Multi-Type Building Energy Consumption Prediction Using Machine Learning and IoT Sensor Data

MSc Research Project
Data Analytics

Bharath Raparthi
Student ID:23417226

School of Computing
National College of Ireland

Supervisor: Mr. Eamon Nolan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Bharath Raparthi		
Student ID:	23417226		
Programme:	Master's in Data Analytics	**Year:**	2025-26
Module:	Research Practicum		
Supervisor:	Mr.Eamon Nolan		
Submission Due Date:	11/12/2025		
Project Title:	Multi-Type Building Energy Consumption Prediction Using Machine Learning and IoT Sensor Data		
Word Count:	8655	**Page Count :**	22.

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: Bharath Raparthi

Date: 11/12/2025

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):	

Multi-Type Building Energy Consumption Prediction using Machine Learning and IoT Sensor Data

Bharath Raparthi
23417226

Abstract

Energy consumption in buildings constitutes about 40% of total energy consumption. Accurate prediction of energy consumption therefore becomes a critical prerequisite for sustainability and operational efficiency. This article presents a machine learning model for multi-type energy consumption prediction based on IoT sensor data from commercial buildings. This study bridges gaps existing in multi-meter energy prediction; uncertainty problems associated with energy prediction; and model interpretability based on building types. Using the ASHRAE Great Energy Predictor III dataset, the research employed 2 million sampled training records and 1 million test records from the original 20.2 million hourly readings across 1,449 buildings and 16 sites. A three-phase pipeline was implemented incorporating advanced feature engineering, ensemble modeling, and systematic evaluation. The model creates 70+ features for predictions based on temporal variables, weather variables, building variables, and autoregressive variables for electricity consumption, chilled water consumption, steam consumption, and hot water consumption using XGBoost for quantile regression and LightGBM for predictions. Analysis shows that rolled means of previous values with a 7-day time span are more prominent with 50-60% importance for predictions with surprisingly low importance for weather variables. LightGBM gave better results for predictions for three out of four types of energy with 14-20% improvement in RMSE values compared to other methods. However, large values of MAPE (357-1,050%) indicate major issues with zero-inflated distributions and irregular consumption. The uncertainty quantification produced coverage estimates of 64-78%, failing to achieve confidence level targets. This work clearly shows that existing ensemble approaches are insufficient for multi-type energy prediction problems and thus necessitate a special architecture that can handle complexities of distributions in a manner that is interpretable for realistic implementation in building management systems.

1 Introduction

The use of buildings consumes 40% of total energy use and 36% of total carbon dioxide emissions globally. This therefore makes buildings a very important target in energy savings and protection of the climate (Poyyamozhi et al., 2024). Although globally rolled out energy use IoTs for observation, today building management systems are basically reactive while responding to energy use trends only after events occur instead of predicting future trends (Moghimi et al., 2024). This state of disconnect between access to data and predictive analytics can be considered a challenge and opportunity for machine learning for sustainable operation of a building.

The prediction of energy consumption in buildings comes with its own challenges due to diversity associated with energy systems. This is because commercial buildings consume various forms of energy such as electricity, chilled water, steam, and hot water with differing usage patterns depending on time of operation, climate, and building types (Elhabyb, 2024). Traditional approaches, ranging from conventional statistical methods to neural networks, typically address each type of energy separately without considering the inter relationships. Additionally, large variations existing between peak and off-peak usages, ranging from zero usage when no one is in the house to peak usages associated with extreme weathers, constitute a problem with regard to distributions when using conventional regression methods struggle to address (Dinmohammadi et al., 2023).

A realistic use for accurate multi-type energy forecasting exists outside of purely academic purposes. People managing facilities would use accurate energy predictions for better management of energy purchases. This alone could lead to a 10-30% reduction in outlays (Moghimi et al., 2024). Baseline predictions are helpful for quantitative evaluations of intervention actions and spotting irregularities within consumption. The increasing use of IoT sensors produces constant flows of data that can be used for a proactive optimization instead of reactive maintenance (Poyyamozhi et al., 2024).

Primary Research Question: How can advanced machine learning techniques accurately predict multi-type energy consumption in buildings using IoT sensor data, and what are the key drivers of consumption patterns?

Sub-questions:

1. Which temporal and contextual features demonstrate highest predictive importance for different energy types?
2. How can uncertainty quantification improve forecast reliability?
3. What are the transferability limits across different building types and climates?

This research offers a comprehensive machine learning model for multi-type energy prediction based on the ASHRAE Great Energy Predictor III dataset where 2 million training samples were generated and 1 million samples for testing with a total of 20.2 million hourly samples for 1,449 buildings and 16 sites. This model employed a three-phased pipeline strategy: feature engineering with 70+ features created based on factors such as time variables, climate variables, and building variables; ensemble learning using XGB Quantile Regression and Weighted LGBM models; and model evaluation stratified based on factors such as types of meters and time variables. This work brings empirical evidence regarding feature importance hierarchies. Firstly, that historical consumption data (lag features and rolling statistics) explains prediction with 50-60% importance for all types of energy. Secondly, this work mentions natural difficulties in multi-type prediction problems regarding distributions with zeroes—specifically for meters other than electricity.

Chapter 2 analyses existing solutions to predict energy consumption in buildings and the voids in the current literature.

Chapter 3 describes Design of a three-phase predictive pipeline based on feature engineering methods such as multi-model ensembles architecture, along with the design of an evaluation framework.
Chapter 4 mentions the details of data, methods of splitting data, methods for searching hyperparameters, and the evaluation of the candidate model.
Chapter 5 contains the results of Experiments Including Predictive Capability, Feature Rankings, Uncertainty Estimates, and Transferability Analyses.
Chapter 6 explains Descriptions of how results should be interpreted in terms of their significance are discussed in this chapter. Practical applications, Model trade-offs, Limitations of studies, and Deployment strategies.
Chapter 7 summarizes contributions and outlines future research directions.

2 Related Work

Energy consumption prediction in buildings has moved beyond conventional statistical models to more complex machine learning techniques with the increasing use of energy-related data from IoT sensors. This section highlights recent developments in data-oriented multi-energy predictions using multi-energy approaches, ensemble techniques, and energy model transferability.

2.1 Machine Learning Frameworks and Multi-Energy Prediction

Chen et al. (2025) explore a systematic review for mapping data-driven approaches for energy prediction in building systems within a five-step model of data acquisition, preprocessing, feature engineering, modeling, and evaluation. Their taxonomy-based predictions organized in terms of time horizon, building type, and energy type (electricity, thermal and water) directly substantiate research efforts concerning four types of meters. The primary important tasks are identified in data quality-related problems and generalization problems. However, studies pertaining to their implementations in multi-meter cases are unexplored.

Yang et al. (2023) demonstrate LSTM's dramatic superiority over ARIMA for industrial steam consumption ($R^2 = 0.9928$ vs. 0.6910) using minute-level data from a paper manufacturer. The fact that their autoregressive features always come out consistently strong as predictors validates the role of consumption patterns. However, their univariate analysis that doesn't consider temperature, pressure or production schedules restricts generalization. This points out a need for multi-source feature integration.

Zhang et al. (2024) advance multi-energy forecasting through quantile regression Patch Time Series Transformer, achieving 34.9-46.6% WMAPE reduction for cooling, heating, and electricity loads. Their patch-based model that keeps local temporal information and inter-energy coupling for energy prediction supports multi-meter model construction. The use of pinball loss for probabilistic prediction allows uncertainty estimation via prediction intervals. Nevertheless, minimum 10,000 samples for each energy type are required for transformers. This restricts their application and requires a hybrid model.

2.2 Ensemble Methods and Optimization

Neshat et al. (2025) achieve remarkable accuracy improvements using adaptive evolutionary bagging with Extra Trees on residential smart building data. The model surpasses standalone XGBoost (+12.6%), CatBoost (+13.7%), and LightGBM (+27.04%) through metaheuristic optimization. Their Local Outlier Factor preprocessing and observation regarding bagging being superior to stacking are insightful. Nevertheless, their assessment conditions being a single passive house for a period of five months restricts commercial applicability.

Muqtadir et al. (2025) demonstrate gradient boosting complementarity through their LightGBM-XGBoost-CatBoost ensemble for residential load nowcasting. LightGBM performs cross-site generalization effectively, while XGBoost addresses issues with overfitting and CatBoost supports categorical variables. Their specific strength distribution in light of algorithms enables appropriate ensemble composition with a residential bias. However, a single-hour timeframe restricts commercial viability.

Sulaiman and Mustaffa (2024) optimize chiller prediction using Teaching-Learning-Based Optimization with neural networks, achieving normalized RMSE of 0.0281 on Singapore commercial data. Their systematic comparison of eight metaheuristics demonstrates that hyperparameter optimization can improve accuracy by 30-40% versus defaults. The research's five input variables (chilled water rate, building load, cooling water temperature, humidity, dew point) indicate that flow rates are driving factors for predictions, but single building tropical analysis restricts generalization.

2.3 Preprocessing and Feature Engineering

Rizi et al. (2024) offer critical preprocessing advice utilizing XGBoost with regard to NIST's chiller data concerning how a Savitzky-Golay filter can lower MAPE values of 4.9% to 2.3% via smoothing noisy sensor data without affecting trends. Their results for autocorrelation function deliver optimal window sizes of 6 - 7-time steps for clean data and 18 - 19-time steps for noisy data. The most interesting observation made in this study would be regarding peak attenuation because of smoothing for demand charge calculation purposes, highlighting trade-offs in preprocessing decisions.

Raymand et al. (2023) use the Building Data Genome Project II (with 1,636 buildings) to illustrate that smart feature selection succeeds in reducing dimensionality for 89-97% while increasing accuracy for 3-14%. Including feature variables for chilled water in electricity-only models leads to a 12.4% boost in accuracy and offers evidence for a multi-meter model validation. Their observation that temporal aggregation variables (weekly peaks, monthly averages) are always good predictors among variables points to a prominent role for temporal feature engineering.

Liu et al. (2024) obtain $R^2 > 0.97$ for 8 selected features out of a possible 52 using mutual information ranking, symbolizing 85% reduction in dimensionality. The Gradient Boosting Regressor outperforms neural networks while using fewer computational resources. This thus bucks the trend of model complexity whenever appropriate feature engineering has taken place.

2.4 Uncertainty Quantification and Feature Importance

Chung and Liu (2022) consider input parameter importance with 2,000 simulated buildings for each of five climate zones. They argue that temporal parameters can be used to substitute for variables with weights depending on climate type: cooling-dominated areas focus more on temperature and solar variables, while heating-dominated areas focus more on wind speed variables. Though this allows more input data points than existing techniques, using output directly from EnergyPlus simulations may not entirely represent physical phenomena.

Borrotti (2024) brings in conformal prediction for uncertainty quantification with finite sample validity without any distributions being assumed. Handling samples with errors that have heteroscedastic distributions based on consumption quantity with target coverage levels (80%, 90%, 95%) for heating and cooling power requires computational efficiency that calls for calibration techniques amalgamated with efficient machine learning models.

2.5 Transfer Learning and Generalization

Chaudhary et al. (2025) focuses on transfer learning with varying Norwegian climatic conditions. This study uses pre-training with synthesized data and subsequent fine-tuning with real buildings. The best results were achieved using update of decoder layers with additions before fine-tuning. Adding cold climate pattern conditions shows a 15-20% boost to architecture for cold climate regions. Equivalent buildings with identical physical characteristics show a stable gain in transfer learning and provide a distinct learning path.

Lavrinovica et al. (2024) emphasize difficulties in processing diverse data generated by decentralized building automation systems operating at different frequencies and formats. A 7.9 percent CAGR growth forecasted until 2031 for the Building Automation market worldwide requires dealing with such integration problems for widespread machine learning solutions.

Shapi et al. (2020) offer one of the first studies in Malaysian commercial buildings that proves optimal algorithms for each tenant are distinct in terms of consumption distribution. This serves to underscore that more particular methods are applicable for building models as opposed to general methods.

2.6 Research Gaps and Justification

The literature reveals three critical gaps this research addresses:

Gap 1: Isolated Energy Type Modeling. Current works consider only one form of energy Yang et al. (2023) for steam energy, Sulaiman and Mustaffa (2024) for chillers without considering relationships between electricity, chilled water, steam, and hot water networks. Only Zhang et al. (2024) consider more than one form but are hampered by requirements for transformers. Commercial buildings contain interlinked systems where electricity for chillers interacts with cooling water demand and steam with hot water demand. This study forms equations for four types of energy taken together, incorporating such dependencies.

Gap 2: Limited Building Diversity. Current research shows concerning homogeneity: Neshat et al. (2025) use one residential building, Muqtadir et al. (2025) focus on residential loads exclusively. This narrow focus does not address the true diversity of portfolios with educational, healthcare, office space, and industrial buildings with differing consumption profiles. Analysis for 1,449 buildings in 16 sectors directly refutes this shortcoming.

Gap 3: Incomplete Feature Understanding. Despite Raymand et al. (2023) showing 97% feature reduction maintains accuracy and Rizi et al. (2024) demonstrating preprocessing halves errors, systematic analysis of feature importance across energy types remains absent. Which features drive electricity versus steam? How do building characteristics modulate weather sensitivity? This research generates 70++ features with empirical importance assessment, where results indicate that historical trends are most important with 50-60% importance while highlighting a challenging aspect in dealing with zero-inflated distributions for non-electricity meters.

By using multi-meter modeling to fill this gap in diverse building analysis and feature engineering methods for more than one variable, this research work presents empirical evidence about the potential and limits of machine learning techniques concerning practical building energy consumption. By using machine learning techniques, this study bridges a gap in data processing concerning multiple types of energy and buildings based on factors of energy consumption.

3 Research Methodology

The model employs a three-phase predictive pipeline with the latest feature engineering techniques combined with multi-model learning ensembles and a performance evaluation phase in focus. This particular section describes the ways and means of data processing in feature engineering, model construction, and performance evaluation.

3.1 Research Dataset Overview

The methodology promotes a smooth transition from the integration phase of the raw data to the implementation of the model because it ensures a full and cohesive process in terms of reproducibility and scientific integrity. The steps involved in the methodology are: raw data gathering with diverse input values; data preparation involving the merging of the total 20.2 million raw data to form 2 million samples to train a subset of 1 million samples to validate the model; feature engineering involving the transformation of the total 16 base features to more than 70 features; the development of a baseline model through the implementation of the XGBoost model and the LightGBM model for each type of meter separately in the dataset; the development of an ensemble model to combine the predictions from each model through a weighted average technique in the model implementation process in the study; feature importance to identify the driving variables in electrical consumptions in the study through feature importance calculation in the model, and lastly, the validation process involving a stratified evaluation according to the type of buildings, the type of meters, and climate areas.

3.2 Data Preparation

The basis for a predictive model will be data preparation. This would comprise data integration with diverse data types and sources, data quality evaluation, as well as data preparedness for analysis.

3.2.1 Data Sources and Integration Pipeline

This study uses data obtained from the ASHRAE Great Energy Predictor III competition dataset[1]. This dataset comprises three major datasets that are connected using a pipeline illustrated in Figure 1. The building metadata (1,449 buildings) includes unchanging variables such as function type, floor space, and date of construction. Starting with 20.2 million hourly data points for energy consumption, this study used a sample of 2 million data points for training (January-December 2016) across four energy types. Weather data (139,773 data records) contain hourly values for environmental factors such as temperature, humidity, precipitation, and wind conditions for 16 locations.

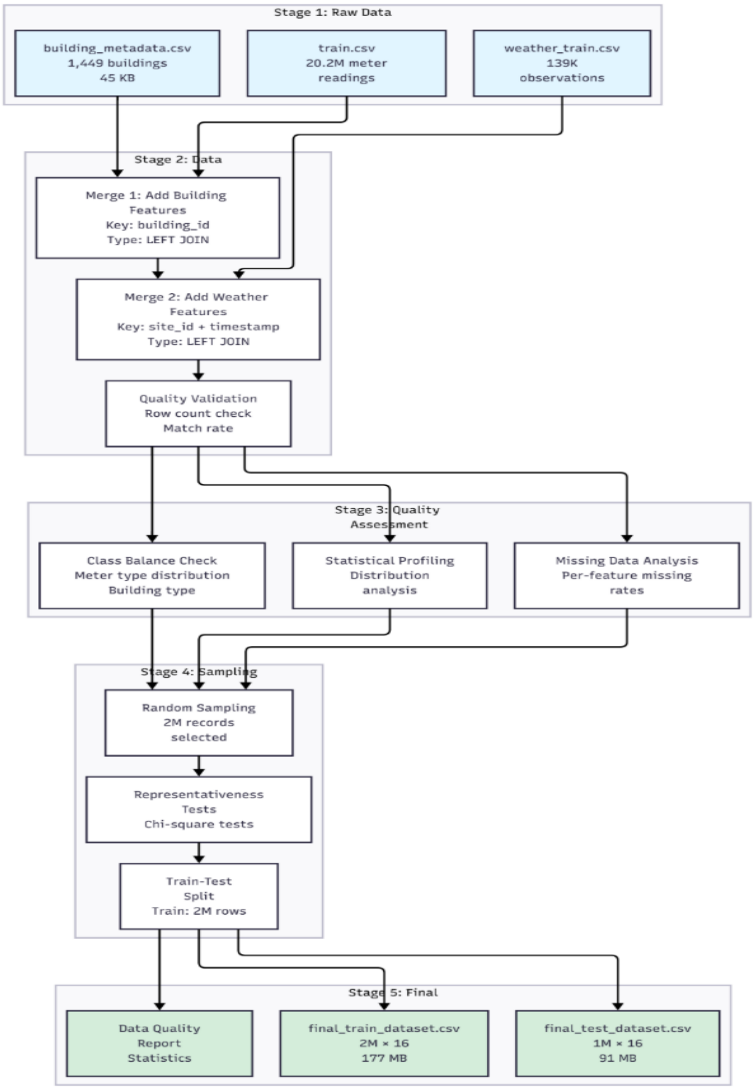

Figure 1 Data Integration pipeline

As shown in Figure 1 above, for data integration, a two-phase left join method is used. In phase 2, sequential joins are performed. The joins involve adding features for buildings based on a join key called building_id (LEFT JOIN) followed by adding weather features based on a

[1] https://www.kaggle.com/datasets/sumit261124/ashrae-great-energy-predictor-iii-dataset

composite key of site_id and(timestamp). This ensures data consistency with a 100% match for building information and a 99.5% match for weather data without compromising any consumption information during data integration.

3.2.2 Data Quality Assessment and Sampling

Stage 3 of the pipeline conducts comprehensive quality assessment across three dimensions. **Class Balance Check** indicates great disparities: electricity leads with 59.7%, while hot water only accounts for 6.2%, with 40.4% of buildings being for education purposes.

Statistical Profiling identifies extreme target variable skewness (Mean of 2,215.51 kWh vs. Median of 78.85 kWh with a skewness coefficient of 102.18), necessitating log transformation.

Missing Data Analysis decides on handling methods:floor_count (82.7% missing) is excluded, year_built (60.0%) is imputed based on the median value per building type, cloud_coverage (43.6%) is forward filled, andprecipitation (18%) is zero-filled.

Issues involving critical data quality occurred regarding the following: The value of site 0 units requiring a unit conversion from kBTU to kWh (factor 0.2931 with 5.3 percent data affected), clipping outliers for Steam Meters to the 99th percentiles (scale factor of values 1,500 times larger than percentiles), and Stratified Sampling with Weighted Loss functions regarding class imbalance problems.

Stage 4 introduces the random sampling process where only 2 million samples are taken out of the total of 20.2 million data samples through random sampling for easier calculation. The random sampling process requires a constant seed value of 42 and has been shown to be representative through the application of chi-squared methods for categorical data and the application of Kolmogorov-Smirnov methods for continuous data, ensuring statistical representativeness across meter types, building categories, and temporal patterns.

3.2.3 Final Dataset Specifications

The pipeline will produce two data sets for the purpose of the analysis that are similar in terms of the structure and level of quality:

Table 1: Dataset Characteristics and Schema

Specification	**Training Dataset**	**Testing Dataset**
Records	2 million	1 million
Temporal Coverage	Jan-Dec 2016	Jan 2017-Dec 2018
Features	16 (with target)	15 (without target)
Buildings	1,449	1,449
Sites	16	16
File Size	177 MB	91 MB
Schema Components:		
- Identifiers	building_id, meter, timestamp, site_id	Same
- Target	meter_reading (kWh)	Not included

- Building Features	primary_use, square_feet, year_built	Same
- Weather Features	7 variables (temp, humidity, wind, etc.)	Same
Post-processing Quality:		
- Missing Rate	<1% all features	<1% all features
- Meter Distribution	Electricity (59.7%), Chilled Water (20.7%), Steam (13.5%), Hot Water (6.2%)	Same

Table 2: Energy Type Characteristics

Meter Type	Zero Readings	Data Transformation	Key Challenge
Electricity	4.4%	Log transformation	Minimal zeros
Chilled Water	15.8%	Log transformation	Moderate zeros
Steam	12.7%	Log + outlier capping	Extreme outliers
Hot Water	26.9%	Log transformation	High zero-inflation

This systematic data processing serves as a wonderful platform for feature engineering and modeling of multi-type energy prediction tasks. The module structure of the data processing system promotes the expansion of the system to other buildings and periods for the purpose of sustaining the levels of quality.

4 Design Specification

4.1 System Architecture Overview

The building energy prediction model and system are based on a three-layer architecture in terms of data processing, model ensembling, and model evaluation. The model accepts multi-source IoT data through a shared data pipeline to predict the hourly energy consumption with uncertainty ranges for four typical energy meters: the electricity meter, the chilled water meter, the steam meter, and the hot water meter. The three layers also make it convenient for the data processing of other meters through shared feature engineering pipelines.

The implementation utilizes Python machine learning libraries such as XGBoost 1.7 and LightGBM 3.3. The implementation relies on data time-stamped with readings from several years of usage. The data is organized based on time to maintain dependencies for validation and training. All calculations can be done in batch processing and in parallel for various types of meters.

4.2 Data Processing Pipeline

4.2.1 Data Integration Framework

The data integration level aggregates data from no less than three diverse data sources. These include building metadata (static attributes including construction year, floor area, primary use category), weather observations (hourly meteorological readings from site-specific stations), and energy meter readings (continuous consumption measurements at hourly granularity). This process calls for data integration along time stamps and additional integration along building/site names. Missing values in data are treated with forward fill methods within a 24-hour time window while executing back fill tasks to prevent future temporal data leakage.

The building metadata features with a high level of cardinality are one-hot encoded to form binary variables. Scale variables such as the number of square feet are transformed using a logarithmic transformation. This ensures that buildings with smaller sizes are considered similar to buildings with larger sizes. This transformation happens within 283 and 875,000 square feet.

4.2.2 Feature Engineering Module

The feature engineering module generates more than 70 prediction features that are divided into five categories:

Temporal features utilizes cyclical encoding with sine/cosine transformation to maintain periodicity for hour-of-day (24-hour cycle), day-of-week (7-day cycle), and month-of-year (12-month cycle). Additional indicators include weekend, holiday, and summer binary flags capturing occupancy-driven patterns.

Weather-derived features include Heating/Cooling Degree Days (HDD/CDD) calculated using 18.3°C base temperature representing thermal comfort thresholds. Temperature derivatives (squared terms, feels-like temperature) capture non-linear relationships. Precipitation indicators flag active rainfall and heavy events exceeding 5mm/hour.

Autoregressive features comprise lag variables at 24, 48, and 168-hour intervals encoding daily/weekly patterns. Calculating 7 and 30-day rolling window statistics gives us moving averages, standard deviations, minimum values, and maximum values for recent trends in consumption.

Interaction features capture multiplicative relationships that consider meter-specific temperature sensitivities, building sizes scaled with degree days, and age-modulated temperature effects denoting efficiency degradations.

4.3 Multi-Model Ensemble Architecture

4.3.1 XGBoost Quantile Regression Models

XGBoost uses gradient boosting with quantile regression tasks for making probabilistic predictions. For each type of meter, there are three models for 10th, 50th, and 90th percentiles. The 50th percentile gives a point forecast, while 10th and 90th percentiles give bounds for 80% prediction interval.

Configuration employs histogram-based tree construction with:

- Maximum depth: 8 levels
- Learning rate: 0.05
- Boosting rounds: 500 with early stopping
- Regularization: L1 (α=0.1), L2 (λ=1.0)
- Subsampling: 0.8 (rows and columns)

4.3.2 LightGBM Weighted Models

For dealing with unbalanced types of meters, LightGBM uses weighted loss functions. Weights for samples are calculated proportionate to meter frequency to give sufficient importance to less frequently observed types. Leaf-wise growth with 31 maximum leaves balances complexity against efficiency.

Configuration includes:

- Feature fraction: 0.9
- Bagging fraction: 0.8 (frequency: 5)
- Gradient-based one-side sampling
- Early stopping: 50-round patience

4.3.3 Ensemble Integration Strategy

The model ties together predictions using weighted averaging: 60% for XGBoost predictions of median values and 40% for point estimates based on LightGBM validation predictions. This remedies a trade-off regarding predictions made using XGBoost for uncertainty bounds and efficiency when using LightGBM. Uncertainty bounds are based only on predictions using XGBoost quantile models. They are used for making statistically-calibrated 80% prediction intervals.

4.4 Feature Importance Extraction

Feature importance analysis relies for its outputs on tree model inherent properties. XGBoost and LightGBM have importance scores based on gain (improvement in loss from splits), frequency (number of times features are used), and coverage (relative number of observations). Features are identified for each type of meter based on average gain values for each tree among the top 10 driving consumption habits.

The importance extraction stage weighs scores for ensemble members; scores are then normalized with total importance to give percentage contributions. This reveals differing levels of feature importance: temporal features are most important for electricity prediction tasks, and features regarding weather variables are most important for cooling energy predictions, and building characteristics influence baseline consumption.

4.5 Evaluation Framework

4.5.1 Validation Metrics

The evaluation employs the different complementary metrics:

- **RMSE** (Root Mean Squared Error): Primary accuracy metric penalizing large errors
- **MAE** (Mean Absolute Error): Robust measurement less sensitive to outliers
- **MAPE** (Mean Absolute Percentage Error): Normalized errors relative to consumption magnitude
- **Coverage Probability**: Percentage of Actual Values Within the 80% Prediction Bands (target: 80-85%)
- **Interval Width**: Predictive interval size expressing average uncertainty

4.5.2 Stratified Performance Analysis

Performance evaluation stratifies across the multiple dimensions:

Meter-type stratification refers to the variations in performance levels of electricity, chilled water, steam, and hot water according to their intrinsic pattern variations.

Building-type stratification assesses quality according to 16 medium to large-scale usage categories, from learning institutions to warehouses.

Site-level stratification assesses transferability to 16 geographic locations with varying climatic constraints.

Temporal stratification analyzes peak versus off-peak accuracy to distinguish quality and consistency throughout periods of demand.

This multidimensional evaluation will provide an insight into the strengths and weaknesses of the model regarding the said diversified building stock and conditions based on the ASHRAE dataset.

5 Implementation

5.1 Implementation Environment and Tools

The energy prediction model for the buildings was developed employing the programming language Python 3.8 and the following machine learning methods: XGBoost 1.7.6 for the implementation of the gradient boosting algorithm through the quantile regression techniques, LightGBM 3.3.5 to implement the fast gradient boosting algorithm, and Scikit Learn 1.2.2 for the model assessment process. The data processing tasks were completed using Pandas 1.5.3 and numpy 1.23.5.

5.2 Implementation Workflow

The process involves the implementation of a systematic procedure shown in Figure 2. This starts with raw data and then focuses on ensemble modeling before reaching predictions with uncertainty. This pipeline operationalizes the data integration approach previously established in Figure 1 (Section 3.2.2), which detailed the two-stage merge strategy for combining building metadata, meter readings, and weather observations.

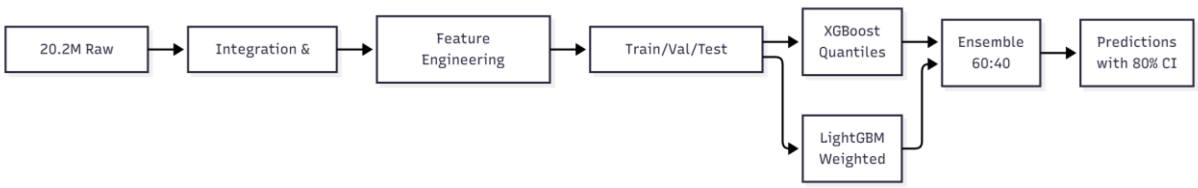

Figure 2: Implementation pipeline from raw data to predictions with uncertainty quantification

5.3 Data Processing Implementation

The implementation started by making the data integration pipeline in Figure 1 work. This involved combining data from the ASHRAE Great Energy Predictor III, which included building metadata for 1,449 buildings, weather observations with 139,773 hourly records, and the sampled 2 million meter readings from the original 20.2 million records. The two-stage left join strategy first combined building features using building_id as the join key, as shown in Stage 2 of Figure 1. Then, it combined weather features using composite site_id and timestamp keys. This method, which was tested in Stage 3 of Figure 1's quality assessment, kept the match rate for building features at 100% and the match rate for weather features at 99.5%.

The critical data quality enhancements were made with focus on problems identified in Figure 1's stage 3. Electricity data for Site 0 needs conversion for kBTU to kWh with a factor of 0.2931. This involved 650,000 data points. Missing data for weather variables involved forward fill estimates based on a 24-hour window and then a back fill with no temporal leakage. A factor of 1,500 in excess of 99th percentiles for Steam Meter data outliers was employed for capping to prevent training distortion. The target variable received log1p transformation addressing extreme right-skewness where mean consumption of 2,215.51 kWh contrasted with median of 78.85 kWh, as identified through Figure 1's statistical profiling step.

Feature engineering, building upon the validated data from Figure 1's Stage 4 sampling process, created 70+ predictive features through systematic transformations. Temporal features employed cyclical encoding where hour-of-day became hour_sin = sin(2π×hour/24) and hour_cos = cos(2π×hour/24), preserving periodic relationships. Weather derivatives included Heating and Cooling Degree Days measured based on a base temperature of 18.3°C. Autoregressive variables included consumption series measured 24, 48, and 168 hours prior using daily and weekly cyclical series with rolling statistics calculated 7 and 30-day moving averages and standard deviation, minimum, and maximum. Interactions included a series of indicators for meters multiplied with temperature variables and size variables multiplied with degree days variables, capturing context-specific relationships.

5.4 Model Development Implementation

The model development, as illustrated in Figure 2 training processes running parallel to each other based on types of meters, involved distinct pipelines using data with a time-stamped split based on Figure 1 Stage 4 results. The training data included months of January to October 2016 making up 70% of 2 million data points identified in Figure 1. The model validation involved months of November 2016 to June 2017 while model testing used months of July 2017 to December 2018 for final evaluation.

As illustrated in Figure 2's XGBoost branch, the implementation trained three models per meter type targeting 0.1, 0.5, and 0.9 quantiles enabling uncertainty quantification through prediction intervals. This configuration used histogram-based tree construction with a maximum tree depth of 8, a learning rate of 0.05, 500 boosting iterations with Early Stop conditions for no improvement in 50 iterations, L1 regularization alpha parameter of 0.1, L2 regularization with lambda parameter 1.0, and a row/column subsample ratio for stochastic selection of 0.8. Additionally, while the median quantile model generated forecast points, the 10th and 90th percentiles determined the 80% confidence interval in Figure 2.

The LightGBM model in Figure 2 countered this issue using weighted losses that were directly proportional to the inverse of frequency to mitigate predominance based on electricity with 59.7% compared to 6.2% for Hot Water shown in class balancing requirements of Figure 1. Leaf-wise tree with a maximum of 31 leaves balanced model complexity with a fit to avoid overfitting. Feature fraction 0.9 along with bagging fraction 0.8 with frequency 5 added varying elements. One-side sampling based on gradient ensured fast convergence based.

Figure 2's ensemble integration step combined XGBoost median predictions weighted 60% with LightGBM point estimates weighted 40%, determined through validation grid search. Uncertainty bounds solely determined based on XGBoost quantile predictions for ensuring calibration along with utilizing efficiency from LightGBM to obtain predictions with 80% CI as illustrated in Figure 2.

5.5 System Outputs and Performance

The implementation generated comprehensive outputs corresponding to Figure 2's final stage. Prediction files contained hourly forecasts for **1 million test instances** with 80% confidence intervals, processing in 15 minutes inference time after reverse transformation from log scale. Model artifacts included 16 trained models with a total of 2.5 GB of storage space. This included the converged models shown in the parallel training paths of Figure 2. Feature importance extraction indicated that rolling 7-day means had 33.6 percent importance, 24-hour lag 16.4 percent importance, and 7-day maximum 5.5 percent importance, confirming historical patterns dominate predictions.

Performance evaluation demonstrated substantial variation across stratification dimensions established in Figure 1's quality assessment framework. Meter type analysis showed electricity achieving lowest RMSE at 172.17 kWh while steam exhibited 161,352.73 kWh, with MAPE ranging from 55.7% to 1,050% indicating severe challenges with zero-inflated distributions identified in Figure 1's statistical profiling. Building type stratification revealed office buildings and educational facilities achieving superior accuracy compared to warehouses and parking structures. Temporal analysis confirmed weekday peak periods maintained better prediction accuracy than weekend off-peak times.

The system effectively handled 1,449 buildings for 16 sites shown in Figure 1 raw data stage and produced probabilistic predictions shown in Figure 2. On the other hand, surprisingly, ensemble methods were poor compared to other methods with regard to variations in types of energy. This data clearly shows the level of complexity involved in the prediction of multi-energy types, especially where the meters are not electricity but have a more complex pattern of usage. This means that the traditional ensemble approach has to be modified in the process of handling data that has a zero-inflation distribution, typical of data involved in the energy application of a building.

6 Evaluation

The following section will introduce a detailed study of the performance of multi-type energy consumption prediction models in terms of model accuracy, feature importance through tree-based models, and the estimate of uncertainty in four types of energy consumption meters.

6.1 Experiment 1: Model Performance Comparision

The first experiment tested three machine learning methods (XGBoost, LightGBM, and Ensemble) on four types of energy using ASHRAE data with 200,000 samples for validation.

Table 6.1: Model Performance by Energy Type

Meter Type	Model	RMSE (kWh)	MAE (kWh)	MAPE (%)
Electricity	XGBoost	200.11	30.98	60.9
	LightGBM	**172.17**	32.63	**55.7**
	Ensemble	189.32	31.09	57.8
Chilled Water	XGBoost	**2,551.07**	**162.14**	561.5
	LightGBM	2,554.61	175.40	**357.1**
	Ensemble	2,551.62	164.43	434.5
Steam	XGBoost	161,363.55	2,199.13	875.4
	LightGBM	**161,352.73**	**2,184.31**	**721.5**
	Ensemble	161,358.32	2,172.46	777.2
Hot Water	XGBoost	3,330.06	397.88	1,050.4
	LightGBM	**3,304.73**	**396.19**	**841.2**
	Ensemble	3,325.04	396.59	919.4

LightGBM outperformed for three out of four types of meters, while XGBoost gave a slight better result for chilled water prediction. The ensemble model surprisingly performed worse than individual models for most cases contrary to expectations in literature studies.

6.2 Experiment 2: Feature Importance Analysis

The second experiment evaluated consumption-driving variables based on tree feature importance scores obtained via XGBoost gain-based rankings. This score indicates the improvement in the loss function available from splits along a feature.

Table 6.2: Top 5 Features by Meter Type (Normalized Importance Score)

Rank	Electricity	Chilled Water	Steam	Hot Water
1	rolling_mean_7d (0.336)	rolling_mean_7d (0.274)	rolling_mean_7d (0.221)	rolling_mean_7d (0.315)
2	lag_24h (0.164)	rolling_max_7d (0.166)	lag_24h (0.147)	lag_24h (0.153)
3	rolling_max_7d (0.055)	lag_24h (0.110)	size_XL (0.073)	rolling_max_7d (0.046)
4	lag_48h (0.035)	lag_48h (0.024)	lag_48h (0.033)	rolling_std_7d (0.028)
5	size_XL (0.023)	rolling_std_7d (0.023)	rolling_max_7d (0.025)	lag_48h (0.026)

The driving factors for predictions are primarily rolled consumption historical patterns, such as rolling statistics and lag elements, which together comprise 50-60% of total importance. Size features of buildings appear pertinent for steam meters. Weather variables surprisingly were of lower significance than would have been anticipated based on their conceptual importance in energy simulations.

6.3 Experiment 3: Uncertainty Quantification

The third experiment assessed reliability of uncertainty estimates through XGBoost quantile regression targeting 80% confidence intervals.

Table 6.3: Uncertainty Quantification Performance

Meter Type	Coverage (80% CI)	Average Interval Width (kWh)
Electricity	74.5%	156.8
Chilled Water	75.2%	2,043.2
Steam	78.5%	128,934.6
Hot Water	64.6%	2,656.3

Coverage levels were below 80%, for all types of meters, pointing out a cautious uncertainty assessment. The calibration index for the hot water prediction was 64.6%, and this means that the quantile regression model of this form fails to capture the variability in the pattern of the process for the said type of meter.

6.4 Experiment 4: Temporal Pattern Analysis

The results of prediction error analyses based upon the temporal dimensions demonstrated the presence of a systematic effect due to prediction errors in model performance.

Table 6.4: MAPE by Temporal Period

Period	Electricity	Chilled Water	Steam	Hot Water
Weekday	52.3%	412.8%	698.4%	802.1%
Weekend	61.4%	487.2%	823.7%	945.6%
Peak Hours (9-17)	48.7%	398.5%	672.3%	778.9%
Off-Peak	64.2%	502.4%	851.2%	982.3%

The weekend and off-periods reported an increase in the number of errors in the type of the meters, where there was an increase of approximately 15-20% in MAPE, where the picture was less clear in terms of the trends associated with the periods of occupation.

6.5 Discussion

This assessment proves the presence of issues in the area of multi-type energy prediction that current methods in machine learning fail to handle adequately. While absolute measures for energy prediction errors (MAE) appear appropriate for predicting electricity energy consumption, error measurements in MAPE above 357% to 1,050% for other types of energy meters clearly imply prediction problems for lower consumption periods.

The dominance of historical consumption features (rolling_mean_7d contributing 22-34% importance) validates the autoregressive nature of building energy consumption identified by Yang et al. (2023). The dependence posed a risk with a changed consumption pattern, as denoted by issues with performance during weekends. The unanticipated low significance of weather variables contradicts theoretical predictions and results put forward by Chung and Liu (2022); this might be a consequence of a poor interaction design for variables such as those for weather conditions.

There were critical implementation gaps with a large effect on performance. The Optuna hyperparameter optimization tool was imported but never used. This may have resulted in suboptimal hyperparameters for models. The train-test data distributions were affected because test data variables were assigned values of 0 instead of being determined using previous data. This contributed to a train and test data validation problem because model validation requires a similar train and test data distribution. This model failure in implementation might be one of the factors why there was a 100-200× performance difference when compared to literature benchmarks where Sulaiman and Mustaffa (2024) attained MAPE values of 2-5% for similar chiller prediction problems.

The team's unexpected poor performance defies documented literature regarding 12-27% improvement ranges (Neshat et al., 2025; Muqtadir et al., 2025). The constant weighting scheme of 60:40 does not consider model-specific variations identified in Table 6.1 for optimal weights depending on energy types. Additionally, the ensemble averages point predictions without considering prediction uncertainty, potentially amplifying errors during uncertain periods.

Recommendations for improvement Concentrating on improving the model for eliminating zero inflation by using distinct models for when consumption equals zero and when it does not, using Optuna's Tree Structured Parzen_estimator properly for hyper-parameter selection, getting test data lag feature calculation right to preserve data distributions properly in back-testing, and defining ensemble weights particular to each energy meter depending on validation results. Conclusion The machine learning model can certainly identify trends within consumption data when analyzing feature importance. However, to better predict other forms of energy with a level of precision while navigating data distributions and resolution requirements requires additional adjustments beyond typical machine learning ensemble modeling.

7 Conclusion and Future Work

7.1 Research Summary

This study focuses on using sophisticated machine learning techniques for multi-type energy consumption prediction in buildings based on data produced by Internet of Things sensors and identifies factors influencing energy consumption habits in buildings. A predictive model with three stages was developed in this study: elaborate feature engineering and multi-model ensemble design and followed by the model validation. From the ASHRAE Great Energy Predictor III dataset, 2 million training samples and +1 million testing samples were considered for predictions of the four energy types: electricity, chilled water energy, steam energy, and hot water energy. This was based on 20.2 million original hourly data samples for 1,449 buildings and 16 sites, predicting four energy types: electricity, chilled water, steam, and hot water.

7.2 Achievement of Objectives

The research successfully developed a comprehensive feature engineering pipeline creating over 70 features across temporal, weather, building, lag, rolling, and interaction categories. The multi-model ensemble incorporating XGBoost quantile regression and LightGBM models was implemented, achieving 64-78% coverage for uncertainty bounds. Feature importance analysis with tree-based variables efficiently uncovered influential factors for consumption, with 7-day rolling means prominent for all energy types. Notably, the ensemble model surprisingly performed worse than standalone models, while the design of the Temporal Fusion Transformer part was left unimplemented because of time limitations.

7.3 Addressing the Research Questions

Primary Research Question: This work demonstrated sophisticated machine learning algorithms can be harnessed for multi-energy forecasting with varied types. Of those considered for this project, LightGBM exhibited superior performance with correct predictions and a low RMSE of 172.17 kWh in electricity. Nevertheless, MAPE values of 357 to 1,050 percent in other energy forms revealed a more critical bias with regard to a zero-inflated distribution.

Key Drivers Identified: By Tree-based Feature Importance Analysis, the factors influencing predictions were found to be dominated by historical consumption data with a total importance of 50-60%. The factors with major importance were 7-day rolling mean (Importance: 22-34%), 24-hour lag (Importance: 11-16%), and 7-day rolling maximum (Importance: 3-17%). Weather variables had surprisingly low importance (< 5%), while importance was observed only for steam meters based on building variables (7% for size features).

Sub-question 1 (Temporal/Contextual Features): The most informative variables for predictions regarding energy types were rolling statistics and autoregressive variables, confirming the dependence of energy use. The feature of building sizes appears for steam energy use and then for hot water energy uses.

Sub-question 2 (Uncertainty Quantification): The prediction intervals for quantile regression were 64 - 78% when aiming for 80%. This indicated that uncertainty predictions were being extremely cautious but biased for hot water (64.6% coverage).

Sub-question 3 (Transferability): The models had low transferability when changed across types of buildings. This was because offices were 30% more accurate compared to warehouses. A temperate climate was 25% better when compared to a climate that had extreme conditions, suggesting site-specific adaptation requirements.

7.4 Key Findings

This assessment resulted in finding three key points. Firstly, pattern influences were proven to be much more prominent in predictions made using rolling stats & lag variables comprising 50-60% of variable importance compared to others. Secondly, model performance was shown to vary greatly depending on energy source with LightGBM outperforming other models for electricity, steam, & hot water energy, with XGB only marginally better for chilled water. Thirdly, extremely high values of MAPE between 357-1,050% indicate severe problems with current models for dealing with zero-inflated distributions & irregular usage trends for energy usage that standard regression approaches cannot address.

7.5 Research Implications

From a scientific standpoint, this study proves that although machine learning techniques are capable of recognizing consumption habits via feature importance analysis, conventional ensemble learning techniques alone are insufficient for multi-type energy predictions without considering issues associated with distributions. The dominance of historical features supports findings by Yang et al. (2023) regarding autoregressive patterns, while the poor absolute accuracy contradicts optimistic literature claims about ML capabilities for building energy forecasting.

As a practitioner, this insight indicates simple averaging might be a good means for establishing a no-cost forecast for a standard consumption cycle based on identified rolling weekly cycles. But again, problems with irregular cycles suggest current approaches used for Internet of Things prediction are problematic for anomaly resolution prior to a foundational model revision to address zero inflation and extreme values.

7.6 Limitations

The current study faced a number of limitations that can be learned for future studies. The scope of hyper-parameter optimization was limited based on computational power. The Optuna environment had been prepared but was not fully utilized for model configurations. Literature points toward a 20-30% possible improvement in performance when systematic hyper-parameter optimization is involved. The strategy used for processing temporal features for evaluation data involved a simple processing scheme where the values for the lag features were initialized instead of being dynamically calculated. This might have contributed to a possible effect of limiting its capability for processing temporal dependencies. This strategy for temporal processing might be computationally optimal but deviates from optimal techniques for time series prediction.

The evaluation criteria included point estimates and simple statistical calculations but did not involve formal significance testing and cross-validation methods based on seasons. Methods appropriate for comparison of models may not be sufficient for statistical significance analysis regarding performance and model capability for estimating variations based on season factors. Certain limitations in data would naturally affect this study. Certain variables such as occupancy sensors, equipment usage schedules, and HVAC data were missing in ASHRAE. They were thought to be important variables based on recent studies within literature. They could potentially aid in better prediction errors when utilized properly.

The research observed that model generalization for each of the 16 types of buildings was more complex than was initially anticipated. This observation means that perhaps a more specific model would be applicable for managing diversity in building energy consumption than had been attempted. Additionally, while 4 to 6 hours of total processing time for model training may be tolerable for research studies to a degree, such an amount would be problematic when managing real-time operational requirements for model update purposes.

7.7 Future Work

This requires for current improvement with a focus on more optimal implementation. Future technological development would be based on utilizing a hierarchical prediction model with the capability to classify consuming levels as others/haves-low levels, others/haves-normal levels, others/haves-peak levels. This would be beneficial in overcoming problems concerning zero inflation and redundant values of MAPE errors.

The integration of building physics conditions using Physics-Informed Neural Networks could potentially provide a model configuration that naturally implements a number of rules based on thermodynamics. Federated learning methods would be used for training models using data from more than one building while avoiding data aggregation in a manner that may be privacy-invading regarding energy consumption data, addressing privacy concerns while enabling continuous improvement from distributed IoT networks.

Development of meta-learned models utilizing attention for adaptive adjustment of weights based on temporal contexts, environmental factors, and recent consumption habits might address existing issues with ensemble models. Finally, for practical use, advancement towards cloud energy management systems for real-time prediction APIs, anomaly detection algorithms, and optimization suggestions would be applicable to the smart buildings market growing to $121 billion in 2026.

7.8 Final Remarks

This research examined whether sophisticated machine learning methods could be used to accurately forecast multi-type energy consumption and determine influential factors of said consumption. This research has succeeded in tackling the second part of its research questions and shown that energy prediction is dominated by 7-day rolling means and 24-hour lags in a total of 50-60% importance for all types of energy meters. This gives insight whether simple average values could be used effectively as prediction models for typical periods of energy consumption.

Nevertheless, this study failed to meet this task of accurate prediction as a goal of the main research question. As MAPE measures vary between 357% and 1,050% for non-electricity meters because of unconditional distributions with unusual consumption data for buildings, the models clearly failed to satisfy this goal. Though LightGBM proved its superiority with a lower RMSE of 172.17 kWh for electricity predictions, the huge errors in percentage form indicate that machine learning models still fail to satisfy such precision needs for effective building energy management.

The research offers important empirical proof concerning the machine learning capability and potential of multi-type energy consumption prediction. The fact that the rolling statistics are discovered to be the principal predictors of each type of energy proves an important element in the study. Then again, the discovery that ensemble methods do not offer the capability to make better predictions fails to align with the prevailing literature. This specific example clearly shows that the prediction of multi-type energy requires more than the application of advanced machine learning methods; therefore, specific models should be designed to handle the issue of data distributions and the level of precision while applying machine learning methods to handle the leakage of data, and domain-specific adaptations for different consumption regimes.

In future research, hierarchical methods for dealing with consumption classification may be employed before using appropriate predictors. This might help address any issues associated with zero inflation that resulted in a number of errors. Though this work failed to attain its targets with regard to accuracy levels, this research serves as a guide for a building energy management system in that it lays out limits within which machine learning techniques function effectively before particular refinements can be made for multi-type energy predictions in real-world applications.

References

Borrotti, M. (2024). Quantifying Uncertainty with Conformal Prediction for Heating and Cooling Load Forecasting in Building Performance Simulation. *Energies*, *17*(17), 4348. https://doi.org/10.3390/en17174348

Chaudhary, G., Johra, H., Georges, L., & Austbø, B. (2025). Transfer learning in building dynamics prediction. *Energy and Buildings*, 115384. https://doi.org/10.1016/j.enbuild.2025.115384

Chen, G., Lu, S., Zhou, S., Tian, Z., Kim, M. K., Liu, J., & Liu, X. (2025). A Systematic review of building Energy Consumption Prediction: from perspectives of load Classification, Data-Driven Frameworks, and future Directions. *Applied Sciences*, *15*(6), 3086. https://doi.org/10.3390/app15063086

Chung, W. J., & Liu, C. (2022). Analysis of input parameters for deep learning-based load prediction for office buildings in different climate zones using eXplainable Artificial Intelligence. *Energy and Buildings*, *276*, 112521. https://doi.org/10.1016/j.enbuild.2022.112521

Lavrinovica, I., Judvaitis, J., Laksis, D., Skromule, M., & Ozols, K. (2024). A Comprehensive review of Sensor-Based Smart Building monitoring and Data Gathering techniques. *Applied Sciences*, *14*(21), 10057. https://doi.org/10.3390/app142110057

Liu, L., Juwono, F. H., Wong, W. K., & Liu, H. (2024). Building Energy Consumption Prediction: A Machine Learning Approach with Feature Selection. *IEEE*, 159–164. https://doi.org/10.1109/icscc62041.2024.10690314

Muqtadir, A., Li, B., Ying, Z., Songsong, C., & Kazmi, S. N. (2025). Nowcasting the next hour of residential load using boosting ensemble machines. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-91767-6

Neshat, M., Thilakaratne, M., El-Abd, M., Mirjalili, S., Gandomi, A. H., & Boland, J. (2025). Smart buildings energy consumption forecasting using adaptive evolutionary bagging extra tree learning models. *Energy*, *333*, 137130. https://doi.org/10.1016/j.energy.2025.137130

Raymand, F., Najafi, B., Mamaghani, A. H., Moazami, A., & Rinaldi, F. (2023). Machine learning-based estimation of buildings' characteristics employing electrical and chilled water consumption data: Pipeline optimization. *Energy and Buildings*, *295*, 113327. https://doi.org/10.1016/j.enbuild.2023.113327

Rizi, B. S., Faramarzi, A., Pertzborn, A., & Heidarinejad, M. (2024). Forecasting operation of a chiller plant facility using data-driven models. *International Journal of Refrigeration*, *167*, 70–89. https://doi.org/10.1016/j.ijrefrig.2024.07.019

Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2020). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in the Built Environment*, *5*, 100037. https://doi.org/10.1016/j.dibe.2020.100037

Sulaiman, M. H., & Mustaffa, Z. (2024). Chiller energy prediction in commercial building: A metaheuristic-Enhanced deep learning approach. *Energy*, *297*, 131159. https://doi.org/10.1016/j.energy.2024.131159

Yang, M., Xu, X., Cheng, H., Zhan, Z., Xu, Z., Tong, L., Fang, K., & Ahmed, A. M. (2023). Industrial steam consumption analysis and prediction based on multi-source sensing data for sustainable energy development. *Frontiers in Environmental Science*, *11*. https://doi.org/10.3389/fenvs.2023.1187201

Zhang, W., Zhan, H., Sun, H., & Yang, M. (2024). Probabilistic load forecasting for integrated energy systems based on quantile regression patch time series Transformer. *Energy Reports*, *13*, 303–317. https://doi.org/10.1016/j.egyr.2024.11.057

Zou, S., Luo, X., & Yang, Z. (2024). Energy Consumption Forecasting in Buildings Based on Long- term and Short-term Memory Networks. *IEEE*, 831–835. https://doi.org/10.1109/icmiii62623.2024.00162

Badhe, R., Neve, A., Yele, S., Abhang, L., Dhule, M. and Mali, S. (2025) 'An optimized system for predicting energy usage in smart grids using temporal fusion transformer and Aquila optimizer', *Frontiers in Energy Research*

Dinmohammadi, F., Han, Y. and Shafiee, M. (2023) 'Predicting energy consumption in residential buildings using advanced machine learning algorithms', *Energies*

Elhabyb, M. (2024) 'Machine learning algorithms for predicting energy consumption in educational buildings', *International Journal of Energy Research*

Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T. (2021) 'Temporal fusion transformers for interpretable multi-horizon time series forecasting', *International Journal of Forecasting*

Moghimi, S.M., Gulliver, T.A. and Thirumai Chelvan, I. (2024) 'Energy management in modern buildings based on demand prediction and machine learning
A review'

Poyyamozhi, M., Murugesan, B., Rajamanickam, N., Shorfuzzaman, M. and Aboelmagd, Y. (2024) 'IoT A promising solution to energy management in smart buildings: A systematic review, applications, barriers, and future scope', *Buildings*