International Journal of Education, Science, Technology and Engineering
p-ISSN: 2685-1458, e-ISSN: 2684-9844

Original Research Paper

Anomaly Detection Using Autoencoders for Household Electricity Meters

Nattaporn Wongsuwan[1*], Somchai Srisawat[1], Thanakorn Kittisak[1], Anongrat Boonmee[1], Mirella Sanna[1]

[1] *Department of Information Engineering, Electrical Engineering and Applied Mathematics, Faculty of Engineering, University of Salerno. Fisciano, Italy.*
[2] *Department of Electrical Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi. Pathum Thani, Thailand.*

Article History
Received:
27.11.2024

Revised:
19.06.2025

Accepted:
03.07.2025

***Corresponding Author:**
Nattaporn Wongsuwan
Email:
wongsu.wan72@yahoo.com

Abstract: Household electricity consumption often exhibits sudden and unexplained spikes that typically go unnoticed until the monthly bill arrives. These anomalies may stem from equipment malfunction, inefficient appliance usage, or irregular electrical patterns that households cannot easily observe. This study proposes an unsupervised anomaly detection framework based on autoencoders to identify abnormal consumption behavior from high resolution household electricity meter data. The model learns normal consumption patterns through reconstruction and flags anomalies using a dynamic threshold derived from reconstruction error distribution. Experimental results demonstrate strong detection capability, particularly for sudden spikes, achieving a precision of 0.92, recall of 0.88, and F1 score of 0.90. The findings highlight the potential of deep learning–based unsupervised methods to support real time, edge deployable solutions for energy efficiency and early fault detection in residential environments.

Keywords: Adaptive Thresholding, Anomaly Detection Household, Autoencoder, Edge Computing, Energy Consumption.

1. Introduction

Electricity consumption within residential households represents a complex interplay of behavioral routines, appliance characteristics, and infrastructural conditions. Despite the central role of household electricity usage in national energy systems, most consumers possess limited visibility into their consumption dynamics [1]. This lack of granularity often results in unnoticed anomalies, which may manifest as unexpected surges in monthly bills and remain unexplained without advanced monitoring tools. Given the rising emphasis on sustainability and efficient energy use, detecting such irregularities has become both a practical and societal necessity [2] [3].

Sudden spikes in household electricity usage can emerge from a variety of sources, including malfunctioning appliances, poorly maintained electrical installations, or behavioral deviations from typical daily routines. These anomalies are often subtle in real time but accumulate substantial energy waste when undetected. Conventional monitoring methods typically limited to monthly billing statements provide insufficient temporal resolution for consumers to recognize abnormal usage patterns early [4]. This gap highlights the need for automated, real-time anomaly detection solutions.

Traditional statistical and rule-based methods, such as moving averages or fixed thresholds, struggle to adapt to the highly nonlinear and irregular nature of household electricity consumption. Residential patterns fluctuate according to lifestyle shifts, weather changes, occupancy levels, and appliance usage schedules [5] [6]. These diverse influences produce a degree of complexity that renders rigid models inadequate. Thus, an approach capable of learning from the underlying structure of normal behavior is essential for reliable anomaly detection [7].

Deep learning models, particularly autoencoders, offer a promising alternative by learning compact representations of normal consumption patterns without requiring labeled anomaly data. By compressing and reconstructing input signals, autoencoders detect anomalies by measuring deviations in reconstruction error. This unsupervised capability is critical in household energy contexts, where collecting labeled anomalies is highly impractical due to the unpredictable and varied nature of abnormal events [8].

The emergence of smart meters and IoT-enabled monitoring devices has significantly expanded the availability of high-resolution electricity data. These technologies enable measurements at intervals as fine as minutes or seconds, providing rich temporal information suitable for deep learning models [9] [10]. Leveraging such granular datasets allows autoencoders to capture the nuanced rhythms of household energy dynamics, improving sensitivity to both sudden and subtle anomalies [11].

The household environment presents unique challenges that differentiate it from industrial or commercial settings. Noise, irregular usage patterns, incomplete data, and inconsistent appliance behavior introduce complexities that may affect model performance. Existing studies often focus on structured environments where consumption patterns are more predictable. As a result, there remains a need for dedicated research addressing anomaly detection within the inherently unpredictable household domain [12].

Beyond technical considerations, anomaly detection in household electricity consumption carries broader socio-economic implications. Early identification of irregular usage can reduce household expenditures, promote energy-efficient behaviors, and assist utility providers in minimizing operational inefficiencies. In this sense, anomaly detection systems can serve as tools for both consumer empowerment and infrastructural optimization [13] [14].

Energy sustainability goals further reinforce the importance of accurate anomaly detection. Governments, regulatory bodies, and energy providers increasingly advocate for digital transformation initiatives aimed at promoting transparency and reducing waste. Incorporating AI-based monitoring systems within residential contexts aligns with these policy objectives by supporting data-driven decision-making and fostering more responsible energy consumption practices [15].

This study addresses these opportunities and challenges by proposing an autoencoder-based framework optimized for household electricity consumption. The model is designed to learn normal usage behavior, detect deviations through reconstruction error analysis, and operate effectively under realistic residential conditions. By using an adaptive thresholding mechanism, the framework enhances detection sensitivity while minimizing false alarms [16] [17].

Overall, this research contributes to the growing field of AI-driven energy analytics by demonstrating the viability of unsupervised autoencoder models for detecting anomalies in household electricity data. Its findings provide a foundation for developing practical, real-time solutions capable of enhancing energy efficiency, reducing waste, and supporting the evolution of intelligent residential energy systems.

2. Literature Review

2.1. Traditional Approaches to Energy Anomaly Detection

Traditional approaches to energy anomaly detection rely heavily on statistical modeling techniques that assume regularity and stationarity in consumption patterns. Methods such as Moving Average, Exponential Smoothing, and Seasonal-Trend Decomposition of Time Series (STL) are widely used to approximate expected values and identify deviations. These approaches are effective in controlled environments but tend to struggle with chaotic or highly variable household electricity usage [18].

One of the earliest techniques employed in anomaly detection is the use of rule-based thresholding, where consumption exceeding predefined limits is flagged as abnormal. Although simple to implement, threshold-based systems lack adaptability, especially when user behavior or environmental conditions change over time. This often leads to high rates of false alarms or missed anomalies [19] [20].

Autoregressive Integrated Moving Average (ARIMA) models have been used extensively for predicting energy consumption due to their strength in modeling linear temporal dependencies. However, ARIMA assumes stationarity and struggles with non-linear dynamics that are common in residential energy usage, making it insufficient for detecting anomalies arising from abrupt behavioral changes [21].

Seasonal ARIMA (SARIMA) extends ARIMA by incorporating seasonal patterns, making it more suitable for cyclical electricity consumption data. Despite this improvement, SARIMA models are highly sensitive to noise and require careful parameter tuning, limiting their practicality for real-time applications [22] [23].

Statistical hypothesis testing has also been explored, where observed consumption is compared against expected distributions. While conceptually sound, these methods depend on strong assumptions about data distribution and independence, both of which are rarely satisfied in real-world household environments [24].

Clustering techniques such as K-Means and DBSCAN have been applied to detect unusual consumption patterns by grouping similar behavior and identifying outliers. These methods can capture non-linearity but require careful selection of features and cluster counts, and they often fail when data exhibits continuous evolution rather than distinct clusters [25].

Principal Component Analysis (PCA) provides another classical solution by reducing dimensionality and identifying anomalies in residual variance. However, PCA assumes linear correlations and is ineffective when dealing with complex, multi-modal household consumption patterns [26].

Time-series decomposition techniques attempt to isolate trend, seasonal, and residual components to facilitate anomaly detection. Although useful for understanding aggregate patterns, decomposition methods often fail to capture sudden local spikes characteristic of household anomalies [27].

Hybrid methods combining statistical forecasting and thresholding have been proposed to improve detection accuracy. Yet, their performance remains limited by the underlying linearity assumptions and lack of ability to capture diverse behavioral influences in a home setting.

Another challenge with traditional methods is their reliance on long-term historical data to model normal patterns. In households with inconsistent user behavior, this requirement becomes difficult to satisfy, leading to inaccurate models.

Moreover, computational limitations of some statistical methods hinder their scalability when applied to high-resolution smart meter data, where consumption is recorded at minute-level intervals [28]. Overall, while traditional methods provide foundational insights, their inability to model non-linear, noisy, and irregular residential consumption renders them insufficient for modern anomaly detection needs.

2.2. **Autoencoder-Based Approaches for Anomaly Detection in Energy Systems**

Autoencoder architectures have emerged as one of the most widely used deep learning frameworks for anomaly detection in energy systems due to their ability to learn compact latent representations of normal consumption behavior. Fundamentally, an autoencoder is trained to reconstruct input data that follow typical patterns; anomalies are identified when reconstruction errors exceed a predefined threshold. This paradigm aligns naturally with the characteristics of household electricity consumption, which often exhibits predictable daily and seasonal cycles. By learning these regularities, autoencoders can highlight subtle deviations, including sudden spikes, abnormal load shapes, and unusual temporal correlations [29].

Early applications of autoencoders in the energy domain focused on industrial equipment monitoring, particularly in detecting faults in rotating machinery, compressors, and transformers.

These pioneering studies demonstrated the method's ability to capture high-dimensional relationships that classical statistical models failed to represent. The success of these industrial applications encouraged researchers to extend autoencoder-based anomaly detection to finer-granularity settings, such as smart grid feeders and residential demand profiles, where the complexity and noise levels are significantly higher [30].

One of the appealing characteristics of autoencoders for energy anomaly detection is their flexibility. Variants such as Denoising Autoencoders (DAE) can filter out sensor noise typical in residential smart meters, while Sparse Autoencoders impose constraints that enhance the interpretability of learned features. These variants have been applied to real-world consumption datasets to detect abnormal patterns caused by defective appliances, faulty meters, or unexpected behavioral changes within households [31].

The increasing availability of high-resolution smart-meter data has driven the adoption of Sequence Models, particularly Recurrent Neural Network (RNN)-based autoencoders such as LSTM-AE and GRU-AE. These architectures explicitly capture temporal dependencies in energy data, enabling more accurate modeling of patterns like morning peak loads, nighttime basal consumption, and weekend–weekday differences. Studies demonstrate that LSTM-based autoencoders outperform static models when anomalies manifest as temporal distortions rather than isolated data points [32].

Another important development is the use of Convolutional Autoencoders (CAE), which exploit local temporal correlations by treating household electricity time series as one-dimensional signals. CAEs have shown significant effectiveness in identifying anomalies related to abrupt load changes, sporadic appliance cycling, and irregular consumption bursts. Their ability to extract multi-resolution features also makes them suitable for detecting complex anomalies arising from multi-appliance interactions [33] [34].

Hybrid architectures combining convolutional and recurrent layers have further improved anomaly detection performance in residential environments. These models leverage convolutional layers for localized feature extraction and recurrent layers for long-term temporal modeling. Recent research demonstrates that such architectures reduce false positives in datasets with high variability, making them particularly relevant for urban households with diverse appliance ownership and behavioral heterogeneity [35].

Variational Autoencoders (VAE) represent another direction, enabling probabilistic modeling of household energy consumption. Instead of merely reconstructing inputs, VAEs approximate the underlying distribution of normal load profiles. This probabilistic perspective allows more rigorous anomaly scoring, especially for borderline anomalies or subtle drifts in consumption behavior. VAEs have been increasingly adopted in applications targeting long-term monitoring of household electrification programs and off-grid systems [36].

A growing body of work has explored attention-enhanced autoencoders, where attention mechanisms prioritize relevant segments of the time series. These models help disambiguate anomalies from legitimate variations, such as seasonal changes, occupant behavior shifts, and irregular appliance usage. By focusing computational resources on informative periods, attention-based autoencoders improve both detection accuracy and interpretability, which is critical for applications requiring user trust and actionable insights [37].

The field has also seen substantial progress in multimodal autoencoders that integrate data from multiple sources such as weather information, occupancy sensors, and tariff schedules. Incorporating contextual variables allows the detection system to differentiate between expected increases (e.g., a heatwave) and unexpected anomalies (e.g., malfunctioning air conditioners). Multimodal approaches represent a promising direction for future smart-home deployments where heterogeneous data streams are readily available.

Despite their strengths, autoencoder-based methods face notable challenges. They are highly sensitive to training data quality, and the presence of undetected anomalies in the training set may corrupt the learned representation. Moreover, household electricity consumption can be inherently stochastic, leading to reconstruction errors even during normal operation if the model overfits or fails to generalize. These limitations motivate ongoing research into robust training strategies, anomaly-injection techniques, and uncertainty quantification frameworks [38].

Scalability is another emerging concern, particularly as smart-meter networks expand to millions of users. Many advanced autoencoder variants especially hybrid and attention-based models—require significant computational resources, both during training and real-time inference. Consequently, recent studies emphasize lightweight and edge-deployable autoencoder architectures, enabling on-device processing to reduce latency, preserve privacy, and minimize energy consumption [39].

Overall, autoencoder-based anomaly detection has evolved into a mature yet rapidly advancing research field, offering strong potential for improving the reliability, efficiency, and sustainability of household electricity systems. Its capacity to model complex consumption patterns positions it at the forefront of modern energy analytics. However, continued progress requires addressing methodological limitations, integrating contextual data, and ensuring interpretability for end users and grid operators alike [40].

3. Methodology

The methodology begins with a comprehensive preprocessing pipeline applied to minute-level household electricity consumption data. Missing values are interpolated using linear smoothing, and sensor-induced outliers are removed to prevent false anomaly signatures. The cleaned time series is then normalized using Min-Max scaling to stabilize model convergence. To capture temporal dependencies, a sliding-window segmentation strategy divides the continuous signal into 60-minute windows with a 15-minute stride, enriching the dataset while preserving short-term consumption dynamics.

The anomaly detection framework is built around a symmetric autoencoder designed to learn latent representations of normal consumption behavior. The encoder compresses each window into a low-dimensional latent vector, while the decoder reconstructs the original signal. Hyperparameters including latent dimension size, batch size, activation functions, and learning rate were optimized using grid search to maximize reconstruction fidelity. Dropout and L2 regularization were incorporated to reduce overfitting, especially during high-variability consumption periods.

Model performance was evaluated using two datasets:

1) normal consumption data, and
2) data augmented with synthetic anomalies representing appliance malfunctions and abrupt load surges.

Tabel 1. Workflow Table

Step	Process	Description
1	Data Collection	Minute-level household electricity consumption is collected from smart meters.
2	Data Preprocessing	Missing values interpolated, outliers smoothed, data normalized using Min–Max scaling.
3	Sliding Window Segmentation	Time series segmented into 60-minute windows with 15-minute stride.
4	Model Design	Autoencoder constructed with symmetric encoder–decoder architecture.
5	Model Training	Trained using MSE loss and Adam optimizer; hyperparameters tuned via grid-search.
6	Reconstruction Error Computation	Error between original and reconstructed signals calculated for every window.
7	Adaptive Thresholding	Gaussian-based dynamic threshold applied to determine anomalies.
8	Evaluation	Metrics (Precision, Recall, F1, ROC-AUC) computed using normal + synthetic anomaly datasets.
9	Deployment Design	Edge-device integration proposed for real-time inference and privacy preservation.

Precision, Recall, F1-score, and ROC-AUC were used to quantify detection performance, while temporal localization accuracy assessed how precisely the model marked the onset of anomalies. Baseline comparisons with ARIMA, Holt–Winters, and z-score detection demonstrated that the autoencoder outperforms traditional statistical methods on nonlinear, irregular household patterns.

4. Finding and Discussion

This section presents a comprehensive evaluation of the proposed Autoencoder-based anomaly detection system. The discussion covers the experimental setup, the analysis of reconstruction capabilities, quantitative performance metrics, a comparative study against traditional baselines, and a critical look at the current limitations.

4.1. Experimental Setup

The experiments were conducted in a controlled computational environment designed to handle high-resolution time-series data. We utilized a workstation equipped with an Intel Core i7 processor and an NVIDIA GPU to accelerate the training of the deep neural network. The software stack was built upon Python 3.8, leveraging the PyTorch framework for model architecture and gradient descent optimization. This setup ensured that the training process was both efficient and reproducible, allowing for rapid iteration of hyperparameter tuning.

The dataset employed for this study consists of real-world residential electricity consumption data collected at one-minute intervals. This high granularity is essential for detecting transient anomalies that might be smoothed out in hourly or daily aggregates. The raw data underwent rigorous preprocessing, including the imputation of missing values using linear interpolation to maintain temporal continuity. Furthermore, the data was normalized using Min-Max scaling to map the consumption values to a range of [0, 1], which is crucial for the stability of the sigmoid activation functions used in the output layer of the Autoencoder.

To evaluate the model's robustness, the dataset was partitioned into a training set (70%) and a testing set (30%). The training set consisted exclusively of "normal" consumption patterns to allow the Autoencoder to learn the baseline behavior of the household. This unsupervised approach is particularly advantageous in the energy domain, where labeled anomaly data is scarce. The testing set, however, contained a mix of normal data and synthetic anomalies, providing a realistic scenario for performance evaluation.

Synthetic anomalies were injected into the test set to simulate various types of energy leakage and equipment faults. These included sudden spikes representing short circuits or motor start-up failures, as well as prolonged elevated consumption representing resistive leakage or appliances left running unintentionally. The magnitude and duration of these anomalies were varied stochastically to test the model's sensitivity across a spectrum of fault severities. This injection strategy allows for a more controlled assessment of Recall and Precision compared to relying solely on rare, naturally occurring anomalies.

The Autoencoder architecture itself was tuned through a grid search process. We experimented with different latent vector sizes, finding that a bottleneck dimension of 8 provided the best trade-off between compression and reconstruction quality. The model used a symmetric encoder-decoder structure with three hidden layers on each side. The Rectified Linear Unit (ReLU) activation function was applied to the hidden layers to mitigate the vanishing gradient problem, ensuring efficient learning of complex non-linear patterns.

Optimization was performed using the Adam optimizer, chosen for its adaptive learning rate capabilities, with an initial learning rate set to 0.001. The Mean Squared Error (MSE) served as the loss function, quantifying the difference between the input time-window and the reconstructed output. The model was trained for 100 epochs with a batch size of 64, using early stopping to prevent overfitting. The training loss converged effectively, indicating that the model successfully learned the underlying manifold of the normal consumption data.

We implemented a sliding window approach for data ingestion, with a window size of 60 minutes and a stride of 15 minutes. This segmentation strategy captures the temporal context necessary for the model to understand short-term dependencies. By feeding the model overlapping sequences, we ensured that anomalies occurring at the boundaries of a window were not missed. The reconstruction error was then calculated for each window, forming the basis for the anomaly score.

Finally, the evaluation metrics were carefully selected to provide a holistic view of performance. We utilized Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). These metrics are standard in anomaly detection literature and provide a clear indication of the model's ability to minimize false alarms while maximizing detection rates. The threshold for anomaly detection was determined dynamically, rather than statically, to adapt to the inherent noise in the data.

4.2. Reconstruction Analysis and Thresholding

The fundamental premise of this study is that an Autoencoder trained on normal data will fail to accurately reconstruct anomalous patterns. Our analysis confirms this hypothesis. During normal operation periods, the model's reconstructed output closely mirrored the input signal. The reconstruction error (MSE) remained consistently low, hovering near zero, which demonstrates that the latent representation successfully captured the regular cyclic patterns of household appliances and daily routines.

In contrast, when the model encountered the injected anomalies, the reconstruction error spiked significantly. The Autoencoder attempted to compress the anomalous input into the learned latent space of "normal" behaviors. Since the anomaly did not conform to the learned rules, the decompression (decoding) process resulted in a distorted output that differed largely from the input. This discrepancy generated a high MSE signal, serving as a distinct indicator of an abnormal event.

Visual inspection of the signal plots further validates this behavior. Figure 4 (in the manuscript) displays the overlay of the original and reconstructed signals. It is evident that for routine activities—such as the refrigerator cycling or lights being toggled—the reconstruction overlaps almost perfectly with the ground truth. However, at the timestamps where synthetic spikes were introduced, the reconstructed line remains flat or follows the expected normal trend, leaving a large visible gap (error) between the actual and predicted values.

A critical component of our analysis is the thresholding mechanism. Traditional static thresholds are often insufficient for electricity data due to its high variance. A fixed value might be too high for low-consumption periods (missing anomalies) or too low for peak periods (causing false alarms). To address this, we analyzed the distribution of reconstruction errors from the validation set. The distribution followed a log-normal pattern, prompting the use of statistical thresholding.

We implemented an adaptive threshold defined as the mean of the reconstruction error plus three standard deviations ($\mu + 3\sigma$). This dynamic limit accommodates natural fluctuations in the household's baseload. By calculating this threshold over a rolling window, the system adapts to long-term changes in consumption behavior, such as seasonal shifts from winter to summer, without requiring manual recalibration.

The effectiveness of this Gaussian-based thresholding is evident in the reduction of false positives. In initial tests with a static threshold, normal high-load events (like using an electric oven) were often flagged as anomalies. The adaptive approach correctly identified these as "normal high usage" because their reconstruction error, while higher than baseline, did not statistically deviate enough to breach the dynamic limit. This distinction is vital for user acceptance in a real-world residential product.

We analyzed the sensitivity of the reconstruction error to different types of anomalies. The model showed the highest sensitivity to "point anomalies" (sudden, sharp spikes). The MSE response to these events was immediate and high magnitude. "Contextual anomalies," such as an appliance running at an odd time but with normal wattage, resulted in lower but still detectable error elevations, proving the model learns temporal context, not just amplitude.

Extremely subtle anomalies, such as very low-ampere leakage, sometimes resulted in reconstruction errors that fell within the "noise" margin of the threshold. This indicates a trade-off between sensitivity and false alarm suppression. While tightening the threshold ($\mu + 2\sigma$) increases detection of subtle faults, it exponentially increases false positives. The chosen $\mu + 3\sigma$ parameter represents the optimal balance for general residential monitoring.

In summary, the reconstruction analysis confirms that the Autoencoder acts as an effective filter for normalcy. The error signal provides a high-fidelity proxy for abnormality. The combination of deep learning-based reconstruction and statistical adaptive thresholding creates a robust detection mechanism that filters out the noise of daily life while highlighting genuine irregularities in energy consumption.

4.3. Quantitative Performance

The quantitative evaluation of the proposed model yields promising results, substantiated by rigorous metric analysis. The model achieved an overall Precision of 0.92. This high precision score is particularly significant in the context of residential monitoring. It implies that when the system alerts a user to a potential leak or fault, there is a 92% probability that the alert is genuine, minimizing the "boy who cried wolf" effect that often plagues home security and monitoring systems.

The Recall (Sensitivity) of the model was recorded at 0.88. This indicates that the system successfully identified 88% of all injected anomalies. While slightly lower than the precision, this is a strong result given the diversity of the anomalies introduced. The missed cases (false negatives) were primarily low-magnitude anomalies that blended into the background noise of the sensor data. In a practical safety context, capturing nearly 90% of faults without calibration is a substantial improvement over unmonitored system.

The F1-Score, which creates a harmonic mean between Precision and Recall, stood at 0.90. This balanced score confirms that the model does not disproportionately favor one metric over the other. In many anomaly detection tasks, it is easy to achieve 100% recall by flagging everything as an anomaly, or 100% precision by only flagging the most obvious extreme events. An F1-Score of 0.90 demonstrates that our model maintains a healthy equilibrium, making it reliable for continuous deployment.

The Area Under the Curve (ROC-AUC) score was 0.94, indicating excellent separability between the positive (anomalous) and negative (normal) classes. A score closer to 1.0 represents a perfect classifier. The 0.94 result suggests that the reconstruction error distributions for normal and anomalous data are distinct and well-separated, validating the choice of Autoencoder architecture for feature extraction.

We also analyzed the False Positive Rate (FPR), which remained below 5% throughout the testing phase. Keeping the FPR low is critical for scalability. If deployed across thousands of smart meters, a high FPR would overwhelm utility providers with support calls. The low FPR achieved here is largely attributed to the robust feature learning of the encoder, which effectively ignores sensor noise that would typically trigger simple threshold-based alarms.

Temporal localization accuracy was another aspect of our quantitative assessment. We measured the time delay between the onset of an anomaly and its detection. The average detection latency was found to be within the 15-minute stride of the sliding window. While not instantaneous, this near-real-time performance is sufficient for residential energy management, where the goal is to stop persistent leakage or identify faulty appliances before the next billing cycle.

We further dissected the performance based on anomaly duration. The model performed best on short-duration, high-intensity anomalies (Precision > 0.95). Performance dropped slightly for long-duration, low-intensity anomalies (Precision ~ 0.85). This variance highlights the model's reliance on "change" or "deviation" from the norm; anomalies that slowly become the new norm are harder to detect without a longer memory mechanism.

Overall, the quantitative data confirms that the Deep Learning approach provides a high degree of reliability. The metrics surpass the typical performance of consumer-grade energy monitors which rely on simple power limits. The statistical validation provides a strong confidence level for the feasibility of deploying this algorithm in commercial smart metering infrastructure.

4.3. Comparative Analysis

To benchmark the effectiveness of our Deep Learning framework, we compared it against two established traditional methods: ARIMA (Autoregressive Integrated Moving Average) and a Z-Score based statistical method. These baselines represent the standard approaches currently used in many industrial time-series applications. The comparison highlights the specific advantages of using neural networks for the complex domain of residential energy.

The ARIMA model, while powerful for linear forecasting, showed significant limitations in this study. It achieved a Precision of only 0.74 and a Recall of 0.68. The primary failure mode of ARIMA was its inability to handle the non-linear, abrupt changes typical of human behavior (e.g., turning on multiple appliances simultaneously). ARIMA assumes a level of stationarity and linear correlation that simply does not exist in minute-level household data, leading to a high rate of false alarms during peak usage hours.

The Z-Score method, which relies on simple statistical deviations from the mean, performed even more poorly with an F1-Score of 0.65. This method assumes that the data follows a normal distribution. However, household electricity consumption is often multi-modal (having multiple "normal" states, such as "nobody home," "evening activity," and "night sleep"). The Z-Score method struggled to differentiate between a "normal high peak" and an "anomaly," proving it is too simplistic for this application.

In contrast, our Autoencoder outperformed both baselines significantly across all metrics. The key differentiator is the Autoencoder's ability to learn a non-linear latent representation. It essentially memorizes the complex "shape" of daily routines rather than just the mathematical average. This allows it to understand that a spike at 7:00 PM is normal (cooking dinner), whereas a similar spike at 3:00 AM is anomalous, a context that Z-Score methods completely miss.

Another comparative advantage is adaptability. The ARIMA model required frequent re-training and parameter tuning (p, d, q values) to maintain performance as seasons changed. The Autoencoder, once trained on a representative dataset, showed greater generalization capabilities. It remained robust even when minor shifts in daily schedules occurred, whereas the rigid mathematical structure of ARIMA flagged these shifts as errors.

In terms of computational efficiency during inference, the Z-Score method is naturally the fastest, followed by the Autoencoder, with ARIMA being the most computationally intensive due to the iterative nature of its forecasting steps. Although the Autoencoder requires heavy computation during training, its inference (forward pass) is essentially a matrix multiplication operation, making it surprisingly efficient and faster than ARIMA for real-time detection on edge devices.

We also compared the robustness to noise. When we added Gaussian noise to the test data to simulate sensor degradation, the performance of the Z-Score and ARIMA models degraded linearly. The Autoencoder, however, acted as a "Denoising Autoencoder" to some extent. Its performance degraded much more slowly, proving that the neural network focuses on the structural patterns of the data rather than the raw pixel-level values.

The comparative analysis clearly establishes the superiority of the Autoencoder for this specific domain. While statistical methods are sufficient for aggregated, smooth data (like monthly grid load), the chaotic and granular nature of individual household data requires the feature extraction capabilities of Deep Learning. The 20-25% improvement in F1-Score over traditional methods justifies the increased complexity of the neural network approach.

Ultimately, this comparison validates the shift towards AI-driven analytics in the smart grid sector. The failure of linear models to capture the intricacies of human energy consumption underscores why previous generations of smart meters failed to provide actionable anomaly insights, and why deep learning represents the necessary evolution of the technology.

Despite the strong performance demonstrated in the experiments, several limitations inherent to the proposed methodology must be acknowledged. First and foremost is the issue of "Concept Drift." Household energy patterns are not static; they evolve over time due to family growth, new appliance purchases, or lifestyle changes. A model trained on data from 2023 may flag normal behavior in 2024 as anomalous if the household buys an electric vehicle. The current model lacks an online learning module to update itself automatically, requiring periodic retraining.

Secondly, the detection of "Gradual Drifts" remains a weakness. The model excels at identifying sudden spikes or drops (point anomalies). However, a resistive fault that causes energy usage to creep up by 1% per week might go undetected for a long time. The reconstruction error for such a slow change would increase incrementally, potentially staying within the adaptive threshold'

5. Conclusion

This study demonstrates that autoencoder-based anomaly detection provides an effective and scalable method for identifying irregularities in household electricity consumption. By learning the latent structure of normal load patterns, the model can reliably distinguish abnormal deviations such as sudden spikes, equipment malfunctions, or unexpected usage behaviors. The reconstruction-error framework proves particularly suitable for residential environments where consumption patterns vary yet remain rhythmically structured over time.

The adaptive thresholding mechanism significantly strengthens the practical deployment potential of the method. Household energy data often contain noise, daily fluctuations, and occasional irregularities that are not true anomalies. The adaptive threshold accounts for these natural variations while minimizing false positives, thereby enhancing the reliability of anomaly detection in real-world settings. Combined with appropriate preprocessing—such as smoothing and handling missing values—the system maintains stable performance across diverse household profiles.

Overall, the findings contribute both technical and practical insights into the development of intelligent energy-monitoring systems. The proposed autoencoder model is lightweight, unsupervised, and suitable for integration into smart meters or edge-computing devices. Future research may incorporate explainable AI to interpret anomaly types, develop seasonally aware detection models, and optimize architectures for low-power embedded hardware. These advancements will support the evolution of smarter, safer, and more energy-efficient households.

References

[1] J. Chew, A. Sharma, D. S. Kumar, W. Zhang, N. Anant, and J. Dong,

"Unveiling the dynamics of residential energy consumption: A quantitative study of demographic and personality influences in Singapore using machine learning approaches," *Sustainability*, vol. 16, no. 14, 2024.

[2] C. P. Paneru and A. K. M. Tarigan, "Reviewing the impacts of smart energy applications on energy behaviours in Norwegian households," *Renewable and Sustainable Energy Reviews*, vol. 183, 113511, Sep. 2023.

[3] M. M. Naseer, A. I. Hunjra, A. Palma, and T. Bagh, "Sustainable development goals and environmental performance: Exploring the contribution of governance, energy, and growth," *Research in International Business and Finance*, vol. 73, no. B, 102646, Jan. 2025.

[4] R. N. Muniz, C. T. da Costa Júnior, W. G. Buratto, A. Nied, and G. V. González, "The sustainability concept: A review focusing on energy," *Sustainability*, vol. 15, no. 19, 2023.

[5] L. N. Tran, Q. Wu, and H. T. Hoang, "The effect of changing heat use patterns on residential energy efficiency in a Japanese smart community," *Scientific Reports*, vol. 15, no. 20404, 2025.

[6] D. Novianto, M. D. Koerniawan, M. Munawir, and D. Sekartaji, "Impact of lifestyle changes on home energy consumption during pandemic COVID-19 in Indonesia," *Sustainable Cities and Society*, vol. 83, Aug. 2022.

[7] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, "A comprehensive survey of machine learning methods for surveillance videos anomaly detection," *IEEE Access*, vol. 11, pp. 105046–105072, Oct. 2023.

[8] Y. Song, S. Kuang, J. Huang, and D. Zhang, "Unsupervised anomaly detection of industrial building energy consumption," *Energy and Built Environment*, in press, Dec. 2024.

[9] C. Yang, H. Li, Y. Ma, Y. Huang, and X. Chu, "Enhanced TSMixer model for the prediction and control of particulate matter," *Sustainability*, vol. 17, no. 7, 2025.

[10] K. Hu, C. Shen, T. Wang, *et al.*, "Overview of temporal action detection based on deep learning," *Artificial Intelligence Review*, vol. 57, no. 26, 2024.

[11] G. Wang, S. Sun, S. Fan, Y. Liu, S. Cao, and R. Guan, "A spatial–temporal data-driven deep learning framework for enhancing ultra-short-term prediction of distributed photovoltaic power generation," *International Journal of Electrical Power & Energy Systems*, vol. 160, 110125, Sep. 2024.

[12] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, "Deep learning for anomaly detection in log data: A survey," *Machine Learning with Applications*, vol. 12, 100470, Jun. 2023.

[13] H. Park and H. Jang, "Enhancing time series anomaly detection: A knowledge distillation approach with image transformation," *Sensors*, vol. 24, no. 24, 2024.

[14] S. Kumari, C. Prabha, A. Karim, M. M. Hassan, and S. Azam, "A comprehensive investigation of anomaly detection methods in deep learning and machine learning: 2019–2023," *IET Information Security*, vol. 2024, no. 1, 8821891, Nov. 2024.

[15] M. R. Shadi, H. Mirshekali, and H. R. Shaker, "Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects," *Renewable and Sustainable Energy Reviews*, vol. 216, 115668, Jul. 2025.

[16] M. Orabi, K. P. Tran, P. Egger, and S. Thomassey, "Anomaly detection in smart manufacturing: An Adaptive Adversarial Transformer-based model," *Journal of Manufacturing Systems*, vol. 77, pp. 591–611, Dec. 2024.

[17] S. Chen and T. Nakachi, "Multi-Channel Gaussian Process Regression for Network Traffic Anomaly Detection via Inter-Channel Correlation," in *Proc. 2025 Int. Tech. Conf. Circuits/Syst., Comput., Commun. (ITC-CSCC)*, pp. 1–6, 2025.

[18] E. Estrada-Patiño, G. Castilla-Valdez, J. Frausto-Solis, *et al.*, "A Novel Approach for Temperature Forecasting in Climate Change Using Ensemble Decomposition of Time Series," *Int. J. Comput. Intell. Syst.*, vol. 17, p. 253, 2024.

[19] S. Karamolegkos and D. E. Koulouriotis, "Advancing short-term load forecasting with decomposed Fourier ARIMA: A case study on the Greek energy market," *Energy*, vol. 325, p. 135854, Jun. 2025.

[20] S. B. Chudo and G. Terdik, "Modeling and forecasting time-series data with multiple seasonal periods using periodograms," *Econometrics*, vol. 13, no. 2, 2025.

[21] C. Prakash, B. Dhyani, A. Chauhan *et al.*, "ARIMA based forecasting of solar and hydro energy consumption with implications for grid stability and renewable policy," *Discover*

Sustainability, vol. 6, no. 663, 2025.
[22] M. L. Hossain, S. M. N. Shams, and S. M. Ullah, "Time-series and deep learning approaches for renewable energy forecasting in Dhaka: a comparative study of ARIMA, SARIMA, and LSTM models," *Discover Sustainability*, vol. 6, no. 775, 2025.
[23] S.-C. Necula, I. Hauer, D. Fotache, and L. Hurbean, "Advanced hybrid models for air pollution forecasting: Combining SARIMA and BiLSTM architectures," *Electronics*, vol. 14, no. 3, 2025.
[24] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, p. 160, 2021.
[25] P. Koukaras and C. Tjortjis, "Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices," *AI*, vol. 6, no. 10, 2025.
[26] G. Murariu, L. Dinca, and D. Munteanu, "Trends and applications of principal component analysis in forestry research: A literature and bibliometric review," *Forests*, vol. 16, no. 7, Art. no. 1155, Jul. 2025.
[27] M. Taktak and F. Derbel, "Evaluating the impact of frequency decomposition techniques on LSTM-based household energy consumption forecasting," *Energies*, vol. 18, no. 10, 2025.
[28] M. A. Yahya, A. R. Moya, and S. Ventura, "Deep learning for multivariate time series anomaly detection: an evaluation of reconstruction-based methods," *Artificial Intelligence Review*, vol. 58, 2025.
[29] Y. Ghazi, M. Tabaa, M. Ennaji, and G. Zaz, "An explainable Markov chain–machine learning sequential-aware anomaly detection framework for industrial IoT systems based on OPC UA," *Sensors*, vol. 25, no. 19, 2025.
[30] E. Stracqualursi, A. Rosato, G. Di Lorenzo, M. Panella, and R. Araneo, "Systematic review of energy theft practices and autonomous detection through artificial intelligence methods," *Renewable and Sustainable Energy Reviews*, vol. 184, 113544, Sep. 2023.
[31] F. Harrou, B. Bouyeddou, A. Dairi, and Y. Sun, "Exploiting autoencoder-based anomaly detection to enhance cybersecurity in power grids," *Future Internet*, vol. 16, no. 6, 2024.
[32] H. Ishfaq, S. Kanwal, S. Anwar, M. Abdussalam, and W. Amin, "Enhancing smart grid security and efficiency: AI, energy routing, and T&D innovations (A review)," *Energies*, vol. 18, no. 17, 2025.
[33] J. Feng, T. Yu, K. Zhang, and L. Cheng, "Integration of multi-agent systems and artificial intelligence in self-healing subway power supply systems: Advancements in fault diagnosis, isolation, and recovery," *Processes*, vol. 13, no. 4, 2025.
[34] X. He, T. Du, T. Long, *et al.*, "Signaling cascades in the failing heart and emerging therapeutic strategies," *Signal Transduction and Targeted Therapy*, vol. 7, no. 134, 2022.
[35] R. Lin, S. Chen, Z. He, B. Wu, H. Zou, X. Zhao, and Q. Li, "Electricity behavior modeling and anomaly detection services based on a deep variational autoencoder network," *Energies*, vol. 17, no. 16, 2024.
[36] R. Liao, M. Manfren, and B. Nastasi, "Off-grid PV systems modelling and optimisation for rural communities—leveraging understandability and interpretability of modelling tools," *Energy*, vol. 324, 135948, Jun. 2025.
[37] N. H. A. Mutalib, A. Q. M. Sabri, A. W. A. Wahab, *et al.*, "Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review," *Artificial Intelligence Review*, vol. 57, p. 297, 2024.
[38] R. J. L. Taloma, F. Cuomo, D. Comminiello, *et al.*, "Machine learning for smart water distribution systems: exploring applications, challenges and future perspectives," *Artificial Intelligence Review*, vol. 58, p. 120, 2025.
[39] G. Chen, S. Lu, S. Zhou, Z. Tian, M. K. Kim, J. Liu, and X. Liu, "A systematic review of building energy consumption prediction: From perspectives of load classification, data-driven frameworks, and future directions," *Applied Sciences*, vol. 15, no. 6, 2025.
[40] K. O. Ukoba, K. O. Olatunji, E. Adeoye, T.-C. Jen, and D. M. Madyira, "Optimizing renewable energy systems through artificial intelligence: Review and future prospects," *Energy & Environment*, vol. 35, no. 7, May 2024.