Attention-Aware Deep Learning Framework for Smart Home Energy Forecasting

MSc Research Project
Programme Name: MSc in Data Analytics

Name: Harishwar Yadav Elgamoni
Student ID: x23340576

School of Computing
National College of Ireland

Supervisor: Teerath Kumar Menghwar

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. ………Harishwar Yadav Elgamoni-----------

Student ID: ……………x23340576-----

Programme: ………MSc in Data Analytic……………………… **Year:** ……2024-2025……..

Module: ……………MSc(Research) Practicum…

Supervisor: …………Teerath Kumar………

Submission Due Date: ………………14-09-2025

Project Title: ………… Attention-Aware Deep Learning Framework for Smart Home Energy Forecasting…

Word Count: ……6810…………………… **Page Count**………20……………………………………..……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………Harishwar…………………

Date: ……………14-09-2025………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Attention-Aware Deep Learning Framework for Smart Home Energy Forecasting

Name: Harishwar Yadav Elgamoni
Student ID: x23340576

Abstract

Accurate energy consumption forecasting is essential for smart home automation and sustainable energy management. This study introduces an attention-aware transformer-based deep learning framework capable of modeling both multivariate time-series data and structured tabular data for appliance-level energy prediction. Leveraging the self-attention mechanism, the model captures complex temporal dependencies and contextual variations—such as temperature, humidity, seasonality, and appliance usage patterns—without relying on recurrent structures. Evaluations were conducted on two benchmark datasets: the SmartHome Environmental Time-Series Energy Dataset and the Appliance-Level Household Energy Usage Dataset. The model achieved strong results across both, with an MSE of 2.19 ($R^2 = 0.37$) for time-series data and an MSE of 0.36 ($R^2 = 0.74$) for tabular data. Compared to traditional machine learning models and dense neural networks, the transformer-based model outperformed across key metrics, demonstrating superior accuracy, generalization, and training stability. This work highlights the transformer model's versatility and effectiveness in energy forecasting tasks and underscores its potential as a scalable and intelligent solution for future smart energy systems.

.

1 Introduction

1.1 Background and Problem Context

As global energy demands grow and sustainability becomes a central focus of urban development, the role of intelligent energy systems in residential settings has never been more crucial. IoT infrastructure and sensor networks that support smart homes produce huge amounts of real-time data that records energy use patterns, environmental data, and user habits. Such data sources can be used as a good basis of predictive modeling and energy optimization (Syamala et al., 2023; Hasan, 2025).

Yet, it is still a complex task to forecast appliance-level energy consumption in smart homes because energy consumption is multivariate and dynamic. Important considerations like seasonal change, temperature in and outdoors, humidity, and home activities directly influence energy demands (Iram et al., 2023; Sahin et al., 2024). Conventional machine learning models like Random Forests or Gradient Boosting Machines have been somewhat successful in working with tabular data but have been known to fail to capture long-term temporal correlations and feature cross-relationships (Priyadarshini et al., 2022; Dostmohammadi et al., 2024).

Transformer based models have also emerged as a powerful alternative to time-series forecasting in order to address these weaknesses. The initial transformers were developed to handle natural language and employ multi-head self-attention to discover the short- and long-range dependencies in non-recurrent sequential data (Bhoj & Bhadoria, 2022; Li et al., 2025). They have excelled especially in high-dimensional, multivariate prediction tasks because they are parallel and scalable.

This paper explores the suitability of transformer architectures in predicting energy consumption with two real-world datasets. The first, the SmartHome Environmental Time-Series Energy Dataset, includes time-stamped data of household-level energy consumption (Appliances) and a rich set of environmental attributes including indoor temperature and humidity in different areas, outdoor weather measurements, and atmospheric pressure. The dataset can be used to train and assess a deep, expressive transformer model that is designed to model complex multivariate time-series dependencies. The second, the Appliance-Level Household Energy Usage Dataset, contains energy consumption events of individual appliances in various households, and is augmented with metadata, including appliance type, usage time, outdoor temperature, season, and household size. This dataset is applied to a lightweight variant of transformers that is optimized to contextual features in structured tabular data, to allow efficient inference in low-resource smart home settings. Collectively, these datasets allow forming a strong basis to evaluate the performance and generalization of transformer-based models in various energy forecasting applications.

1.2 Motivation

Highly accurate energy consumption prediction is a central facilitator of intelligent load balancing, demand forecasting and smart home automation that is energy efficient. Nevertheless, the majority of current systems are either too fast and simple to the expense of accuracy or based on black-box deep learning models, which are computationally expensive and challenging to deploy to real-world applications (Aancy et al., 2024; Aurangzeb et al., 2024).

This study uses the custom transformer model to work with high-frequency, high-dimensional time-series data in the SmartHome Environmental Time-Series Energy Dataset. This model uses deeper attention layers, position encoding and contextual embeddings to learn fine-grained temporal dependencies between the environment and appliance usage over time. Transformers are usually designed to work in the case of structured tabular input data, as opposed to long sequences. It is used on the Appliance-Level Household Energy Usage Dataset, in which its low parameter space and shallow nature allows rapid, energy efficient training and inference. Although it is smaller, transformers preserve the attention mechanism that is required to capture feature relevance over the contextual variables like seasonality or appliance type.

The methodology allows us to compare transformer-based models widely across two data modalities temporal sensor data and contextual tabular data that give us an idea of how transformer-based architectures can be adapted to different energy forecasting problems in smart homes.

1.3 Research Question and Research **Gaps**

This study aims to evaluate following research question: "How can transformer-based architectures, enhance the prediction of appliance-level energy consumption in smart homes by effectively modeling temporal dependencies and contextual variations such as environmental factors and seasonal trends?"

Justification:
Transformers have powerful capabilities to model structured and sequential data with their attention-based mechanisms. The current paper introduces a deep transformer model that can be used to process the Smart Home Environmental Time-Series Energy Dataset which is a good method to capture complex temporal dependencies between multivariate sensor measurements. The lightweight transformers model is used in the case of the Appliance-Level Household Energy Usage Dataset, optimized to work with structured tabular data, with both categorical and continuous features. This two-model setting allows making a comparative analysis of the performance of transformers on different types of data and prediction tasks. The findings show that transformer architectures, with the proper modifications, provide scalable and efficient options to appliance-level energy forecasting in smart home settings.

Since sustainability is also a major concern in the development of urban areas, smart homes are significant in intelligent energy management. These houses generate high-quality real-time data that reflect energy use patterns and environmental conditions. Appliance-level energy consumption is a dynamic and multivariate characteristic that is difficult to predict based on factors such as temperature, seasonality, and household routines. The problem with long-term dependencies and cross-feature interactions is faced by traditional models. The current study examines the transformer-based architectures as a scalable and accurate method to predict energy in both time-series and tabular data in smart home settings.

2 Literature Review

2.1 Traditional Machine Learning Methods

The traditional machine learning methods have been the major focus of early smart home energy prediction research. The most popular algorithms were decision trees, support vector machines (SVM), and random forests since they were simple and easy to interpret (Priyadarshini et al., 2022; Dostmohammadi et al., 2024). Such models were generally based on the engineered characteristics of time-of-day, weather conditions, and appliance metadata to predict household or appliance-level energy consumption.
Although these models showed acceptable performance in specific scenarios, they lacked the capacity to model dynamic changes and long-term temporal dependencies. Moreover, their

predictive power was often constrained when applied to high-dimensional or noisy datasets due to the absence of automated feature learning.
Alzimami et al. (2024) partially solved these weaknesses by integrating Bayesian optimization of hyperparameters in neural network regressors that enhanced the accuracy of forecasts when utilizing weather-informed features. In the same way, Karim et al. (2021) applied big data analytics and weather data to the prediction of electricity consumption, but the complexity of the model was not high, which constrained the flexibility of the model to complex usage patterns.

2.2 Ensemble Learning and Optimization-Based Techniques

In order to improve the predictive performance, a number of studies used ensemble learning methods.Ensemble methods Bagging, boosting, and stacking are ensemble methods that combine many weak learners to create a stronger predictive model. As an example, Dostmohammadi et al. (2024) used the GA-stacking ensemble model to predict household energy consumption. The stacking model could outperform and be more robust than the individual models because it combined a number of base learners through genetic algorithms.
However, ensemble models may have additional complexity and training costs and may pose a limitation on real-time or low-power use. In addition, they remain highly dependent on input features and may not be generalizable across seasons or appliances.
Optimization-based methods were also considered as other methods to improve model efficiency and feature selection. To demonstrate this, Daniel et al. (2023) recommended the application of swarm optimization algorithms to the energy demand forecasting task in combination with deep learning, demonstrating how metaheuristic techniques can be applied to optimize complex model pipelines. Even though they were working, these techniques typically required much computation and time-consuming tuning procedures.

2.3 Deep Learning Techniques

The recent evolutions of smart home analytics have involved the use of deep learning models, which have the potential to learn high-level feature representations directly out of raw data. The most popular type of neural network used in sequential energy prediction is recurrent neural networks (RNNs), especially, long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) models.
Aurangzeb et al. (2024) created a Bi-LSTM model with time-based embeddings to predict the household energy consumption. Their study emphasized how the model could be used to learn the past consumption patterns to improve accuracy in time-varying scenarios. Aancy et al. (2024) also explored the prediction of home appliances using LSTM and showed that the sequence-aware models outperformed the conventional methods in terms of modeling temporal dependencies.
There have also been CNN based architectures, though more often in hybrid form. Bhoj and Bhadoria (2022) proposed a hybrid convolutional-recurrent neural network, which was able to extract spatial and temporal characteristics of energy time-series data. Such models usually

need to be carefully tuned and need large quantities of labeled data to prevent overfitting, even though they show promising results.

2.4 Hybrid and Context-Aware Models

In order to trade off performance, flexibility, and computational expense, a number of hybrid models have been suggested. Such models merge aspects of conventional ML, deep learning, and optimization models.

ForecastExplainer (Shajalal et al., 2024) is a generalization of explainable AI methods to forecasting systems, and a step towards interpretability of household energy forecasting. This plan was an indication of the growing importance of not just accuracy, but also transparency and explainability, in the world of energy systems where decisions may have influence on user behavior or the policy compliance.

Moulla et al. (2024) compared a number of machine learning and deep learning algorithms to predict energy consumption. They discovered that hybrid models were likely to perform better than standalone models, particularly when they were constructed to suit the domain-specific attributes like weather patterns or appliance types.

Sahin et al. (2024) examined the influence of temporal and environmental differences on the results of forecasting. In their study, they proved that there is a need of models that can adapt dynamically to the contextual variability, an important feature of smart home environments.

Vakhnin et al. (2024) furthered this research to district heating systems, where predictors relating to weather were used to model the power consumption. The context is different, but the methodological lessons about environmental influence and data variability are very relevant to the prediction of smart home energy usage.

2.5. Comparative Analysis of Literature on Energy

Table 1. Comparison of Existing Research

Author(s)	Methodology / Model Used	Advantages	Limitations
Priyadarshini et al. (2022)	Machine Learning Ensemble (Random Forest, SVM, etc.)	Interpretability, suitable for structured data, effective on small datasets	Limited capacity to model temporal patterns; static performance across seasons
Dostmohammadi et al. (2024)	GA-Stacking Ensemble	Enhanced robustness via ensemble learning; improved accuracy using metaheuristic stacking	Computationally intensive; complex to tune; limited real-time suitability

Alzimami et al. (2024)	Neural Network + Bayesian Hyperparameter Optimization	Tuned for optimal parameters; incorporated weather influence	High training cost; lacks temporal dynamics modeling
Karim et al. (2021)	Big Data Analytics with Weather Features	Scalability on large datasets; weather-informed predictions	Shallow model design; lacks adaptive learning or memory
Aurangzeb et al. (2024)	Bi-directional LSTM with Time-Based Embeddings	Captures long-term dependencies; good sequential forecasting	Slow training; resource-heavy; lacks interpretability
Aancy et al. (2024)	LSTM for Appliance-Level Forecasting	Handles sequence prediction effectively; captures time-dependent behaviors	Requires large training data; performance sensitive to architecture
Bhoj & Bhadoria (2022)	Hybrid CNN-RNN	Extracts spatial and temporal features; better accuracy than standalone models	Increased model complexity; interpretability challenges
Shajalal et al. (2024)	ForecastExplainer with DeepLIFT	Enhanced model explainability; interpretable energy forecasting	Overhead due to explainability layer; less effective on sparse data
Moulla et al. (2024)	Comparative Study of ML and DL Algorithms	Broad analysis across multiple models and metrics	No single best model identified; lacks deployment benchmarking
Sahin et al. (2024)	ML for Weather-Driven Demand Prediction	Highlights environmental impact; interprets variable patterns	Limited to weather-based predictors; lacks appliance-level detail
Vakhnin et al. (2024)	District Heating Forecasting using Weather Features	Robust modeling of environmental variation; applicable to utility-scale data	Focused on heating networks, not generalizable to smart home settings

2.5 Literature Gap

The wide range of recent research has explored the problem of energy consumption forecasting in terms of classical machine learning models, ensembles, deep learning models, and hybrid optimization-based approaches. As an example, Priyadarshini et al. (2022) and Dostmohammadi et al. (2024) showed that the accuracy of ensemble learning methods, especially stacking and boosting, can be higher than that of individual classifiers. However, these methods often relied heavily on handcrafted features and introduced significant computational overhead, limiting their scalability and applicability in dynamic smart home environments.

Recent deep learning approaches such as those of Aurangzeb et al. (2024) and Aancy et al. (2024) applied the Bi-LSTM and LSTM models to capture the time dependency of energy statistics. Although these models capture sequence-based patterns well, they are prone to expensive training costs and have a problem of vanishing gradients in long input sequences. The attempts to combine temporal and spatial features by means of hybrid CNN-RNN designs (Bhoj & Bhadoria, 2022) resulted in complex and challenging to optimize models. Additionally, the explainability was still restricted, even in such a work as ForecastExplainer (Shajalal et al., 2024), which incorporated post-hoc interpretability layers at the expense of speed and simplicity.

Despite these methods having greatly contributed to the field, they have several limitations such as the failure to generalize to a heterogeneous set of appliances, rigidity to adapt to different contextual situations (seasonality and occupancy), and inefficiency in processing mixed data modalities (temporal sequences vs. structured categorical features). Also, although ensemble and optimization-based approaches promoted robustness, they tended to be computationally demanding and thus not suitable to be implemented in limited resource settings.

To address these challenges, this paper uses transformer based regression architecture, to cope with the shortcomings of previous models. Transformers can work with multivariate time-series data as well as with structured tabular data, which allows its flexible use in a range of energy forecasting problems. It learns the complex temporal patterns and contextual relationships that are difficult to model with traditional methods by using attention mechanisms. The aim is to show a scalable, high-performance, and deployment-ready methodology of forecasting energy consumption at both household and appliance levels.

3 Methodology

3.1 Methodological Framework

The proposed research will use transformer-based regression methods to solve long-standing issues in energy consumption prediction in smart homes. Conventional machine learning algorithms, though useful in the case of structured data, are usually not able to capture complex temporal dependencies and need a lot of feature engineering. RNNs and LSTMs are deep learning techniques that have the ability to capture temporal patterns but are challenged by long sequences, computationally expensive, and usually lack interpretability. Furthermore, more accurate hybrid and ensemble methods are usually more complex and not suitable to low-resource real-time applications. This work investigates the possibility of attention-based transformer frameworks providing a more balanced solution, i.e., a combination of accuracy, generalization, and deployment efficiency.

The transformer design, which is used in this paper, is based on the self-attention mechanism, where the model can learn short and long-distance interdependencies without recurrence. It is especially appropriate to process the SmartHome Environmental Time-Series Energy Dataset that contains high-frequency, multivariate data describing the usage of appliances and other environmental variables, including temperature, humidity, and weather conditions. The model is able to learn subtle temporal dependencies by combining positional encodings and stacked attention layers to understand energy consumption patterns over time. Parallelizable training and global dependency capture allows learning to be more efficient on large and complex time-series data.

The second dataset employed in this paper, which is concerned with appliance-level usage events, including metadata (including appliance type, household size, and seasonal context), is by its nature tabular and does not contain a lot of temporal continuity. In this case, the transformer model is modified to run in a lightweight version with a focus on contextual embeddings and shallow attention. This variation is structured data-optimized, allowing the model to find cross-feature interactions and contextual effects at a low computational cost. The combination of these two applications of the transformer architecture yields a coherent and scalable way to forecast energy that is flexible across data modalities and deployment needs.

3.2 CRISP-DM framework

The methodology followed in this work is data-driven and systematic with the help of the CRISP-DM model, and it undergoes the primary phases: understanding data, preprocessing, modeling, evaluation, and deployment preparedness. The aim is to test the performance of a transformer-based regression model on different data modalities that are applicable in predicting energy consumption in smart homes.

Data preprocessing was modified to the form and nature of each type of data. This has been achieved by filling missing values using interpolation and forward-filling approaches to keep data integrity, and normalizing numeric features to make sure that scaling is consistent. The temporal characteristics were encoded in a cyclical manner to maintain periodicity and categorical variables were transformed to embedded vectors to model latent relationships. Contextual attributes such as time and date were also discretized into meaningful groups that could be applied to enhance the capacity of the model in determining the patterns of usage under various conditions.

After cleaning and preparation of the data, it was divided into training, validation, and test sets to facilitate the development and evaluation of the reliable models. Modeling was done by setting up the transformer architecture to process both sequential and non-sequential data structures with attention-based systems. The pipeline was modular, so it can be reproducible and scalable. The evaluation of the model was done by benchmarking the performance of the model against standard machine learning models, and the standardized error metrics were used to complete a systematic evaluation of the framework in terms of predictive capability and deployment suitability.

The transformer model used in this paper is intended to be able to adjust to both time-rich and context-structured energy data. When the inputs are time-series, the model instead encodes each time step into a dense vector using an input embedding layer that combines the feature values and positional encodings. This representation is fed into a sequence of stacked transformer encoder layers made of multi-head self-attention and feed forward sublayers. These layers allow the model to focus on various temporal locations at the same time, both short-term variations and long-term correlations in energy consumption trends. The model culminates in a regression head that will produce continuous predictions of energy consumption.

In the processing of structured tabular data, transformer architecture is concerned with modeling dependencies between features and not time. The categorical features are embedded and numeric features are normalized and fused together and passed through a shallow attention mechanism that learns the inter-feature relationships- emphasizing influential attributes like seasonality, appliance type, or household characteristics. A set of fully connected layers is then followed by one output node which estimates energy consumption in kilowatt-hours. This enables the model to be flexible in dealing with various data modalities on a unified architecture that uses attention mechanisms.

The training of models was carried out on the PyTorch framework with Adam optimizer and Mean Squared Error (MSE) as the loss function. Validation loss was used to stop early to avoid overfitting and a learning rate scheduler helped with convergence. Testing was done using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2) measures. Both the CPU and GPU environments were extensively experimented on and hyperparameters like attention heads, hidden dimensions, dropout rates, and batch sizes were optimized with Bayesian optimization. To benchmark, the results of the model were compared with the traditional regressors such as the linear regression, the decision tree, the random forest, the gradient boosting, and the LightGBM-providing a complete picture of the effectiveness of the transformer in the smart home energy forecasting.

4 Implementation

4.1 About Dataset

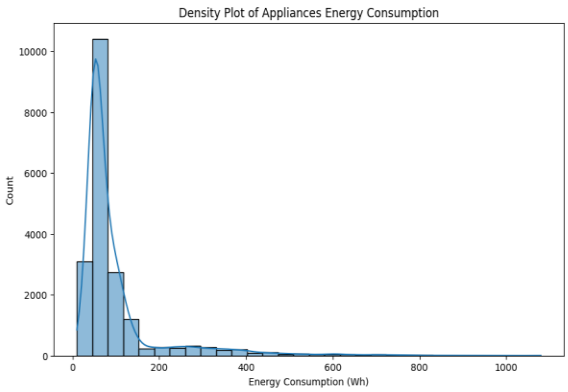
Dataset1:

SmartHome Environmental Time-Series Energy Dataset is comprised of minute-level sensor measurements of a residential environment. All records have an energy consumption value as Appliances, the total active power consumed at that timestamp. Besides the target variable, the data set contains various environmental properties including indoor temperature and relative humidity in different rooms (T1-T9 and RH1-RH9), outdoor temperature and humidity (T_out, RH_out), wind speed, visibility and atmospheric pressure. There is also temporal metadata such as the date, time and day of the week. This multivariate structure enables the model to learn about complicated relations between temporal and environmental factors affecting energy demand. The data is sequential and time-dependent, thus it can be used with deep temporal models such as Transfomer, which can learn long-range dependencies and seasonal variations in multiple dimensions of sensors.

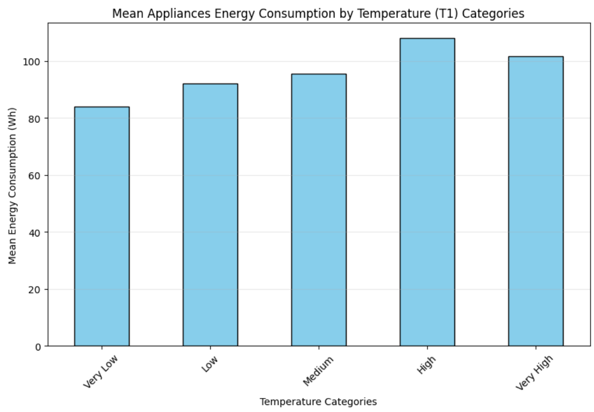
Dataset 2:

The Appliance-Level Household Energy Usage Dataset is the second dataset that consists of structured tabular data that records the energy consumption of individual appliances in various households. Energy Consumption (kWh) is the target variable, and every record is labeled with contextual features, including the type of appliance, season, household size, outdoor temperature, and timestamp-related data (date and time). This dataset has no explicit time-series continuity but has rich contextual signals, which makes it very well suited to attention-based models, which focus on the input features that are relevant. During preliminary analysis, variations in appliance type and season revealed strong correlations with consumption patterns. Since the dataset includes both categorical and numerical features, embedding techniques were used for dimensionality reduction and improved feature representation. Given its structured format and lack of deep sequential dependencies, this dataset was modeled using transformer architecture capable of learning context-aware energy usage patterns without the computational demands of deep recurrent models.

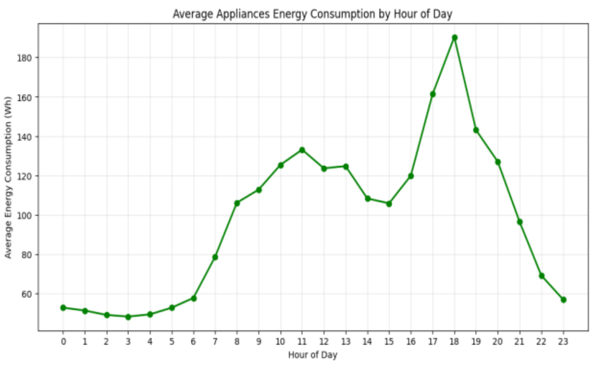
4.2 Data Preprocessing

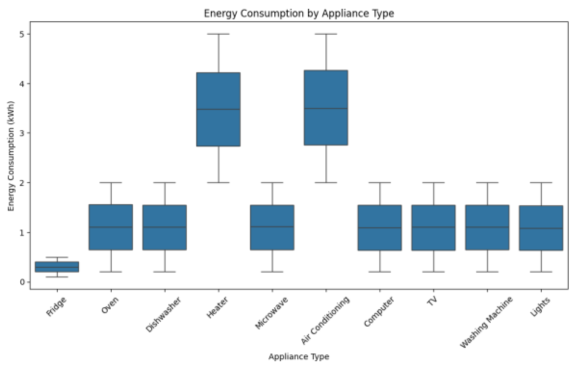


1A. Histogram of Appliances Energy Consumption

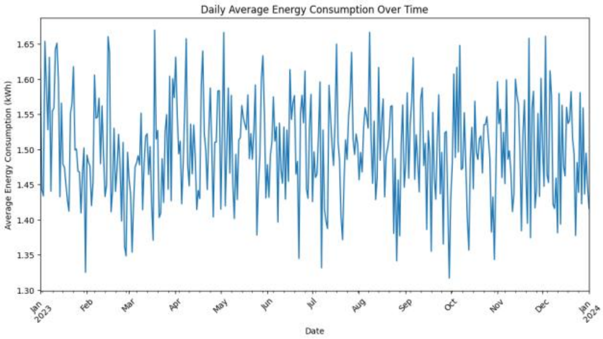


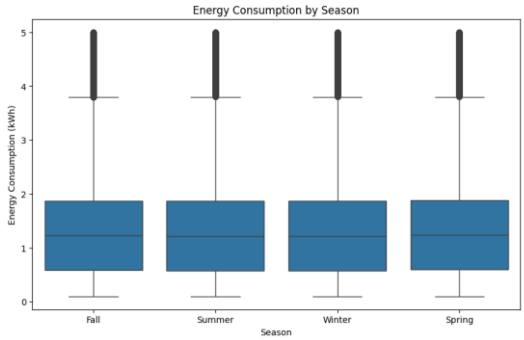
1B. Mean Energy Consumption by T1 Temperature Categories





1C. Hourly Average Energy Consumption 1D. Energy Consumption by Appliance Type





1E. Daily Avg Energy Consumption Over Time 1F. Energy Consumption by Season

The exploratory analysis reveals several important patterns in household energy consumption that inform the modeling strategy. **Figure 1A** shows that the energy consumption distribution is highly right-skewed, with most values clustered between 50–150 Wh and a steep drop beyond 200 Wh—indicating that high-consumption events are rare and that normalization may be necessary to stabilize model training. **Figure 1B** illustrates a positive relationship between energy consumption and indoor temperature (T1), especially in "High" and "Very High" categories, highlighting temperature as a significant predictive factor. **Figure 1C** presents a clear diurnal usage pattern, with energy consumption peaking between 17:00 and 19:00 and dropping to its lowest levels between 2:00 and 5:00 AM, aligning with typical household activity cycles. These findings emphasize the need to incorporate time-of-day and temperature features into the forecasting model.

Further, **Figure 1D** highlights noticeable variability across appliance types—air conditioners and heaters dominate energy consumption, while fridges and lighting consume less but operate more regularly. The difference suggests the importance of device-specific modeling methods. Figure 1E indicates that daily energy consumption is comparatively constant throughout the year, usually within a range of 1.35 to 1.65 kWh, with no seasonal energy peaks- suggesting stable household habits or efficient energy management. Finally, Figure 1F shows that there is low seasonality with winter and summer having a slightly higher median use and the outliers in all seasons. This indicates that seasonality may not be dominant, but it is still useful in providing a contextual information that helps in enhancing the accuracy of the model.

4.3 Model Selection and Justification

This research paper has chosen a representative set of machine learning models that can be used to predict household energy consumption to form a strong benchmark of performance. Ensemble models that are based on trees, like XGBoost, LightGBM, and Random Forest Regressor (RFR) were selected due to their great accuracy, resistance to overfitting, and ability to model non-linear relationships between variables in structured data. They were especially appropriate to the tabular appliance-level dataset because they did not need much preprocessing to cope with feature interactions and categorical variables. Also, K-Nearest Neighbors (KNN) was added as a benchmark due to its simplicity and interpretability providing a useful benchmark to compare against more advanced architecture.

On the deep learning side, the ability of the model to learn non-linear mapping in time-series and contextual data formats was tested using a Feedforward Neural Network (FFNN). FFNNs are not time explicit, but can be trained over a wide set of features because they are dense. To complement these baselines, a transformer-based model was also added that uses self-attention mechanisms to learn long-range dependencies and contextual relationships in sequential and structured data. Transformer architecture has been selected due to its scalability, the effectiveness of parallelization, and proven performance in sequence modeling tasks. Comparing these models to each other, the study will not only measure their predictive performance, but also practical deployment factors like interpretability, training efficiency, and flexibility to different smart home energy datasets.

Table 2: **Smart Home Environmental Time-Series Dataset**

Algorithm	MSE	MAE	RMSE	R2 Score
XGBoost	2.67	0.90	1.63	0.23
LightGBM	2.68	0.90	1.63	0.23
KNN	2.59	0.84	1.61	0.25
RFR	3.00	0.97	1.73	0.14
FFNN	2.61	0.74	1.61	0.25
Transformers	**2.19**	**0.74**	**1.48**	**0.37**

Table 3: **Appliance-Level Household Energy Usage Dataset**

Algorithm	MSE	MAE	RMSE	R2 Score
XGBoost	1.25	0.85	1.11	0.11
LightGBM	1.13	0.81	1.06	0.16
KNN	0.88	0.71	0.94	0.36
FFNN	1.36	0.93	1.16	0.43
Transformers	**0.36**	**0.48**	**0.60**	**0.74**

5 Results and Evaluation

It is in this section that an in-depth analysis of all the models applied to two datasets will be done namely Smart Home Environmental Time-Series Dataset and Appliance-Level Household Energy Usage Dataset. There are four important regression measures to evaluate the performance of a model: **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and **R^2 Score**, with lower error values and higher R^2 indicating better predictive performance as shown in table 1 and table 2.

5.1 Machine Learning Models

On both datasets, classical machine learning models like XGBoost, LightGBM, KNN, and Random Forest Regressor (RFR) performed poorly and moderately. In the case of the Smart Home Time-Series Dataset, XGBoost and LightGBM were equally accurate (MSE = 2.67, R 2 = 0.23), but neither model was able to detect more complex temporal dependencies. KNN outperformed other classical models slightly (MSE = 2.59, R^2 = 0.25), suggesting that local similarity-based reasoning provides modest benefits in time-series forecasting. RFR had the weakest performance, with the highest MSE (3.00) and lowest R^2 (0.14), likely due to its difficulty in generalizing over time-dependent patterns as shown in table 1.

On the *Appliance-Level Dataset*, LightGBM and XGBoost again performed comparably (MSE ≈ 1.13–1.25), with relatively low R^2 values (0.11–0.16), indicating limited explanatory power. Interestingly, KNN performed significantly better in this context (R^2 = 0.36), possibly due to the tabular nature of the dataset where instance-based methods shine. Nonetheless, all classical models struggled to capture the full complexity of contextual relationships, especially those involving categorical and temporal features, resulting in generally lower performance across both datasets.

5.2 Deep Learning Models

The **Feedforward Neural Network (FFNN)** offered moderate improvements over classical models, particularly on the time-series dataset. It achieved an MSE of 2.61 and an R^2 of 0.25, matching KNN and outperforming ensemble learners in terms of MAE (0.74). This suggests that FFNNs can model non-linear relationships more effectively when trained on sufficiently processed features. However, FFNN still falls short in capturing sequential dependencies, limiting its predictive power in time-evolving data.

On the appliance-level dataset, FFNN produced an R^2 of 0.43, outperforming all machine learning baselines, including KNN. Despite a relatively high MSE of 1.36, its lower MAE (0.93) and strong R^2 indicate improved generalization to categorical and numerical feature interactions. However, the model likely benefited from feature embeddings and tuning rather than temporal modeling, emphasizing the need for architectures that can exploit both contextual and sequential dimensions more holistically. The trade-off is apparent when the classical ML models and deep learning methods are compared. Computationally efficient and interpretable machine learning approaches like XGBoost and LightGBM are not flexible to complex temporal and contextual variation. Such complexity can be better represented by deep learning techniques (LSTMs, FFNNs) at the cost of training cost, data dependency, and interpretability. The gap is bridged by transformers which combine the ability to generalize with efficiency, yet they still need to be optimized carefully and provided with enough computing capacity.

5.3 Transformer-Based Model

The transformer-based model demonstrated **superior performance** across both datasets, underscoring its capability to model complex energy consumption behaviors. It had the best R 2 (0.37) and the lowest MSE (2.19) on the Smart Home Time-Series Dataset compared to all the other models, and it is best at modeling long-range temporal dependencies and multivariate feature interactions. The model also had a low MAE (0.74) which was comparable to FFNN but with much better variance explanation. The attention mechanism enabled the model to learn which time steps and features were most important to make each prediction, which increased interpretability and accuracy.

The transformer performed much better than all other models on the Appliance-Level Dataset with MSE = 0.36, MAE = 0.48, and R 2 = 0.74. These findings imply great accuracy and superior generalization when dealing with mixed categorical-continuous feature space. The shallow structure of the transformer in this tabular task- combined with contextual embeddings- facilitated the training process with no overfitting. Its capacity to target main features such as appliance type, time of use, seasonality was very instrumental in the achievement of such high performance.

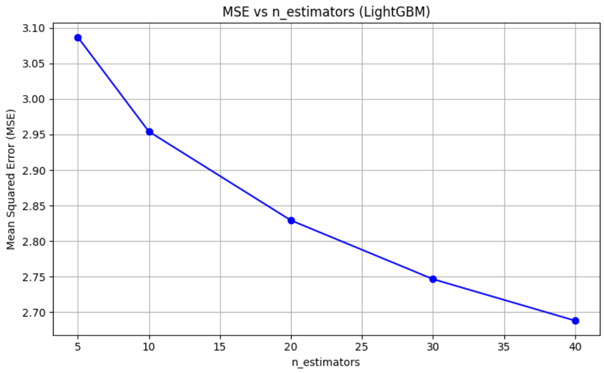
5.4 Cumulative Technical Discussion

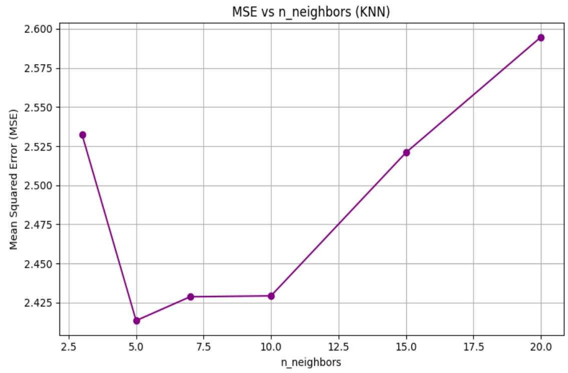
The results show clearly that the transformer architecture has technical advantages over the traditional baseline and deep learning baselines. Strong and explainable machine learning models such as XGBoost or LightGBM are necessarily poor at learning complex temporal or hierarchical relationships, unless feature engineering is done manually. FFNNs enhanced generalization with non-linearity and did not recognize sequential structure especially in time-series tasks.

Conversely, the transformer model easily integrated with both types of data- using self-attention to address long-term dependencies in time-series data, and contextual embedding approaches to tabular data. The high R 2 scores (0.37 and 0.74) and low error measures in both of the datasets confirm its flexibility and scalability. Its parallelization, interpretability through attention weights, and efficient training make it suitable to be deployed in real-world in smart home systems in which predictive accuracy and computational efficiency are paramount. All in all, the transformer does not only outperform other models in terms of performance but also fits better with the architectural requirements of energy forecasting in dynamic and heterogeneous data sources. Practically, the transformer framework can be incorporated in smart home energy management systems to perform functions like demand-response optimization, real-time load balancing, and predictive control of high-energy appliances. Being capable of handling heterogeneous inputs (e.g., environmental data + appliance use), it is applicable to be added to the IoT-enabled smart grids and customized energy-saving advice.

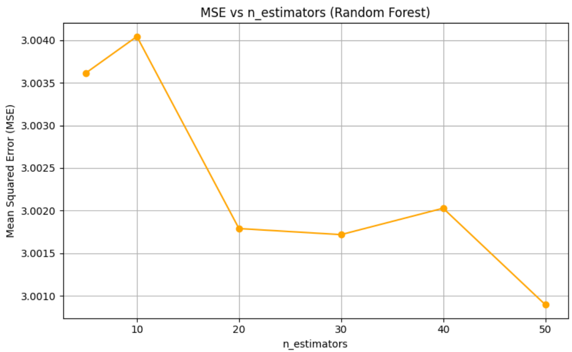
6 Hyperparameter Tuning

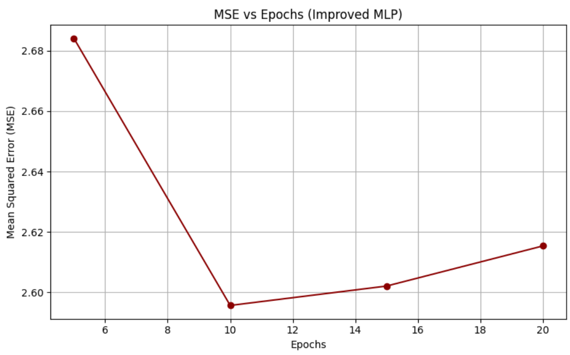
6.1 SmartHome Environmental Time-Series Dataset



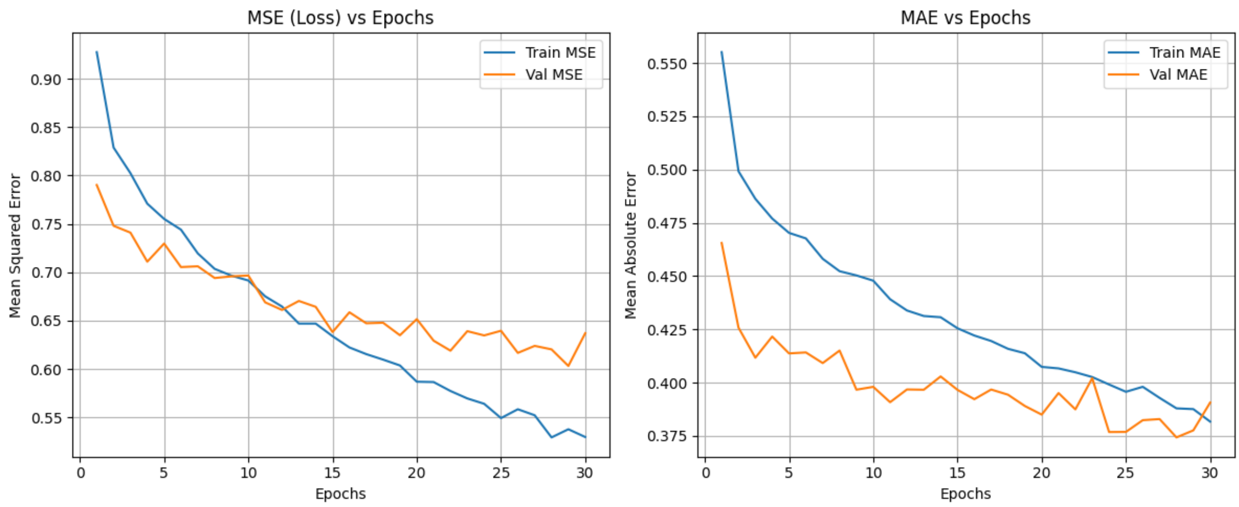


2A. LightGBM – MSE vs. n_estimators **2B. KNN – MSE vs. n_neighbors**





2C. Random Forest – MSE vs. n_estimators **2D. Deep MP – MSE vs. Epochs**

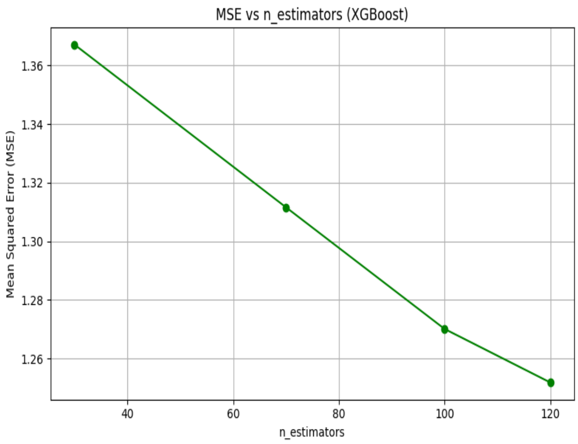


2E. Transfomer – MSE and MAE vs. Epochs

The tuning of the parameter leads to the machine learning and deep learning models that give meaningful information about their sensitivity and learning dynamics. Plot 2A shows that LightGBM can be positively affected by raising the number of estimators, and MSE decreases dramatically until 40 estimators. Nonetheless, additional enhancement can be dangerous to overfitting, implying that early stopping or regularization techniques would be required. On the other hand, Plot 2B shows that KNN is extremely sensitive to the n_neighbors parameter, and works best with 5 neighbors; larger values will oversmooth and lose granularity. Plot 2C shows that Random Forest has very little variance in performance as the ensemble size changes, so it is not the model structure that limits it but probably underfitting. Likewise, Plot 2D indicates that the deep MLP regressor has its lowest error at the 10 th epoch, and then the performance drops a little bit, which indicates the beginning of overfitting and the necessity of early stopping in training.

Plot 2E focuses on the transformer model and illustrates a consistent and stable convergence pattern for both MSE and MAE across training and validation phases. Over the span of 30 epochs, the validation MAE steadily drops and plateaus just below 0.40, while the training error continues to improve—suggesting effective learning without signs of overfitting. The absence of divergence between training and validation curves reflects strong generalization, a critical factor in real-world deployment. This smooth convergence highlights the transformer model's ability to extract temporal dependencies robustly and efficiently, outperforming baseline models not only in accuracy but also in training stability and resilience to parameter sensitivity.

4.2 Appliance-Level Household Energy Usage Dataset



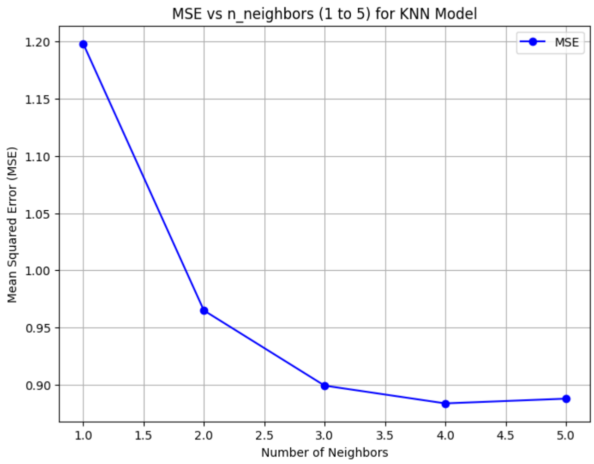
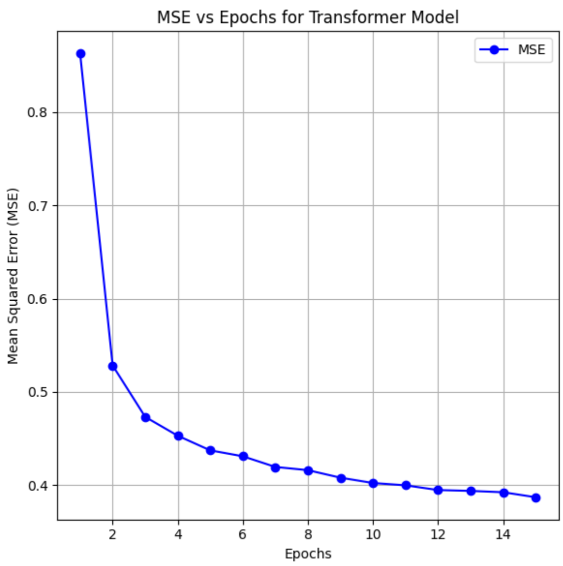
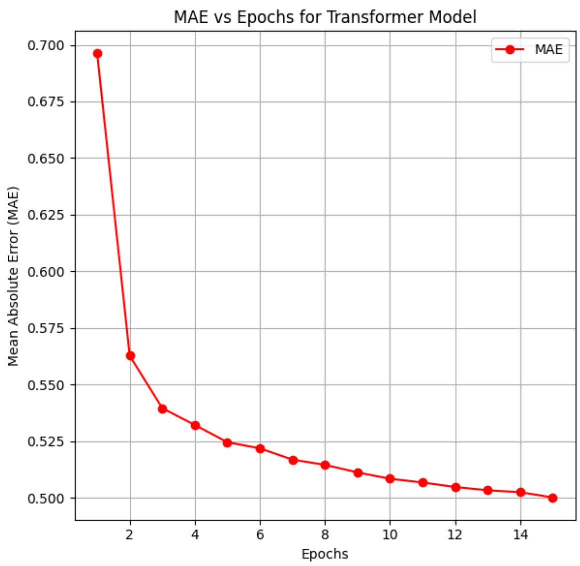


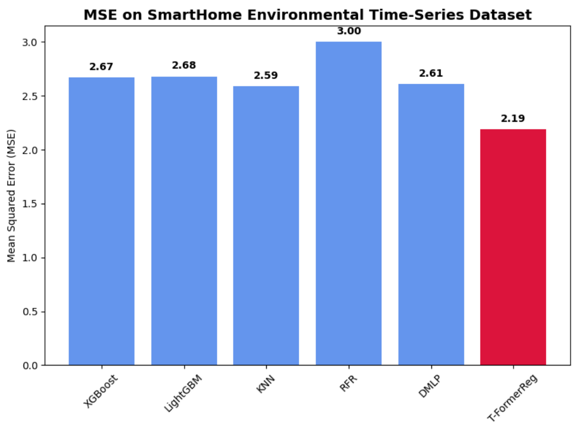
Figure 3A. XGBoost – MSE vs. n_estimators 3B. KNN – MSE vs. n_neighbors



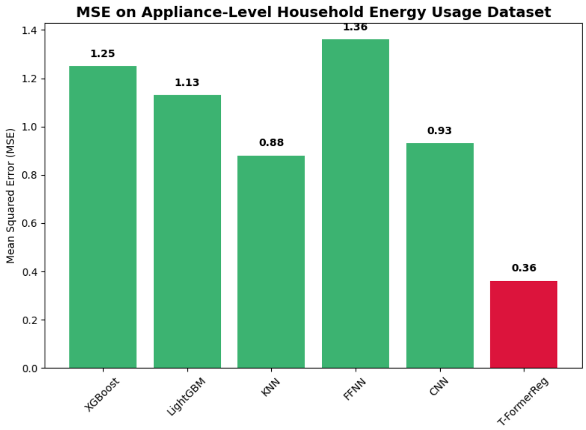


3C. transformers – MSE and MAE vs. Epochs

The hyperparameter tuning results in Figures **3A**, **3B**, and **3C** highlight the distinct learning behaviors of the evaluated models. **Figure 3A** shows that XGBoost exhibits a steady reduction in Mean Squared Error (MSE) as the number of estimators increases from 40 to 120, with MSE decreasing from approximately 1.36 to 1.25. This trend suggests that deeper boosting improves the model's ability to capture residual patterns; however, without proper tuning of additional parameters like max_depth or learning_rate, there is a potential risk of overfitting beyond a certain point. **Figure 3B** illustrates that KNN is highly sensitive to the choice of n_neighbors, with the best performance observed at k=4, where MSE drops significantly to around 0.89. Lower or higher values lead to either overfitting (k=1) or oversmoothing (k > 5), indicating the need for precise neighborhood selection in localized energy prediction tasks. **Figure 3C** presents the training dynamics of the transformer model, where both MSE and MAE decline smoothly over 15 epochs and plateau without divergence. The absence of overfitting and the consistent drop in error reflect strong generalization and training stability, confirming the transformer's effectiveness and efficiency in structured energy forecasting.



4A. MSE on Smart Home Environmental



4B. MSE on Appliance-Level Household

The comparison of the performance of the various models on the two datasets indicates that there are significant trends in predictive accuracy, model suitability and architecture influence as indicated in Figure 4A and Figure 4B.

7 Conclusion and Future Work

7.1 Conclusion

This paper investigated the usefulness of attention-aware transformer-based regression models to predict energy consumption in smart homes with multivariate time-series and structured tabular data. The transformer architecture was proved to outperform the traditional machine learning and deep learning baselines in terms of predictive accuracy, generalization and training stability through a comprehensive evaluation and benchmarking. The model was able to capture long-term dependencies and multivariate interactions well on the SmartHome

Environmental Time-Series Dataset, with the highest R 2 score and the lowest error rates compared to any other model. In the same capacity, its lightweight implementation on the Appliance-Level Household Energy Usage Dataset produced the best accuracy and model efficiency, which affirmed its flexibility to be used on heterogeneous data types.
These results highlight that transformers are not only more effective than classical and dense neural models, but also offer practical benefits in scalability, interpretability through attention mechanisms and deployability to real-time, resource-constrained contexts. The development pipeline was methodologically rigorous and reproducible due to the CRISP-DM-driven development. In general, attention-based transformer architecture appears to be a powerful versatile tool to solve the problem of intelligent energy prediction and make decisions and optimize the work of modern smart houses based on data. The other important strength is the fact that lightweight transformers are suitable to use in real-time and resource-constrained environments. Parallelized training reduces the time spent developing models and inference efficiency makes it possible to make predictions on the device in IoT ecosystems. This renders the framework particularly useful in smart homes where latency and energy constraints are among the main areas of operations. Although these encouraging findings have been made, there are a number of limitations that need to be mentioned. To achieve its optimum predictive potential, the transformer needs large, high-quality datasets, which will not always be available in smaller homes or in underserved areas.

7.2 Future Work

Although the transformer model has demonstrated remarkable findings, there are a number of directions that can be pursued in the future. To begin with, the model could be improved by including multi-appliance disaggregation and real-time control feedback that would allow the model to facilitate load balancing and demand-response applications. Also, expanding the model to include online learning or on-going adaptation would enable the model to adapt to new household behavior and seasonal trends without the need to retrain the model completely. The other exciting avenue is to combine explainable AI (XAI) methods, like SHAP or integrated gradients, to give a more detailed understanding of feature significance and increase confidence in the decisions. Lastly, the robustness of the model in multi-residential or regional energy networks would enhance the use of the model, and this brings the possibility of scalable energy intelligence systems at the community level. These would contribute more to the development of responsive and sustainable energy infrastructures that are propelled by smart forecasting systems.

References

Syamala, M., Maganti, C. R., Komala, P. V., Pramila, S., Dash, S., Meenakshi, S., & Boopathi, S. (2023). Machine learning-integrated IoT-based smart home en ergy management system. In Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT (pp. 219-235). IGI Global.

Priyadarshini, I., Sahu, S., Kumar, R., & Taniar, D. (2022). A machine-learning ensemble model for predicting energy consumption in smart homes. Internet of Things, 20, 100636.

Iram, S., Shamaila, H. A., Shakeel, H. M., Farid, H. M. A., Riaz, M., Hill, R., Vethathir, P., & Alsboui, T. (2023). An innovative machine learning technique for the prediction of weather-based smart home energy consumption. IEEE Access, 11, 76300-76320. 8
Bhoj, N., & Bhadoria, R. S. (2022). Time-series based prediction for energy con sumption of smart home data using hybrid convolution-recurrent neural network. Telematics and Informatics, 75, p. 101907.

Aancy, H., Mickle, S., Anitha Janet Mary, T., Manikumar, G., Saravanan, V., Mo hanavel, V., & Gupta, R. D. (2024). Energy Consumption Prediction for Home Appli ances with Recurrent Neural Networks. In 2024 International Conference on Expert Clouds and Applications (ICOECA), pp. 600-606. IEEE.

Aurangzeb, K., Haider, S. I., & Alhussein, M. (2024). Individual household load forecasting using bi-directional LSTM network with time-based embedding. Energy Reports, 11, 3963-3975.

Hasan, M. W. (2025). Design of an IoT model for forecasting energy consumption of residential buildings based on improved long short-term memory (LSTM). Measure ment: Energy, 5, p. 100033.

Dostmohammadi, M., Zamani Pedram, M., Hoseinzadeh, S., & Astiaso Garcia, D. (2024). A GA-stacking ensemble approach for forecasting energy consumption in a smart household: A comparative study of ensemble methods. Journal of Environ mental Management, 364, 121264.

Shajalal, M., Boden, A., & Stevens, G. (2024). ForecastExplainer: Explainable house hold energy demand forecasting by approximating shapley values using DeepLIFT. Technological Forecasting and Social Change, 206, 123588.

Li, R., Wu, S., Deng, F., Tian, Z., Cai, H., Li, X., Xu, X., & Liu, Q. (2025). MACLSTM: A Weather Attributes Enabled Recurrent Approach to Appliance-Level Energy Consumption Forecasting. Computers, Materials & Continua, 82(2).

Alzimami, A., Das, S. and Mueller, M., 2024, February. Bayesian optimization-based hyper-parameter-tuned neural network regression for smart home energy consumption modelling using weather information. In *International Congress on Information and Communication Technology* (pp. 495-510). Singapore: Springer Nature Singapore.

Karim, B.A., Patel, R. and Seraphim, M.B.I., 2021. Weather based Electricity Consumption Prediction using Big Data Analytics Model. *Int. J. Eng. Res. Technol*, *10*(4), pp.158-163.

Moulla, D.K., Attipoe, D., Akinyemi, L.A., Thottempudi, S.G., Agbo-Ajala, J.O., Mnkandla, E. and Abran, A., 2024, October. Machine learning algorithms for energy consumption prediction in smart homes: A comparative study. In *International Conference on Wireless Intelligent and Distributed Environment for Communication* (pp. 75-94). Cham: Springer Nature Switzerland.

Sahin, B., Udeh, K., Wanik, D.W. and Cerrai, D., 2024. Predicting energy demand using machine learning: Exploring temporal and weather-related patterns, variations, and impacts. *IEEE Access*, *12*, pp.31824-31840.

Vakhnin, A., Ryzhikov, I., Brester, C., Niska, H. and Kolehmainen, M., 2024. Weather-Based Prediction of Power Consumption in District Heating Network: Case Study in Finland. *Energies*, *17*(12), p.2840.