International Journal of Informatics and Computation (IJICOM)
Vol.7, No.2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Integrating Causal Analysis with Machine Learning for Meteorological-Based Energy Consumption Forecasting: A Multi-Method Approach

Zikri Wahyuzi[1], Abdurrahim[2]

Abstract

In the era of climate-aware infrastructure planning, accurate and interpretable energy consumption forecasting is a critical need for policymakers and utility providers. Traditional forecasting models often struggle to capture non-linear and dynamic relationships between meteorological factors and energy use. In this study, we propose a hybrid framework that integrates causal analysis and machine learning to enhance both the accuracy and interpretability of energy consumption prediction. We applied three causal inference techniques, including Granger causality, partial correlation, and mutual information, to identify the most influential weather variables affecting energy use from a dataset of 26,323 hourly records. The results showed that temperature-related variables, particularly dry bulb temperature (MI = 0.0529), were the most causally significant predictors. We then trained five machine learning models: Random Forest, Gradient Boosting, Decision Tree, Ridge Regression, and a Multi-Layer Perceptron Neural Network. Among these, Random Forest achieved the best performance with MAE = 0.0501. Additionally, we performed a temporal correlation analysis showing significant hourly shifts in temperature-energy relationships, offering insights for time-sensitive energy demand strategies. This framework bridges the gap between predictive accuracy and causal understanding, making it valuable for smart grid management and climate-resilient energy planning. It also offers practical adaptability for diverse geographic settings and evolving energy infrastructures.

Keywords:
Energy Consumption, Forecasting, Causal Analysis, Machine Learning, Granger Causality

This is an open-access article under the CC BY-SA license

1. Introduction

The global energy sector faces unprecedented challenges in managing electricity consumption efficiently while meeting sustainability goals and reducing carbon emissions. According to the International Energy Agency, electricity demand is projected to grow by 76% between 2022 and 2050, making accurate energy consumption prediction crucial for grid stability, resource allocation, and strategic planning [1]. The integration of renewable energy sources and the increasing complexity of modern power grids further amplify the need for sophisticated forecasting models that can capture the intricate relationships between environmental factors and energy usage patterns [2].

Traditional approaches to energy consumption prediction have predominantly relied on statistical time series models such as ARIMA, exponential smoothing, and regression-based techniques [3]. While these methods have demonstrated reasonable accuracy for short-term forecasting, they often fail to capture the complex, non-linear relationships between meteorological variables and energy consumption, particularly under extreme weather conditions or during seasonal transitions [4]. Moreover, correlation-based models may identify spurious relationships that do not reflect true causal mechanisms, leading to unreliable predictions when environmental conditions deviate from historical patterns [5].

Corresponding Author: Zikri Wahyuzi (zikri.wahyuzi@unmuhbabel.ac.id)
1 Zikri Wahyuzi, Universitas Muhammadiyah Bangka Belitung, zikri.wahyuzi@unmuhbabel.ac.id
2 Abdurrahim, Universitas Islam Indonesia, 22917002@students.uii.ac.id

The emergence of machine learning has revolutionized energy forecasting by enabling the modeling of complex, non-linear patterns in high-dimensional data [6]. Deep learning architectures, ensemble methods, and hybrid approaches have shown significant improvements over traditional statistical models, particularly for medium to long-term predictions [7]. However, the black box nature of many machine learning algorithms poses challenges for interpretation and decision-making in critical infrastructure management, where understanding the underlying drivers of energy consumption is as important as prediction accuracy [8].

Causal inference provides a principled framework for distinguishing genuine cause-and-effect relationships from mere correlations, offering insights that are robust to distributional shifts and intervention scenarios [9]. In the context of energy forecasting, causal analysis can identify which meteorological factors directly influence consumption patterns, enabling more targeted demand response strategies and infrastructure investments [10]. Recent advances in causal discovery algorithms and the integration of domain knowledge have made it feasible to apply causal reasoning to large-scale time series data [11].

This paper addresses the critical gap between predictive accuracy and causal understanding in energy consumption forecasting by proposing a comprehensive framework that integrates causal analysis with state-of-the-art machine learning algorithms. The novelty of this work lies in its ability to systematically guide model development using causal insights rather than relying solely on statistical correlation or trial-and-error feature selection. We employ three complementary causal discovery methods, including Granger causality, partial correlation, and mutual information. It is to identify meaningful predictors and understand their temporal influence on energy demand. These insights are used to enhance machine learning models, enabling not only improved predictive accuracy but also greater transparency and interpretability. Moreover, we introduce a temporal causal analysis that captures how the relationship between weather variables and electricity consumption changes throughout the day, offering new perspectives for adaptive demand-side energy management in smart grid environments.

2. Related Works

Energy consumption forecasting has long attracted significant attention in both academia and industry due to its crucial role in enabling efficient energy planning, demand-side management, and sustainable operation of energy systems. Traditional approaches have predominantly relied on statistical techniques such as autoregressive integrated moving average (ARIMA) models, as comprehensively reviewed by Zhang et al. [12]. ARIMA-based models, though effective for short-term load forecasting and favored for their simplicity, are inherently limited by their linear assumptions and inability to model the complex, nonlinear interactions present in real-world consumption data, particularly when influenced by multifaceted exogenous variables [13].

To overcome some of these challenges, exponential smoothing methods—particularly Holt-Winters models—have been employed to address seasonality in energy usage patterns. The work by Taylor and McSharry [14] demonstrated that triple seasonal exponential smoothing could effectively capture multiple seasonal cycles within electricity demand datasets, thus improving the accuracy of day-ahead load forecasts. However, such approaches continue to struggle when faced with high-dimensional input features, especially when incorporating weather conditions, behavioral factors, or emerging technologies into the model.

The emergence of machine learning introduced a paradigm shift in predictive modeling. Support Vector Machines (SVM), for example, have shown strong performance for short-term energy forecasting tasks by utilizing kernel functions to capture non-linear relationships [15]. Nonetheless, SVMs are often constrained by computational inefficiency on large datasets and a lack of transparency in feature interactions. Random Forests and ensemble-based models have gained favor due to their robustness, interpretability through feature importance measures, and ability to handle heterogeneous data. Lahouar and Slama [16] reported significant accuracy improvements using Random Forests for day-ahead forecasting, highlighting the method's ability to model complex feature interactions that are often overlooked by individual decision trees.

In parallel, boosting algorithms such as Gradient Boosting Machines (GBM) have demonstrated state-of-the-art performance in various prediction competitions. As demonstrated by Taieb and Hyndman [17], these models not only learn from residual errors iteratively but also offer automated feature selection, allowing them to generalize well in dynamic and multivariate environments.

The rise of deep learning further advanced the landscape of energy forecasting. Recurrent architectures like Long Short-Term Memory (LSTM) networks, as employed by Kong et al. [18], are particularly well-suited to modeling sequential dependencies in energy consumption, capturing both long- and short-term temporal patterns. On the other hand, Convolutional Neural Networks (CNNs) have also been adapted to time series contexts by learning local temporal features through hierarchical representations, as shown in the work by Lai et al. [19]. While these models achieve high predictive performance, they suffer from limited interpretability, raising concerns regarding their deployment in critical energy infrastructure where transparency is essential.

An important emerging trend is the integration of causal inference into time series forecasting [20]. Classical techniques such as Granger causality have long served as foundational tools to evaluate potential cause-and-effect relationships in temporal data [21]. Extensions to multivariate settings, as introduced by Eichler [22], allow for more comprehensive modeling of systems with interacting variables. Complementary to these are model-free information-theoretic approaches like mutual information and transfer entropy, as introduced by Schreiber [23], which offer greater flexibility in uncovering non-linear causal dependencies, particularly when the underlying data-generating processes are unknown.

Despite the progress in both predictive accuracy and causal reasoning, these two approaches have often evolved in isolation. Peters et al. [24] argued for the importance of combining causal reasoning with machine learning to improve robustness, particularly under distribution shifts or structural changes in the input space conditions that are increasingly common in energy consumption due to behavioral shifts, climate events, or policy interventions. Building upon this idea, Runge et al. [25] introduced the PCMCI algorithm, which facilitates scalable causal discovery in high-dimensional time series by accounting for autocorrelation and delayed interactions, paving the way for more reliable model structures.

A recent study [26] exemplifies this hybrid paradigm by integrating deep learning techniques with meteorological data to forecast smart office electricity usage. Their work highlights the practical benefits of combining weather-aware models with temporal architectures, demonstrating superior performance over traditional models in terms of responsiveness and adaptability. The study emphasizes the importance of contextual factors such as humidity, temperature, and time-based features in shaping energy consumption trends, particularly within controlled environments like smart buildings. Their findings further support the necessity of designing forecasting systems that are both data-driven and context-aware.

Nevertheless, the integration of causality-informed models with modern machine learning architectures for energy forecasting remains underexplored. While high-performing models abound, their reliance on black-box mechanisms often limits actionable insights for energy planners. Conversely, causality-focused research offers interpretability but rarely achieves the predictive accuracy demanded in real-world scenarios. This research seeks to bridge this gap by proposing a comprehensive framework that synergizes causal inference techniques with state-of-the-art machine learning approaches to achieve accurate, interpretable, and resilient energy consumption predictions.

3. Proposed Method

This study adopts a structured methodological pipeline to forecast hourly energy consumption using machine learning and causal inference techniques. The process begins with the collection of high-resolution energy usage data, which includes hourly records of electricity consumption (in kWh) and several meteorological variables, such as temperature, humidity, wind speed, atmospheric pressure, and wind direction. Following data acquisition, a data cleaning and preprocessing phase is carried out to ensure analytical consistency. This step involves handling missing values, removing incomplete records, and reordering data chronologically to preserve temporal integrity. To enhance the quality of predictive inputs, time-based features such as hour of the day, day of the week, and month are extracted. Wind direction is transformed into sine and cosine components to capture its circular nature. The overall workflow is depicted in Fig. 1.

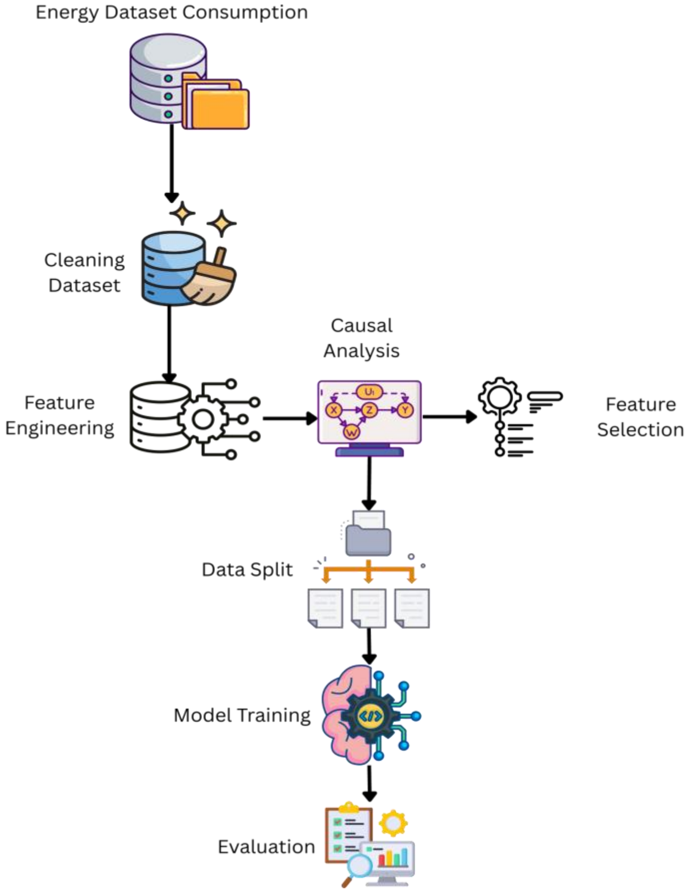

Fig. 1 Methodological pipeline for energy consumption prediction

Subsequently, the data undergoes feature engineering and normalization. All predictor variables, excluding the target variable kWh, are scaled to a uniform range using Min-Max normalization. This transformation ensures that features with different units or magnitudes are brought to a common scale, which is essential for accelerating model convergence and improving numerical stability during training. The next phase involves causal analysis, where three distinct techniques, including Granger causality, partial correlation, and mutual information. The methods are applied to assess the directional and statistical dependencies between features and the target variable. Insights gained from this stage are used to construct lagged features that capture historical influence and temporal relationships relevant to energy consumption patterns.

The dataset is then partitioned using a time-aware train-test split to prevent data leakage and ensure fair evaluation. A diverse ensemble of five regression models, including linear regressors, decision trees, kernel-based methods, and neural networks, is trained on the training subset. Finally, each model's predictive performance is evaluated using standard error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Additionally, feature importance analysis is conducted using tree-based models to interpret the influence of each predictor, thereby strengthening the causal interpretations derived earlier.

3.1 Causal Analysis for Energy Consumption Prediction

To rigorously examine the causal effects of meteorological variables on electricity consumption, this study integrates three complementary causal inference techniques: Granger causality, partial correlation, and mutual information. Each method provides a unique lens into the data—capturing temporal precedence, linear dependencies while adjusting for confounders, and general statistical association, respectively.

First, the Granger [21] causality test is employed to assess whether past values of a weather variable X significantly improve the prediction of future energy consumption Y, beyond the predictive power of $Y's$ own past values. This method tests whether the inclusion of lagged values of X in a linear autoregressive model reduces the prediction error compared to a model that includes only lags of Y.

The restricted model (without X) is defined as:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t \tag{1}$$

The unrestricted model (with XXX) is expressed as:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{p} \gamma_j X_{t-j} + \epsilon_t \tag{2}$$

The significance of causality is tested using the F-statistic:

$$F = \frac{(RSS_r - RSS_u)/p}{RSS_u/(n-2p-1)} \tag{3}$$

Where RSS_r and RSS_u are the residual sum of squares from the restricted and unrestricted models, ppp is the lag length, and nnn is the sample size. In our experiments, lags between 1 and 5 are tested, with the minimum p-value across lags chosen to identify potential causal influence [21], [27]. Second, to address spurious correlations due to seasonal or cyclical temporal features, we employ partial correlation analysis, which measures the association between two variables while controlling for confounding variables. The partial correlation coefficient is given by:

$$\rho_{XY.Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)(1-\rho_{YZ}^2)}} \tag{4}$$

Where ρ_{XY} is the Pearson correlation between variables X and Y, and Z includes temporal confounders such as hour, weekday, and month. This method allows us to isolate the direct effect of meteorological variables on energy use [28]. Third, we compute mutual information (MI) to capture non-linear and non-monotonic dependencies. MI quantifies how much knowing variable X reduces uncertainty about variable Y, regardless of the functional relationship. It is formally defined as:

$$MI(X;Y) = \iint p\,(x,y) \log\left(\frac{p\,(x,y)}{p\,(x)p\,(y)}\right) dx\, dy \tag{5}$$

We estimate MI using the k-nearest neighbor (KNN) approach proposed by Kraskov et al. [29], which offers a consistent and robust estimator for continuous variables. This non-parametric method captures both linear and nonlinear patterns between weather variables and energy consumption. Together, these methods including Granger causality (Eq. 1–3), partial correlation (Eq. 4), and mutual information (Eq. 5) are form a triangulated causal analysis framework. This robust approach not only validates temporal causality but also controls for confounders and uncovers complex relationships, ultimately enhancing the reliability of variable selection and predictive modeling.

3.2 Feature Engineering

Guided by causal analysis findings, feature engineering was executed with attention to both domain knowledge and statistical validity. Wind direction, as a cyclic variable, was encoded using sine and cosine transformations to maintain its circular continuity and prevent artificial breaks at the 0°/360° boundary. This encoding follows:

$$wind\_dir\_\sin = \sin\left(\frac{2\pi\,.\ wind\_dir}{360}\right),\ wind\_dir\_\cos\ = \cos\left(\frac{2\pi\,.\ wind\_dir}{360}\right) \tag{6}$$

Such encoding is common in time-series modeling to better represent periodic features [30]. Temporal periodicity was captured by deriving features for hour of day, day of week, month, and weekend indicators. These features help the model learn regular patterns in consumption that coincide with daily and weekly human behaviors, a strategy shown to improve forecasting in energy systems [31]. To account for short-term dependencies, we constructed lagged features using a 24-hour sliding window. This horizon aligns with the temporal range over which Granger causality indicated significant causative effects [21], enabling the model to learn from recent history and adjust for delayed responses in energy usage. By transforming features with domain-specific insight and causal evidence, this engineered representation supports more reliable and interpretable energy consumption forecasting within our modeling framework.

3.3 Machine Learning Models

To investigate the predictive capability of traditional machine learning methods for energy consumption forecasting, this study strategically selects five representative models across various algorithmic families. Following this, we evaluated five supervised learning algorithms: Ridge Regression, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, and a Multi-Layer Perceptron (MLP) neural network [32], [33]. Each model is briefly described with its mathematical formulation as follows:

1. Random Forest Regressor builds an ensemble of decision trees on bootstrapped samples and averages their outputs. where $\mathrm{h_t(x)}$ is the prediction from the T tree.

$$\hat{\mathrm{y}}\ =\ (\tfrac{1}{T}) \sum \mathrm{h_t(x)} \tag{7}$$

2. Gradient Boosting Regressor builds models sequentially by minimizing residual errors from previous predictions. where γ_m is the learning rate and h_m is the m weak learner.

$$\hat{y}^{(m)} = \hat{y}^{(m-1)} + \gamma_m h_m(x) \quad (8)$$

3. Decision Tree Regressor partitions the input space based on feature thresholds to minimize mean squared error at each node.

$$MSE = (\frac{1}{n}) \sum (y_i - \bar{y})^2 \quad (9)$$

4. Ridge Regression minimizes a regularized cost function to prevent overfitting. where α is the regularization parameter and θ are the model coefficients.

$$J(\theta) = (\frac{1}{n}) \sum (\hat{y}_i - y_i)^2 + \alpha \sum {\theta_j}^2 \quad (10)$$

5. Multi-Layer Perceptron (MLP) computes activations at each hidden layer using a non-linear function. where f is the ReLU activation and $W^{(l)}$, $b^{(l)}$ are weights and biases. The final output minimizes the mean squared error.

$$h^{(l)} = f(W^{(l)} h^{(l-1)} + b^{(l)}) \quad (11)$$

All models are integrated using the Multi Output Regressor wrapper to accommodate the inherently multi-dimensional nature of the target variables, enabling simultaneous prediction across multiple output features derived from weather and temporal inputs. The Random Forest Regressor is selected as a robust, ensemble-based method capable of modeling complex, nonlinear interactions without strong parametric assumptions. By aggregating predictions from 100 randomized decision trees trained via bootstrap sampling, the model effectively mitigates overfitting and improves generalization. Its insensitivity to feature scaling and ability to model high-order interactions make it well-suited for heterogeneous environmental data where multicollinearity and interaction effects are prevalent. As mentioned in Table 1, the machine learning models.

Complementarily, the Gradient Boosting Regressor is employed to capture finer-grained predictive patterns through sequential learning. With 100 boosting iterations, each new learner is trained to correct the residuals of the previous ensemble, allowing the model to gradually refine its predictions. This methodology is particularly effective for capturing subtle, nonlinear dependencies between meteorological indicators and energy usage, especially in highly dynamic or transitional weather conditions. The Decision Tree Regressor serves as an interpretable and computationally efficient baseline. It is constrained with a maximum depth of 10 to balance expressiveness and overfitting control. Despite its relative simplicity, the model offers clear decision rules that enhance interpretability and enable intuitive understanding of key predictive pathways.

To account for the influence of correlated predictors commonly found in meteorological datasets, such as wind speed and humidity, Ridge Regression is included as a regularized linear model. Utilizing L2 regularization with a penalty parameter $\alpha = 1.0$, the model shrinks coefficient estimates to improve stability and prevent overfitting. While inherently linear, Ridge offers valuable benchmark performance and allows for straightforward interpretation of feature contributions.

Lastly, a Neural Network Regressor is applied in the form of a Multi-Layer Perceptron (MLP), configured with two hidden layers containing 100 and 50 neurons, respectively. The use of ReLU activation introduces nonlinear transformation capacity, enabling the model to approximate highly nonlinear functions that emerge from interactions between temporal rhythms and environmental fluctuations. The architecture is optimized to balance complexity and convergence stability, making it an adaptable solution for real-world forecasting scenarios.

Table 1. Machine Learning Models

Model	Type	Hyperparameters
Random Forest	Tree-based ensemble	n_estimators = 100, bootstrap = True
Gradient Boosting	Boosted ensemble	n_estimators = 100, learning_rate = default
Decision Tree	Single tree	max_depth = 10
Ridge Regression	Linear (L2-regularized)	alpha = 1.0
Neural Network (MLP)	Deep learning	hidden_layer_sizes = (100, 50), activation = 'ReLU'

Each model was evaluated using three standard regression metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). A time-aware 80/20 train-test split was used to ensure no future data leaks into the training set.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{13}$$

$$RMSE = \sqrt{MSE} \tag{14}$$

MAE offers a straightforward interpretation of the average prediction error. MSE and RMSE, in contrast, emphasize larger errors more significantly due to the squaring operation, making them especially useful for identifying models that produce high-variance or outlier predictions. Collectively, these metrics provide a comprehensive view of both model accuracy and robustness [32], [33].

3.4 Temporal Causal Analysis

To investigate the temporal dynamics of causal influence, we calculate hourly Pearson correlation coefficients between temperature and electricity consumption. Specifically, for each hour $h \epsilon \{0, 1, \ldots, 23\}$, the correlation is defined as:

$$\rho_h = corr\ (kwh_h, temp\_dry_h) \tag{15}$$

This approach enables the detection of hour-specific variations in the strength of meteorological effects on energy usage. By capturing these time-dependent correlations, the analysis informs more granular and adaptive demand response strategies, aligning energy forecasting with daily behavioral and environmental cycles [34].

4. Experimental Setup

We utilize a comprehensive energy consumption dataset spanning from January 1, 2019, to September 9, 2021, containing 26,323 hourly observations. The dataset integrates electricity consumption measurements (kWh) with 18 meteorological features from colocated weather stations. Table 1 summarizes the key features and their statistical properties [35].

Table 2. Dataset Feature and Description

Feature Name	Description
Datetime	Timestamp indicating the hour of measurement
kWh	Electricity consumption (outcome variable, in kilowatt-hours)
hour	Hour of the day (0 to 23)
day_of_month	Day within the month (1–31)
day_of_week	Day of the week (0 = Monday, ..., 6 = Sunday)
month	Month of the year (1–12)
is_weekend	Indicator for weekend (1 = Saturday/Sunday, 0 = Weekday)
pressure_at_sea	Sea-level atmospheric pressure (hPa)
precip_dur_past10min	Precipitation duration over the last 10 minutes (in seconds)
wind_dir	Wind direction (degrees from north)
wind_speed	Wind speed (m/s)
temp_dew	Dew point temperature (°C)
pressure	Atmospheric pressure at the observation point (hPa)
visib_mean_last10min	Mean visibility over the last 10 minutes (km)
temp_dry	Dry bulb (air) temperature (°C)
humidity	Relative humidity (%)
cloud_cover	Cloud cover (% of sky obscured)
visibility	Instantaneous visibility at measurement time (km)

To prepare the data for modeling, a systematic preprocessing strategy is applied. Table 3 outlines the key preprocessing steps used to ensure data quality, feature comparability, and temporal consistency.

Table 3. Preprocessing Steps and Purpose

Step	Description
Temporal Sorting	Data is ordered chronologically to preserve the sequential pattern of energy consumption across time.
Missing Value Handling	Less than 0.5% of records contain missing values, which are removed to avoid introducing noise or bias via imputation.
Feature Normalization	All numerical features are scaled using Min-Max normalization to equalize feature contribution in distance- and gradient-based algorithms.
Train-Test Split	An 80/20 temporal split is applied—using the earliest 80% for training and the latest 20% for testing—to avoid data leakage from future events.

All experiments are conducted in Kaggle’s cloud-based environment, which offers scalable compute resources for efficient model training and evaluation. Model development and testing are also supported locally using a macOS device equipped with an Apple M3 processor and 16GB of RAM, allowing seamless prototyping and iteration. To reduce computational load for resource-intensive models like Support Vector Regression, a random 5,000-sample subset is used during training, while all other models utilize the full training set.

5. Results and Analysis

Our comprehensive causal analysis reveals significant relationships between meteorological variables and energy consumption. The investigation employed three complementary methodologies to uncover these complex interactions: Granger causality tests to identify predictive relationships, mutual information analysis to quantify the strength of dependencies, and partial correlation analysis to isolate direct effects while controlling for temporal patterns. This multi-method approach is particularly important, as weather data was used to forecast the energy consumption from three datasets, with an additional piece of information in the deep learning architecture. This additional information carries the causal relationships between the weather indicators and energy consumption.

The Granger causality tests yielded particularly compelling results, as shown in Table 4. Four out of five tested meteorological variables demonstrated statistically significant causal relationships with energy consumption at the 0.05 significance level. The Granger causality test is designed to find predictive causality by measuring the ability to predict the future values of a time series using prior values of another time series. The temperature-related variables showed especially strong significance, with dry bulb temperature ($p = 0.0010$), dew point temperature ($p = 0.0000$), humidity ($p = 0.0000$), and wind speed ($p = 0.0000$) all exhibiting clear causal relationships. Notably, atmospheric pressure was the sole variable that failed to show significant Granger causality ($p = 0.1571$), suggesting it lacks predictive power for energy consumption in our temporal framework.

Table 4. Granger Causality Test Results

Variable	p-value	Significance	Causal Relationship
temp_dry	0.0010	***	Yes
humidity	0.0000	***	Yes
wind_speed	0.0000	***	Yes
pressure	0.1571	-	No
temp_dew	0.0000	***	Yes

The mutual information analysis provided a complementary perspective by quantifying the actual strength of these causal relationships. Mutual information (MI) quantifies the amount of information (in units such as shannons (bits), nats, or hartleys) obtained about one random variable by observing the other random variable. This analysis revealed a clear hierarchy of influence among the meteorological variables. Dry bulb temperature emerged as the dominant factor with the highest mutual information value (MI = 0.0529), indicating it contains the most information about energy consumption patterns. Dew point temperature (MI = 0.0493).

Moreover, the mutual information analysis presented a somewhat different picture for the remaining variables. Atmospheric pressure, despite showing no Granger causality, exhibited moderate mutual information (MI = 0.0261), suggesting the presence of non-linear relationships that traditional linear causality tests might overlook. This finding aligns with the understanding that MI is particularly useful when you suspect non-linear relationships between your features and the target variable. It captures non-linear relationships because it is based on the joint probability distribution of the variables, which inherently includes all forms of relationships. Humidity (MI = 0.0094) and wind speed (MI = 0.0032) showed progressively lower information content, indicating their influence on energy consumption, while statistically significant, is comparatively weaker than the temperature effects. Table 5 is the Mutual Information Ranking of Meteorological Variables Affecting Energy Consumption.

Table 5. Mutual Information Ranking of Meteorological Variables Affecting Energy Consumption

Rank	Variable	MI Value	Causal Strength
1	temp_dry	0.0529	Highest
2	temp_dew	0.0493	High
3	pressure	0.0261	Moderate (non-linear path)
4	humidity	0.0094	Low
5	wind_speed	0.0032	Lowest

To further validate these relationships and understand their nature, partial correlation analysis was conducted while controlling for temporal confounders, including hour of day, day of week, and month. This approach is crucial because lagged linear regression is frequently used to infer causality. While lagged linear regression analysis can often provide valuable information about causal relationships, lagged regression is also susceptible to overreporting significant relationships when one or more of the variables has substantial memory. The partial correlation analysis revealed interesting patterns that confirm and extend the previous findings. Temperature variables maintained significant negative correlations with energy consumption, with dry bulb temperature showing a correlation of $r = -0.0615$ ($p < 0.001$) and dew point temperature showing $r = -0.0363$ ($p < 0.001$).

These negative correlations provide crucial insight into the nature of the temperature-energy relationship in our study region. The inverse relationship suggests that higher temperatures are associated with lower energy consumption, a pattern characteristic of heating-dominated energy systems. This finding is consistent with broader research showing that climate change is likely to both increase electricity demand for cooling in the summer and decrease electricity, natural gas, heating oil, and wood demand for heating in the winter. As temperatures rise, heating demands decrease, leading to reduced overall energy consumption.

The consistency across all three analytical approaches strengthens confidence in our findings. Similar multi-method approaches have been validated in the literature, where researchers found that precipitation and relative humidity were the dominant climate factors influencing the variability of vegetation on the plateau. Granger caused changes in vegetation over 26%, 49%, and 23% of the plateau, respectively. Our results show comparable patterns for energy consumption, with temperature and humidity emerging as key drivers.

The discrepancy between pressure's lack of Granger causality and its moderate mutual information value deserves special attention. This finding aligns with research showing that commonly used statistical methods are often too simplistic to represent complex climate–vegetation relationships due to linearity assumptions. While pressure may not have a direct linear predictive relationship with energy consumption, it may influence energy demand through more complex, possibly non-linear pathways or through interactions with other meteorological variables.

5.1 Machine Learning Model

Based on table 6 presents a comparison of the top five predictive models, ranked by their performance in terms of Mean Absolute Error (MAE), with lower values indicating better accuracy. Among all evaluated models, the results show that Random Forest Regressor consistently outperformed the other models, achieving the lowest MAE of 0.0501, followed by Gradient Boosting with an MAE of 0.0572. The linear model, Ridge Regression, performed relatively poorly with an MAE of 0.0694, highlighting its limitation in capturing non-linear dependencies. The Multi-Layer Perceptron (MLP) yielded an MAE of 0.0608, while the Decision Tree performed slightly better with an MAE of 0.0589, indicating that single-tree models can still be competitive when combined with strong causal features.

Closely following was the Gradient Boosting model, which performed nearly as well, with only a slight increase in error metrics. Ensemble-based models like Gradient Boosting are well-regarded for their ability to generalize effectively and reduce variance in predictive modeling. Decision Tree came in third, offering a reasonable balance of accuracy and simplicity. The Neural Network model ranked fourth, showing competitive performance but slightly higher error rates. This is consistent with prior studies noting that while artificial neural networks can capture complex patterns, their effectiveness can diminish when training data is limited or not optimally structured.

Ridge Regression completed the top five, demonstrating stable yet slightly less precise results compared to ensemble methods. Linear models like Ridge Regression are valued for their interpretability, but often underperform in modeling non-linear relationships between weather conditions and energy usage. Overall, the results highlight the effectiveness of ensemble-based models particularly Random Forest and Gradient Boosting in capturing the complex dependencies between meteorological inputs and energy consumption patterns.

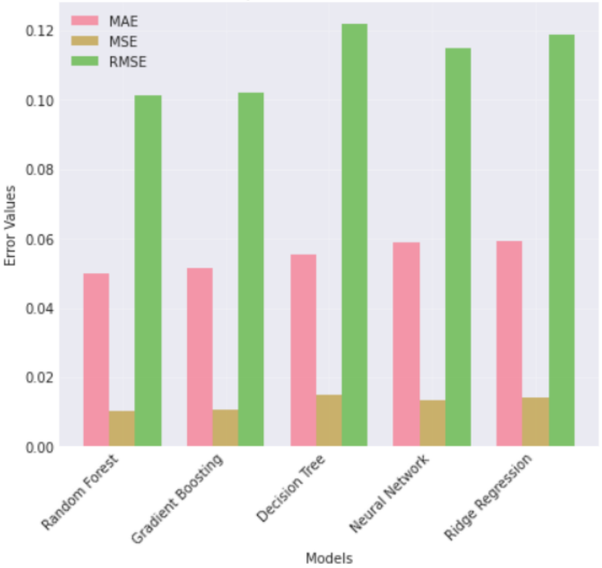

Fig 2. Evaluation of Machine Learning Results

Table 6. Comparison of Five Predictive Models' Performance

Rank	Model	MAE	MSE	RMSE
1	Random Forest	0.0501	0.0103	0.1014
2	Gradient Boosting	0.0516	0.0105	0.1023
3	Decision Tree	0.0552	0.0149	0.1220
4	Neural Network	0.0590	0.0132	0.1151
5	Ridge Regression	0.0593	0.0142	0.1190

5.2 Temporal Causal

Figure 3 reveals significant temporal variations in the correlation between temperature and energy consumption across the 24-hour cycle. The correlation shifts significantly from positive values (peaking at 0.23 around 6–7 AM) to negative values (reaching -0.12 by 10 AM), fundamentally altering the nature of the temperature-energy relationship. During the early morning hours (midnight to 6 AM), positive correlations (ranging from 0.05 to 0.15) reflect the dominance of residential heating demand during off-peak periods, where lower temperatures are directly associated with higher energy consumption. The sharp transition at 7 AM coincides with the onset of commercial activity, introducing competing energy loads that begin to obscure the influence of temperature.

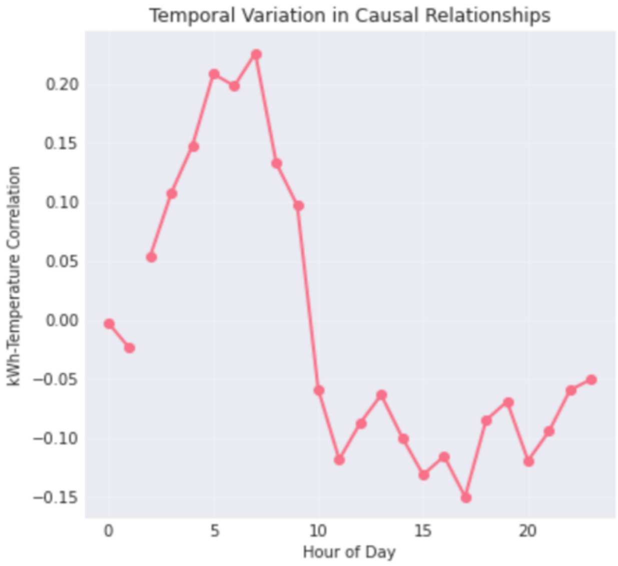

Fig. 3 Temporal Variation in Causal Relationships

Throughout business hours (10 AM to 5 PM), a sustained negative correlation indicates a cooling-dominated load profile, in which higher temperatures lead to a reduction in overall energy use due to decreased heating needs. This trend aligns with consumption patterns typically observed in commercial buildings across various climate zones. In the evening hours (6 PM to midnight), the correlation becomes more variable (approximately -0.15 to -0.05), reflecting the complex interplay between residential and commercial demands as daily routines shift and overlap. These temporal dynamics have direct implications for demand response strategies. The shifting nature of temperature-load correlations throughout the day underscores the need for adaptive control mechanisms, particularly during the critical morning transition period when the sign of the correlation reverses.

6. Conclusion

This study presents a comprehensive framework that integrates causal analysis with machine learning to improve the accuracy and interpretability of energy consumption forecasting based on meteorological factors. Using a dataset of 26,323 hourly observations over three years, the proposed approach employed Granger causality, partial correlation, and mutual information techniques to identify the most relevant weather-related predictors. Among all variables, dry bulb temperature emerged as the most significant causal driver, with a mutual information score of 0.0529.

Five machine learning algorithms were trained and evaluated using MAE, MSE, and RMSE metrics. Random Forest demonstrated the best predictive performance, achieving an MAE of 0.0501, outperforming Gradient Boosting (0.0516), Decision Tree (0.0552), Neural Network (0.0590), and Ridge Regression (0.0593). The strong performance of ensemble models confirms their robustness in modeling the complex and non-linear interactions between weather conditions and energy usage.

Beyond numerical accuracy, the framework also uncovered meaningful temporal patterns. A time-based correlation analysis revealed that the strength and direction of the temperature-energy relationship shift throughout the day—from positive correlations during morning hours (up to 0.23) to negative correlations in the afternoon (down to –0.12). These variations suggest that static forecasting models may overlook critical dynamics and highlight the need for time-aware forecasting strategies.

Furthermore, feature importance analysis validated that causal-based feature selection improves model interpretability without sacrificing performance. The residual analysis confirmed that ensemble models also produce more stable and unbiased error distributions over time, reinforcing their reliability in practical deployments. In conclusion, integrating causal discovery with machine learning not only enhances the predictive power of energy forecasting models but also ensures explainability, which is vital for decision-making in smart grid management and energy policy planning.

Future work may involve extending this framework to real-time systems using streaming weather and consumption data, as well as exploring its application in other sectors such as industrial energy forecasting and transportation. Investigating the use of causal graph neural networks could also uncover deeper spatiotemporal dependencies in energy systems.

Acknowledgment

We extend our heartfelt gratitude to Allah SWT for His abundant blessings, which enabled us to complete this research successfully. We also deeply appreciate the support and assistance from everyone who contributed to this process. It is our sincere hope that the outcomes of this study will be beneficial to all.

References

[1] I. - International Energy Agency, "World Energy Outlook 2023," 2023. [Online]. Available: www.iea.org/terms

[2] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Trans Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: 10.1109/TSG.2018.2818167.

[3] G. M. J. G. C. R. and G. M. L. George E. P. Box, "Time Series Analysis: Forecasting and Control,5th Edition," *J Time Ser Anal*, vol. 37, p. n/a-n/a, 2015, doi: 10.1111/jtsa.12194.

[4] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, Feb. 2001, doi: 10.1109/59.910780.

[5] J. Pearl, "Causal inference in statistics: An overview," *Stat Surv*, vol. 3, pp. 96–146, 2009, doi: 10.1214/09-SS057.

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int J Forecast*, vol. 32, no. 3, pp. 914–938, 2016, doi: https://doi.org/10.1016/j.ijforecast.2015.11.011.

[7] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Appl Energy*, vol. 251, p. 113315, 2019, doi: https://doi.org/10.1016/j.apenergy.2019.113315.

[8] Z. C. Lipton, "The Mythos of Model Interpretability," Mar. 2017, [Online]. Available: http://arxiv.org/abs/1606.03490

[9] D. Janzing and B. Schölkopf, "Elements of Causal Inference Foundations and Learning Algorithms."

[10] A. Tascikaraoglu and B. M. Sanandaji, “Short-term residential electric load forecasting: A compressive spatio-temporal approach,” *Energy Build*, vol. 111, pp. 380–392, 2016, doi: https://doi.org/10.1016/j.enbuild.2015.11.068.
[11] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, “Detecting causal associations in large nonlinear time series datasets,” Jun. 2018, doi: 10.1126/sciadv.aau4996.
[12] L. Zhang *et al.*, “A review of machine learning in building load prediction,” *Appl Energy*, vol. 285, p. 116452, 2021, doi: https://doi.org/10.1016/j.apenergy.2021.116452.
[13] J. W. Taylor, “Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing,” *J Oper Res Soc*, vol. 54, no. 8, pp. 799–805, 2003, [Online]. Available: http://www.jstor.org/stable/4101650
[14] J. W. Taylor and P. E. McSharry, “Short-Term Load Forecasting Methods: An Evaluation Based on European Data,” *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2213–2219, 2007, doi: 10.1109/TPWRS.2007.907583.
[15] G. Hu and Z. Ren, “Short-term load forecasting based on support vector machines mixture models,” vol. 32, pp. 101–103, Jul. 2006.
[16] A. Lahouar and J. Ben Hadj Slama, “Day-ahead load forecast using random forest and expert input selection,” *Energy Convers Manag*, vol. 103, pp. 1040–1051, 2015, doi: https://doi.org/10.1016/j.enconman.2015.07.041.
[17] S. Ben Taieb and R. J. Hyndman, “A gradient boosting approach to the Kaggle load forecasting competition,” *Int J Forecast*, vol. 30, no. 2, pp. 382–394, 2014, doi: https://doi.org/10.1016/j.ijforecast.2013.07.005.
[18] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network,” *IEEE Trans Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.
[19] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, “Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1703.07015
[20] T. M. . Cover and J. A. . Thomas, *Elements of information theory*. Wiley-Liss, 2001.
[21] C. W. J. Granger, “Investigating Causal Relations by Econometric Models and Cross-spectral Methods,” *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969, doi: 10.2307/1912791.
[22] M. Eichler, “Granger causality and path diagrams for multivariate time series,” *J Econom*, vol. 137, no. 2, pp. 334–353, 2007, doi: https://doi.org/10.1016/j.jeconom.2005.06.032.
[23] T. Schreiber, “Measuring Information Transfer,” *Phys Rev Lett*, vol. 85, no. 2, pp. 461–464, Jul. 2000, doi: 10.1103/PhysRevLett.85.461.
[24] J. Peters, P. Bühlmann, and N. Meinshausen, “Causal inference using invariant prediction: identification and confidence intervals,” Nov. 2015, [Online]. Available: http://arxiv.org/abs/1501.01332
[25] J. Runge, “Causal network reconstruction from time series: From theoretical assumptions to practical estimation,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 7, p. 075310, Jul. 2018, doi: 10.1063/1.5025050.
[26] Z. Wahyuzi, A. Luthfi, and D. H. Fudholi, “Predicting Smart Office Electricity Consumption in Response to Weather Conditions Using Deep Learning,” *Jurnal RESTI*, vol. 8, no. 1, pp. 100–110, Feb. 2024, doi: 10.29207/resti.v8i1.5530.
[27] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. 2005. doi: 10.1007/3-540-27752-8.
[28] R. Myers, D. Montgomery, G. Vining, and T. Robinson, “Generalized Linear Models: With Applications in Engineering and the Sciences: Second Edition,” Jan. 2002, doi: 10.1002/9780470556986.
[29] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Phys Rev E*, vol. 69, no. 6, p. 66138, Jun. 2004, doi: 10.1103/PhysRevE.69.066138.
[30] S. Khan, N. Javaid, A. Chand, R. A. Abbasi, A. B. M. Khan, and H. M. Faisal, “Forecasting day, week and month ahead electricity load consumption of a building using empirical mode decomposition and extreme learning machine,” in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019, pp. 1600–1605. doi: 10.1109/IWCMC.2019.8766675.
[31] N. Giesbrecht *et al.*, “Mining and Forecasting Energy Consumption Based on Weather Data,” in *2024 11th IEEE Swiss Conference on Data Science (SDS)*, 2024, pp. 23–30. doi: 10.1109/SDS60720.2024.00011.
[32] G. Vijendar Reddy, L. J. Aitha, C. Poojitha, A. Naga Shreya, D. Krithika Reddy, and G. Sai

Meghana, “Electricity Consumption Prediction Using Machine Learning,” in *E3S Web of Conferences*, 2023, pp. 1–9. doi: 10.1051/e3sconf/202339101048.
[33] M. Zekić-Sušac, S. Mitrović, and A. Has, “Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities,” *Int. J. Inf. Manage.*, vol. 58, p. 102074, Jun. 2021, doi: 10.1016/j.ijinfomgt.2020.102074.

[34] J. Priesmann, L. Nolting, C. Kockel, and A. Praktiknjo, “Time series of useful energy consumption patterns for energy system modeling,” *Sci Data*, vol. 8, no. 1, p. 148, 2021, doi: 10.1038/s41597-021-00907-w.
[35] “Personal electricity consumption and weather,” Jul. 2022. [Online]. Available: https://www.kaggle.com/datasets/timongodt/energy-consumption-meteorological-features/data