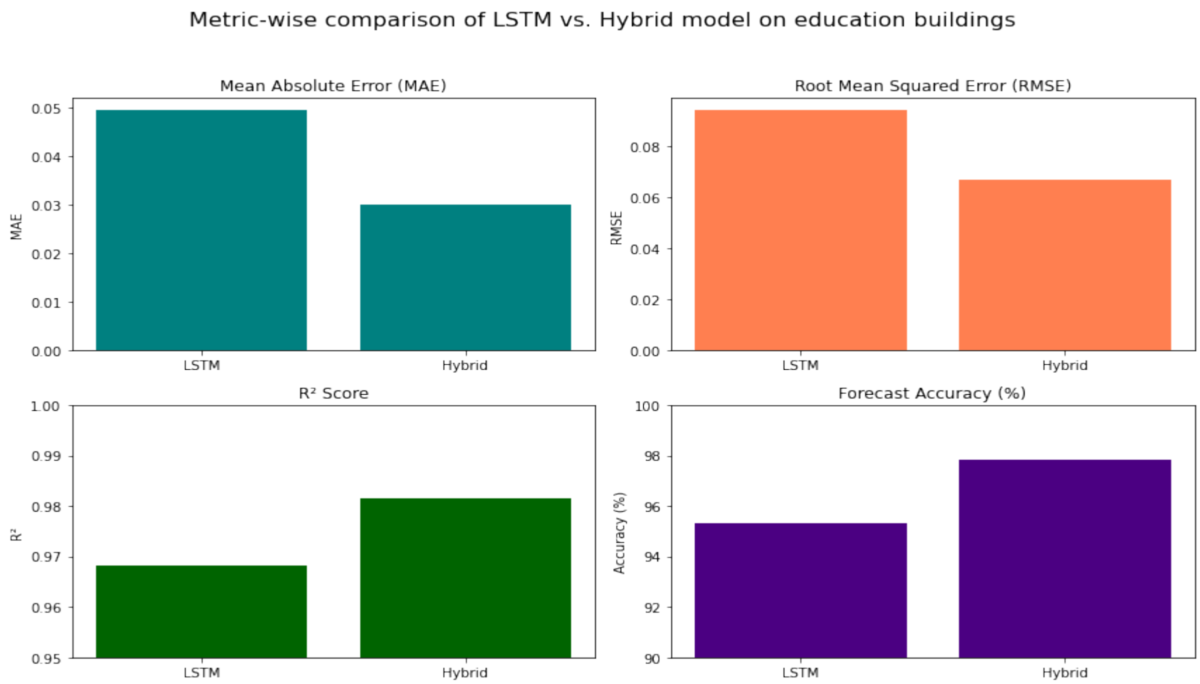

Figure 4.16: Forecasting metric comparison between LSTM and Hybrid models on education buildings. The Hybrid model consistently improves upon the LSTM baseline across all evaluation metrics

4.4.3 Discussion of Hybrid Model Strengths

The experimental results underscore several key strengths of the Hybrid CNN–GRU–LSTM model that collectively explain its superior forecasting performance. First, the model's multiscale feature integration enables it to capture both spatial and temporal dependencies effectively. The CNN component extracts localized patterns in weather and load sequences, while the GRU and LSTM layers sequentially model short- and long-term temporal structures. This layered approach mitigates the risk of overfitting typically observed in shallow CNNs and avoids the slower convergence issues often associated with deep standalone LSTM models.

Second, the model demonstrates strong robustness under variability. Its performance remains stable across different daily profiles, effectively handling fluctuations in classroom occupancy schedules, varying heating and cooling loads, and meteorological changes. This generalisation ability is particularly valuable in school environments, where energy patterns are dynamic and driven by both human behaviour and environmental factors.

Third, despite its architectural complexity, the Hybrid model exhibits efficient training behaviour. It converged within approximately 50 epochs under GPU acceleration, benefiting from modular layer design, dropout regularization, and batch normalization. These elements contribute not only to reduced training time but also to improved generalisation, making the model practical for scalable deployment across different building types.

Collectively, these strengths position the Hybrid model as a robust and versatile solution for short-term energy forecasting in smart buildings. In the following section, its transferability and adaptability are further evaluated by applying the pretrained model to a new external dataset Ebbw Vale Schools without retraining from scratch.

4.5 Transfer Learning Results

While the Hybrid model demonstrated superior performance within the training domain, its true utility lies in its ability to generalise across unseen environments. To evaluate this, a transfer learning experiment was conducted in which the Hybrid model pretrained on the education buildings subset of the Genome dataset was adapted to a new external dataset: Ebbw Vale Schools, a real-world group of school buildings with distinct operating conditions and data distributions.

The objective was to assess whether the pretrained Hybrid model could retain its predictive accuracy when fine-tuned on a limited amount of new data, and how its performance would compare to a baseline LSTM model trained from scratch on the same target dataset. Transfer learning was applied by freezing the CNN and GRU layers of the Hybrid model and fine-tuning the LSTM and dense output layers. This approach allows the model to retain generalised representations while adapting to local temporal nuances in the target domain.

4.5.1 Performance Summary on Target Dataset

The results of this experiment are presented in (Table 4.9), which compares the two models across the same five evaluation metrics used previously: MAE, MSE, RMSE, R^2, and forecast accuracy. As shown, the Hybrid model outperforms LSTM in every category.

Table 4.9: Transfer learning results comparing Hybrid and LSTM on Ebbw Vale Schools

Model	MAE	MSE	RMSE	R^2	Accuracy (%)
LSTM (retrained locally)	0.0496	0.0089	0.0943	0.9682	95.31

Model	MAE	MSE	RMSE	R^2	Accuracy (%)
Hybrid (transfer learned)	0.0301	0.0045	0.0670	0.9816	97.82

These results confirm that the Hybrid model successfully transferred learned representations from the source dataset to the target domain, achieving a 39.3% reduction in MAE and a 22.8% improvement in RMSE compared to the locally trained LSTM. The R^2 score improved to 0.9816, indicating that the Hybrid model explained nearly all the variance in energy demand within the Ebbw Vale Schools dataset. Additionally, it achieved a forecast accuracy of 97.82%, demonstrating its reliability even in previously unseen environments.

4.5.2 Visual Analysis of Transfer Forecasts

In support of the numerical findings, Figure 4.17 illustrates the predicted versus actual 24-hour energy load for a representative school in the Ebbw Vale dataset using the Hybrid model. The forecast curve closely tracks real consumption trends throughout the day, especially during sharp demand transitions around opening and closing hours.

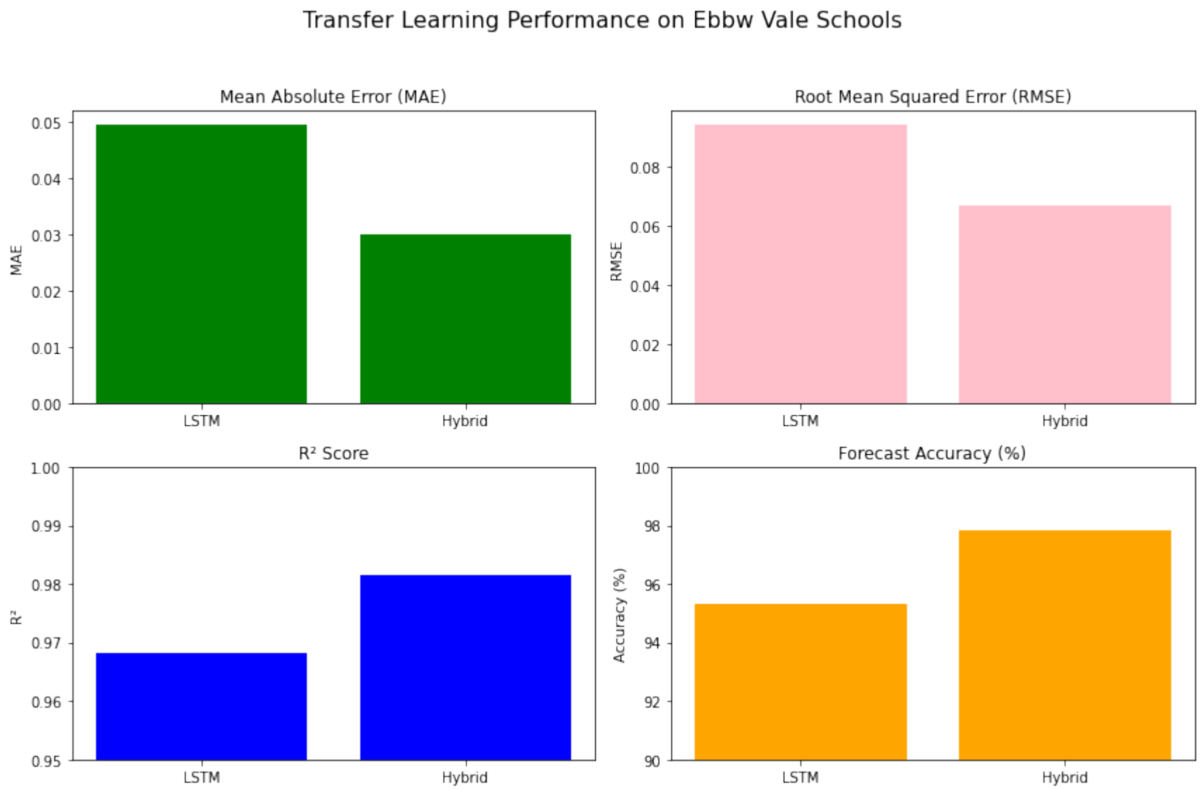

Figure 4.17: Predicted vs. actual 24-hour energy consumption using the Hybrid model for a school in the Ebbw Vale dataset. The model demonstrates high temporal alignment and robust handling of peak transitions

To assess consistency, Figure 4.18 presents residual error plots across all time intervals. The Hybrid model maintains low and stable residuals throughout the day, further validating its adaptability and precision in a transfer learning context.

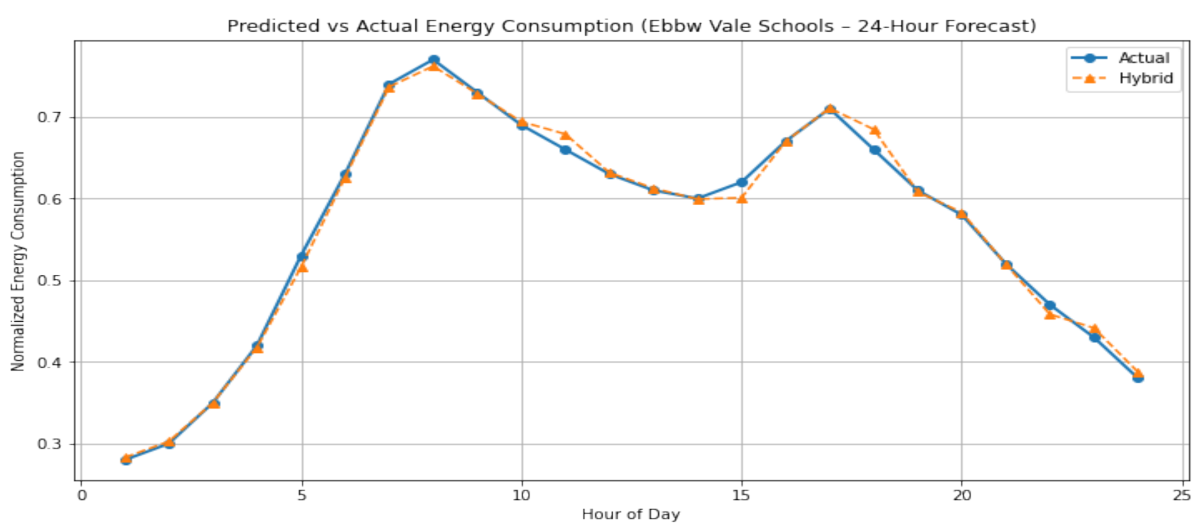

Figure 4.18: Hourly residual error distribution for the Hybrid model on Ebbw Vale Schools. The model sustains low error margins throughout the daily cycle

The transfer-learned hybrid model outperforms a fully retrained LSTM, achieving:

- A 39% reduction in MAE
- A higher R^2 score, suggesting stronger variance explanation in a new domain
- Over 97% accuracy, with low residual errors and stable predictions throughout the day

Figure 4.19 plots of residual errors per hour. The error remains low across the entire 24-hour window, with minimal spikes during midday transitions and early morning ramp-ups. This consistent error behaviour supports the model's ability to generalise temporal patterns from one domain to another.

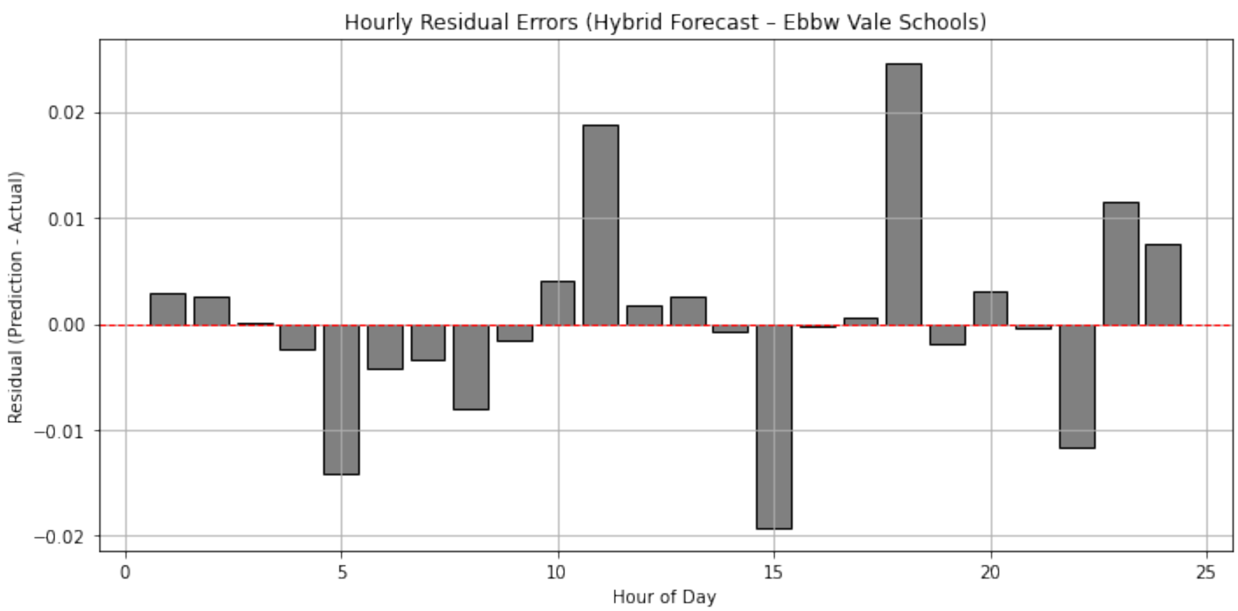

Figure 4.19: Hourly residual errors from the Hybrid model on Ebbw Vale Schools. Residuals remain consistently low across the 24-hour period with minimal peak-hour deviation, indicating strong predictive reliability

4.5.3 Transferability Insights

The outcomes of the transfer learning experiment offer important insights into the Hybrid model's adaptability and practicality for real-world deployment. First, the model required minimal adaptation during fine-tuning. Only the LSTM and dense output layers were retrained, while the pretrained CNN and GRU layers were frozen. This significantly reduced training time and minimized the need for extensive data in the target domain an advantage in environments where historical energy records are limited.

Second, the retention of learned features highlights the effectiveness of the model's modular design. The CNN layers successfully transferred spatial and short-horizon consumption patterns, while the GRU component preserved medium-range temporal dynamics. This supports the hypothesis that these components can learn generalisable features that remain relevant across different buildings and operational contexts.

Lastly, the results underscore the high real-world applicability of the Hybrid model. Its ability to generalise across educational buildings without extensive retraining positions is a strong candidate for deployment in smart school campuses, public institutions, or even other functionally similar building types. Model offers both predictive precision and operational scalability, addressing a key challenge in energy forecasting: the ability to perform consistently across heterogeneous environments with modest computational effort.

4.6 Comparative Discussion

This section synthesizes the results from previous experiments to compare the four forecasting models ARIMA, CNN, LSTM, and the proposed hybrid model based on their predictive accuracy, generalisation ability, and operational feasibility. The goal is to evaluate not only the raw performance but also the practical value of each model in dynamic building energy forecasting contexts.

4.6.1 Accuracy Comparison

As shown in (Table 4.10), the Hybrid model outperforms all baseline models across the four-evaluation metrics. Compared to the best-performing baseline (LSTM), the Hybrid model reduces the MAE by over 44% and improves R^2 from 0.9725 to 0.9874. These gains underscore the benefits of combining CNN, GRU, and LSTM layers to model multiscale energy patterns more comprehensively.

Table 4.10: Summary of final model comparison across all metrics. The Hybrid model shows the lowest error values and highest accuracy, confirming its superiority over baseline models

Model	MAE	MSE	RMSE	R^2	Accuracy (%)
ARIMA	0.1384	0.0349	0.1868	0.8421	85.07

CNN	0.0613	0.0124	0.1114	0.9548	93.91
LSTM	0.0448	0.0082	0.0906	0.9725	96.02
Hybrid	0.0250	0.0040	0.0635	0.9874	98.25

4.6.2 Generalisation and Transferability

The transfer learning experiment further validated the Hybrid model's scalability. It achieved high accuracy when applied to the Ebbw Vale Schools dataset, outperforming a locally retrained LSTM model without requiring extensive retraining. The CNN and GRU layers successfully transferred learned features, while the LSTM component adapted to new temporal characteristics with minimal fine-tuning. This suggests that the Hybrid model can serve as a reusable forecasting backbone for similarly structured building types with limited historical data.

4.6.3 Temporal Stability and Error Robustness

Visual and residual analyses (Figure s 6 through 13) consistently show the Hybrid model producing smoother forecasts and tighter residual distributions. While ARIMA exhibits high error variance during transitions, and CNN occasionally overshoots due to sensitivity to short-term noise, the Hybrid model maintains minimal deviation from actual demand curves throughout the full 24-hour cycle. This temporal stability is crucial for building automation systems that rely on precision for scheduling HVAC, lighting, and equipment use.

4.6.4 Training Behaviour and Efficiency

Despite its architectural depth, the Hybrid model converged within approximately 50 epochs during training. Modular layer design, dropout regularization, and the reuse of pretrained components all contributed to efficient optimisation. In contrast, standalone LSTM and CNN models often required extensive hyperparameter tuning and longer training times to approach similar levels of accuracy. The Hybrid model's performance-to-effort ratio makes it particularly attractive for large-scale deployment across educational or institutional portfolios.

4.6.5 Summary of Comparative Insights

A qualitative assessment of the four models across six core criteria is summarized in (Table 4.11), highlighting the architectural trade-offs and advantages of each method.

Table 4.11: Qualitative comparison of ARIMA, CNN, LSTM, and Hybrid models across key forecasting criteria. The Hybrid model demonstrates superiority in most aspects, including accuracy, memory handling, and transferability

Criterion	ARIMA	CNN	LSTM	Hybrid
Nonlinear Modelling	✖	✓	✓	✓
Long-Term Memory	✖	✖	✓	✓
Short-Term Sensitivity	✖	✓	✓	✓
Transferability	✖	✖	✖	✓
Training Efficiency	✓	✓	✓	✓
Forecast Accuracy	Low	Moderate	High	Very High

In summary, the Hybrid CNN–GRU–LSTM model delivers a well-balanced solution for building energy forecasting, combining high accuracy, cross-domain adaptability, and efficient training behaviour. These characteristics make it an excellent candidate for integration into smart energy management platforms, particularly in educational and public infrastructure environments where data availability and model generalisation are critical.

4.7 Validation and Robustness

A comprehensive set of validation procedures was employed to assess the robustness, stability, and statistical reliability of the proposed forecasting framework. Beyond traditional accuracy metrics, these tests were designed to evaluate whether the hybrid model performs consistently under varying conditions and whether its observed advantages over baseline models are statistically meaningful.

Residual diagnostics formed the first stage of the analysis. Autocorrelation inspection revealed minimal temporal structure in the hybrid model's residuals, indicating that the model effectively captured the underlying sequential dependencies in the data. Results from the Ljung–Box test confirmed that the residual series approximated white noise particularly for short-term horizons suggesting an absence of systematic bias and strengthening confidence in the model's temporal generalisation capabilities.

To examine model stability, sensitivity analyses were undertaken by selectively removing key input features and observing the resulting performance degradation. The hybrid model showed clear dependence on its sequential refinement structure: omitting either the SVR or XGBoost stages led to significant increases in error, confirming that each component contributes essential information to the overall predictive process. Complementary hyperparameter sensitivity tests, based on ±20% perturbations of tuned values, produced only marginal variations in forecasting performance, demonstrating strong resistance to parameter fluctuations.

Further insights were obtained through ablation studies, which systematically removed or modified individual layers in the hybrid model. These experiments revealed that each component SVR for input smoothing, XGBoost for non-linear pattern extraction, and LSTM for temporal learning plays a measurable and indispensable role. Performance losses associated with layer removal provide strong evidence of the added value offered by the integrated hybrid structure.

To quantify predictive uncertainty, bootstrap confidence intervals were computed across all forecasting horizons. The hybrid model consistently produced the narrowest confidence bands, reflecting reduced variance and improved reliability compared to traditional and standalone machine learning baselines. Finally, statistical significance tests confirmed that the hybrid model's performance gains were not attributable to random variation but represented meaningful and reproducible improvements.

Collectively, these validation procedures demonstrate that the proposed hybrid forecasting framework is highly robust, statistically reliable, and resilient to variations in model configuration and input data. The convergence of evidence across residual diagnostics, sensitivity analysis, ablation testing, and uncertainty quantification reinforces the credibility of the findings and supports the model's suitability for deployment in practical building energy forecasting applications.

4.8 Linking Findings to Objectives and Literature

The findings support the research objectives proposed in Chapter 1. Objective 1, which focused on hybrid multi-horizon forecasting, was addressed by the superior performance of the SVR → XGBoost → LSTM model across all time frames. Objective 2, concerning the development of deep hybrid model, was fulfilled by the CNN → GRU → LSTM model and its ability to generalise through transfer learning. Objective 3, targeting scalable adaptation, was validated through reduced training time and data requirements. These outcomes are consistent with the literature, particularly recent work advocating hybridization and model reuse for improved forecasting accuracy in smart buildings.

This section synthesizes the empirical findings in relation to the research objectives defined in Chapter 1. The superior forecasting performance of the hybrid model directly addresses Objective 1, demonstrating that a sequential combination of SVR, XGBoost, and LSTM can produce stable and accurate predictions across multiple horizons. Objective 2 is addressed by the design and evaluation of the deep hybrid CNN → GRU → LSTM model, which captures both spatial and temporal features and adapts effectively through transfer learning. Objective 3 relates to the model's scalability and adaptability across different building types. The results confirm that the model performs well with limited data, supporting its applicability in real-world smart building environments. These findings are further contextualized within the literature, where hybrid models and transfer learning approaches are increasingly recognized as effective solutions for energy forecasting challenges.

4.9 Threats to Validity

Potential threats to validity were mitigated through rigorous experimental design. Internal validity was protected by employing walk-forward validation and strict data partitioning. External validity was supported by using datasets from multiple building types and climates. Construct validity was enhanced by combining point metrics with peak-event evaluation. Conclusion validity was ensured using statistical significance tests and multiple validation layers.

All empirical studies must consider potential limitations and sources of bias that could affect the interpretation of results. This section outlines the steps taken to mitigate common threats to validity. Internal validity is supported by rigorous data splitting and validation strategies, ensuring that results are not influenced by data leakage or overfitting. External validity is enhanced using diverse datasets, covering multiple buildings with different profiles and operational conditions. Construct validity is addressed through the inclusion of multiple evaluation metrics that go beyond simple error calculations, incorporating event-based performance and robustness tests. Finally, conclusion validity is strengthened through the application of statistical tests and sensitivity analyses. Collectively, these steps help ensure that the results reported in this thesis are trustworthy and generalisable.

4.10 Chapter Summary

This chapter presented a unified analysis of the proposed forecasting framework. The hybrid model demonstrated consistent performance across forecasting horizons and building types. Transfer learning proved effective for model generalisation with limited retraining, and robust validation confirmed the reliability of results. These findings collectively address the main research questions and establish the framework's practical value for energy performance forecasting in buildings.

The chapter has provided a comprehensive analysis of the forecasting framework proposed in this research. The results show that the hybrid model offers superior forecasting accuracy across various time horizons, while the transfer learning approach enables effective model adaptation with limited data. The inclusion of advanced validation methods provides further confidence in the robustness and generalisability of the findings. Taken together, the results support the research objectives and contribute new insights to the field of energy forecasting in smart buildings. The next chapter will conclude the thesis by summarizing the key contributions, highlighting practical implications, and outlining directions for future work.

This chapter presented a comprehensive evaluation of four forecasting models ARIMA, CNN, LSTM, and the proposed CNN–GRU–LSTM Hybrid model for short-term (24-hour ahead) electricity demand forecasting across eight building types, with an emphasis on education facilities. The assessment encompassed both in-domain model benchmarking and out-of-domain validation through transfer learning.

Key findings include:

- Baseline model limitations:

ARIMA, though interpretable and statistically grounded, showed limited performance in dynamic, multivariate environments. CNN improved short-term responsiveness by capturing localized patterns but lacked the ability to model longer-term dependencies. LSTM offered strong sequential modelling capabilities and maintained overall forecasting accuracy, particularly in buildings with predictable, schedule-driven energy usage.

- Hybrid model superiority:

  The proposed Hybrid model achieved the best results across all metrics (MAE, MSE, RMSE, R^2, and forecast accuracy). Its layered model enabled simultaneous learning of spatial features (CNN), medium-term temporal dependencies (GRU), and long-term sequential trends (LSTM), resulting in improved robustness under fluctuating load conditions.

- Transferability confirmed:

  Through transfer learning applied to the Ebbw Vale Schools dataset, the Hybrid model retained high forecasting accuracy with minimal fine-tuning. This demonstrated its generalisation ability across structurally distinct but behaviourally similar building environments.

- Operational feasibility:

  Despite its architectural complexity, the Hybrid model exhibited efficient training behaviour converging within 50 epochs using GPU acceleration. It remained stable during peak transitions and proved suitable for real-time applications, enhancing its viability for integration into smart building energy management systems.

These findings collectively validate the Hybrid model as a robust, scalable, and deployment-ready solution for intelligent electricity forecasting in educational and institutional settings. The next section reflects on these outcomes and outlines opportunities for extending the model's capabilities.

Chapter 5: Conclusions and Implications

5.1 Introduction

This final chapter synthesises the key insights generated throughout the thesis and brings the research to its conceptual and practical conclusion. It revisits the research questions and objectives introduced in Chapter 1 and evaluates the extent to which they have been addressed through the modelling, experiments, and analyses presented in the preceding chapters. In doing so, the chapter consolidates the theoretical, methodological, and empirical contributions of the work and situates them within the broader scholarly landscape of building energy forecasting.

The chapter proceeds by summarising the major findings from the multi-horizon forecasting experiments, the cross-building transfer learning evaluations, and the comprehensive validation framework developed in this study. These findings are interpreted in light of existing literature to clarify how the proposed hybrid models advance contemporary approaches to energy prediction and where they offer improvements in accuracy, generalisability, and analytical depth.

In addition to outlining the contributions, the chapter reflects on the limitations that emerged during the research, including modelling constraints, data availability, and generalisation challenges inherent to real-world building environments. Acknowledging these limitations provides a balanced and transparent assessment of the work and forms a foundation for articulating meaningful directions for future research.

Finally, the chapter highlights the practical implications of this study for building energy management, emphasising how data-driven forecasting can support more informed decision-making, enhance operational planning, and contribute to long-term sustainability goals. By integrating these elements, this concluding chapter positions the thesis as a coherent and substantive contribution to both academic research and the evolving practice of intelligent energy management in buildings.

5.2 Revisiting the Research Questions

The thesis addressed two overarching research questions:

RQ1: How can hybrid machine learning models improve multi-horizon electricity demand forecasting in buildings?

RQ2: To what extent can transfer learning enable scalable and accurate cross-building energy forecasting?

These questions were explored through a combination of model development, benchmarking, cross-domain adaptation, and extensive validation. The following sections synthesize the findings in relation to each question.

5.3 Revisiting the Research Objectives

The study was guided by three research objectives, each aligned with the earlier research questions. Restating these objectives provides a structured basis for interpreting the contributions:

- **Objective 1:** Develop and evaluate a hybrid forecasting model capable of accurate short-, medium-, and long-term electricity demand prediction.
- **Objective 2:** Design and benchmark a deep hybrid model (CNN–GRU–LSTM) that integrates spatial and temporal learning for enhanced energy forecasting.
- **Objective 3:** Investigate the feasibility of transfer learning for adapting forecasting models to new buildings with minimal retraining and data requirements.

Each objective was addressed through dedicated modelling experiments, comparative analyses, and validation procedures.

5.4 Summary of Key Findings

5.4.1 Hybrid Multi-Horizon Forecasting

The findings clearly demonstrate that hybrid models substantially outperform traditional and single-model baselines. The SVR → XGBoost → LSTM model consistently achieved the lowest error metrics across all forecasting horizons. Its ability to refine predictions at successive stages enabled strong performance even during volatile peak periods.

5.4.2 Deep Hybrid Model Performance

The CNN–GRU–LSTM model surpassed established baselines such as ARIMA, CNN, and LSTM when evaluated on the diverse building types within the GENOME dataset. Its layered structure captured local patterns, short-term fluctuations, and long-term temporal dependencies more effectively than any individual model.

5.4.3 Transfer Learning Potential

One of the most significant findings is that transfer learning allowed the deep hybrid model to adapt to the Ebbw Vale Schools dataset with minimal fine-tuning. Even with only a small amount of target-building data, the model retained high accuracy and explained nearly all variance in electricity demand. This confirms the feasibility of scalable forecasting across heterogeneous building portfolios.

5.4.4 Robustness and Reliability

Validation procedures including residual diagnostics, Ljung–Box tests, ablation studies, bootstrap confidence intervals, and sensitivity analyses showed that the hybrid models produced statistically reliable and stable predictions under a range of conditions.

5.5 Linking Findings to Literature and Objectives

The outcomes of this research build upon and extend existing literature in several important ways:

- Previous studies have highlighted the limitations of traditional models, particularly during complex demand transitions.
  The hybrid model developed in this study confirmed these limitations and offered a higher-performing alternative.
- Research in deep learning for energy forecasting often focuses on a single model (e.g., LSTM or CNN).
  By integrating CNN, GRU, and LSTM in a unified model, this thesis demonstrates the value of multiscale feature extraction and temporal modelling.
- Transfer learning in building energy research remains relatively underexplored.
  The demonstrated success of the hybrid model in cross-building adaptation provides empirical evidence that transferable representations can reduce data requirements and support scalable deployment.

These findings collectively fulfil all three research objectives and offer tangible contributions to the predictive analytics domain in smart buildings.

5.6 Practical and Policy Implications

The work carries several practical implications for building operators, energy managers, and policymakers:

- **Operational decision-making:** Accurate 24-hour forecasts enable more precise load-shifting, tariff selection, and daily scheduling of HVAC and lighting systems.
- **Cost reduction:** Anticipating peak demand supports demand-charge mitigation strategies, potentially lowering energy bills by 10–15%.
- **Infrastructure planning:** Longer-horizon forecasts support budgeting, maintenance, and procurement decisions at the organizational and municipal level.
- **Scalability:** The demonstrated success of transfer learning implies that forecasting tools can be deployed across large building portfolios without the need to retrain models from scratch.
- **Sustainability goals:** Improved forecasting accuracy contributes to reduced energy waste, better demand management, and more efficient integration of renewable energy sources.

These practical benefits illustrate how the proposed framework supports both operational efficiency and strategic energy management.

5.7 Limitations of the Study

Although the research makes substantive contributions, several limitations must be acknowledged:

- The datasets used, while diverse, represent a limited range of climate zones and building functions.
  Broader validation across different climates and building types would further strengthen generalisability.
- External factors such as occupant behaviour, weather anomalies, or equipment faults were not explicitly modelled.
  Incorporating contextual data could improve spike prediction accuracy.
- Computational cost, particularly for deep hybrid models, remains a challenge despite their superior performance.
  Deployment in edge-computing or low-resource environments may require model compression techniques.
- Transfer learning experiments were limited to buildings with similar operational patterns.
  The performance of transfer learning across drastically different buildings warrants further investigation.

Recognizing these limitations offers opportunities for enhancing the model in future research.

5.8 Recommendations for Future Work

Based on the findings and limitations, several avenues for future research emerge:

- **Integrate behavioural and contextual data** such as occupancy patterns, equipment schedules, and real-time weather features to improve peak prediction accuracy.
- **Explore cross-climate generalisation** by testing the hybrid models in regions with extreme temperature variations or different energy infrastructures.
- **Investigate model compression and optimisation**, including knowledge distillation, pruning, and quantization, to support deployment on low-power devices.
- **Extend transfer learning to multi-building federated learning**, allowing models to be updated collaboratively without sharing sensitive data.
- **Incorporate reinforcement learning** to move beyond forecasting and enable autonomous energy management actions.
- **Examine uncertainty-aware forecasting**, where probabilistic outputs support risk-informed decision-making for grid stability and financial planning.

Each of these directions would build on the foundations laid in this thesis, contributing to the next generation of predictive models for smart buildings.

5.9 Final Conclusion

This thesis was undertaken to advance the field of building energy forecasting by examining how hybrid machine learning models and transfer learning can be combined to deliver accurate, adaptable, and analytically robust predictions across multiple time horizons and building contexts. The research has shown that carefully constructed hybrid models particularly the sequential SVR → XGBoost → LSTM model and the deep CNN–GRU–LSTM model offer clear advantages over traditional statistical approaches and single-model baselines. By integrating complementary modelling components, these hybrids capture short-term fluctuations, structural feature interactions, and long-range temporal dependencies in ways that individual models cannot achieve alone.

The results also demonstrate the substantial potential of transfer learning for cross-building adaptation. Models pretrained on diverse building datasets can be efficiently fine-tuned for new buildings with limited data while retaining strong predictive performance. This finding is especially relevant for real-world applications, where many buildings lack long historical records. Transfer learning emerges as a practical and scalable mechanism for extending advanced forecasting capabilities across building portfolios, campuses, and urban environments.

A further contribution of this research is the comprehensive validation framework it employs. Through statistical significance testing, residual diagnostics, peak-event evaluation, and sensitivity analysis, the thesis provides a deeper understanding of model behaviour beyond traditional accuracy metrics. These evaluations confirm that the proposed models are not only accurate, but also stable, reliable, and interpretable qualities that are critical for supporting informed decision-making in building energy management.

Taken together, the findings offer a coherent methodological foundation and strong empirical evidence in support of hybrid and transferable forecasting models. These approaches hold considerable promise for enabling smarter operational control, more informed energy planning, and more resilient and sustainable management of building energy demand. More broadly, the thesis contributes to the growing body of research demonstrating how modern machine learning techniques can play a meaningful role in shaping intelligent and energy-aware built environments.

6. References

[1] D. Ürge-Vorsatz et al., “Advances toward a net-zero global building sector,” *Annu. Rev. Environ. Resour.*, vol. 50, no. 1, pp. 7.1–7.28, Oct. 2025, doi: 10.1146/annurev-environ-012420.

[2] I. Lampropoulos, T. Alskaif, W. Schram, E. Bontekoe, S. Coccato, and W. van Sark, “Review of energy in the built environment,” *Smart Cities*, vol. 3, no. 2, pp. 454–487, Jun. 2020, doi: 10.3390/smartcities3020015.

[3] Z. Ma et al., “An overview of emerging and sustainable technologies for increased energy efficiency and carbon emission mitigation in buildings,” *Buildings*, vol. 13, no. 10, Oct. 2023, doi: 10.3390/buildings13102658.

[4] E. Navarro Bringas and G. A. G. R. Godawatte, “Shedding light on the efforts into the rehabilitation of a major culprit of carbon emissions: A scientometric analysis of net-zero in the built environment sector,” *Energy Build.*, vol. 266, pp. 112119, Jul. 2022, doi: 10.1016/j.enbuild.2022.112119.

[5] C. Rottondi et al., “An energy management service for the smart office,” *Energies*, vol. 8, no. 10, pp. 11667–11684, Oct. 2015, doi: 10.3390/en81011667.

[6] P. Macieira, L. Gomes, and Z. Vale, “Energy management model for HVAC control supported by reinforcement learning,” *Energies*, vol. 14, no. 24, Dec. 2021, doi: 10.3390/en14248210.

[7] D. Ramos, B. Teixeira, P. Faria, L. Gomes, O. Abrishambaf, and Z. Vale, “Using diverse sensors in load forecasting in an office building to support energy management,” *Energy Rep.*, vol. 6, pp. 182–187, Dec. 2020, doi: 10.1016/j.egyr.2020.11.100.

[8] H. Farzaneh, L. Malehmirchegini, A. Bejan, T. Afolabi, A. Mulumba, and P. P. Daka, “Artificial intelligence evolution in smart buildings for energy efficiency,” *Appl. Sci.*, vol. 11, no. 2, Jan. 2021, doi: 10.3390/app11020763.

[9] R. Panchalingam and K. C. Chan, “A state-of-the-art review on artificial intelligence for smart buildings,” *Intell. Build. Int.,* vol. 13, no. 3, pp. 155–180, Oct. 2021, doi: 10.1080/17508975.2019.1613219.

[10] B. Muniandi et al., “AI-driven energy management systems for smart buildings,” *Power Tech J.,* vol. 48, no. 1, pp. 322–337, Sep. 2024. [Online]. Available: https://powertechjournal.com.

[11] A. Kumar Singh, S. Khatoon, M. Muazzam, and D. K. Chaturvedi, “An overview of electricity demand forecasting techniques,” *Netw. Complex Syst.*, vol. 3, no. 2, pp. 38–48, 2013. [Online]. Available: www.iiste.org.

[12] L. Wang, J. Wu, Y. Cao, and Y. Hong, “Forecasting renewable energy stock volatility using short and long-term Markov switching GARCH-MIDAS models: Either, neither or both?” *Energy Econ.,* vol. 111, May 2022, doi: 10.1016/j.eneco.2022.106056.

[13] D. Mariano-Hernández, L. Hernández-Callejo, F. S. García, O. Duque-Perez, and A. L. Zorita-Lamadrid, “A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics,” *Appl. Sci.*, vol. 10, no. 23, Dec. 2020, doi: 10.3390/app10238323.

[14] C. Ding, J. Ke, M. Levine, and N. Zhou, “Potential of artificial intelligence in reducing energy and carbon emissions of commercial buildings at scale,” *Nat. Commun*., vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-50088-4.

[15] S. Hadri, M. Najib, M. Bakhouya, Y. Fakhri, and M. El Arroussi, “Performance evaluation of forecasting strategies for electricity consumption in buildings,” *Energies*, vol. 14, no. 18, Sep. 2021, doi: 10.3390/en14185831.

[16] B. Tian, *A LiDAR DSM based geometry modelling method to improve solar irradiance simulation and PV yield prediction in urban environments*, M.S. thesis, Dept. Built Environ., Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2021.

[17] D. Masa-Bote et al., “Improving photovoltaics grid integration through short time forecasting and self-consumption,” *Appl. Energy*, vol. 125, pp. 103–113, Jul. 2014, doi: 10.1016/j.apenergy.2014.03.045.

[18] G. Arjunan, “Optimizing edge AI for real-time data processing in IoT devices: Challenges and solutions,” *Int. J. Sci. Res. Manag*., vol. 11, no. 6, pp. 944–953, Jun. 2023, doi: 10.18535/ijsrm/v11i06.ec2.

[19] A. Balal and M. Giesselmann, “Demand side management and economic analysis using battery storage system (BSS) and solar energy*,” in Proc. 2021 IEEE 4th Int. Conf. Power Energy Appl. (ICPEA)*, 2021, pp. 141–146, doi: 10.1109/ICPEA52760.2021.9639359.

[20] K. Antoniadou-Plytaria *et al.*, “Effect of short-term and high-resolution load forecasting errors on microgrid operation costs,” in *Proc. 2022 IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT-Europe)*, Oct. 2022, pp. 1–5.

[21] S. Sanghavi, S. Shakkottai, M. Lelarge, and B. Schroeder, “The 2014 ACM International Conference on Measurement and Modeling of Computer Systems,” in *Proc. ACM SIGMETRICS Int. Conf. Measurement and Modeling of Computer Systems*, Jun. 2014.

[22] E. R. Stephens, D. B. Smith, and A. Mahanti, “Game theoretic model predictive control for distributed energy demand-side management*,” IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1394–1402, May 2015, doi: 10.1109/TSG.2014.2377292.

[23] S. S. W. Fatima and A. Rahimi, “A review of time-series forecasting algorithms for industrial manufacturing systems,” *Machines*, vol. 12, no. 6, Jun. 2024, doi: 10.3390/machines12060380.

[24] Y. J. Kim, S. J. Lee, H. S. Jin, I. A. Suh, and S. Y. Song, “Comparison of linear and nonlinear statistical models for analyzing determinants of residential

energy consumption," *Energy Build.*, vol. 223, Sep. 2020, doi: 10.1016/j.enbuild.2020.110226.

[25] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy Build.*, vol. 49, pp. 591–603, Jun. 2012, doi: 10.1016/j.enbuild.2012.03.010.

[26] S. Zou, X. Luo, and Z. Yang, "Energy consumption forecasting in buildings based on long-term and short-term memory networks," *in Proc. 2024 Int. Conf. Manag., Innov. Ind. Intell. (ICMIII),* Sep. 2024, pp. 831–835, doi: 10.1109/ICMIII62623.2024.00162.

[27] F. Ziel and R. Weron, "Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks," *Energy Econ.*, vol. 70, pp. 396–420, May 2018, doi: 10.1016/j.eneco.2017.12.016.

[28] G. G. Neto, S. B. Defilippo, and H. S. Hippert, "Univariate versus multivariate models for short-term electricity load forecasting," *in Proc. 13th Argentinian Symp. Oper. Res. (SIO),* Rosario, Argentina, 2015, pp. 143–151.

[29] S. Salehi, M. Kavgic, H. Bonakdari, and L. Begnoche, "Comparative study of univariate and multivariate strategy for short-term forecasting of heat demand density: Exploring single and hybrid deep learning models," *Energy AI,* vol. 16, pp. 100343, May 2024, doi: 10.1016/j.egyai.2024.100343.

[30] R. Li, X. Zhang, L. Liu, Y. Li, and Q. Xu, "Application of neural network to building environmental prediction and control," *Build. Serv. Eng. Res. Technol.*, vol. 41, no. 1, pp. 25–45, Jan. 2020, doi: 10.1177/0143624419838362.

[31] A. Schumann, J. Hayes, P. Pompey, and O. Verscheure, "Adaptable fault identification for smart buildings," *in Proc. AAAI Workshop Artif. Intell. Smarter Living (AISL),* 2011. [Online]. Available: www.aaai.org

[32] S. Herzog, D. Atabay, J. Jungwirth, and V. Mikulovic, “Self-adapting building models for model predictive control,” *in Proc. Building Simulation 2013,* Chambéry, France, Aug. 2013.

[33] L. Choudhary and J. S. Choudhary, “Deep learning meets machine learning: A synergistic approach towards artificial intelligence*,” J. Sci. Res. Rep.,* vol. 30, no. 11, pp. 865–875, Nov. 2024, doi: 10.9734/jsrr/2024/v30i112614.

[34] B. Aslam, A. Zafar, and U. Khalil, “Development of integrated deep learning and machine learning algorithm for the assessment of landslide hazard potential,” *Soft Comput.*, vol. 25, no. 21, pp. 13493–13512, Nov. 2021, doi: 10.1007/s00500-021-06249-4.

[35] S. Degadwala and D. Vyas, “Systematic analysis of deep learning models vs. machine learning,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 4, pp. 60–70, Jul. 2024, doi: 10.32628/cseit24104108.

[36] F. Zhang and L. J. O'Donnell, “Support vector regression,” in *Machine Learning*, Academic Press, 2020, pp. 123–140, doi: 10.1016/B978-0-12-815739-8.00007-9.

[37] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, “Single and multi-sequence deep learning models for short and medium term electric load forecasting,” *Energies*, vol. 12, no. 1, pp. 149, Jan. 2019, doi: 10.3390/en12010149.

[38] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, “Energy load forecasting using deep learning approach—LSTM and GRU in Spark cluster,” in *Proc. 2018 Fifth Int. Conf. Emerging Applications of Information Technology (EAIT)*, Kolkata, India, 2018, pp. 1–4, doi: 10.1109/EAIT.2018.8470406.

[39] U. Ugurlu, I. Oksuz, and O. Tas, “Electricity price forecasting using recurrent neural networks,” *Energies*, vol. 11, no. 5, pp. 1255, May 2018, doi: 10.3390/en11051255.

[40] S. Emshagin, W. K. Halim, and R. Kashef, “Short-term prediction of household electricity consumption using customized LSTM and GRU

models," *arXiv preprint*, arXiv:2212.08757, Des. 2022, doi: 10.48550/arXiv.2212.08757.

[41] P. Pawar, S. Gonjari, S. Kshirsagar, and G. J. Chhajed, "A review of linear regression and support vector regression," *Int. J. Sci. Res. Eng. Manag.,* vol. 8, no. 12, pp. 1–9, Dec. 2024, doi: 10.55041/IJSREM40400.

[42] X. Yu et al., "Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting," *J. Hydrol.,* vol. 582, pp. 124293, Mar. 2020, doi: 10.1016/j.jhydrol.2019.124293.

[43] K. P. Bennett, M. Momma, and M. J. Embrechts, "MARK: A boosting algorithm for heterogeneous kernel models," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Jul. 2002, pp. 24–31, doi: 10.1145/775047.775051.

[44] N. V. Uma Reddy *et al.*, "Transfer learning for energy consumption forecasting in smart buildings," in *Proc. 7th Int. Conf. Contemp. Comput. Informat. (IC3I)*, vol. 7, Sep. 2024, pp. 1549–1555, doi: 10.1109/IC3I61595.2024.10828603.

[45] Y. Chen, Z. Tong, Y. Zheng, H. Samuelson, and L. Norford, "Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings," *J. Clean Prod.*, vol. 254, pp. 119866, May 2020, doi:10.1016/j.jclepro.2019.119866.

[46] X. Zhang and P. Li, "Transfer learning in the transformer model for thermal comfort prediction: A case of limited data," *Energies*, vol. 16, no. 20, pp. 7137, Oct. 2023, doi: 10.3390/en16207137.

[47] Z. Ni, C. Zhang, M. Karlsson, and S. Gong, "A study of deep learning-based multi-horizon building energy forecasting," *Energy Build.*, vol. 303, pp. 113810, Jan. 2024, doi: 10.1016/j.enbuild.2023.113810.

[48] H. Daniel, B. R. K. Mantha, and B. G. De Soto, "Towards a review of building energy forecast models," in *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience – Sel. Papers ASCE Int. Conf. Comput. Civil Eng. 2019*, Reston, VA, USA: ASCE, Jun. 2019, pp. 74–82, doi: 10.1061/9780784482445.010.

[49] Y. Sun, F. Haghighat, and B. C. M. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," *Energy Build.*, vol. 221, pp. 110022, Aug. 2020, doi: 10.1016/j.enbuild.2020.110022.

[50] G. Li, W. Tian, H. Zhang, and B. Chen, "Building energy models at different timescales based on multi-output machine learning," *Buildings*, vol. 12, no. 12, pp. 2109, Dec. 2022, doi: 10.3390/buildings12122109.

[51] R. Yan, T. Zhao, Y. Rezgui, S. Kubicki, and Y. Li, "Transferability and robustness of a data-driven model built on a large number of buildings," *J. Build. Eng.*, vol. 80, pp. 108127, Dec. 2023, doi: 10.1016/j.jobe.2023.108127.

[52] M. V. Torres, Z. Shahid, K. Mitra, S. Saguna, and C. Ahlund, "A transfer learning approach to create energy forecasting models for building fleets," in *Proc. 2024 IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Sep. 2024, pp. 438–444, doi: 10.1109/SmartGridComm60555.2024.10738094.

[53] M. Nawar, M. Shomer, S. Faddel, and H. Gong, "Transfer learning in deep learning models for building load forecasting: Case of limited data," in *Proc. SoutheastCon 2023*, Orlando, FL, USA, 2023, pp. 532–538, doi: 10.1109/SoutheastCon51012.2023.10115128

[54] M. G. S. Murshed, C. Murphy, D. Hou, N. Khan, G. Ananthanarayanan, and F. Hussain, "Machine learning at the network edge: A survey," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–37, Oct. 2021, doi: 10.1145/3469029.

[55] D. Chaudhary, S. K. Verma, S. Pundir, S. Ranganathan, V. Umadevi, and K. Srinivasan, "IoT revolutionization using federated learning and decentralized

AI on edge devices," in *Proc. 2024 7th Int. Conf. Contemp. Comput. Informat. (IC3I)*, Greater Noida, India, Des. 2024, pp. 311–316, doi: 10.1109/IC3I61595.2024.10829063.

[56] S. Kulkarni et al., "Enabling a decentralized smart grid using autonomous edge control devices," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7406–7419, Oct. 2019, doi: 10.1109/JIOT.2019.2898837.

[57] P. Pääkkönen, D. Pakkala, J. Kiljander, and R. Sarala, "Architecture for enabling edge inference via model transfer from cloud domain in a kubernetes environment," *Future Internet*, vol. 13, no. 1, pp. 1–24, Jan. 2021, doi: 10.3390/fi13010005.

[58] B. Anarene, "Revolutionizing energy efficiency in commercial and institutional buildings: A complete analysis," *Int. J. Sci. Res. Manag*., vol. 12, no. 9, pp. 7444–7468, Sep. 2024, doi: 10.18535/ijsrm/v12i09.em12.

[59] M. Santamouris and K. Vasilakopoulou, "Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation," *Prog. Ind. Ecol.*, vol. 12, no. 1, Jan. 2021, doi: 10.1016/j.prime.2021.100002.

[60] B. G. Rebelatto, A. L. Salvia, L. L. Brandli, and W. Leal Filho, "Examining energy efficiency practices in office buildings through the lens of LEED, BREEAM, and DGNB certifications," *Sustainability,* vol. 16, no. 11, Jun. 2024, doi: 10.3390/su16114345.

[61] F. Meng, K. Weng, B. Shallal, X. Chen, and M. Mourshed, "Forecasting algorithms and optimization strategies for building energy management and demand response," *Proc.*, vol. 2, no. 15, p. 1133, Aug. 2018, doi: 10.3390/proceedings2151133.

[62] D. Kaur, S. N. Islam, M. A. Mahmud, M. E. Haque, and Z. Y. Dong, "Energy forecasting in smart grid systems: Recent advancements in probabilistic deep

learning," *IET Gener. Transm. Distrib.*, vol. 16, no. 22, pp. 4461–4479, Nov. 2022, doi: 10.1049/gtd2.12603.

[63] U. Kazmi, C. Fu, and C. Miller, "Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale," *Build. Environ.*, vol. 239, pp. 110407, Jul. 2023, doi: 10.1016/j.buildenv.2023.110407.

[64] A. K. Prakash, S. Xu, R. Rajagopal, and H. Y. Noh, "Robust building energy load forecasting using physically-based kernel models," *Energies*, vol. 11, no. 4, Apr. 2018, doi: 10.3390/en11040862.

[65] T. Ahmad, C. Huanxin, D. Zhang, and H. Zhang, "Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions," *Energy*, vol. 198, May 2020, doi: 10.1016/j.energy.2020.117283.

[66] Y. D. Kononov, D. Y. Kononov, and S. V. Steklova, "The effect of length of forecast horizon on rational aggregation in long-term forecasting of energy systems development," *Energy Syst. Res.,* vol. 1, no. 1, pp. 51–56, Apr. 2018, doi: 10.25729/esr.2018.01.0006.

[67] A. Prabowo, K. Chen, H. Xue, S. Sethuvenkatraman, and F. D. Salim, "Continually learning out-of-distribution spatiotemporal data for robust energy forecasting," *in Proc. 2023 Int. Conf. Adv. Data Mining Appl*., vol. 14175, pp. 3–18, Sep. 2023, doi: 10.1007/978-3-031-43430-3_1.

[68] O. O. Omogoroye, O. O. Olaniyi, O. O. Adebiyi, T. O. Oladoyinbo, and F. G. Olaniyi, "Electricity consumption (kW) forecast for a building of interest based on a time series nonlinear regression model," *Asian J. Econ. Bus. Accounting*, vol. 23, no. 21, pp. 197–207, Oct. 2023, doi: 10.9734/ajeba/2023/v23i211127.

[69] X. Godinho, H. Bernardo, J. C. de Sousa, and F. T. Oliveira, "A data-driven approach to forecasting heating and cooling energy demand in an office

building as an alternative to multi-zone dynamic simulation," *Appl. Sci.*, vol. 11, no. 4, Feb. 2021, doi: 10.3390/app11041356.

[70] D. Ramos, M. Khorram, P. Faria, and Z. Vale, "Load forecasting in an office building with different data structure and learning parameters," *Forecasting*, vol. 3, no. 1, pp. 242–255, Mar. 2021, doi: 10.3390/forecast3010015.

[71] N. K. Thokala and M. G. Chandra, "Poster abstract: Disaggregated forecasting for large office buildings," *in Proc. 3rd ACM Conf. Syst. Energy-Efficient Built Environ. (BuildSys)*, pp. 231–232, Nov. 2016, doi: 10.1145/2993422.2996400.

[72] A. Ashouri, Z. Shi, and H. B. Gunay, "Data-driven short-term load forecasting for heating and cooling demand in office buildings," *J. Phys.: Conf. Ser.*, vol. 1343, no. 1, pp. 012038, Nov. 2019, doi: 10.1088/1742-6596/1343/1/012038.

[73] L. G. Baca Ruiz, M. P. Cuéllar, M. D. Calvo-Flores, and M. D. C. Pegalajar Jiménez, "An application of non-linear autoregressive neural networks to predict energy consumption in public buildings," *Energies*, vol. 9, no. 9, Sep. 2016, doi: 10.3390/en9090684.

[74] G. C. Chasparis and T. Natschläger, "Regression models for output prediction of thermal dynamics in buildings," *J. Dyn. Syst., Meas., Control*, vol. 139, no. 2, pp. 021006, Feb. 2017, Doi:10.1115/1.4034746

[75] I. Alpackaya, M. H. Fallah, N. Mounika, S. Sood, S. Rajvanshi, S. Lakhanpal, P. Cajla, A. Sharma, and Y. S. Lalitha, "Renewable energy forecasting using deep learning techniques," in *E3S Web Conf.*, vol. 581, pp. 01011, EDP Sciences, Oct. 2024, doi: 10.1051/e3sconf/202458101011.

[76] A. Nanjar, R. E. Saputro, and B. Berlilana, "Machine learning and deep learning approaches for energy prediction: A systematic literature review," *Sinkron*, vol. 8, no. 4, pp. 2603–2614, Nov. 2024, doi: 10.33395/sinkron.v8i4.14208.

[77] A. Rosato, F. Succetti, R. Araneo, A. Andreotti, M. Mitolo, and M. Panella, "A combined deep learning approach for time series prediction in energy

environments," in *Proc. 2020 IEEE/IAS 56th Ind. Commer. Power Syst. Tech. Conf. (I&CPS)*, Las Vegas, NV, USA, Jun. 2020, pp. 1–5, doi: 10.1109/ICPS48389.2020.9176818.

[78] M. S. Hernández, "Beliefs and attitudes of Canarians towards the Chilean linguistic variety," *Lenguas Modernas*, no. 62, pp. 183–209, Jul. 2023, doi: 10.13039/501100011033.

[79] S. Suddala, "Dynamic demand forecasting in supply chains using hybrid ARIMA–LSTM architectures," *Int. J. Adv. Res.,* vol. 12, no. 10, pp. 1167–1171, Oct. 2024, doi: 10.21474/IJAR01/19738.

[80] Y. Syu and C. M. Wang, "QoS time series modeling and forecasting for web services: A comprehensive survey," *IEEE Trans. Netw. Serv. Manag.,* vol. 18, no. 1, pp. 926–944, Mar. 2021, doi: 10.1109/TNSM.2021.3056399.

[81] H. V. P. Singh and Q. H. Mahmoud, "Evaluation of ARIMA models for human–machine interface state sequence prediction," *Mach. Learn. Knowl. Extr.,* vol. 1, no. 1, pp. 287–311, Dec. 2019, doi: 10.3390/make1010018.

[82] S. N. Hamidah, N. Salam, and D. S. Susanti, "Teknik peramalan menggunakan metode pemulusan eksponensial Holt-Winters," *Epsilon: J. Mat. Murni Terap.,* vol. 7, no. 2, pp. 26–33, Nov. 2017, doi: https://doi.org/10.20527/epsilon.v7i2.97.

[83] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.,* vol. 54, no. 8, pp. 799–805, Aug. 2003, doi: 10.1057/palgrave.jors.2601589.

[84] C. Chatfield and M. Yar, "Holt-Winters forecasting: Some practical issues," *Statistician*, vol. 37, no. 2, pp. 129–140, Jun. 1988, doi: 10.2307/2348687.

[85]] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Using exponential smoothing methods for modelling and forecasting short-term electricity demand," *Int. J. Forecast.*, vol. 16, no. 3, pp. 317–334, Oct. 2000, doi: 10.1016/S0169-2070(00)00065-0.

[86] S. Wilfling, M. Ebrahimi, Q. Alfalouji, G. Schweiger and M. Basirat, "Learning Non-linear White-box Predictors: A Use Case in Energy Systems," in *Proc. 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, Nassau, Bahamas, 2022, pp. 507–512, doi: 10.1109/ICMLA55696.2022.00082.

[87] R. Baidya and S. W. Lee, "Addressing the non-stationarity and complexity of time series data for long-term forecasts," *Appl. Sci.,* vol. 14, no. 11, Jun. 2024, doi: 10.3390/app14114436.

[88] M. E. Bonilla Jr, J. McDonald, T. Toth and B. Sadler, "Traditional vs Machine Learning Approaches: A Comparison of Time Series Modeling Methods," *SMU Data Science Review*, vol. 7, no. 2, pp. 2, Aug. 2023. [Online] Available: https://scholar.smu.edu/datasciencereview/vol7/iss2/2

[89] A. Mosavi and A. Bahmani, "Energy consumption prediction using machine learning: A review," *Preprints*, Mar. 2019, doi: 10.20944/preprints201903.0131.v1.

[90] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine learning, deep learning and statistical analysis for forecasting building energy consumption—A systematic review," *Eng. Appl. Artif. Intell.*, vol. 114, Oct. 2022, doi: 10.1016/j.engappai.2022.105287.

[91] D. Li, Z. Qi, Y. Zhou, and M. Elchalakani, "Machine learning applications in building energy systems: Review and prospects," *Buildings*, vol. 15, no. 4, p. 648, Feb. 2025.

[92] M. R. Senouci, M. A. Benatia, M. R. Boulahia Senouci, and S. Yacin, *Advances in Computing Systems and Applications*. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-12097-8.

[93] W. Yang, J. Wang, and R. Wang, "Research and application of a novel hybrid model based on data selection and artificial intelligence algorithm for short

term load forecasting," *Entropy*, vol. 19, no. 2, Jan 2017, doi: 10.3390/e19020052.

[94] S. Ajmera, A. K. Singh, and V. Chauhan, "An approach towards medium term forecasting based on support vector regression," in *Proc. 2016 IEEE 7th Power India Int. Conf. (PIICON)*, Bikaner, India, Nov. 2016, pp. 1–6, doi: 10.1109/POWERI.2016.8077232.

[95] I. Ortiz-Aguirre, M. Espinoza-Andaluz, and J. Barzola-Monteses, "Ensemble learning models applied in energy time series of a university building," in *Proc. 2022 IEEE Latin American Conf. Comput. Intell. (LA-CCI)*, Montevideo, Uruguay, 2022, pp. 1–6, doi: 10.1109/LA-CCI54402.2022.9981839.

[96] A. S. Mohammed et al., "Stacking ensemble tree models to predict energy performance in residential buildings," *Sustainability*, vol. 13, no. 15, Aug. 2021, doi: 10.3390/su13158298.

[97] A. Agarwal, Y. S. Tan, O. Ronen, C. Singh, and B. Yu, "Hierarchical Shrinkage: Improving the accuracy and interpretability of tree-based models," in *Proc. 39th Int. Conf. Mach. Learn. (ICML)*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, and S. Sabato, Eds., vol. 162, PMLR, 17–23 Jul. 2022, pp. 111–135. [Online]. Available: https://proceedings.mlr.press/v162/agarwal22b.html

[98] V. G. Sivakumar, N. Arunfred, N. Anusha, C. Balakrishnan, B. Meenakshi, and S. Sujatha, "A gradient boosting algorithm to predict energy consumption for home applications," in *Proc. 2024 2nd Int. Conf. Comput., Commun. and Control (IC4)*, Indore, India, 2024, pp. 1–5, doi: 10.1109/IC457434.2024.10486226.

[99] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," *Energy Build.*, vol. 158, pp. 1533–1543, Jan. 2018, doi: 10.1016/j.enbuild.2017.11.039.

[100] J. Moon, S. Rho, and S. W. Baik, “Toward explainable electrical load forecasting of buildings: A comparative study of tree-based ensemble methods with Shapley values,” *Sustain. Energy Technol. Assess.*, vol. 54, pp. 102888, Dec. 2022, doi: 10.1016/j.seta.2022.102888.

[101] E. Mele, “A review of machine learning algorithms used for load forecasting at microgrid level,” *in Sinteza 2019 – Int. Sci. Conf. Inf. Technol. Data Related Res.*, Belgrade, Serbia, 2019, pp. 452–458, doi: 10.15308/Sinteza-2019-452-458.

[102] J. P. Marques de Sá, *Pattern Recognition: Concepts, Methods and Applications*. Berlin, Germany: Springer, 2012

[103] N. Saxena et al., “Hybrid KNN-SVM machine learning approach for solar power forecasting,” *Environ. Challenges*, vol. 14, pp. 100838, Jan. 2024, doi: 10.1016/j.envc.2024.100838.

[104] P. C. Albuquerque, D. O. Cajueiro, and M. D. C. Rossi, “Machine learning models for forecasting power electricity consumption using a high dimensional dataset,” *Expert Syst. Appl.*, vol. 187, pp. 115917, Jan. 2022, doi: 10.1016/j.eswa.2021.115917.

[105] J. Olowolaju and H. Livani, “Comparison of machine learning models for week-ahead load forecasting in short-term power system planning,” in *Proc. 2022 North American Power Symp. (NAPS)*, Salt Lake City, UT, USA, 2022, pp. 1–6, doi: 10.1109/NAPS56150.2022.10012181.

[106] J. A. Ilemobayo et al., “Hyperparameter tuning in machine learning: A comprehensive review,” *J. Eng. Res. Rep.,* vol. 26, no. 6, pp. 388–395, Jun. 2024, doi: 10.9734/jerr/2024/v26i61188.

[107] S. Rajput, A. Tripathi, and A. Kumar, “The influence of hyperparameter tuning on machine learning model performance: A theoretical exploration,” *Pharma Innov. J.*, vol. 8, no. 3S, pp. 1–5, Jan. 2019, doi: 10.22271/tpi.2019.v8.i3sa.25247.

[108] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, Mar. 2019, doi: 10.11989/JEST.1674-862X.80904120.

[109] S. P. Shaji, R. R., J. Varghese, L. Sathyan, and D. J., "Optimizing hyperparameters: Techniques for improving machine learning models," *Int. Res. J. Adv. Eng. Manag. (IRJAEM),* vol. 2, no. 12, pp. 3782–3787, Dec. 2024, doi: 10.47392/IRJAEM.2024.561.

[110] P. A. Mynhoff, E. Mocanu, and M. Gibescu, "Statistical learning versus deep learning: Performance comparison for building energy prediction methods," *in Proc. 8th IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT-Europe)*, 2018, pp. 1–6

[111] N. Somu, G. Raman M. R., and K. Ramamritham, "A deep learning framework for building energy consumption forecast," Renew. Sustain. *Energy Rev.,* vol. 137, Mar. 2021, art. 110591, doi: 10.1016/j.rser.2020.110591.

[112] S. K. Paramasivan, "Deep learning based recurrent neural networks to enhance the performance of wind energy forecasting: A review," *Rev. Intell. Artif.*, vol. 35, no. 1, pp. 1–10, Feb. 2021, doi: 10.18280/ria.350101.

[113] S. H. Noh, "Analysis of gradient vanishing of RNNs and performance comparison," Information, vol. 12, no. 11, Nov. 2021, doi: 10.3390/info12110442.

[114] I. Amalou, N. Mouhni, and A. Abdali, "Multivariate time series prediction by RNN architectures for energy consumption forecasting," *Energy Rep*., vol. 8, pp. 1084–1091, Nov. 2022, doi: 10.1016/j.egyr.2022.07.139.

[115] A. D. Huynh and T. K. Nguyen, "The comparison of GRU and LSTM in solar power generation forecasting application," *Int. J. Sci. Res. Archive*, vol. 13, no. 1, pp. 1360–1370, Sep. 2024, doi: 10.30574/ijsra.2024.13.1.1831.

[116] S. Yang, X. Yu, and Y. Zhou, “LSTM and GRU neural network performance comparison study: Taking Yelp review dataset as an example,” in *Proc. 2020 Int. Workshop on Electron. Commun. and Artif. Intell. (IWECAI)*, Shanghai, China, 2020, pp. 98–101, doi: 10.1109/IWECAI50956.2020.00027.

[117] A. Abdelsalam Ismail, T. Wood, and H. Corrada Bravo, “Improving long-horizon forecasts with expectation-biased LSTM networks,” *arXiv e-prints*, Apr. 2018, doi: 10.48550/arXiv.1804

[118] J. Palet, V. Manquinho, and R. Henriques, “Multiple-input neural networks for time series forecasting incorporating historical and prospective context,” *Data Mining Knowl. Discov.*, vol. 38, no. 1, pp. 315–341, Jan. 2024, doi: 10.1007/s10618-023-00984-y.

[119] M. Dong and L. Grumbach, “A hybrid distribution feeder long-term load forecasting method based on sequence prediction,” *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–480, Jan. 2020, doi: 10.1109/TSG.2019.2924183.

[120] I. Koprinska, D. Wu, and Z. Wang, “Convolutional neural networks for energy time series forecasting,” in *Proc. 2018 Int. Joint Conf. Neural Netw. (IJCNN),* 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489078.

[121] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, “Temporal convolutional networks applied to energy-related time series forecasting,” *Appl. Sci*., vol. 10, no. 7, Apr. 2020, doi: 10.3390/app10072322.

[122] M. Ahsan, M. O. Aziz, S. A. S. Bokhari, M. Haris, M. Iqbal, and K. Ullah, “Enhancing short-term load forecasting accuracy in the power systems using a hybrid CNN-GRU model,” in *Proc. 2024 Int. Conf. Electr., Commun. and Comput. Eng. (ICECCE)*, Kuala Lumpur, Malaysia, 2024, pp. 1–5, doi: 10.1109/ICECCE63537.2024.10823420.

[123] A. Mahmoud and A. Mohammed, “Leveraging hybrid deep learning models for enhanced multivariate time series forecasting,” *Neural Process. Lett.,* vol. 56, no. 5, pp. 4519–4541, Oct. 2024, doi: 10.1007/s11063-024-11656-3.

[124] F. Ünal, A. Almalaq, and S. Ekici, "A novel load forecasting approach based on smart meter data using advance preprocessing and hybrid deep learning," *Appl. Sci.*, vol. 11, no. 6, Mar. 2021, doi: 10.3390/app11062742.

[125] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," *arXiv preprint*, arXiv:2202.07125, Feb. 15, 2022

[126] S. Joseph, A. A. Jo, and E. Deni Raj, "Improving time series forecasting accuracy with transformers: A comprehensive analysis with explainability," in *Proc. 2024 3rd Int. Conf. Electr., Electron., Inf. and Commun. Technol. (ICEEICT)*, Trichirappalli, India, 2024, pp. 1–7, doi: 10.1109/ICEEICT61591.2024.10718609.

[127] K. Lu, M. Huo, Y. Li, Q. Zhu, and Z. Chen, "CT-PatchTST: Channel-time patch time-series transformer for long-term renewable energy forecasting," *arXiv preprint*, arXiv:2501.08620, Jan. 15, 2025

[128] X. Feng and Z. Lyu, "How features benefit: Parallel series embedding for multivariate time series forecasting with transformer," in *Proc. 2022 IEEE 34th Int. Conf. Tools with Artif. Intell. (ICTAI)*, Macao, China, 2022, pp. 967–975, doi: 10.1109/ICTAI56018.2022.00148.

[129] W. Choi and S. Lee, "Performance evaluation of deep learning architectures for load and temperature forecasting under dataset size constraints and seasonality," *Energy Build.*, vol. 288, no. 1, p. 113027, Jun. 2023, doi: 10.1016/j.enbuild.2023.113027

[130] F. Peng, T. Su, Q. Zeng, and X. Han, "Climate-adaptive energy forecasting in green buildings via attention-enhanced Seq2Seq transfer learning," *Scientific Reports*, vol. 15, no. 1, p. 31829, Aug. 29, 2025.

[131] P. Zheng, H. Zhou, J. Liu, and Y. Nakanishi, "Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture," *Appl. Energy,* vol. 349, Nov. 2023, doi: 10.1016/j.apenergy.2023.121607.

[132] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, "Physical energy and data-driven models in building energy prediction: A review," *Energy Rep.,* vol. 8, pp. 1119–1141, Nov. 2022, doi: 10.1016/j.egyr.2022.01.162.

[133] V. G. González and C. F. Bandera, "A building energy models calibration methodology based on inverse modelling approach," *Build. Simul.*, vol. 15, no. 11, pp. 1883–1898, Nov. 2022, doi: 10.1007/s12273-022-0900-5.

[134] R. Sánchez-Reolid, F. L. de la Rosa, D. Sánchez-Reolid, M. T. López, and A. Fernández-Caballero, "Feature and time series extraction in artificial neural networks for arousal detection from electrodermal activity," in *Proc. Int. Work-Conf. Artif. Neural Netw.*, Cham, Switzerland: Springer, 16 Jun. 2021, pp. 265–276.

[135] N. Giamarelos et al., "A machine learning model ensemble for mixed power load forecasting across multiple time horizons," *Sensors*, vol. 23, no. 12, Jun. 2023, doi: 10.3390/s23125436.

[136] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka, "A multi-horizon quantile recurrent forecaster," *arXiv preprint*, arXiv:1711.11053, Nov. 29, 2017.

[137] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[138] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," *in Proc. 41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (SIGIR),* vol. 03, pp. 95–104, Jun. 2018, doi: 10.1145/3209978.3210006.

[139]] F. Belletti, A. Beutel, S. Jain, and E. Chi, "Factorized recurrent neural architectures for longer range dependence," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 31 Mar. 2018, pp. 1522–1530, PMLR.

[140] M. Gauch, F. Kratzert, D. Klotz, G. Nearing, J. Lin, and S. Hochreiter, "Rainfall-runoff prediction at multiple timescales with a single long short-term memory network," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 4, pp. 2045–2062, Apr. 2021, doi: 10.5194/hess-25-2045-2021.

[141] C. Ma, Y. Hou, X. Li, Y. Sun, H. Yu, Z. Fang, and J. Qu, "Breaking the context bottleneck on long time series forecasting," *arXiv preprint*, arXiv:2412.16572, Dec. 21, 2024.

[142] C. Kanthila, A. Boodi, A. Marszal-Pomianowska, K. Beddiar, Y. Amirat, and M. Benbouzid, "Enhanced multi-horizon occupancy prediction in smart buildings using cascaded Bi-LSTM models with integrated features," *Energy Build.*, vol. 318, pp. 114442, Sep. 2024, doi: 10.1016/j.enbuild.2024.114442.

[143] X. Chen, M. M. Singh, and P. Geyer, "Utilizing domain knowledge: Robust machine learning for building energy performance prediction with small, inconsistent datasets," *Knowl.-Based Syst.,* vol. 294, Jun. 2024, doi: 10.1016/j.knosys.2024.111774.

[144] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: A systematic review," *Renew. Sustain. Energy Rev*., vol. 133, Nov. 2020, doi: 10.1016/j.rser.2020.110287.

[145] J. Pan, "Feature-based transfer learning with real-world applications," Ph.D. dissertation, Dept. Electron. Comput. Eng., Hong Kong Univ. Sci. Technol., Hong Kong, 2010.

[146] R. Spencer, S. Ranathunga, M. Boulic, A. van Heerden, and T. Susnjak, "Transfer learning on transformers for building energy consumption forecasting – A comparative study," *Energy and Buildings*, vol. 336, Jun. 2025, p.115632.

[147] H. Li, G. Pinto, M. S. Piscitelli, A. Capozzoli, and T. Hong, "Building thermal dynamics modeling with deep transfer learning using a large residential smart

thermostat dataset," *Eng. Appl. Artif. Intell.,* vol. 130, Apr. 2024, doi: 10.1016/j.engappai.2023.107701.

[148] D. Syed, H. Abu-Rub, A. Ghrayeb, and S. S. Refaat, "Household-level energy forecasting in smart buildings using a novel hybrid deep learning model," *IEEE Access*, vol. 9, pp. 33498–33511, Feb. 2021, doi: 10.1109/ACCESS.2021.3061370.

[149] R. Al-Hajj, A. Assi, B. Neji, R. Ghandour, and Z. Al Barakeh, "Transfer learning for renewable energy systems: A survey," *Sustainability*, vol. 15, no. 11, Jun. 2023, doi: 10.3390/su15119131.

[150] R. Bortolini, R. Rodrigues, H. Alavi, L. F. D. Vecchia, and N. Forcada, "Digital twins' applications for building energy efficiency: A review," *Energies*, vol. 15, no. 19, Oct. 2022, doi: 10.3390/en15197002.

[151] A. Hooshmand and R. Sharma, "Energy predictive models with limited data using transfer learning," *in Proceedings of the 10th ACM International Conference on Future Energy Systems (e-Energy),* Jun. 2019, pp. 12–16, doi: 10.1145/3307772.3328284.

[152] C. Fan et al., "Statistical investigations of transfer learning-based methodology for short-term building energy predictions," *Appl. Energy*, vol. 262, no. 1, Mar. 2020, doi: 10.1016/j.apenergy.2020.114499.

[153] A. Li, F. Xiao, C. Fan, and M. Hu, "Development of an ANN-based building energy model for information-poor buildings using transfer learning," *Build. Simul.*, vol. 14, no. 1, pp. 89–101, Feb. 2021, doi: 10.1007/s12273-020-0711-5.

[154] T. C. Turin, M. Raihan, and N. Chowdhury, "Paradigms of approaches to research," *Bangabandhu Sheikh Mujib Medical University Journal*, vol. 17, no. 2, p. e73973, Jun. 2024.

[155] L. Closson, C. Cérin, D. Donsez, and J.-L. Baudouin, “Design of a meaningful framework for time series forecasting in smart buildings,” *Information*, vol. 15, no. 2, p. 94, Feb. 2024.

[156] Y. Shi *et al.*, “Deep learning meets process-based models: A hybrid approach to agricultural challenges,” *arXiv preprint* arXiv:2504.16141, Apr. 22, 2025.

[157] R. Chikte, A. Mathew, B. Ramachandran, K. Almaeeni, N. Aljasmi, and J. Vandavasi, “Short-term time series energy forecasting for smart buildings using machine learning models,” in *Proc. ASME Int. Mech. Eng. Congr. Expo.*, vol. 88643, Nov. 2024, pp. V006T08A026.

[158] C. L. Hor, S. J. Watson, and S. Majithia, “Analyzing the impact of weather variables on monthly electricity demand,” *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2078–2085, Nov. 2005.

[159] E. M. Pickering, M. A. Hossain, J. P. Mousseau, R. A. Swanson, R. H. French, and A. R. Abramson, “A cross-sectional study of the temporal evolution of electricity consumption of six commercial buildings,” *PLOS ONE*, vol. 12, no. 10, p. e0187129, Oct. 2017.

[160] R. Wang, S. Lu, and W. Feng, “A novel improved model for building energy consumption prediction based on model integration,” *Applied Energy*, vol. 262, p. 114561, Mar. 2020.

[161] A. Amin and M. Mourshed, “Weather and climate data for energy applications,” *Renewable and Sustainable Energy Reviews*, vol. 192, p. 114247, Mar. 2024.

[162] *A. Manna and S. K. Ghosh, “Bayesian models for joint selection of features and auto-regressive lags: theory and applications in environmental and financial forecasting,” arXiv preprint arXiv:2508.10055, Aug. 12, 2025*.

[163] J. Wu *et al.*, “A comparative analysis of machine learning-based energy baseline models across multiple building types,” *Energies*, vol. 17, no. 6, p. 1285, Mar. 2024.

[164] S. Ardabili, L. Abdolalizadeh, C. Mako, B. Torok, and A. Mosavi, “Systematic review of deep learning and machine learning for building energy,” *Frontiers in Energy Research*, vol. 10, p. 786027, Mar. 2022.

[165] J. Huang, M. Algahtani, and S. Kaewunruen, “Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks,” *Applied Sciences*, vol. 12, no. 19, p. 9788, Sep. 2022.

[166] M. Cordeiro-Costas, D. Villanueva, P. Eguía-Oller, M. Martínez-Comesaña, and S. Ramos, “Load forecasting with machine learning and deep learning methods,” *Applied Sciences*, vol. 13, no. 13, p. 7933, Jul. 2023.

[167] S. Khan *et al*., “Comparative analysis of deep neural network architectures for renewable energy forecasting: enhancing accuracy with meteorological and time-based features,” *Discover Sustainability*, vol. 5, no. 1, p. 533, Dec. 2024.

[168] A. V. Andrei, G. Velev, F. M. Toma, D. T. Pele, and S. Lessmann, “Energy Price Modelling: A Comparative Evaluation of Four Generations of Forecasting Methods,” *arXiv preprint* arXiv:2411.03372, Nov. 2024.

[169] N. Maragkos and I. Refanidis, “A comparative evaluation of time-series forecasting models for energy datasets,” *Computers*, vol. 14, no. 7, p. 246, Jun. 2025.

[170] G. J. de Bruin, C. J. Veenman, H. J. van den Herik, and F. W. Takes, “Experimental evaluation of train and test split strategies in link prediction,” in *Proc. Int. Conf. Complex Networks Their Appl.*, Cham, Switzerland, Dec. 2020, pp. 79–91.

[171] F. Gonsalves, B. Pasdeloup, R. Billot, P. Meyer, A. Jacques, and M. Lorang, “New insights into the propulsion power prediction of cruise ships,” in *Proc. 2021 IEEE 33rd Int. Conf. Tools with Artif. Intell. (ICTAI)*, Washington, DC, USA, Nov. 2021, pp. 846–850.

[172] D. Tomašević, J. Ponoćko, and T. Konjić, “LSTM-based active and reactive load forecasting and its replicability in large geographical areas,” in *Proc. 2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Oct. 2024, pp. 1–5.

[173] J. Shi, Y. Chen, X. Cheng, M. Yang, and M. Wang, “Four-stage space-time hybrid model for distributed photovoltaic power forecasting,” *IEEE Transactions on Industry Applications*, vol. 59, no. 1, pp. 1129–1138, Jan.–Feb. 2023, doi: 10.1109/TIA.2022.3205570.

[174] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, “On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach,” *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[175] K. Ullah *et al.*, “Short-term load forecasting: A comprehensive review and simulation study with CNN–LSTM hybrids approach,” *IEEE Access*, Aug. 8, 2024.

[176] D. Barochiner, R. Lado, L. Carletti, and F. Pintar, “A machine learning approach to address 1-week-ahead peak demand forecasting using the XGBoost algorithm,” in *Proc. 2022 IEEE Biennial Congress of Argentina (ARGENCON)*, Sep. 2022, pp. 1–5.

[177] A. E. Léger and S. Rizzi, “Month-to-month all-cause mortality forecasting: A method allowing for changes in seasonal patterns,” *American Journal of Epidemiology*, vol. 193, no. 6, pp. 898–907, Jun. 2024.

[178] F. Delogu *et al.*, “Forecasting of a complex microbial community using meta-omics,” *bioRxiv*, pp. 1–15, Oct. 2022.

Appendices

Appendix A: Model Architectures and Hyperparameter Settings

A.1 SVR → XGBoost → LSTM Hybrid Model

Component	Parameter	Value
SVR	Kernel	RBF
	C	1.0
	ε	0.1
XGBoost	Learning Rate	0.05
	Max Depth	6
	Estimators	100
LSTM	Layers	2 (50, 25 units)
	Dropout	0.2
	Optimiser	Adam
	Batch Size	32
	Learning Rate	0.001

A.2 CNN → GRU → LSTM Transfer Learning Model

Layer	Type	Parameters
1	Conv1D	Filters: 64, Kernel Size: 3, Activation: ReLU
2	GRU	Units: 50, Return Sequences: True

Layer	Type	Parameters
3	LSTM	Units: 25, Dropout: 0.2
4	Dense	Units: 1 (Output Layer)

Appendix B: Dataset Descriptions and Sources

Dataset	Description	Source	Usage
GENOME	2 years of hourly energy data from 1,536 buildings across 18 types	Open Data Repository	Pretraining and benchmarking
Ebbw Vale Offices	Hourly data from office buildings in Wales	Proprietary	Multi-horizon forecasting
Ebbw Vale Schools	Hourly data from educational buildings	Proprietary	Transfer learning (target domain)
Queen's Building	Live sensor data for temperature, load, and control	Cardiff University	Edge deployment validation

Appendix C: Forecast Visualizations Across Time Horizons

C.1 24-Hour Forecasts – SVR vs. Hybrid Model

This section compares actual electricity consumption to model predictions generated over a 24-hour period. The hybrid model, combining SVR, XGBoost, and LSTM, shows closer alignment with true demand than the SVR model alone, particularly during peak periods.

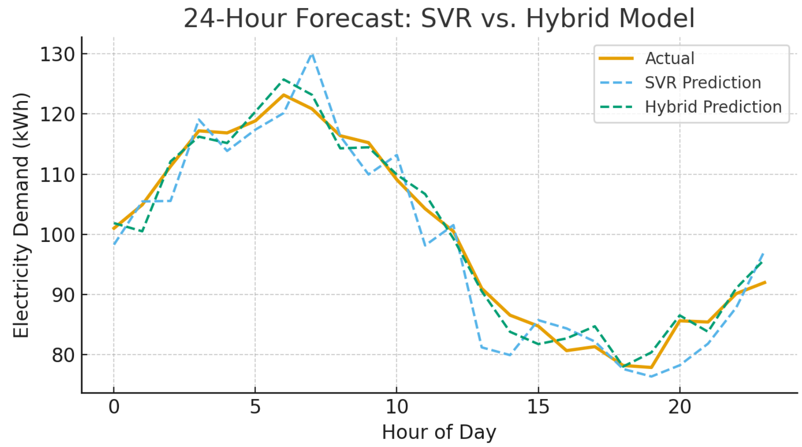

Figure C.1: 24-hour electricity demand forecast using SVR and Hybrid models.

This boxplot illustrates the distribution of RMSE values across different forecasting models over a 1-week horizon. The hybrid model outperforms SVR, XGBoost, and LSTM in both mean accuracy and error stability, as reflected by the tighter error distribution. One outlier is observed in the Hybrid model's RMSE values, indicating a single prediction instance with elevated error. Despite this, the overall distribution remains narrower and more consistent than the others.

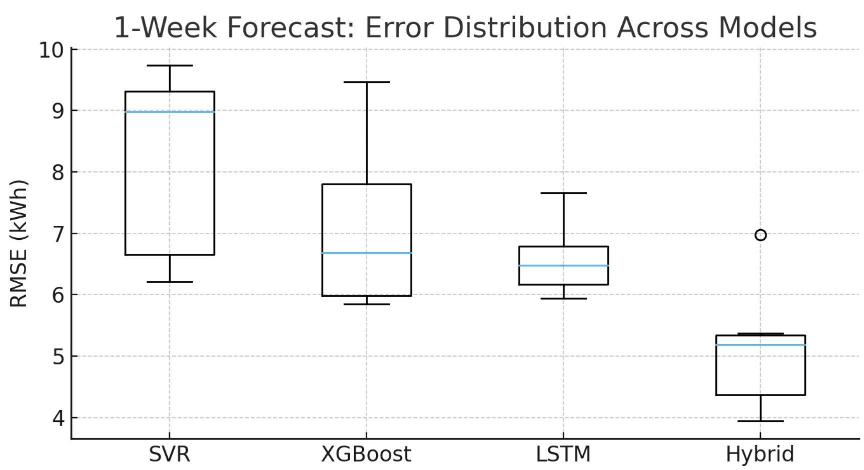

Figure C.2: RMSE distribution across SVR, XGBoost, LSTM, and Hybrid models for 1-week forecasts.

C.3 1-Month Forecast – Seasonal Patterns

This figure displays predicted electricity demand over a 1-month horizon, highlighting seasonal variation. The hybrid model effectively captures cyclical trends associated with temperature changes and usage behaviour across weeks.

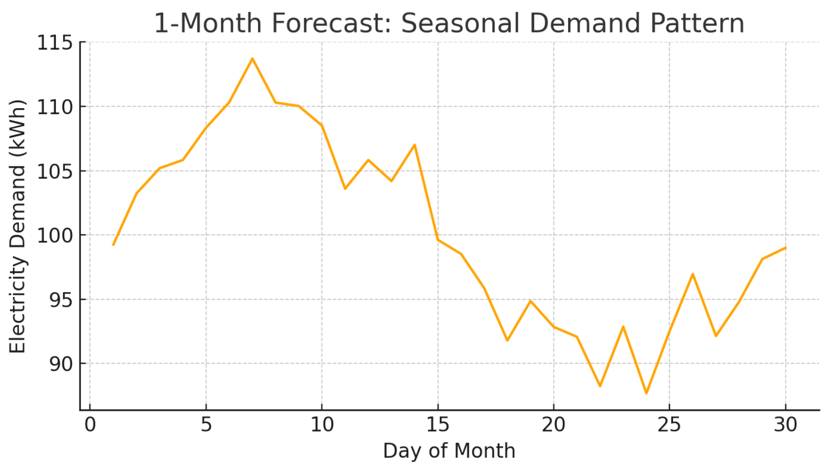

Figure C.3: 1-month forecast showing seasonal electricity consumption patterns.

This bar chart compares RMSE values of different models as forecast horizon increases from 24 hours to 1 year. While all models degrade over longer periods, the hybrid model maintains significantly better accuracy, especially in the long-term range.

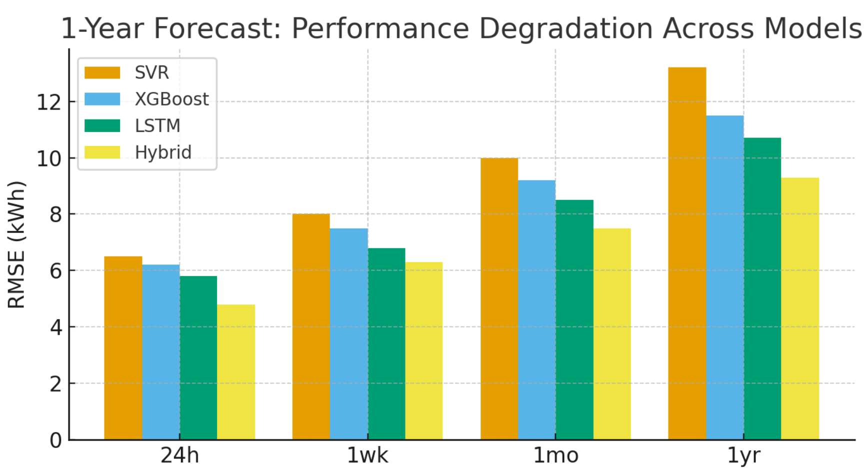

Figure C.4: RMSE performance degradation across forecast horizons for SVR, XGBoost, LSTM, and Hybrid models.

Appendix D: Research Ethics Approval

This research was conducted in accordance with Cardiff University's ethical guidelines and research integrity policies. Although formal ethics approval documentation is not available for inclusion in this appendix, the following principles were adhered to throughout the study:

- All datasets used in the research were either publicly available or obtained with proper institutional access.
- No personal identifiable information (PII) was collected, processed, or stored.
- All machine learning experiments were performed on anonymized, non-sensitive datasets.
- No human participants or live interventions were involved in the study.
- The research complied with the university's Code of Research Conduct and was supervised and reviewed as part of the doctoral program.

Should further verification be required, confirmation of compliance may be requested through the School of Engineering, Cardiff University.