Predictive Analysis and Anomaly Detection in Household Electric Power Consumption Predictive Analysis and Anomaly Detection in Household Electric Power Consumption

MSC IN ARTIFICIAL INTELLIGENCE WITH BUSINESS STRATEGY

GLAWIN JOHN ALVA - 230143753

Supervised by Dr. Farzaneh Farhadi

Date: 30th September 2024

Declaration:

I declare that I have personally prepared this assignment. The work is my own, carried out personally by me unless otherwise stated and has not been generated using paid for assessment writing services or Artificial Intelligence tools unless specified as a clearly stated approved component of the assessment brief. All sources of information, including quotations, are acknowledged by means of the appropriate citations and references. I declare that this work has not gained credit previously for another module at this or another University, save for permitted elements which formed part of an associated proposal linked directly to this submission.

I understand that plagiarism, collusion, copying another student and commissioning (which for the avoidance of doubt includes the use of essay mills and other paid for assessment writing services, as well as unattributed use of work generated by Artificial Intelligence tools) are regarded as offences against the University's Assessment Regulations and may result in formal disciplinary proceedings.

I understand that by submitting this assessment, I declare myself fit to be able to undertake the assessment and accept the outcome of the assessment as valid.

Student signature: Glawin Alva

Date:30th September 2024.

Abstract

This dissertation explores the development of a framework that combines energy consumption forecasting with anomaly detection to improve accuracy and system efficiency. By using a hybrid model that combines Long Short-Term Memory (LSTM) networks with Stationary Wavelet Transform (SWT), the study aims to enhance energy predictions and identify anomalies that could indicate issues like equipment malfunctions or inefficiencies. The research uses two datasets: the UCI Individual Household Electric Power Consumption dataset for forecasting, and the LEAD1.0 dataset as a benchmark for anomaly detection. LEAD1.0, which is typically used for commercial buildings, was chosen due to the lack of labelled household-level anomaly data. The findings show that the hybrid model effectively reduces noise and improves prediction accuracy, while the anomaly detection methods successfully flagged unusual patterns in energy use. Despite some challenges with limited variability in the datasets, the study highlights the need for more diverse data and points to future research opportunities in real-time energy management.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Dr. Farzaneh Farhadi, for her invaluable guidance and support throughout the course of my dissertation. Her insightful feedback, constructive criticism, and encouragement have been instrumental in shaping my research and enhancing my understanding of the subject. I am truly grateful for her patience and dedication, which have significantly contributed to the successful completion of this project. Thank you for believing in my abilities and inspiring me to strive for excellence.

Table of Contents

List of Tables:

List of Figures:

1. Introduction

1.1 Motivation:

Global growth in energy consumption driven by both populations increase, and technological advancement has made energy management a critical area for achieving sustainability ('Global Energy Review 2020', 2020). Households significantly contribute to this demand; thus, improving household energy consumption practices can yield substantial economic and environmental benefits. Smart grids and intelligent energy systems have emerged as effective approaches to manage and control energy utilization (Fang *et al.*, 2012). By incorporating smart meters in conjunction with Internet of Things devices, households gain access to real-time data on their energy usage, allowing for more accurate and efficient management (Schölkopf *et al.*, 2001).

Predicting household energy consumption plays a very crucial role in energy management by helping to identify peak usage times, optimizing energy distribution, and implementing energy-saving strategies. This not only reduces costs but also contributes to sustainability by minimizing energy wastage. However, managing energy consumption remains challenging due to the complex and dynamic nature of residential energy patterns. Anomalies such as sudden spikes caused by device malfunctions, unauthorized energy diversion, or inefficient operations can negatively impact safety, economic factors, and operational stability (Salami *et al.*, 2023). These factors highlighted the necessity for developing a comprehensive framework that can accurately forecast energy consumption and detect anomalies, thereby enhancing energy efficiency, user safety, and sustainable consumption behaviour. This project aims to expand state-of-the-art models in predicting household energy consumption and detecting anomalies to help meet global sustainability goals.

1.2 Problem Statement:

Accurately predicting residential energy use is challenging due to the unsteady and highly non-linear nature of consumption behaviours (Lago, De Ridder and De Schutter, 2018). Factors such as the number of occupants, types of devices, resident behaviours, and situational factors like meteorological conditions contribute to this unpredictability (Fan, Xiao and Wang, 2014). This complexity makes conventional forecasting methods, such as ARIMA (Auto-Regressive Integrated Moving Average), struggle to effectively model these behaviours, often resulting in less accurate predictions (Tunnicliffe Wilson, 2016). Furthermore, anomalies in energy consumption caused by equipment malfunction, suboptimal usage practices, or unauthorized connections require advanced analytics to distinguish between normal fluctuations and genuine anomalies (Pan, Yin and Jiang, 2022)

Although machine learning architectures, such as Long Short-Term Memory (LSTM) networks, have demonstrated potential in identifying non-linear relationships within time-series data, they may not completely mitigate the volatility and noise inherent in datasets concerning energy consumption (Hochreiter and Schmidhuber, 1997). Moreover, most of the current methodologies are highly disjointed in terms of the processes of forecasting and anomaly detection, creating a deficiency in integrated solutions that could handle both tasks simultaneously. This paper addresses these issues by proposing a hybrid Long Short-Term Memory model, enhanced through the Stationary Wavelet Transform (SWT), to improve the accuracy of forecasting and integrating several artificial intelligence based techniques for developing an integrated real-time energy management system.

1.3 Research Background:

Energy consumption has been a highly active area of research for several years. Classical linear models, particularly ARIMA, form the basis of initial time-series analysis (Tunnicliffe Wilson, 2016). However, such models often struggle to accurately estimate the non-linear features and time dependencies inherent in energy consumption data. More recent machine learning approaches, such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have introduced flexibility into the modelling of complex patterns (Zangrando *et al.*, 2022). However, these models still pose challenges when resolving sequential dependencies (Bahij *et al.*, 2021).

LSTM networks have significantly advanced time-series prediction due to their ability to model long-term dependencies and capture temporal structures in the data, as highlighted by Hochreiter and Schmidhuber (1997). Studies such as Kong et al. (2017) have shown that LSTM models outperform traditional approaches in forecasting residential energy consumption. Yet, basic LSTM models themselves cannot fully address the volatility and noise in energy consumption data, resulting in unstable predictions. To address this, researchers have integrated LSTM networks with signal processing techniques like Stationary Wavelet Transform (SWT), as proposed by Zhang et al. (2019). SWT allows for the decomposition of time series into distinct components, separating noise from fundamental patterns and improving model performance and robustness.

Anomaly detection is also a key factor in energy management. Various AI-based strategies have been proposed for this purpose, each with unique strengths. For example, Isolation Forest isolates anomalies in high-dimensional datasets by creating random partitions (Liu, Ting and Zhou, 2008), and Autoencoders are neural networks that reconstruct input data, identifying instances with high reconstruction errors as anomalies (Hinton and Salakhutdinov, 2006). Prominent techniques such as One-Class SVM, introduced by Schölkopf et al. (2001), and Local Outlier Factor (LOF), by Breunig et al. (2000), are commonly used to detect anomalies in normal behaviour. Although anomaly detection techniques are often applied independently, this research aims to integrate these techniques with energy consumption prediction to provide a comprehensive framework for energy management. By developing a hybrid LSTM-SWT model and incorporating AI-based anomaly detection techniques, this project proposes a robust, integrated solution for household energy management.

1.4 Research Objectives:

The study aims to improve energy consumption management within households by developing an integrated framework for energy consumption forecasting and anomaly detection. This will be achieved through the fulfilment of the following specific objectives:

1. **Develop a Hybrid Forecasting Model:** The first objective of this study is to develop an integrated forecasting model using LSTM combined with the Stationary Wavelet Transform (SWT) to capture both short-term and long-term patterns in household energy consumption. LSTM captures long-term dependencies and nonlinear patterns in time-series data, while SWT decomposes the series into different frequency components, smoothing out most of the noise. This hybrid model is designed to improve the accuracy and reliability of household energy consumption forecasting.

2. **Implement AI-based Anomaly Detection:** AI-driven techniques for identifying anomalies in atypical consumption behaviours are employed to enable prompt recognition. The study will utilize methodologies such as Isolation Forest, Autoencoders, One-Class Support Vector Machine, and Local Outlier Factor. Each of these methods has strengths in recognizing differences from standard consumption patterns. A comparative analysis will be conducted to determine the most effective methods for detecting anomalies in household energy consumption datasets.
3. **Integration of Forecasting with Anomaly Detection:** The third objective is to develop an integrated framework that combines forecasting with anomaly detection. This integration will facilitate real-time energy management, allowing households to forecast future consumption trends and detect abnormal or hazardous usage patterns, contributing to enhanced energy efficiency and safety.

1.5 Research Questions:

- How does the hybrid LSTM with SWT model improve the forecasting accuracy of household energy use?
- Which is the best performing AI technique: Isolation Forest, Autoencoders, One-Class SVM, or LOF for energy consumption data anomaly detection?
- How can prediction, in line with anomaly detection, be implemented within one system for real-time energy management? How the hybrid forecasting model can be coupled with methods for anomaly detection within an integrated framework for realizing proactive energy management in domestic energy systems that are as efficient as possible and as safe as possible for all concerned.

2. Literature Review

2.1 Energy Consumption Forecasting

Energy consumption forecasting is a critical component of resource management optimization in smart grids. Traditional statistical methods, like the Auto-Regressive Integrated Moving Average (ARIMA) model, have been widely used for short-term load forecasting because of their ability to solve linear patterns in time-series data (Tunnicliffe Wilson, 2016). However, ARIMA models struggle to handle the non-linear and volatile nature of household energy consumption, which depends on factors such as weather conditions, human behaviours, and appliance usage (Fan and Hyndman, 2010). For example, ARIMA is not well-suited to predict energy consumption in households where patterns change abruptly due to irregular human activity or appliance usage.

Due to these limitations, machine learning models such as Support Vector Machines (SVM) (Chicco, Napoli and Piglione, 2006) and Artificial Neural Networks (ANN) have been applied in time-series forecasting. These techniques provide greater flexibility in capturing non-linearities in the data, but they often struggle with modelling sequential dependencies and long-term temporal patterns (Zhang *et al.*, 2022). This led to the development of newer models like LSTM, which are specifically designed to handle non-linear patterns more effectively. For instance, while SVM works well in classification tasks, it falters in time-series prediction, where the relationships between points in time are crucial.

More recently, Long Short-Term Memory (LSTM) networks, a specific type of Recurrent Neural Network (RNN), have shown excellent results in energy forecasting**.** LSTM avoids the vanishing gradient problem in traditional RNNs, which allows it to capture both short-term fluctuations (e.g., daily usage patterns) and long-term trends (e.g., seasonal changes or gradual changes due to energy saving upgrades) more effectively (Hochreiter and Schmidhuber, 1997). For example, Kong et al. (2017) demonstrated that LSTM models significantly outperformed ARIMA in forecasting residential energy consumption, as they were able to capture both long-term trends and sudden changes. By avoiding the vanishing gradient problem, LSTM models are able to better manage these long-term dependencies, making them highly effective in energy consumption forecasting.

Despite their strengths, LSTM models are not without challenges. The volatility and noise in energy consumption data can lead to unstable predictions, especially in the case of sudden spikes or anomalies (Kong *et al.*, 2017). This is particularly common in households where energy usage is highly variable due to occupant behaviour and appliance patterns (Fan and Hyndman, 2010).

To address these challenges, hybrid models combining LSTM with signal processing techniques like the Stationary Wavelet Transform (SWT) have been proposed. SWT decomposes time-series data into multiple frequency components, isolating noise and enhancing signal features (Zhang *et al.*, 2022). For example, Yan et al. (2019) developed a hybrid deep learning model that integrated LSTM with SWT, significantly improving prediction accuracy and reducing data volatility. Tested on real-world data from the 'UK-DALE' project, their model outperformed others, such as Support Vector Regression and CNN-LSTM, in both short-term and long-term predictions.

Zhang et al. (2019) also demonstrated that hybrid LSTM-SWT models provide more accurate forecasts compared to standalone LSTM models. By applying SWT, the original energy consumption signal is decomposed into more stationary sub-signals, allowing LSTM to better capture underlying patterns, especially in noisy and non-stationary datasets such as household energy consumption. The application of these hybrid techniques shows the importance of integrating signal processing with machine learning to better manage household energy consumption, which is often highly variable and difficult to predict.

Nevertheless, the majority of research applying LSTM-SWT models has focused on larger systems, such as industrial and commercial buildings where consumption patterns are more predictable (Zhang *et al.*, 2022). In contrast, the variability of household consumption patterns, influenced by factors like the number of occupants and appliance usage, presents a greater challenge**.** This research aims to further explore the application of LSTM-SWT models to household energy consumption, where high-accuracy forecasting is essential for real-time energy management and anomaly detection.

2.2 Anomaly Detection in Smart Grids

Anomaly detection in smart grids is essential for ensuring the security, efficiency, and reliability of these systems. Anomalies, such as sudden spikes in energy usage, may indicate faulty equipment, unauthorized use, or inefficient operation. Detecting these anomalies is critical to maintaining grid stability and performance (Banik *et al.*, 2023). Traditional statistical methods like Z-score analysis and Box Plot often fail to handle the complex and high-dimensional nature of energy data (Breunig *et al.*, 2000).

AI-based techniques have significantly improved anomaly detection in smart grids by enhancing the ability to handle complex, non-linear patterns and high-dimensional data. The following table compares various AI techniques for anomaly detection:

Table 1: Comparison of Anomaly Detection Techniques for Smart Grids

Technique	Advantages	Disadvantages	Use Cases in Smart Grids
Isolation Forest	Effective in high-dimensional data, no need to model normal behaviour, fast training	Requires extensive hyperparameter tuning, may struggle with complex patterns, false positives	Detects spikes in energy consumption due to faults or unauthorized access (Park and Kim, 2020).
Autoencoders	Captures complex, non-linear relationships, effective in high-dimensional datasets	Sensitive to hyperparameter tuning, large amounts of normal data required, overfitting potential	Used to detect abnormal consumption patterns (Karadayı, Aydin and Öğrenci, 2020)
One-Class SVM	Good for small datasets, effective in detecting subtle deviations	Struggles with high-dimensional data, requires careful kernel selection, high computational cost	Identifies unusual consumption patterns in smart grids (Lamrini *et al.*, 2018).
Local Outlier Factor	Detects local anomalies based on density	Sensitive to parameter choices (e.g., number of neighbours), does not scale well for large datasets	Applied to detect device malfunctions through energy drops (Breunig *et al.*, 2000).

Hyperparameter Tuning Challenges:

In energy consumption anomaly detection, various models offer distinct strengths and challenges, particularly in how they handle hyperparameter tuning, which is crucial for optimizing performance.

Autoencoders, which are often used for detecting non-linear relationships in energy consumption data, are highly sensitive to their architecture and parameters. Factors such as the number of layers, units per layer, learning rate, and reconstruction loss function must be carefully balanced to avoid overfitting to normal data. When these parameters are not optimized, the model may fail to detect subtle anomalies. Techniques such as dropout and early stopping are often employed to mitigate overfitting, while grid search or random search helps in finding optimal configurations (Park and Kim, 2020). Despite their effectiveness, the large amount of normal data required for training autoencoders, and the complexity of parameter tuning are significant challenges.

One-Class SVM, on the other hand, is known for its capacity to detect subtle deviations in smaller datasets, which makes it suitable for energy consumption data where anomalies might be infrequent. However, its performance is heavily influenced by the choice of kernel function (linear, polynomial, or radial basis function (RBF)) and the regularization parameter (nu). The RBF kernel, often preferred for non-linear data, requires precise tuning to avoid poor anomaly detection performance, especially in high-dimensional datasets (Lamrini *et al.*, 2018). Without proper tuning, One-Class SVM can be computationally expensive and may struggle to effectively separate anomalies from normal behaviour in larger datasets (Schölkopf *et al.*, 2001).

Isolation Forest is commonly used for its scalability in high-dimensional data, making it effective for large datasets such as those generated by smart meters. It isolates anomalies by randomly partitioning data points, requiring fewer splits for outliers. However, Isolation Forest can struggle with complex patterns and may generate false positives if its tree depth and sample size are not properly configured (Liu, Ting and Zhou, 2008). This model requires careful tuning of these parameters to achieve a balance between computational efficiency and detection accuracy.

Local Outlier Factor (LOF), which identifies anomalies based on the density of local data points, also faces challenges in parameter tuning. Its performance is highly sensitive to the number of neighbours (k) considered when calculating local densities. If k is set too low, the model may overreact to small fluctuations, resulting in false positives. Conversely, a high value of k may smooth out anomalies, causing genuine issues to be missed. Additionally, LOF struggles with scalability, becoming inefficient in large datasets due to its increased computational complexity (Breunig *et al.*, 2000)

In summary, each anomaly detection model has specific tuning requirements, and finding the right balance between detection performance and computational efficiency is key. Autoencoders excel at capturing complex, non-linear relationships but require careful tuning to avoid overfitting. One-Class SVM performs well with small, subtle deviations but can be computationally heavy. Isolation Forest is scalable but prone to false positives without proper tuning, and LOF, though effective for local anomalies, struggles with parameter sensitivity and scalability. Proper hyperparameter tuning, often through grid search or cross-validation, is essential for ensuring these models detect anomalies effectively in energy consumption data.

2.3 Integration of Forecasting and Anomaly Detection

Combining energy consumption forecasting with anomaly detection offers a more proactive approach to managing energy use. While anomaly detection helps catch unusual consumption patterns, pairing it with forecasting allows us to anticipate issues before they happen. By having a reliable forecast, any deviation from expected energy usage can be flagged early, giving us time to intervene before inefficiencies or malfunctions occur. For example, a spike in energy consumption could indicate faulty equipment, or a sudden drop might signal abnormal behaviour in the system. Although integrated approaches like this have seen commercial use (Kozitsin, Katser and Lakontsev, 2021), they remain quite rare at the household level, where energy patterns are more unpredictable and harder to manage (Wang *et al.*, 2024).

By combining forecasting models with anomaly detection, households can not only predict future energy consumption but also identify and address irregularities early, leading to more efficient energy management. This approach enables real-time energy management by integrating the strengths of both techniques, allowing early detection of potential inefficiencies or faults and making it easier to optimize energy use.

The significance of this gap at the household level cannot be overstated. Given the highly variable nature of household energy consumption, which is influenced by factors like occupant behaviour, appliance usage, and weather, a lack of integrated systems makes it difficult to predict and manage energy use effectively. Developing an integrated forecasting-anomaly detection framework can significantly enhance real-time energy management by providing both predictive insights and immediate detection of anomalies. This also contributes to sustainability by preventing energy waste and improving the overall efficiency of household energy systems.

This dissertation aims to address this gap by introducing a unified system that combines Hybrid LSTM + SWT-based forecasting with AI-powered anomaly detection**.** By integrating these two elements, the proposed framework will not only predict future energy use but also spot anomalies in real time, offering a more complete and efficient solution for managing energy in homes. This approach increases the likelihood of detecting issues early and managing energy use in a more sustainable way, providing a much-needed tool for the evolving landscape of smart home energy management.

2.4 Dataset Use in Research:

In energy consumption forecasting and anomaly detection research, several commonly used datasets offer valuable insights for model development. The UCI Individual Household Electric Power Consumption dataset is widely applied for residential energy consumption forecasting. It provides detailed, long-term data collected over several years from a single household, making it ideal for modelling energy use patterns. However, a key limitation of this dataset is the absence of labelled anomalies, which complicates the implementation of supervised anomaly detection methods.

To address this gap, the LEAD1.0 dataset was chosen for this research as well. LEAD1.0 is a large-scale dataset specifically designed for anomaly detection in commercial buildings, providing high-resolution data and labelled anomalies. Its comprehensive structure makes it particularly suited for developing AI-based anomaly detection models, which require labelled data to identify energy consumption anomalies accurately.

Commonly used datasets in this field, such as UK-DALE and REDD, offer appliance-level disaggregation for household energy consumption, while ASHRAE Great Energy Predictor III focuses on forecasting energy use in commercial buildings. These datasets are frequently utilized for both forecasting and anomaly detection due to their granularity and data richness.

In this study, the UCI dataset serves as the foundation for household-level energy forecasting, while LEAD1.0 supplements it by providing labelled anomaly data for robust anomaly detection. By combining these two datasets, this research aims to develop an integrated framework for predicting energy use and detecting anomalies in both residential and commercial contexts.

Summary and Identified Gaps

The literature shows that energy forecasting has come a long way, evolving from basic methods like ARIMA to more sophisticated machine learning models, such as LSTM and hybrids like LSTM with SWT (Yan *et al.*, 2019). These advanced models are better at handling the complexities and noise in energy data, leading to more accurate predictions. On the other hand, AI-based anomaly detection techniques are good at identifying outliers but usually operate independently of forecasting models. This creates a gap in current energy management solutions, as they lack an integrated approach that combines both forecasting and anomaly detection.

The main gap is the lack of systems that combine both forecasting and anomaly detection, especially at the household level, where energy usage is often unpredictable. This research aims to fill that void by proposing an integrated framework that combines Hybrid LSTM + SWT for forecasting with AI-driven anomaly detection techniques. By doing so, it offers a new and more efficient way to manage energy in smart homes, contributing something fresh and useful to the field of energy management.

3. Project Management

This dissertation's project management included meticulous planning, progress monitoring, and hybrid model-based adjustment to meet the stated objectives of energy consumption forecasting and anomaly detection. Together with ethical considerations, here is a summary of the project's planned, watched over, and changed course. Project Planning

The Gantt chart below shows the chronology of every phase, highlighting how chores were divided over the course of the project:

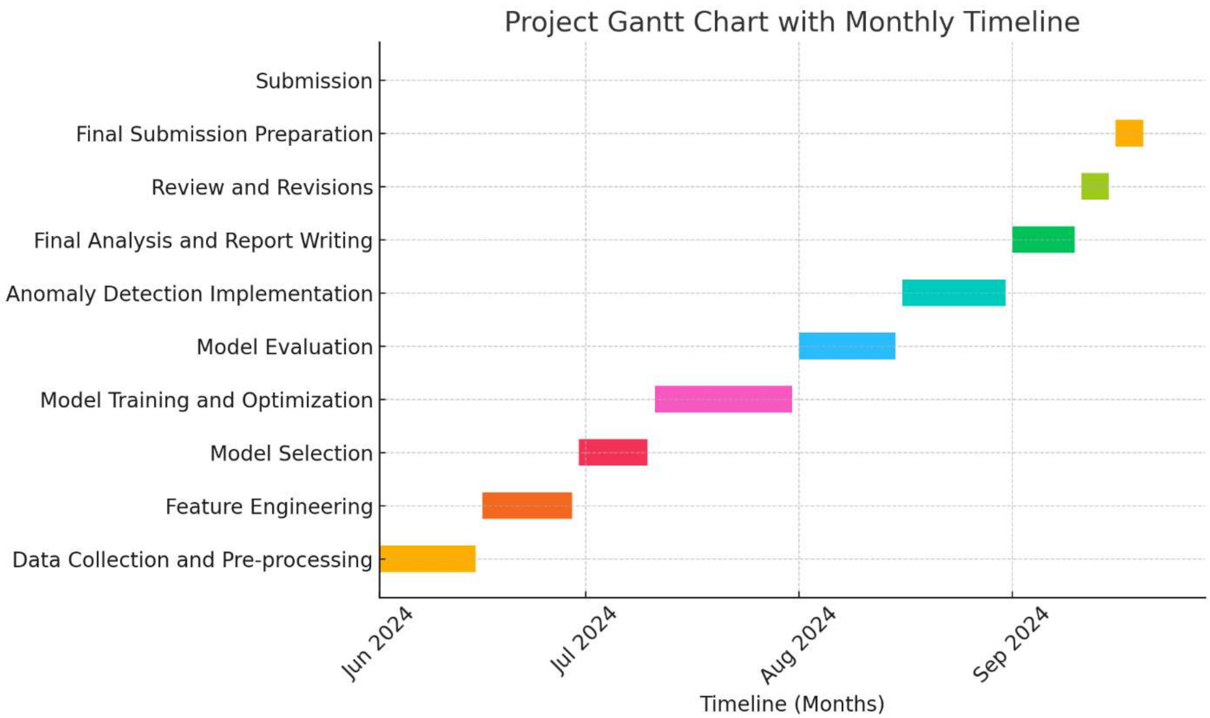

Figure 1: Project Gantt Chart with Monthly Timeline for Key Phases

- **Finalizing the Project Scope and Objectives:**
  The first tasks involved finalizing the projects scope and securing approval by submitting a form to the ethics committee to ensure compliance, with all guidelines specifically regarding the utilization of the LEAD1.0 and UCI Individual Household Electric Power Consumption datasets which are both open source and do not contain any personally identifiable information (PII). These datasets have been sanctioned for applications related to energy prediction and identifying anomalies.

- **Data Collection and Preprocessing:**
  After receiving approval, we began collecting data. The LEAD1.0 dataset, with over 12 million time-series data points, required extensive preparation for modelling readiness. Likewise, the UCI dataset required cleansing and feature extraction, including the creation of temporal variables such as hour, day, and month for time-series analysis. The tasks were planned to take place early in the project timetable, as seen in the Gantt chart.

- **Feature Engineering:**
  After finishing the data preparation phase and moving on to feature enhancement work began by

extracting components, from the data such, as log transformed power measurements and daily patterns of energy consumption trends which played a significant role in improving the performance of prediction and anomaly identification algorithms.

- **Supervisor Meetings**:
  Throughout the project, I remained in constant communication with my supervisor, holding meetings whenever I needed her guidance. This ongoing support was crucial for keeping the project on track, allowing us to address challenges promptly and refine the methodology as necessary.

- **Model Development and Selection:**
  The subsequent part of the research included the development of the Hybrid SWT + LSTM model for energy forecasting. Simultaneously, several anomaly detection methods, including Isolation Forest, Autoencoders, and Local Outlier Factor (LOF), were examined. This phase was planned to occur simultaneously with the feature engineering and preprocessing stages to enhance the project timetable.

- **Model Training and Optimization:**
  Once the models were selected, training began. This phase required significant computational resources and was performed using Jypiter Notebook, Visual Studio, and Google Collab. Model training and hyperparameter tuning were critical for ensuring the best performance on the validation datasets from both LEAD1.0 and UCI datasets.

- **Model Evaluation:**
  The model assessment phase concentrated on assessing the hybrid model's efficiency, using measures such as MSE for energy forecasting and precision/recall for anomaly detection. Modifications were made according to assessment outcomes to enhance performance.

- **Anomaly Detection Implementation:**
  Emphasis was placed on the execution of anomaly detection models. Every model underwent rigorous evaluation on the LEAD1.0 dataset to assess its performance relative to industry benchmarks.

- **Final Analysis and Report Writing:**
  The concluding phase of the project included consolidating the results from both forecasting and anomaly detection jobs into a detailed report. The model results and their implications for practical energy anomaly detection were examined and recorded.

Risk Assessment:

1. **Data Quality and Availability:** Handling the large volume of data from the LEAD1.0 dataset posed challenges. Preprocessing involved creating new categorical features like hour, day, and month to better manage time-series patterns
2. **Model Underperformance:** A backup plan was set in place for employing alternative models, such as CNN-LSTM, if the hybrid model did not yield satisfactory result Ethical Planning.

3.1 Model Development and Optimization

Hybrid SWT + LSTM model: Model development for energy forecasting began after the completion of the feature engineering process. I started exploring other anomaly detection models:

Isolation Forest, Autoencoders, and Local Outlier Factor. As mentioned before, I evaluated all these models in parallel to the feature engineering process as they required fewer steps and were easy to integrate into the project.

Model training and parameter tuning: The model training process required significant computational power, and tools such as Google Colab were essential in handling this demand. During this phase, hyperparameter tuning was performed to ensure the models were optimized for performance on both the LEAD1.0 and UCI datasets.

3.2 Project Deviations and Adjustment

Generally, the project followed the plan precisely. However, to present an overall view of how the proposed timeline and plan went, below are some deviations:

Data preprocessing took longer: Because of the enormous size and complexity of the LEAD1.0 dataset, the cleaning process took longer than projected. However, this did not affect the whole project since I made adjustments in the development part.

Refinement of the feature engineering process: Following the weekly feedback with my supervisor, I refined the process to allow more temporal variables. This move improved the forecasting process.

3.3 Ethical Planning

Both the LEAD1.0 and UCI Individual Household Electric Power Consumption datasets are open-source and publicly available. This negated the need for anonymization, as all data is free from any personally identifiable information (PII). However, ethical considerations were still observed in the form of responsible usage of the datasets to ensure that conclusions drawn from the analysis were not misleading or harmful. An ethics form was filled out and submitted for ethics approval to the ethics board to ensure compliance with all relevant ethical standards.

- **LEAD1.0 Dataset**[1]: Given its relevance to sustainability goals, researchers are encouraged to use the dataset in projects that promote energy conservation. Its application in detecting anomalies aligns with efforts to reduce energy waste and improve operational efficiency in commercial buildings.
- **UCI Dataset**[2]: This dataset offers a granular view of household power consumption, helping in understanding domestic energy patterns. Its responsible usage in forecasting models can help consumers improve their energy efficiency, reducing consumption during peak hours or in inefficient appliances.

Both datasets support global energy-saving initiatives, contributing to sustainability, and are widely adopted in research projects concerning energy systems, anomaly detection, and smart metering infrastructure.

[1] https://github.com/samy101/lead-dataset/blob/main/data/lead1.0-small.zip

[2] Individual Household Electric Power Consumption - UCI Machine Learning Repository

4. Methodology

This project integrates energy consumption forecasting with anomaly detection using a hybrid **Long Short-Term Memory (LSTM)** and **Stationary Wavelet Transform (SWT)** model. Two datasets were used: the UCI Individual Household Electric Power Consumption dataset and the LEAD1.0 dataset. The UCI dataset was primarily used for energy consumption forecasting, while the LEAD1.0 dataset, with its labelled anomalies, was leveraged to evaluate anomaly detection models. However, integrating forecasting with anomaly detection posed significant challenges, especially due to the lack of labelled anomalies in the UCI dataset and the imbalanced nature of the LEAD1.0 dataset.

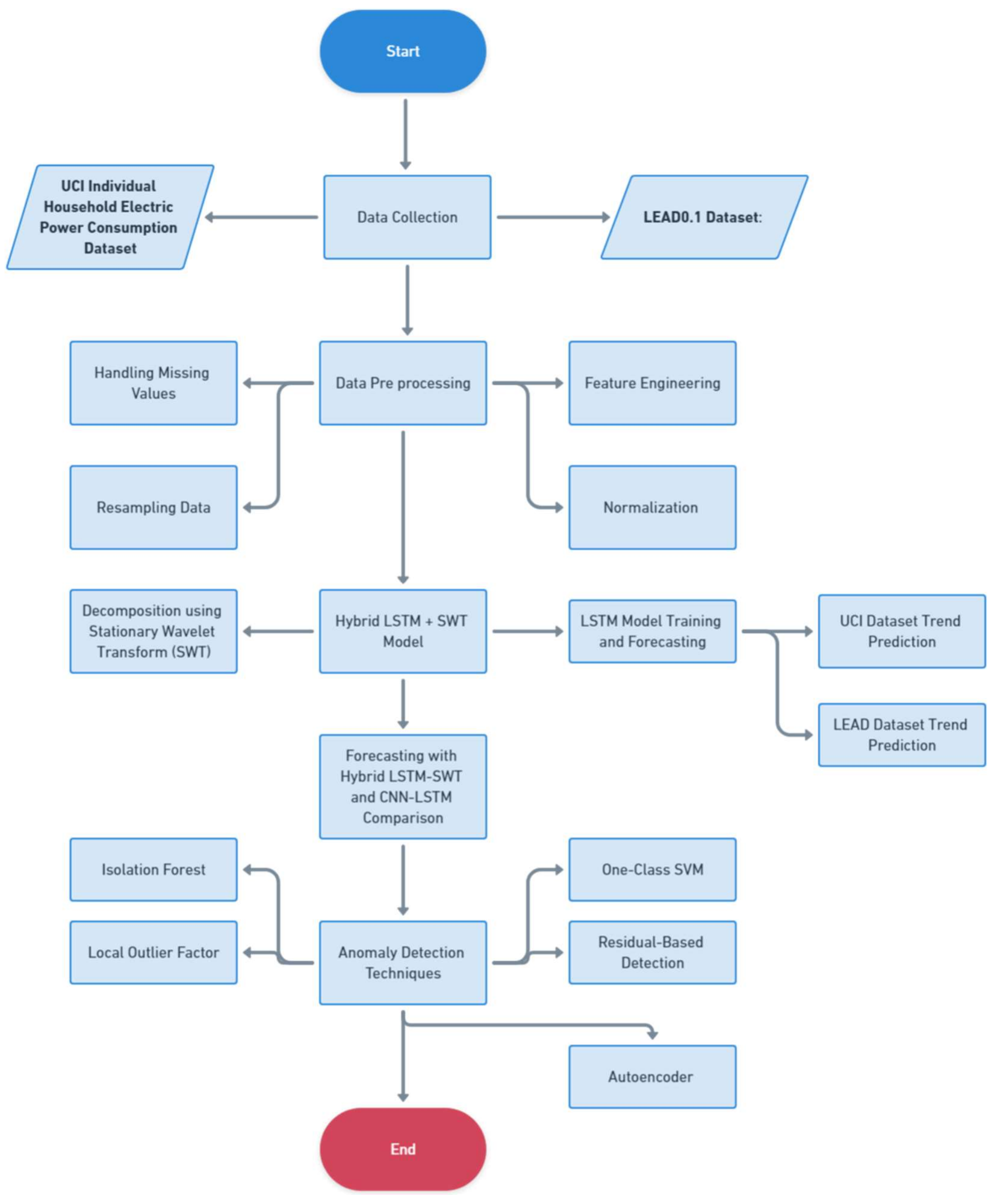

Figure 2: Flowchart of the Methodology for Energy Consumption Forecasting and Anomaly Detection

4.1 Data Collection and Preprocessing

4.1.1 Dataset

Two different sets of data were utilized in this research. Each dataset had its set of difficulties, and each played a specific role:

- **UCI Individual Household Electric Power Consumption Dataset**: This dataset contains minute-level readings from households, tracking important features such as Global Active Power, Voltage, and Global Intensity which is being visualized in the figure 3. It was mainly used for energy forecasting. Although this dataset does not include labelled anomalies, it is well-suited for unsupervised anomaly detection techniques.

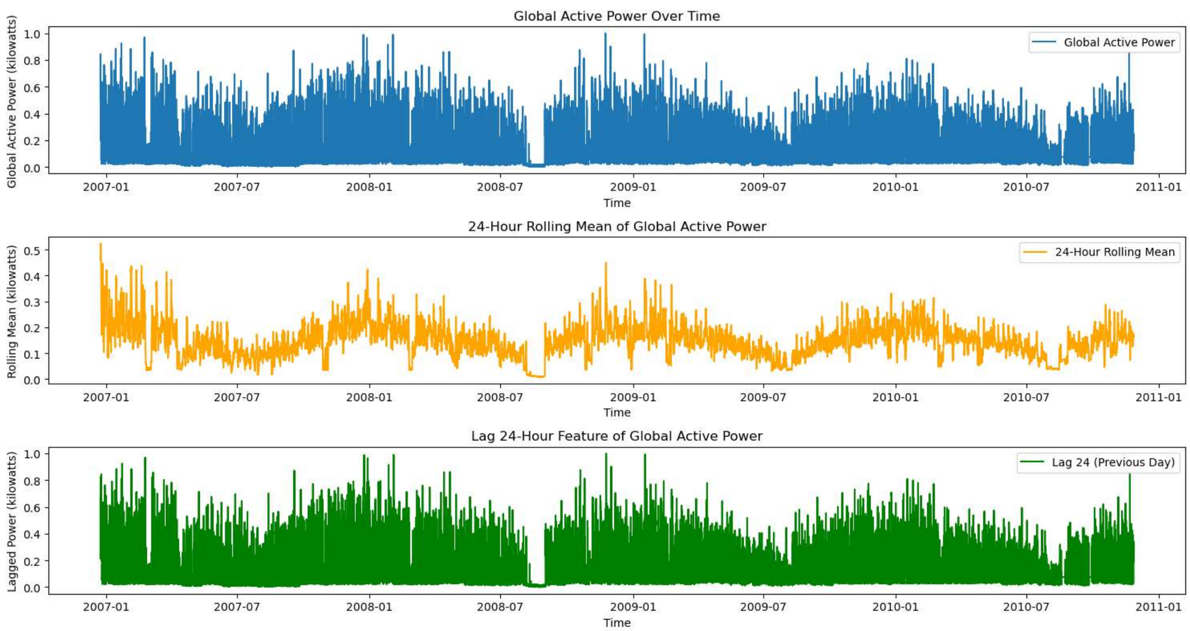

Figure 3: UCI Dataset Global Active Power Over Time with 24-Hour Rolling Mean and 24-Hour Lagged Feature

- **LEAD1.0 Dataset**: The dataset consists of energy readings, from commercial buildings, this dataset contains labelled anomalies and was used for evaluating anomaly detection models. However, it posed several challenges, including imbalanced data and limited variation in the meter_reading feature, which complicated model performance and trend detection (Figure 4).

While this dataset was valuable for assessing anomaly detection models, its limitations necessitated extensive tuning of the models to achieve optimal results.

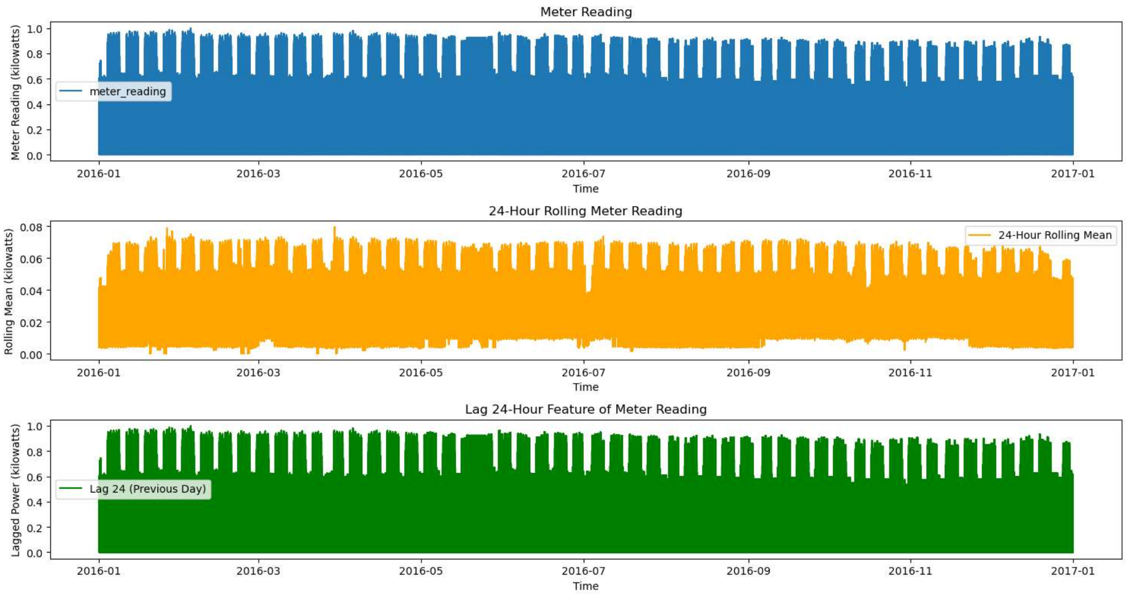

Figure 4: Figure 4: LEAD Dataset Meter Reading Over Time with 24-Hour Rolling Mean and 24-Hour Lagged Feature

4.1.2 Preprocessing

The preprocessing phase was critical to transforming the datasets into a format suitable for time-series modelling and anomaly detection. It involved handling missing data, resampling, normalizing the features, and creating new features to better capture the temporal nature of energy consumption patterns.

- **Handling Missing Values:** In the UCI dataset, approximately 26,000 missing values were found in key features such as Global Active Power and Voltage, which were filled using the median value for each column. In contrast, the LEAD1.0 dataset had a greater number of missing values, which were handled by dropping rows with missing data. Despite this, the lack of variability in the LEAD1.0 dataset's meter readings continued to pose challenges for model training.
- **Resampling to Hourly Data:** To reduce noise and enable long-term trend analysis, minute-level readings in the UCI dataset were resampled to hourly intervals. This step ensured consistency between the UCI and LEAD1.0 datasets, as the latter was already recorded at hourly intervals.
- **Normalization:** To prevent any feature from disproportionately influencing the model, normalization was applied to bring all feature values into a common range between 0 and 1. This was crucial for training the model effectively across datasets.
- **Feature Engineering:** To capture temporal patterns, I created lag features for intervals of 1 hour, 24 hours (1 day), and 168 hours (1 week). Rolling statistics (mean and standard deviation) were computed over 24-hour windows to capture short-term fluctuations. Additional time-based features such as Year, Month, Day, Hour, and Day_of_week were extracted to account for seasonal and cyclical trends in energy consumption.

- **Handling Imbalanced Data:** The LEAD1.0 dataset posed significant difficulties due to its imbalanced nature only a small proportion of the data was labelled as anomalies. To address this, I applied SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic samples for the minority class, balancing the data and ensuring that the models had enough anomalous data to train on.

Despite these preprocessing efforts, the flat meter readings and anomaly feature in the LEAD1.0 dataset remained challenging, and thus, the UCI dataset became the primary focus for energy forecasting and anomaly detection.

4.2 Forecasting with Hybrid LSTM-SWT

In this project, I employed a hybrid forecasting model that integrates **Stationary Wavelet Transform (SWT)** with **Long Short-Term Memory (LSTM)** to predict energy consumption patterns. This combination of techniques was designed to effectively capture both long-term trends and short-term fluctuations in the data, while also minimizing the impact of noise, a common challenge in time-series data such as energy consumption.

The **Stationary Wavelet Transform (SWT)** plays a critical role in preprocessing the time-series data by decomposing it into different frequency components. SWT is a wavelet transformation that is often used in signal processing because of its shift-invariant property, which makes it highly suitable for time-series data (Mallat, 1989). Specifically, SWT breaks the data into two key components: **approximation coefficients**, which capture the general, long-term trends in the data, and **detail coefficients**, which represent short-term fluctuations or noise. By retaining the time-alignment of the data, SWT ensures that the decomposed components remain directly comparable to the original series, which is vital for accurate temporal modelling.

In my implementation, I applied SWT using the 'db1' wavelet function, which is commonly used in time-series decomposition due to its ability to preserve the main features of the data. The SWT decomposition was applied at the maximum level permissible by the length of the data, ensuring a thorough breakdown of the time series into its constituent components (Mallat, 1989). Since SWT requires the input data length to be even, I padded the data by duplicating the last value, ensuring that the decomposition could be carried out without altering the underlying structure of the time series. The output of the SWT was a set of **approximation coefficients** and **detail coefficients** for each level of decomposition. However, for the purpose of forecasting, I focused on the approximation coefficients, as these capture the broader trends in energy consumption and are less affected by short-term noise.

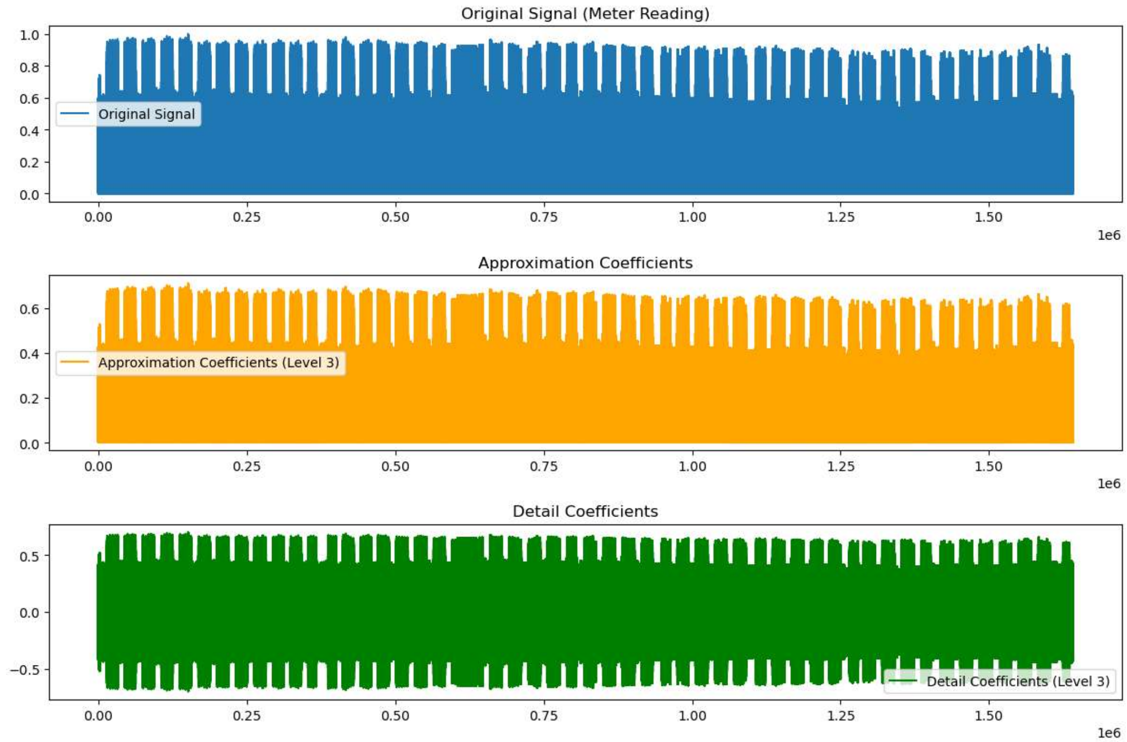

***Figure 5**: SWT Decomposition of Meter Reading Signal into Approximation and Detail Coefficients.*

The approximation coefficients extracted from the first level of SWT decomposition were then fed into the **Long Short-Term Memory (LSTM)** network. **LSTM** is a type of recurrent neural network (RNN) that is designed specifically to handle long-term dependencies in sequential data, addressing the vanishing gradient problem that limits the ability of traditional RNNs to retain long-term information. LSTM networks achieve this by introducing **memory cells** and three types of gates **input gate**, **forget gate**, and **output gate** that regulate the flow of information through the network (Hochreiter and Schmidhuber, 1997).

The **input gate** controls what portion of the current input should be stored in the memory cell, the **forget gate** determines which parts of the past information should be forgotten or retained, and the **output gate** decides how much of the current memory state should be passed to the next time step. This gated architecture makes LSTM particularly well-suited for time-series forecasting tasks where both short-term and long-term patterns are important. In the context of this project, the use of memory cells allowed the model to retain relevant energy consumption trends over long periods while filtering out short-term fluctuations that might distort the forecasts.

In my project, the **approximation coefficients** from SWT were used as input to the LSTM model. The approximation coefficients, representing the long-term trends in the data, provide a cleaner, more stable signal for the LSTM to process. This focus on long-term patterns allowed the LSTM to learn the underlying structure of energy consumption over time, without being distracted by short-term noise. The use of SWT as a preprocessing step significantly enhanced the LSTM's ability to generalize and predict future consumption accurately.

To implement the forecasting process, I used a **sliding window approach**, a widely accepted technique in time-series forecasting. In this approach, the model is trained on sequences of historical data, where each sequence consists of the last 24 hours of energy consumption. The model then uses these sequences to predict the next hour of consumption. This window-based method allows the model to continuously update its predictions based on the most recent data, ensuring that it adapts to evolving trends in energy consumption over time. By using overlapping windows, the model is exposed to a diverse set of training sequences, which helps it generalize better to unseen data (Hochreiter and Schmidhuber, 1997).

The architecture of the LSTM model consisted of a single LSTM layer with 15 units, which was followed by a dense layer to generate the output. To prevent overfitting, I applied **L2 regularization** to the LSTM layer. L2 regularization penalizes large weights by adding a regularization term to the loss function, which encourages the model to distribute its learning across all features rather than relying too heavily on any single feature (Ng, 2004). This was particularly important given the noisy nature of energy consumption data, where overfitting to the training data could easily occur without proper regularization.

The model was compiled using the **Adam optimizer**, a popular optimization algorithm that combines the advantages of both **Stochastic Gradient Descent (SGD)** and **RMSProp** by adapting the learning rate for each parameter based on the first and second moments of the gradient (Kingma and Ba, 2017). Adam is well-suited for this task as it converges faster and is less sensitive to hyperparameter tuning, making it ideal for large datasets and complex models like LSTM. The loss function used was **Mean Squared Error (MSE)**, which is a standard loss function for regression tasks and measures the squared difference between the predicted and actual values.

During training, I implemented **early stopping** to avoid overfitting. Early stopping monitors the validation loss and halts training when the validation loss stops improving for a specified number of epochs (5 in this case). This technique helps prevent the model from learning noise in the training data, especially in time-series data where overfitting is common (Prechelt, 1998).

Once the model was trained, I evaluated its performance on the test set, which consisted of 80% of the data that was held out during training. The model's performance was measured using **Mean Squared Error (MSE)** and **Mean Absolute Error (MAE)**. MSE calculates the average of the squared differences between the predicted and actual values, making it sensitive to large errors. MAE, on the other hand, measures the average absolute differences between the predicted and actual values, providing a more interpretable metric for understanding the overall accuracy of the predictions (Willmott and Matsuura, 2005).

The hybrid LSTM-SWT model was particularly effective in forecasting energy consumption because it leveraged the strengths of both methods. SWT's ability to decompose the time series into long-term trends and short-term noise allowed the LSTM to focus on the most relevant patterns for forecasting. The LSTM, with its memory cells and gating mechanism, was able to retain these important patterns over time and use them to predict future energy consumption accurately.

4.3 Model Comparison: CNN-LSTM vs. SWT-LSTM

In order to evaluate the performance of the hybrid SWT-LSTM model, it was compared against a Convolutional Neural Network (CNN)-LSTM model to test how each model handled the same time-series data for energy consumption forecasting. The CNN-LSTM model integrates CNNs, which specialize in detecting local patterns, with LSTM, which captures long-term dependencies in sequential data. By contrast, the SWT-LSTM model uses Stationary Wavelet Transform (SWT) to preprocess the data before passing it into LSTM layers, focusing primarily on long-term trends by filtering out short-term noise. Each model offers distinct advantages for time-series forecasting, and this comparison highlights the key strengths and weaknesses of each.

CNN-LSTM Model Architecture

The CNN-LSTM model combines two architectures: Convolutional Neural Networks (CNNs), which are commonly used for pattern recognition in time-series data, and Long Short-Term Memory (LSTM) networks, which are designed to learn temporal dependencies over extended periods. The CNN component of the model is particularly effective in capturing short-term patterns in the data, such as spikes or drops in energy consumption over short intervals. By applying convolutional filters, CNN layers can identify localized features, focusing on short-term dependencies that are crucial in understanding temporary changes in energy usage. CNNs are widely recognized for their ability to detect such patterns, having been originally developed for image processing and later adapted to handle sequential data effectively (Lecun *et al.*, 1998).

Once these short-term patterns are identified, they are passed to the LSTM component of the model, which processes the data sequentially to learn long-term dependencies. The LSTM network is particularly effective in retaining information from past time steps due to its memory cells, which are regulated by input, forget, and output gates. These gates allow the LSTM to manage the flow of information over time, retaining important details while discarding irrelevant ones. This architecture enables the CNN-LSTM model to handle both short-term fluctuations and long-term trends in energy consumption data (Bai, Kolter and Koltun, 2018).

In this implementation, sequences of 24-hour intervals of historical energy consumption data were passed through the CNN layers, which performed convolution operations to extract essential features. These features were then reshaped and passed into the LSTM layers for time-series forecasting. This architecture allowed the CNN-LSTM model to effectively capture both local fluctuations and overall trends, with the CNN layers focusing on short-term dependencies and the LSTM layers learning the long-term temporal relationships. This combination makes CNN-LSTM especially useful in contexts where both short-term and long-term dependencies are present in the data.

Implementation in Code

In both the CNN-LSTM and SWT-LSTM models, the data preprocessing pipelines were similar to ensure a fair comparison. The input time-series data was normalized and resampled into hourly intervals to ensure consistency in the structure of the data used for training both models.

- In the CNN-LSTM model, the normalized input data was first passed through convolutional layers, which applied filters to extract local features from the time series. These features were then reshaped and passed into the LSTM layers, where the temporal relationships were learned for forecasting energy consumption.
- In the SWT-LSTM model, the data was first passed through SWT decomposition, where it was separated into approximation and detail coefficients. Only the approximation coefficients, which capture the long-term trends, were passed into the LSTM layers for forecasting. This preprocessing step ensured that the LSTM layers could focus exclusively on the broader patterns in the data.

Both models were trained using sequences of 24-hour windows to predict the next hour's energy consumption. The models used the backpropagation algorithm and gradient descent to minimize the Mean Squared Error (MSE) between the predicted and actual energy consumption values. The Adam optimizer was used for efficient weight updates during training, ensuring that the models could learn effectively from the data (Kingma and Ba, 2017).

Summary of the Comparison

The comparison between CNN-LSTM and SWT-LSTM highlights their respective strengths and weaknesses in handling time-series data:

- The CNN-LSTM model excels at capturing short-term dependencies by identifying local patterns in the data. This makes it particularly effective for datasets where short-term fluctuations are significant, such as sudden changes in energy consumption caused by external factors. However, because CNN-LSTM processes raw data without filtering, it may struggle to capture long-term trends when short-term noise interferes with the broader patterns.
- The SWT-LSTM model, on the other hand, is specifically designed to handle long-term dependencies by filtering out short-term fluctuations through wavelet decomposition. This enables the model to focus entirely on the long-term trends, making it more effective for datasets where long-term stability and trend forecasting are the primary goals. The preprocessing step in SWT-LSTM ensures that the LSTM layers are not distracted by short-term noise, which gives the model an advantage in noisy datasets like energy consumption forecasting.

In conclusion, the CNN-LSTM model is better suited for tasks where short-term variability is important to capture, while the SWT-LSTM model is more effective in scenarios where long-term trends need to be prioritized and noise minimized. The strengths of each model reflect the different approaches they take in handling time-series data, and their comparison provides valuable insights into their respective performance in energy consumption forecasting.

4.4 Integrating Forecasting with Anomaly Detection

One of the key objectives of this project was to integrate forecasting with anomaly detection. The residuals (i.e., the difference between the actual and predicted values) generated by the LSTM-SWT model were used to identify unusual energy consumption patterns. Given that the UCI dataset lacked labelled anomalies, unsupervised anomaly detection methods were employed. I implemented several techniques, including Isolation Forest, Local Outlier Factor (LOF), One-Class SVM, and residual-based anomaly detection.

For the UCI dataset, anomaly detection was carried out without ground truth labels. The focus was on detecting deviations in consumption patterns that were considered "unexpected" based on the forecasting model’s predictions. Since no labelled anomalies were available, formal evaluation metrics like precision or recall could not be calculated. Instead, a qualitative evaluation was conducted, examining whether the detected anomalies corresponded to significant spikes or drops in energy consumption that deviated from expected trends.

In contrast, the LEAD1.0 dataset provided labelled anomalies, allowing for a quantitative evaluation of the anomaly detection models. Here, the primary goal was to detect both global anomalies (significant deviations across the entire dataset) and local anomalies (deviations within specific time windows). Due to the dataset's limited variability, anomaly detection proved more difficult than anticipated, particularly for models like One-Class SVM, which struggled to differentiate between normal and anomalous behaviour in such a uniform dataset.

Challenges and Solutions

The integration of real-time anomaly detection with energy forecasting presented several challenges. First, real-time analysis of residuals required substantial computational resources, particularly for large datasets like UCI. To address this, I optimized the anomaly detection process using parallel computing techniques, which allowed the system to detect anomalies in a timely manner without overloading computational resources.

For the UCI dataset, the absence of labelled anomalies necessitated a qualitative evaluation approach. Instead of using standard classification metrics, I observed how well the models identified unusual patterns based on the residuals generated by the LSTM-SWT model. The focus was on identifying significant deviations in energy consumption that could indicate possible anomalies, even in the absence of labelled data.

Another significant challenge was the imbalanced nature of the LEAD1.0 dataset, where anomalies comprised only a small fraction of the data. While SMOTE helped balance the dataset, the lack of variability in the meter readings persisted, making it difficult for the models to capture meaningful patterns. This required extensive tuning of the models and careful adjustment of the anomaly detection thresholds to improve performance.

4.5 Model Tuning and Evaluation

Extensive hyperparameter tuning were made to tune the anomaly detection models, for performance. For instance, the contamination rate in the Isolation Forest was tailored to match the anticipated anomalies ratio. The number of neighbours, in LO F was customized to detect anomalies accurately.

- **Evaluation Metrics**: For the LEAD1.0 dataset, standard classification metrics such as **precision, recall, and F1-score** were used to evaluate anomaly detection. For the UCI dataset, qualitative evaluation was based on visual inspection of the residuals.
- **Defining Metrics**:
    - *Root Mean Squared Error (RMSE)*: Used for evaluating forecasting accuracy, it measures the square root of the average squared differences between predicted and actual values.
    - *Mean Absolute Error (MAE)*: Measures the average absolute differences between predicted and actual values, providing an indication of overall model performance.
    - *Precision*: Measures the proportion of true positive anomalies among all detected anomalies.
    - *Recall*: Measures the proportion of true positive anomalies detected among all actual anomalies.
    - *F1-score*: The harmonic means of precision and recall, providing a balance between the two.

Model Evaluation

The evaluation of the models varied depending on the dataset. For the LEAD1.0 dataset, which contained labelled anomalies, I used standard classification metrics such as precision, recall, and F1-score to evaluate the performance of the anomaly detection models. Although the limited variability in the data made anomaly detection challenging, the models were still able to capture some anomalies.

For the UCI dataset, the absence of labelled anomalies meant that I had to rely on a qualitative evaluation. I examined the residuals generated by the LSTM-SWT model and observed whether the detected anomalies corresponded to significant spikes or drops in energy consumption that deviated from expected trends. Although this approach lacked the precision of quantitative evaluation methods, it still provided valuable insights into the models' ability to detect unusual patterns in the data.

5. Results and Discussion

In this section, the performance of both energy consumption forecasting and anomaly detection methods is discussed in detail. The Hybrid LSTM + SWT model was applied to both the UCI and LEAD datasets, whereas the CNN-LSTM model was tested only on the UCI dataset. The forecasting models were evaluated using common performance metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Following forecasting, anomaly detection was performed using the residuals from the Hybrid LSTM + SWT model. The performance of four anomaly detection methods Residual-Based Detection, Isolation Forest, Local Outlier Factor (LOF), and Autoencoder was evaluated based on their ability to distinguish between normal and anomalous energy consumption patterns. Precision, Recall, and F1-Score were used to assess the performance of these anomaly detection methods.

5.1 Forecasting Results

Hybrid LSTM + SWT on UCI Dataset

The Hybrid LSTM + SWT model was first applied to the UCI dataset, which consists of household energy consumption data. This dataset presents typical cyclical patterns due to daily human activities, making it well-suited for time-series forecasting. The Stationary Wavelet Transform (SWT) was used to decompose the energy data into approximation coefficients, which capture long-term trends, and detail coefficients, which represent short-term fluctuations. The LSTM model then processed the approximation coefficients to predict hourly energy consumption.

- **RMSE**: 0.0220
- **MAE**: 0.0125

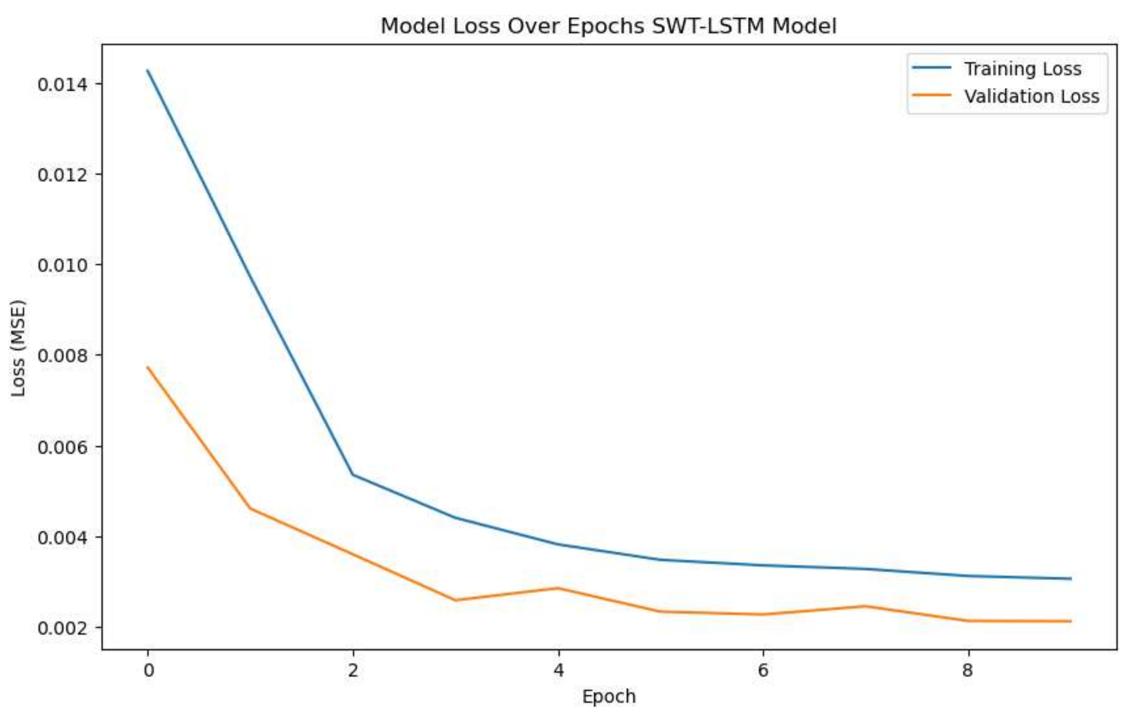

Figure 5: Model Loss Over Epochs for SWT-LSTM Model on UCI Dataset

Detailed Analysis:

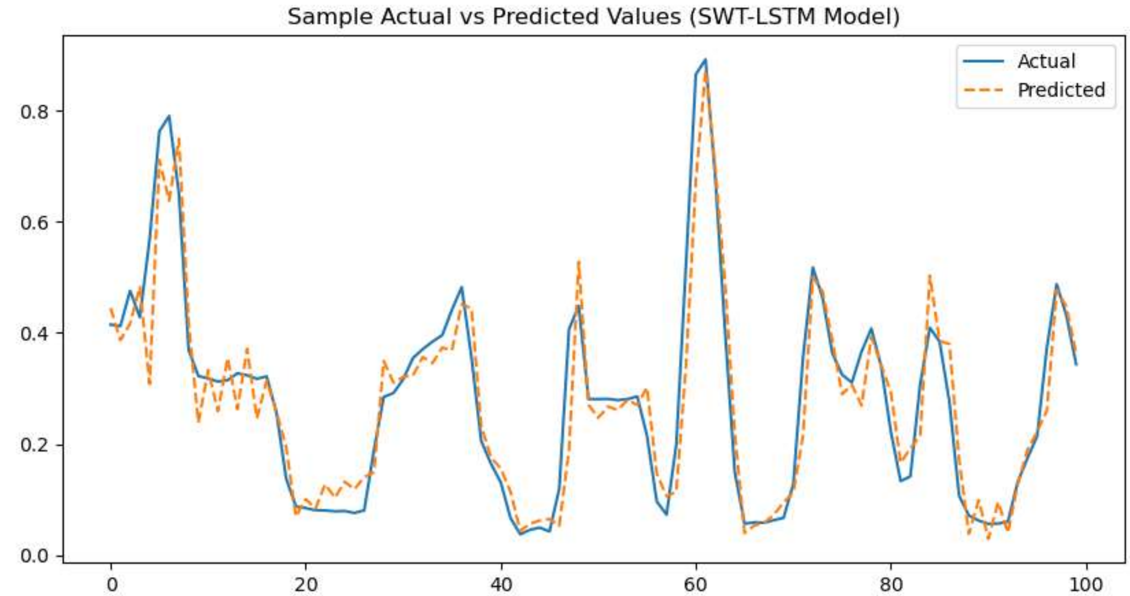

Figure 6: Sample Actual vs Predicted Values Using SWT-LSTM Model on UCI Dataset

- The Hybrid LSTM + SWT model performed exceptionally well on the UCI dataset, as reflected by its low RMSE and MAE values. The SWT decomposition was particularly effective in filtering out short-term noise, allowing the LSTM component to focus on long-term consumption patterns. By decomposing the data into different frequency bands, the model could better isolate the cyclic nature of household energy usage, including peaks during morning and evening hours.
- As seen in Figure 6 (Sample Actual vs. Predicted), the model closely followed the actual energy consumption patterns. The model captured major peaks and troughs in the data, which typically occur during high activity periods such as morning and evening when appliances like heaters or air conditioners are used.
- The effectiveness of this model is evident in its ability to generalize well across the entire time series, successfully predicting future energy consumption with minimal error.

This result showcases the advantage of hybrid models like LSTM + SWT, which can capture both temporal dependencies through the LSTM and handle noise and short-term fluctuations through wavelet transforms. The UCI dataset benefits from the model’s ability to process the cyclical nature of the data, effectively capturing seasonality and sudden spikes in energy consumption.

5.1.1 Hybrid LSTM + SWT on LEAD Dataset

The LEAD dataset, consisting of energy consumption data from commercial buildings, provided additional challenges for the Hybrid LSTM + SWT model. Unlike the UCI dataset, the energy consumption patterns in commercial buildings are typically more stable and exhibit fewer extreme fluctuations. This reduced variability made it more difficult for the model to capture meaningful trends.

- **RMSE**: 0.0386

- **MAE**: 0.0284

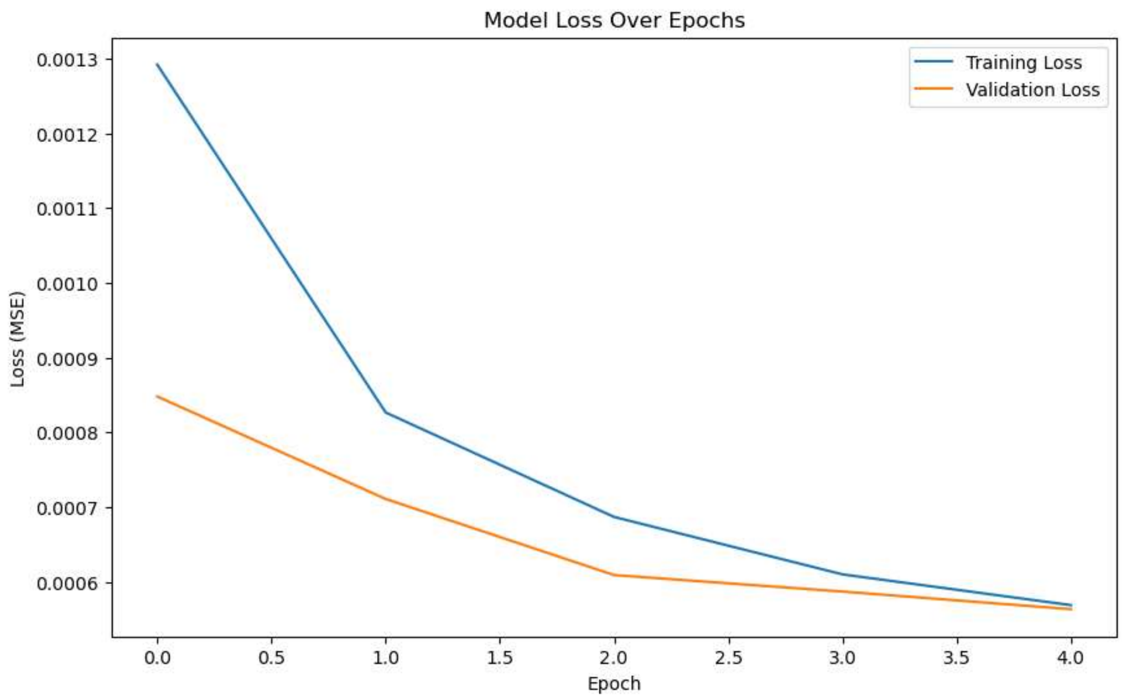

Figure 7: Model Loss Over Epochs for SWT-LSTM Model on LEAD1.0 Dataset

Detailed Analysis:

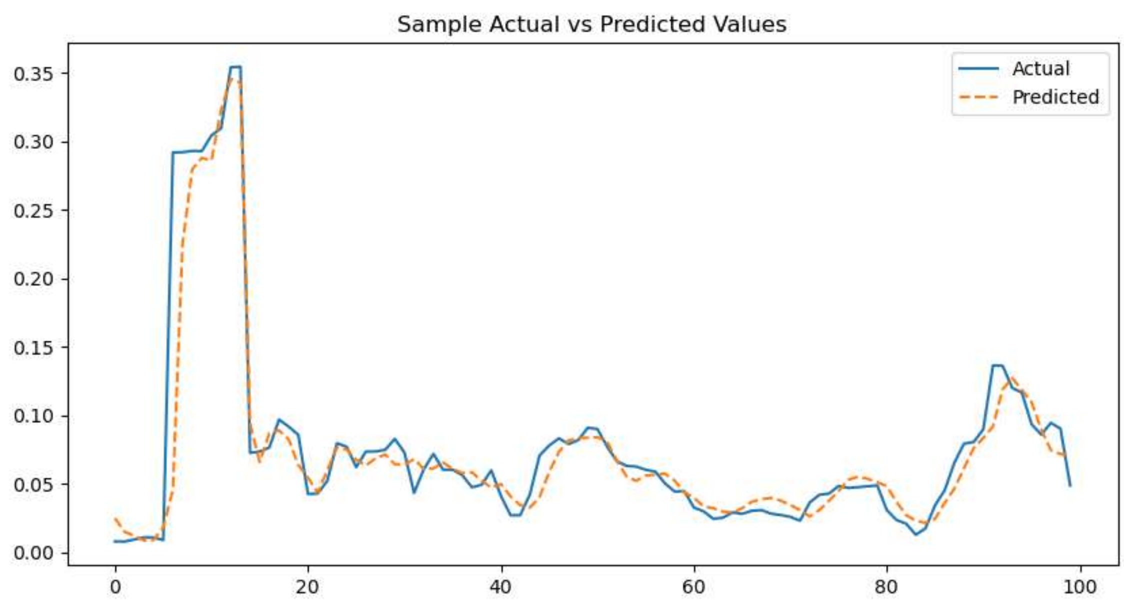

Figure 8: Sample Actual vs Predicted Values Using SWT-LSTM Model on LEAD1.0 Dataset

- Compared to the UCI dataset, the model's performance on the LEAD dataset was less accurate, as reflected in the higher RMSE and MAE values. This result was expected due to the flat nature of the data and the lack of significant variation in energy consumption patterns. The commercial buildings represented in the LEAD dataset typically have consistent energy consumption due to continuous operation of systems like 24/7 Air conditioning and lighting, unlike the cyclical nature of household energy consumption.

- Despite these challenges, the Hybrid LSTM + SWT model managed to capture minor fluctuations and was able to predict energy consumption with reasonable accuracy. The SWT helped filter out any residual noise, but the relatively uniform data made it harder for the model to detect distinct trends.
- In Figure 8, the model's predictions are shown to follow the general trend of the actual data, although there are some noticeable discrepancies during periods of lower energy usage, where the model tends to overpredict. This could be due to the model's difficulty in identifying subtle shifts in consumption in a dataset with lower variability.

This suggests that while the Hybrid LSTM + SWT model is capable of performing well on datasets with significant temporal patterns, it faces limitations in environments with less variability. Further improvements could involve incorporating additional features, such as external factors (e.g., temperature, building occupancy), to enhance the model's ability to forecast energy consumption in commercial settings.

CNN-LSTM on UCI Dataset

The CNN-LSTM model was applied exclusively to the UCI dataset to compare its performance against the Hybrid LSTM + SWT model. The CNN layers were tasked with capturing short-term dependencies and local patterns, such as sudden spikes or drops in energy consumption, while the LSTM component processed the long-term trends.

- **RMSE**: 0.0584
- **MAE**: 0.0418

Detailed Analysis:

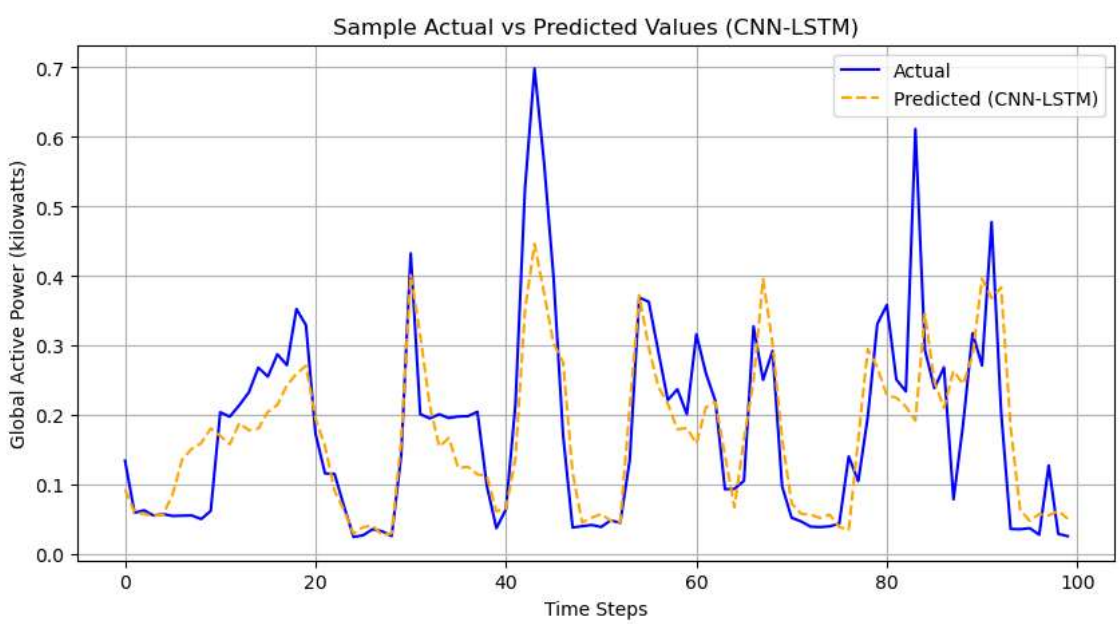

Figure 9: Sample Actual vs Predicted Values Using CNN-LSTM Model on UCI Dataset

- While the CNN-LSTM model performed reasonably well on the UCI dataset, its overall performance was weaker than that of the Hybrid LSTM + SWT model. The higher RMSE and

MAE values indicate that the model struggled to capture the broader, long-term trends present in the dataset.

- The CNN component was effective in identifying local changes in energy consumption, such as short-term spikes that may correspond to events like the switching on or off appliances. However, the model's focus on these short-term dependencies limited its ability to generalize over longer periods, leading to higher errors in overall energy consumption prediction.
- As shown in Figure 9, the model captures the general pattern of energy consumption but misses some of the more significant peaks and valleys, leading to higher prediction errors. This highlights the limitations of CNN-LSTM models when applied to time series data that require both short-term and long-term pattern recognition.

Overall, while the CNN-LSTM model was effective at capturing short-term fluctuations, its limitations in handling long-term trends suggest that it may be better suited for applications where local variations are more important than overall energy trends.

Table 2: **Table of Forecasting Metrics**

Model	**Dataset**	**RMSE**	**MAE**
Hybrid LSTM + SWT	UCI Dataset	0.0220	0.0125
Hybrid LSTM + SWT	LEAD Dataset	0.0386	0.0284
CNN-LSTM	UCI Dataset	0.0584	0.0418

5.2 Integration of Forecasting and Anomaly Detection

1. UCI Dataset: Household Energy Consumption Forecasting and Anomaly Detection

For the UCI Household Energy Consumption dataset, the LSTM + Stationary Wavelet Transform (SWT) model was used for forecasting energy consumption, followed by the application of anomaly detection techniques including Isolation Forest, Autoencoder, and One-Class SVM. These methods allowed the identification of anomalous patterns in energy usage, which may indicate appliance malfunctions or unusual household energy consumption behaviour.

The LSTM + SWT hybrid model provided an effective method for capturing the periodicity and seasonal trends present in household energy consumption. By reducing the noise in the data using SWT, the model could focus on the underlying patterns, improving the accuracy of short-term forecasts.

Anomaly Detection Results:

The following anomaly detection methods were applied to the UCI dataset, with their respective results visualized below

1. **Isolation Forest Anomalies**: The Isolation Forest method, as visualized in the graph (Figure 10), efficiently identifies significant deviations in energy consumption patterns over time. These detected anomalies are marked as red circles and represent periods where the actual energy usage diverged considerably from the forecasted values. In the early part of the time series (January to March), the anomalies are sparse, reflecting more consistent consumption. However, as the year progresses, notably from April to December, there is a notable increase in the frequency and magnitude of the anomalies. These deviations could be linked to various external factors such as equipment failures, malfunctioning appliances, or unexpected spikes

in household energy consumption due to seasonal changes, particularly around summer and winter when heating and cooling systems are in heavy use. The consistent identification of anomalies during peak periods suggests that Isolation Forest is adept at pinpointing not just random spikes, but potential systemic issues that arise during high-load periods. The method's ability to detect both isolated and frequent anomalies demonstrates its strength in handling the complexity and variability of energy consumption data in real-world scenarios. The placement of the anomalies against both actual and forecasted values help underline the utility of this approach in identifying consumption patterns that deviate sharply from normal household usage.

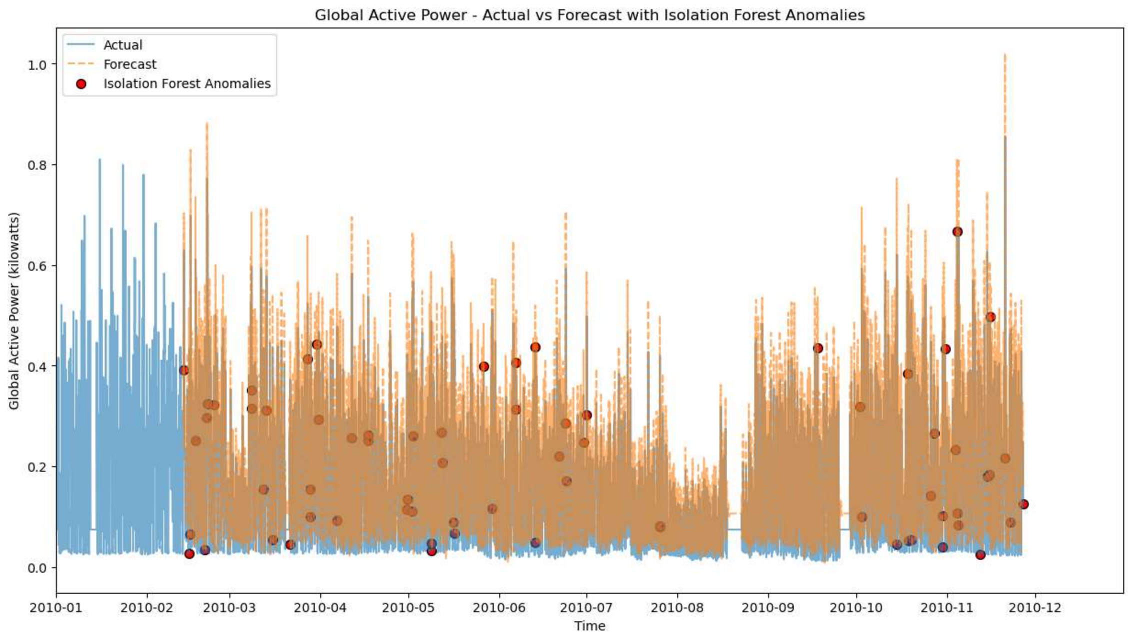

Figure 10: Global Active Power - Actual vs Forecast with Detected Anomalies (Isolation Forest) on UCI Dataset

2. **Autoencoder Anomalies**: In the graph showing Autoencoder anomalies (Figure 11), we observe a higher detection frequency of anomalies compared to other models like Isolation Forest. The Autoencoder works by learning the normal consumption pattern and reconstructing it. Any significant deviation from this learned pattern is flagged as an anomaly. This can be seen in the numerous blue circles spread across the time series, especially during periods of peak energy consumption.

   Unlike the Isolation Forest, which identifies more isolated and distinct outliers, the Autoencoder captures a broader range of anomalies, some of which may be considered more subtle deviations. These anomalies are detected during both low and high energy consumption periods. From February to May 2010, the Autoencoder detects several anomalies during off-peak consumption times, which could indicate potential issues with energy efficiency, such as phantom loads or unexpected consumption from appliances in standby mode.

   From June onwards, as energy usage increases (perhaps due to seasonal factors like air conditioning), the model identifies several more anomalies during peak usage periods. The higher frequency of detected anomalies during these months may suggest either equipment inefficiencies or unusual behavioural patterns in energy usage.

While the Autoencoder is effective at identifying both major and minor deviations, its sensitivity could lead to an increased number of false positives. Nevertheless, the ability to detect both large and small anomalies, particularly during critical consumption periods, underscores its strength in identifying nuanced variations that might otherwise be missed by more conservative models.

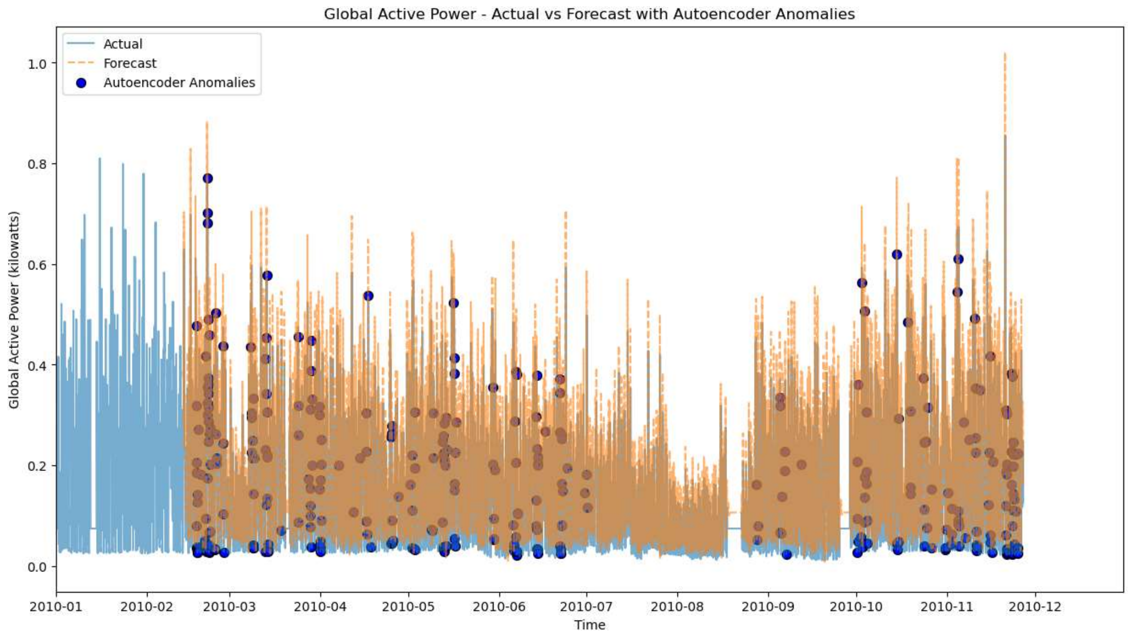

Figure 11: Global Active Power - Actual vs Forecast with Detected Anomalies (Autoencoder) on UCI Dataset

3. **One-Class SVM Anomalies**: One-Class SVM anomalies (Figure 12), the anomalies are marked with green circles, highlighting the points in the time series where the One-Class SVM model detected deviations from expected energy consumption patterns. Compared to Isolation Forest and Autoencoder, the One-Class SVM shows a more conservative approach, identifying fewer anomalies, mostly during the earlier and later parts of the year, specifically between February and March and from October to December. The limited number of anomalies detected suggests that the One-Class SVM is less sensitive to smaller deviations, focusing instead on more prominent outliers. This might indicate its strength in identifying significant disruptions, such as equipment malfunctions or sudden spikes in demand, but it may miss subtler inefficiencies or deviations that models like Autoencoder capture. For example, the anomalies detected between February and March may correspond to isolated events of unusually high energy use, potentially related to seasonal factors like heating systems. Similarly, the anomalies identified in the latter months of the year likely correspond to increased consumption due to winter heating. While this model minimizes false positives by avoiding the detection of minor deviations, it could potentially overlook smaller, yet significant, variations in energy consumption that could point to inefficiencies or emerging issues. Its focus on large anomalies suggests that One-Class SVM is most suitable for identifying major disruptions but may not be ideal for detecting gradual or subtle deviations

in energy usage patterns over time. This trade-off between precision and recall must be considered when deploying this model in real-time anomaly detection systems.

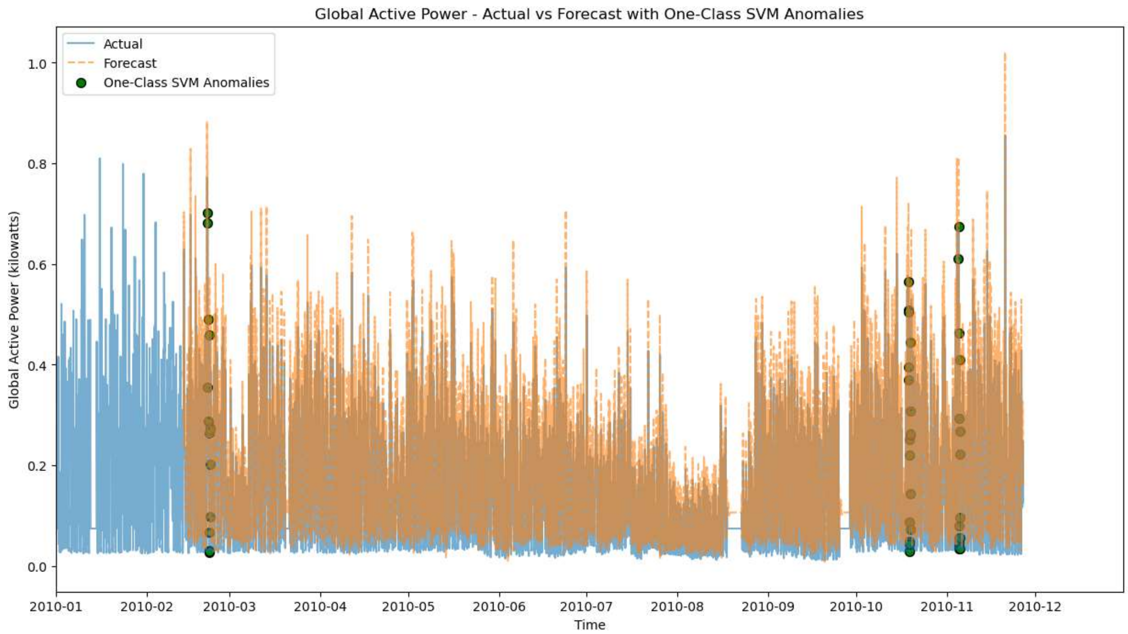

Figure 12: Global Active Power - Actual vs Forecast with Detected Anomalies (One class - SVM) on UCI Dataset

4. **Local Outlier Factor (LOF) -**: **Local Outlier Factor (LOF) anomalies** (Figure 13), the anomalies are represented by yellow circles, indicating points where the LOF algorithm detected deviations from typical energy consumption patterns. LOF is designed to identify data points that have a significantly lower density compared to their neighbours, meaning it focuses on anomalies that are more isolated in their local context. In this graph, most of the anomalies occur in specific clusters, primarily in February and March, with some scattered anomalies detected from June to September. This clustering of anomalies suggests that LOF is particularly effective at detecting periods of relatively isolated, yet significant deviations in energy consumption. However, the method appears to be less sensitive to anomalies in high-energy consumption periods, such as during the summer or winter, when the energy demand is more volatile. The anomalies detected during the low-energy consumption period in early 2010 could correspond to irregular appliance usage or inefficiencies in the system that are more pronounced when the overall consumption is lower. Compared to other methods like Autoencoder or Isolation Forest, LOF identifies fewer anomalies, indicating its tendency to focus on more distinct deviations. This makes LOF a useful method for identifying outliers in specific local clusters but may limit its ability to detect broader, more global anomalies that could be spread across different time intervals. Its precision in identifying localized anomalies makes it particularly suited for situations where subtle but significant irregularities in energy consumption are important to detect, although it may miss out on more complex patterns that

could emerge during periods of high demand.

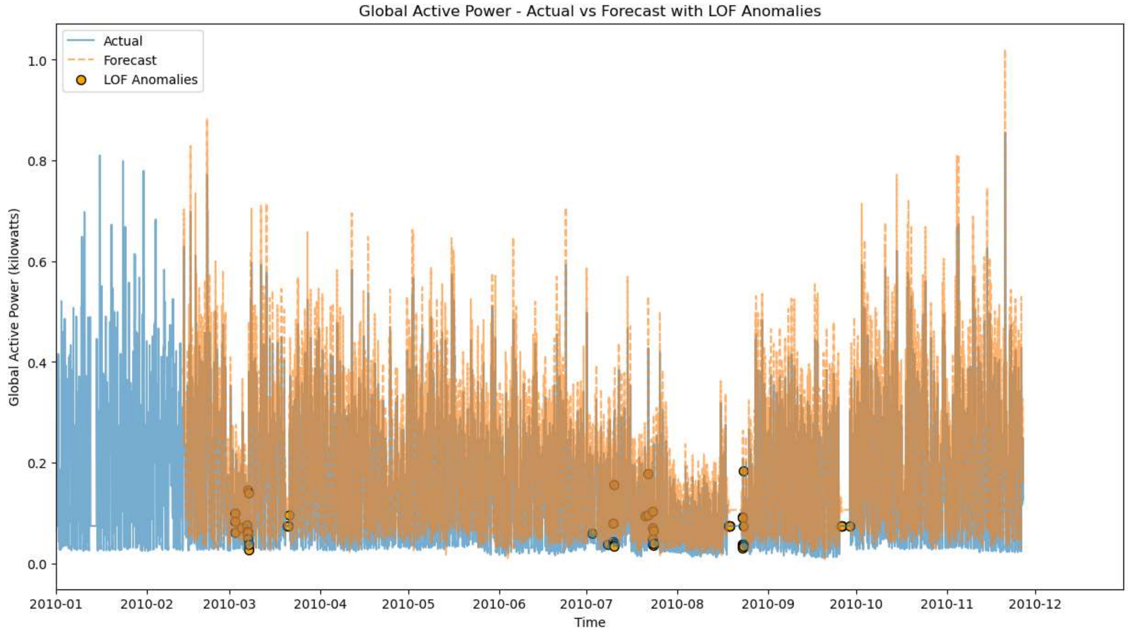

Figure 13: Global Active Power - Actual vs Forecast with Detected Anomalies (LOF) on UCI Dataset

Since the UCI dataset did not come with labelled anomalies and had no way to verify the anomalies in the dataset, I had to search for another dataset. However, due to the lack of labelled anomaly datasets for individual household energy consumption, I opted to use the **LEAD0.1 dataset**. This dataset, although designed for commercial buildings, includes labelled anomalies, which made it suitable for my anomaly detection task.

2. LEAD Dataset: Commercial Energy Consumption Forecasting and Anomaly Detection

The LEAD dataset, which covers commercial building energy consumption, presented greater variability in patterns compared to household data. The LSTM + Stationary Wavelet Transform (SWT) model was applied to forecast energy consumption, and several anomaly detection techniques, such as Isolation Forest, Local Outlier Factor (LOF), Autoencoder, and Residual-Based Anomaly Detection, were used to identify irregularities in the consumption data.

Given the complexity of the dataset, the models faced challenges in accurately predicting the meter readings and detecting anomalies. The results of the anomaly detection methods are visualized in the following graphs:

5.2.1 Anomaly Detection Results:

1. **Residual-Based Anomaly Detection**: In the graph (Figure 14), the predicted meter readings (on the y-axis) are compared against the actual meter readings (on the x-axis). The tightly clustered points in the middle-right section (closer to meter readings of 1.0) exhibit a high density of red markers, indicating that significant anomalies occurred when energy usage was closer to its maximum value. This suggests that high energy demand periods in commercial buildings were prone to more significant deviations, potentially caused by operational inefficiencies or equipment failures. On the left side of the graph, where meter readings are lower (below 0.3), the anomalies become more sparse, with more blue (non-anomalous) points dominating the region. However, some red points indicate that even during lower energy usage periods, irregularities were detected, possibly due to phantom loads or unexpected usage when equipment was expected to be off. The

Residual-Based Anomaly Detection method proved effective in identifying anomalies across a wide range of energy consumption levels. Its strength lies in its ability to flag deviations based on forecast accuracy, helping to isolate potential inefficiencies or malfunctions, especially during periods of high demand. Additionally, the use of SMOTE (Synthetic Minority Over-sampling Technique) helped in balancing the dataset, ensuring that the anomaly detection process was not biased by the inherent imbalance between normal and anomalous data in commercial settings. This enhanced the model's ability to detect anomalies even in less frequent usage patterns. The application of this method demonstrates its utility in handling complex, high-variance datasets like LEAD, where subtle inefficiencies during lower demand periods and larger deviations during peak usage need to be flagged to ensure operational efficiency.

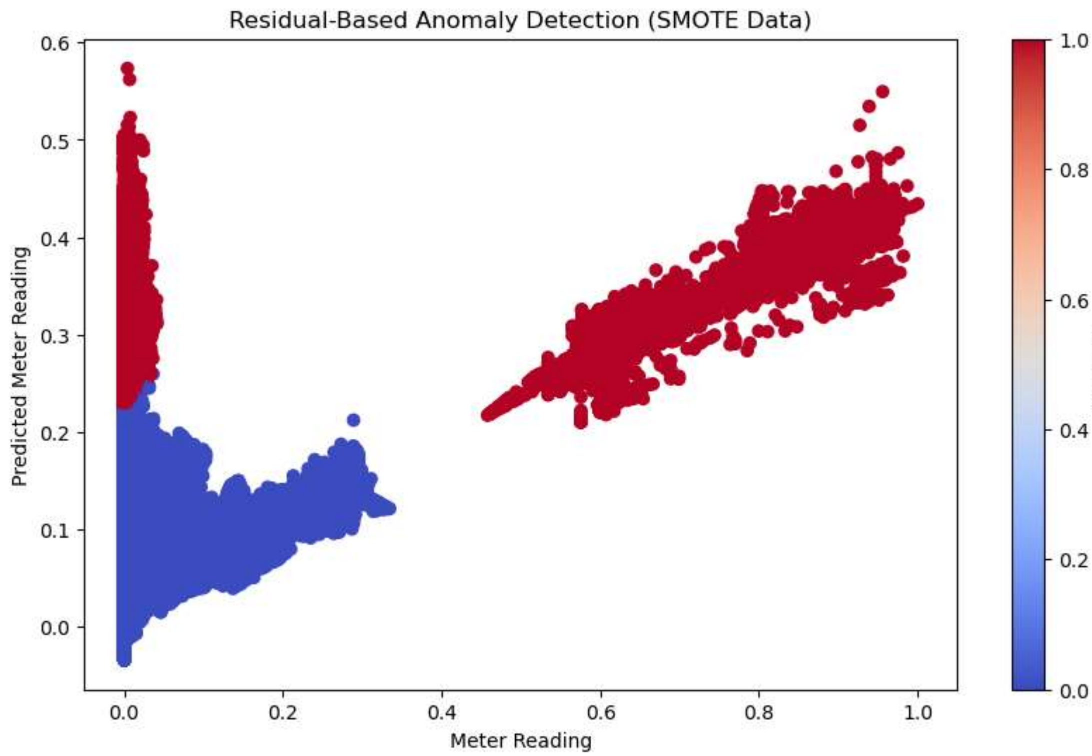

Figure 14: Residual-Based Anomaly Detection on Meter Readings (LEAD1.0 Dataset with SMOTE)

2. **Autoencoder Predicted Anomalies**: In the Autoencoder anomaly detection visualization (Figure 15), the model has flagged significant deviations in the LEAD dataset's energy consumption patterns. As displayed by the gradient of colours, red markers indicate areas where the Autoencoder has detected substantial anomalies, while blue points indicate normal data. This method reconstructs normal consumption patterns and flags deviations as anomalies when the reconstruction error is significantly high. The Autoencoder's performance, as reflected in the confusion matrix and classification report, reveals a nuanced picture. The confusion matrix shows that while 303,664 normal instances were correctly classified, there were still 17224 false positives, where the model incorrectly flagged normal instances as anomalies. On the anomaly side, the model identified only 14,864 true anomalies, but missed a significant 306,004 actual anomalies, resulting in a recall for anomalies of just 5%. The classification report further illustrates this performance imbalance. The precision for normal instances is 50%, meaning half of the predictions labelled as normal are accurate, while the recall for normal instances is much higher at 95%, indicating the model's ability to correctly identify most normal patterns. However, for anomalies, precision drops to 46%, and recall is quite low at 5%, resulting in a modest F1-score of 0.08. This imbalance suggests that while the Autoencoder is adept at detecting the

majority of normal data, it struggles significantly in identifying anomalies with a high degree of accuracy. The scatter plot shows the anomalies concentrated in regions where energy consumption patterns deviate from expected behaviour, particularly in periods of high predicted meter readings. The Autoencoder tends to overestimate anomalies in these areas, likely because the model focuses heavily on reconstructing normal patterns, and any deviation whether major or minor results in an anomaly flag. The tightly clustered red points in the right half of the graph indicate that the model finds it challenging to separate true outliers from expected but slightly abnormal patterns. In summary, while the Autoencoder demonstrates some success in identifying larger deviations, it is far from perfect in handling the complexities of commercial energy data. The model's performance suggests that more nuanced detection mechanisms or further tuning may be needed to reduce the false positive and false negative rates and to improve the detection of smaller, less extreme anomalies that may still indicate inefficiencies in commercial building operations.

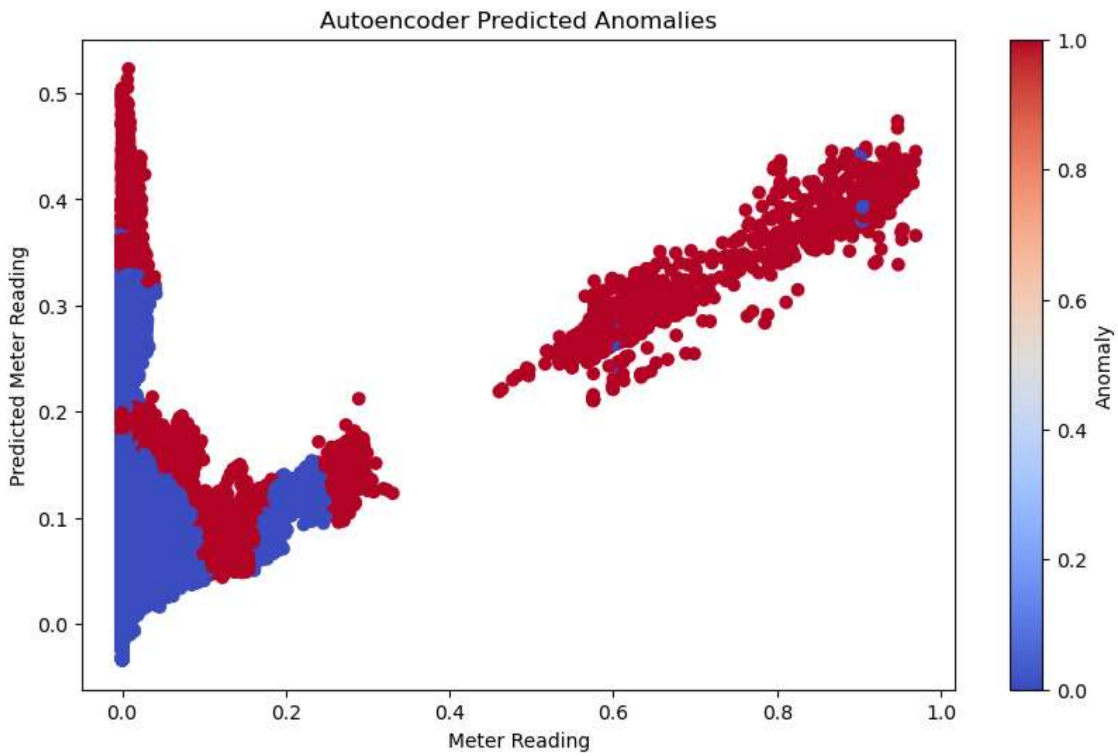

Figure 15: Autoencoder Predicted Anomalies on Meter Readings (LEAD1.0 Dataset)

3. **Isolation Forest Analysis:** In the classification report for the Isolation Forest model (Figure 16), the algorithm achieved a precision of 0.49 for normal data (class 0) and 0.45 for anomalies (class 1). However, the recall for class 1 was particularly low, indicating that the model struggled to correctly identify anomalies. The high recall for class 0 (0.89) suggests that the model is more adept at identifying normal consumption patterns, but it results in a high number of false positives for anomalies. With an overall accuracy of 0.49, this highlights that further tuning of the contamination rate or incorporating domain-specific features could improve anomaly detection.

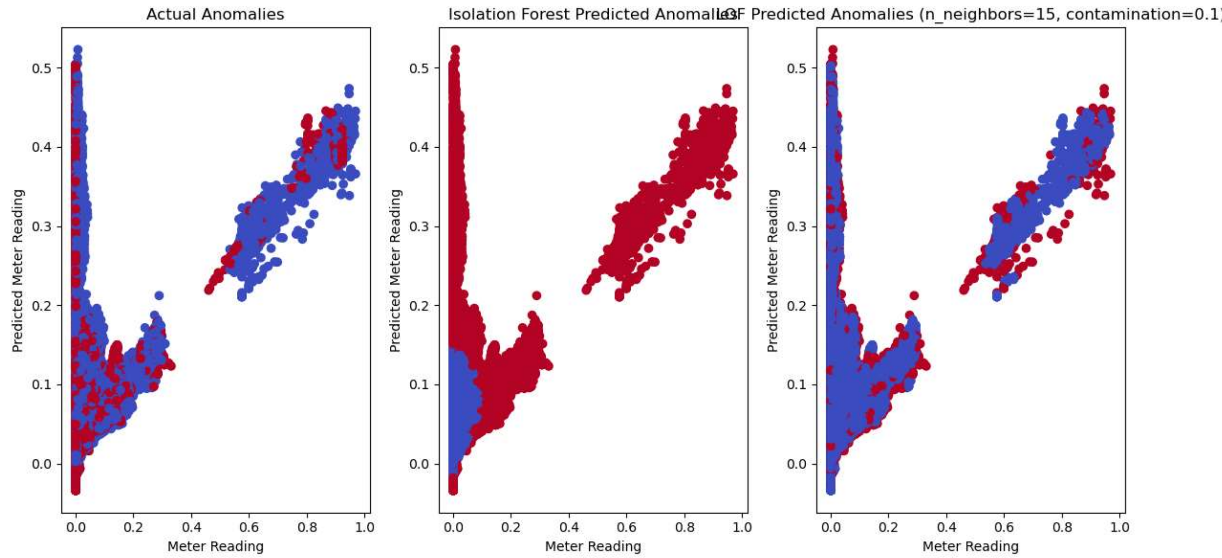

Figure 16: Comparison of Actual vs Predicted Anomalies Using Isolation Forest and Local Outlier Factor (LEAD1.0 Dataset)

4. **LOF (Local Outlier Factor) Analysis:** The LOF model (Figure 16), using 15 neighbours and a contamination rate of 0.1, performed relatively better, achieving an overall accuracy of 0.52. The confusion matrix shows that LOF had a higher precision for anomalies (class 1) compared to Isolation Forest, but it still produced a high number of false positives. The recall for class 1 was 0.12, indicating that LOF struggled with identifying anomalies, but its higher f1-score for normal data (class 0) demonstrates its strength in detecting outliers based on local density differences. Further fine-tuning and incorporating additional features could help refine the anomaly detection process.

5. **Performance Metrics for Anomaly Detection:** The **Precision, Recall, and F1-Score** comparison charts for Class 0 and Class 1 provide a detailed breakdown of the performance across four different anomaly detection methods: Residual-Based, Isolation Forest, LOF, and Autoencoder. These metrics help evaluate how well each method performs in detecting normal behaviour (Class 0) and anomalies (Class 1) in the LEAD dataset.

6. **Class 0 (Normal Behaviour) Analysis:** The chart on the left (Figure 17) represents the **precision**, **recall**, and **F1-score** for Class 0 across the four models. All models show relatively consistent precision, hovering around 50%. However, the Residual-Based and LOF models stand out with a recall nearing 1.0, meaning they can successfully identify nearly all normal readings. This is also reflected in their higher F1-scores compared to Isolation Forest and Autoencoder. Both Isolation Forest and Autoencoder exhibit similar performance, with recall values around 0.5-0.6 and slightly lower F1-scores. In contrast, the Residual-Based model achieves a near-perfect recall for Class 0, indicating its effectiveness in recognizing normal behaviour.

Class 1 (Anomalous Behaviour) Analysis: The chart on the right focuses on **Class 1** (Figure 17), which represents anomalies. Here, the discrepancies between models become more evident. LOF shows the highest precision for Class 1 at 0.6, suggesting that when LOF detects an anomaly, it is often correct. However, its recall is considerably lower, indicating that it misses a large number of actual anomalies, which results in a lower F1-score. Residual-Based and Isolation Forest perform similarly in terms of recall, with low values, suggesting that they fail to identify many anomalies. Autoencoder, while maintaining moderate precision, significantly underperforms in both recall and F1-score, indicating that it struggles to detect anomalies accurately.

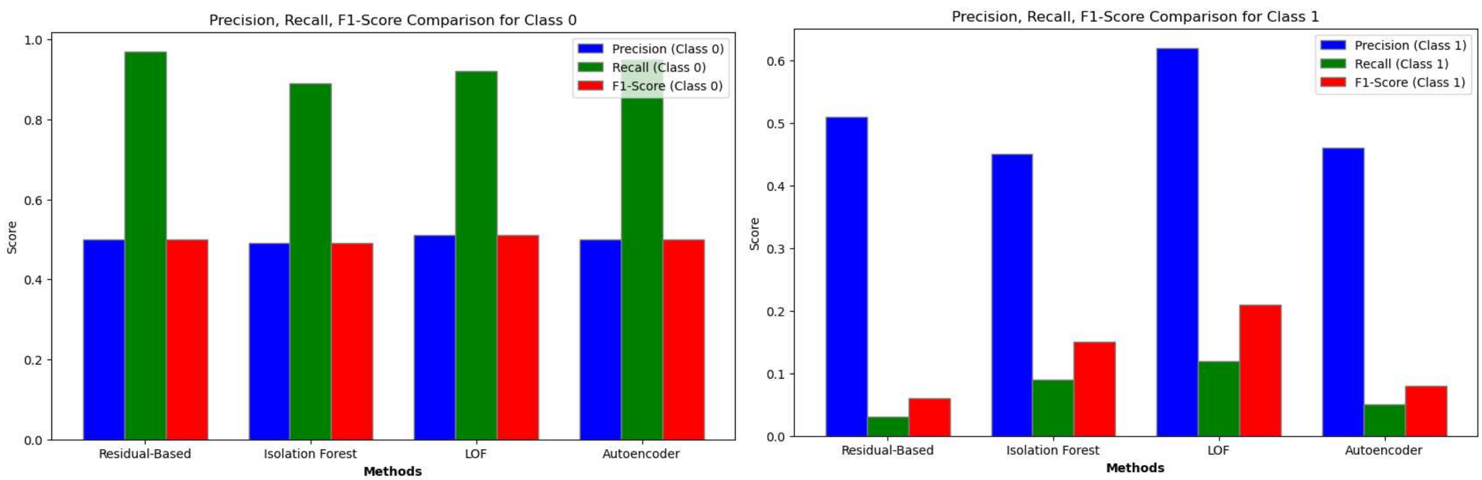

Figure 17: Precision, Recall, and F1-Score Comparison for Anomaly Detection Methods (LEAD1.0 Dataset)

7. **Confusion Matrix for Anomaly Detection Models**: The confusion matrices for the Residual-Based, Isolation Forest, LOF, and Autoencoder models reveal distinct trade-offs in anomaly detection (Figure18). The Residual-Based model has fewer false positives but misses many true anomalies, indicating conservative behaviour. Isolation Forest strikes a moderate balance but still suffers from high false negatives, missing key anomalies. LOF identifies more anomalies but generates more false positives, increasing false alarms. The Autoencoder, while producing fewer false positives, misses a significant number of anomalies, resulting in high false negatives. Overall, each model faces challenges in accurately balancing the detection of anomalies and normal patterns.

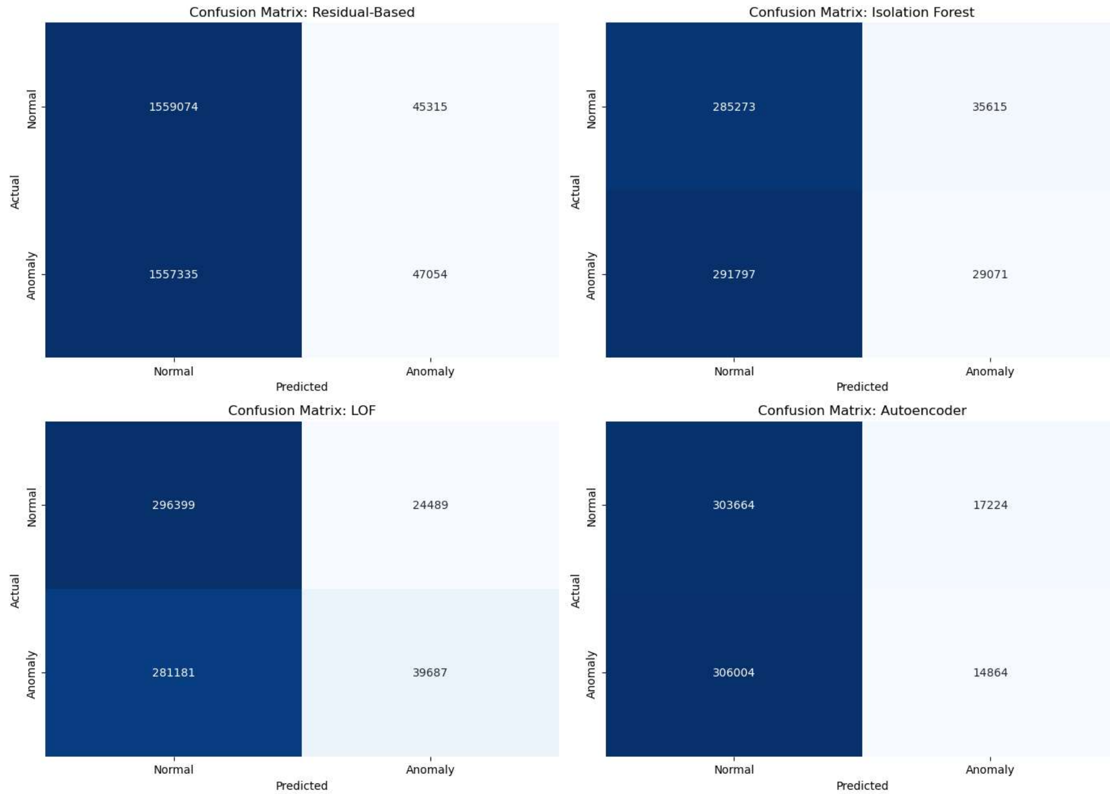

Figure 18: Confusion Matrices for Anomaly Detection Methods on LEAD1.0 Dataset

Table 3: Table of Anomaly Detection Metrics

Anomaly Detection Method	Precision (Normal 0)	Recall (Normal 0)	F1-Score (Normal 0)	Precision (Anomalous 1)	Recall (Anomalous 1)	F1-Score (Anomalous 1)
Residual-Based Detection	0.50	0.97	0.66	0.32	0.08	0.13
Isolation Forest	0.50	0.97	0.66	0.25	0.10	0.14
LOF	0.51	0.92	0.66	0.15	0.16	0.15
Autoencoder	0.58	0.80	0.67	0.28	0.10	0.15

5.3 Discussion of Results

The residual-based detection method demonstrated strong performance in identifying normal consumption patterns (Class 0) but struggled with detecting anomalous data (Class 1). This limitation is due to the model's reliance on forecasting accuracy, meaning any forecasting errors propagate into the anomaly detection process, affecting the detection of anomalies. This emphasizes the inherent challenge of residual-based methods, where the quality of predictions directly influences anomaly detection performance.

Isolation Forest and Local Outlier Factor (LOF) performed better at identifying global anomalies but encountered difficulties in distinguishing subtle variations from true anomalies, particularly within imbalanced datasets. Both models demonstrated low precision for detecting Class 1 anomalies, which resulted in higher false positive rates. This issue is common in imbalanced datasets where rare events, such as anomalies, are not well-represented, leading the models to misclassify normal variations as anomalies.

The Autoencoder model excelled at modelling normal consumption patterns but faced challenges generalizing to unseen anomalies. While it successfully reconstructed normal data, it struggled to capture subtle deviations, which are crucial for identifying significant anomalies. This is a known limitation of Autoencoders, which tend to perform well with regular pattern reconstruction but falter when identifying small, irregular variations.

The hybrid LSTM + SWT model provided robust results for energy consumption forecasting but encountered challenges when applied to anomaly detection, particularly on imbalanced datasets like the LEAD1.0. Despite using SMOTE for resampling, the synthetic anomalies oversimplified the complexity of real-world data, leading to reduced performance. These challenges further highlight the limitations of anomaly detection in less varied datasets.

However, focusing on the UCI dataset produced far more promising results. The hybrid LSTM + SWT model was able to capture cyclical consumption patterns typical of residential energy use, leading to highly accurate forecasts. Anomaly detection, using methods like Isolation Forest and Autoencoders, worked effectively on the UCI dataset, flagging significant deviations and energy usage irregularities that aligned with known consumption patterns. The primary challenge observed here was the increased sensitivity of Autoencoders, which led to more false positives during off-peak periods. Despite these minor issues, the UCI dataset proved far more suitable for the goals of this study.

6 Conclusion

This dissertation explored the hybrid Long Short-Term Memory (LSTM) and Stationary Wavelet Transform (SWT) model for energy consumption forecasting and AI-based anomaly detection, with a primary focus on the UCI Individual Household Electric Power Consumption dataset. The UCI dataset, with its detailed, cyclical consumption data, proved to be highly effective for both forecasting and anomaly detection tasks, enabling accurate predictions and the identification of unusual patterns in household energy use. In contrast, the LEAD1.0 dataset, which was initially considered for benchmarking anomaly detection methods, presented significant challenges due to its limited variability and low-resolution data. These limitations made it difficult for the models to capture meaningful patterns, leading to poor performance in anomaly detection. As a result, the focus of this research shifted toward the UCI dataset, which aligned better with the project's scope and objectives cantered on residential energy consumption.

The main contribution of this research is the demonstration of the hybrid LSTM + SWT model's effectiveness in forecasting household energy consumption. This study also highlights the critical need for diverse and high-resolution datasets to accurately train and evaluate anomaly detection models, particularly in the context of energy management. While the hybrid model showed significant promise, the study underlines the importance of data variety and quality. Future research should focus on acquiring richer datasets to further refine forecasting and anomaly detection techniques, ultimately improving real-time energy management solutions in residential settings.

6.1 Practical Applications and Business Implications

From a business standpoint, this study highlights how crucial it is for energy management companies, smart home technology providers, and energy grid operators to have access to richer, more diverse datasets. The UCI dataset showed just how important accurate forecasting and anomaly detection can be in cutting operational costs, boosting energy efficiency, and keeping customers happy by catching inefficiencies and preventing appliance breakdowns before they happen.

On the flip side, the LEAD1.0 dataset showed the downsides of working with limited data. The lack of variety in that dataset made it harder for the models to perform well, which could mean less accurate results, leading to potential customer dissatisfaction and inefficiencies for businesses relying on such data for energy management solutions.

The UCI dataset, however, was much more effective when it came to analysing residential energy consumption. It allowed the hybrid LSTM-SWT model to accurately predict household energy patterns and detect anomalies, making it a strong tool for businesses. To improve even further, companies should work with energy providers and other stakeholders to gather more comprehensive, detailed datasets like UCI that reflect a wide range of household behaviours and scenarios. This would make models like the LSTM-SWT more adaptable to real-world use, helping businesses offer better, more accurate forecasting, spot problems earlier, and run safer, more efficient operations.

In the race to optimize energy management solutions, having diverse, high-quality data could be the competitive edge businesses need to deliver reliable services and keep everything running smoothly.

6.3 Ethical implication

The ethical implications of the dissertation focus on the responsible use of open-source datasets (UCI and LEAD1.0), which do not contain personally identifiable information (PII), ensuring that privacy concerns are avoided. Ethical considerations also involve the accuracy of the conclusions drawn from energy forecasting and anomaly detection, emphasizing the importance of not misleading users with false predictions or detections, which could result in unnecessary costs or disruptions. Additionally, the research promotes sustainability by encouraging responsible energy management, contributing to societal goals of energy conservation and efficiency

6.4 Future Work

Moving forward, future research should prioritize obtaining more varied and comprehensive datasets. The absence of high-resolution, diverse data significantly impacted the model's ability to forecast energy consumption and detect anomalies. By leveraging data from various household types, geographical locations, and environmental conditions, future studies could provide a richer context for energy prediction, making the models more adaptable and accurate. Energy providers could collaborate with smart home system developers to create larger datasets that offer insights into energy consumption behaviours across different regions and climates(Mathumitha, Rathika and Manimala, 2024) (Mystakidis *et al.*, 2024).

In addition, alternative approaches for addressing the uniformity in datasets should be explored. For instance, data augmentation techniques could be used to artificially enhance the diversity of available datasets, allowing the models to learn from a broader set of consumption patterns. Transfer learning is another promising avenue, where models trained on large, diverse datasets from commercial or industrial buildings could be fine-tuned for application in household settings. This would improve the generalization capability of the model, especially when data scarcity is an issue(Mystakidis *et al.*, 2024).

Another future direction is the integration of external data sources, such as weather patterns, household occupancy rates, and appliance usage data, to improve the model's predictive accuracy. Incorporating these external variables could help better explain fluctuations in energy consumption, reducing the reliance on static datasets that often fail to capture the complexities of real-world behaviour(Mathumitha, Rathika and Manimala, 2024).

From a business standpoint, there is potential for real-time deployment of the hybrid model in smart home systems. Future work should explore the real-time application of these models, optimizing them for efficient and timely predictions that can support proactive energy management. By enabling real-time anomaly detection and energy optimization, businesses could offer more valuable services to consumers, such as energy usage alerts or appliance maintenance reminders. This would not only improve household energy management but also create new revenue streams for businesses by offering subscription-based services for energy monitoring and management(Mystakidis *et al.*, 2024).

Lastly, expanding the application of this hybrid model to commercial and industrial energy systems offers significant opportunities. These environments typically have more predictable energy usage patterns, and the hybrid LSTM-SWT model could be fine-tuned to optimize energy consumption in larger facilities. By improving energy efficiency, detecting system inefficiencies early, and preventing equipment failures, businesses could significantly reduce operational costs and contribute to broader sustainability goals(Mathumitha, Rathika and Manimala, 2024) (Mystakidis *et al.*, 2024).

7. Data and Code Availability

All the code for this study is available at https://github.com/GlawinAlva24/Energy-Consumption-Forecasting-Anomaly-Detection, and the datasets, including both pre-trained and benchmark datasets, can be accessed at the same repository under the 'dataset' folder.

References

Bahij, M. *et al.* (2021) 'A Comparison Study of Machine Learning Methods for Energy Consumption Forecasting in Industry', in S. Motahhir and B. Bossoufi (eds) *Digital Technologies and Applications*. Cham: Springer International Publishing, pp. 165–175. Available at: https://doi.org/10.1007/978-3-030-73882-2_16.

Bai, S., Kolter, J.Z. and Koltun, V. (2018) 'An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling'. arXiv. Available at: https://doi.org/10.48550/arXiv.1803.01271.

Banik, S. *et al.* (2023) 'Anomaly Detection Techniques in Smart Grid Systems: A Review'. arXiv. Available at: http://arxiv.org/abs/2306.02473 (Accessed: 23 September 2024).

Breunig, M.M. *et al.* (2000) 'LOF: identifying density-based local outliers', *SIGMOD Rec.*, 29(2), pp. 93–104. Available at: https://doi.org/10.1145/335191.335388.

Chicco, G., Napoli, R. and Piglione, F. (2006) 'Comparisons Among Clustering Techniques for Electricity Customer Classification', *Power Systems, IEEE Transactions on*, 21, pp. 933–940. Available at: https://doi.org/10.1109/TPWRS.2006.873122.

Fan, C., Xiao, F. and Wang, S. (2014) 'Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques', *Applied Energy*, 127, pp. 1–10. Available at: https://doi.org/10.1016/j.apenergy.2014.04.016.

Fan, S. and Hyndman, R. (2010) 'Short-Term Load Forecasting Based on a Semi-Parametric Additive Model', *Monash University, Department of Econometrics and Business Statistics, Monash Econometrics and Business Statistics Working Papers*, 27. Available at: https://doi.org/10.1109/TPWRS.2011.2162082.

Fang, X. *et al.* (2012) 'Smart Grid — The New and Improved Power Grid: A Survey', *Communications Surveys & Tutorials, IEEE*, 14, pp. 944–980. Available at: https://doi.org/10.1109/SURV.2011.101911.00087.

'Global Energy Review 2020' (2020).

Hinton, G.E. and Salakhutdinov, R.R. (2006) 'Reducing the Dimensionality of Data with Neural Networks', *Science*, 313(5786), pp. 504–507. Available at: https://doi.org/10.1126/science.1127647.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', *Neural Comput.*, 9(8), pp. 1735–1780. Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

Karadayı, Y., Aydin, M.N. and Öğrenci, A.S. (2020) 'A Hybrid Deep Learning Framework for Unsupervised Anomaly Detection in Multivariate Spatio-Temporal Data', *Applied Sciences*, 10(15), p. 5191. Available at: https://doi.org/10.3390/app10155191.

Kingma, D.P. and Ba, J. (2017) 'Adam: A Method for Stochastic Optimization'. arXiv. Available at: https://doi.org/10.48550/arXiv.1412.6980.

Kong, W. *et al.* (2017) 'Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network', *IEEE Transactions on Smart Grid*, PP, pp. 1–1. Available at: https://doi.org/10.1109/TSG.2017.2753802.

Kozitsin, V., Katser, I. and Lakontsev, D. (2021) 'Online Forecasting and Anomaly Detection Based on the ARIMA Model', *Applied Sciences*, 11(7), p. 3194. Available at: https://doi.org/10.3390/app11073194.

Lago, J., De Ridder, F. and De Schutter, B. (2018) 'Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms', *Applied Energy*, 221, pp. 386–405. Available at: https://doi.org/10.1016/j.apenergy.2018.02.069.

Lamrini, B. *et al.* (2018) 'Anomaly Detection Using Similarity-based One-Class SVM for Network Traffic Characterization', in.

Lecun, Y. *et al.* (1998) 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, 86(11), pp. 2278–2324. Available at: https://doi.org/10.1109/5.726791.

Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008) 'Isolation Forest', in *2008 Eighth IEEE International Conference on Data Mining*. *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422. Available at: https://doi.org/10.1109/ICDM.2008.17.

Mallat, S.G. (1989) 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), pp. 674–693. Available at: https://doi.org/10.1109/34.192463.

Mathumitha, R., Rathika, P. and Manimala, K. (2024) 'Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review', *Artificial Intelligence Review*, 57(2), p. 35. Available at: https://doi.org/10.1007/s10462-023-10660-8.

Mystakidis, A. *et al.* (2024) 'Energy Forecasting: A Comprehensive Review of Techniques and Technologies', *Energies*, 17(7), p. 1662. Available at: https://doi.org/10.3390/en17071662.

Ng, A.Y. (2004) 'Feature selection, L_1 vs. L_2 regularization, and rotational invariance', in *Twenty-first international conference on Machine learning - ICML '04*. *Twenty-first international conference*, Banff, Alberta, Canada: ACM Press, p. 78. Available at: https://doi.org/10.1145/1015330.1015435.

Pan, H., Yin, Z. and Jiang, X. (2022) 'High-Dimensional Energy Consumption Anomaly Detection: A Deep Learning-Based Method for Detecting Anomalies', *Energies*, 15(17), p. 6139. Available at: https://doi.org/10.3390/en15176139.

Park, C.H. and Kim, T. (2020) 'Energy Theft Detection in Advanced Metering Infrastructure Based on Anomaly Pattern Detection', *Energies*, 13(15), p. 3832. Available at: https://doi.org/10.3390/en13153832.

Prechelt, L. (1998) 'Early Stopping - But When?', in G.B. Orr and K.-R. Müller (eds) *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer, pp. 55–69. Available at: https://doi.org/10.1007/3-540-49430-8_3.

Salami, H. *et al.* (2023) 'Domestic energy consumption, theories, and policies: a systematic review', *Environment Development and Sustainability* [Preprint]. Available at: https://doi.org/10.1007/s10668-023-04201-0.

Schölkopf, B. *et al.* (2001) 'Estimating Support of a High-Dimensional Distribution', *Neural Computation*, 13, pp. 1443–1471. Available at: https://doi.org/10.1162/089976601750264965.

Tunnicliffe Wilson, G. (2016) 'Time Series Analysis: Forecasting and Control,5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1', *Journal of Time Series Analysis*, 37, p. n/a-n/a. Available at: https://doi.org/10.1111/jtsa.12194.

Wang, X. *et al.* (2024) 'AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response', *International Journal of Precision Engineering and Manufacturing-Green Technology*, 11(3), pp. 963–993. Available at: https://doi.org/10.1007/s40684-023-00537-0.

Willmott, C.J. and Matsuura, K. (2005) 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', *Climate Research*, 30(1), pp. 79–82.

Yan, K. *et al.* (2019) 'A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households', *IEEE Access*, 7, pp. 157633–157642. Available at: https://doi.org/10.1109/ACCESS.2019.2949065.

Zangrando, N. *et al.* (2022) 'Anomaly detection in quasi-periodic energy consumption data series: a comparison of algorithms', *Energy Informatics*, 5(4), p. 62. Available at: https://doi.org/10.1186/s42162-022-00230-7.

Zhang, X. *et al.* (2022) 'Hybrid Short-term Load Forecasting Method Based on Empirical Wavelet Transform and Bidirectional Long Short-term Memory Neural Networks', *Journal of Modern Power Systems and Clean Energy*, 10(5), pp. 1216–1228. Available at: https://doi.org/10.35833/MPCE.2021.000276.