Doctoral Dissertation

Improving Electricity Theft Detection for Smart Homes: Insights from Real and Synthetic Attack Scenarios

Olufemi Abiodun Abraham

Submitted on March 14, 2024

Graduate School of Science and Technology
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Olufemi Abiodun Abraham

Thesis Committee:

Professor Youki Kadobayashi	(Supervisor)
Professor Keiichi Yasumoto	(Co-supervisor)
Professor Yuichi Hayashi	(Co-supervisor)
Associate Professor Yuzo Taenaka	(Co-supervisor)

Improving Electricity Theft Detection for Smart Homes: Insights from Real and Synthetic Attack Scenarios[1]

Olufemi Abiodun Abraham

Abstract

Efforts to improve the security of electricity grids include both physical measures, such as detecting interference, and digital security methods, such as encryption. However, these measures alone are insufficient for addressing the full range of cyberattacks. For example, digital electricity meters, which are essential to new infrastructure, are vulnerable to software and hardware issues, and the digitization of these meters reintroduces concerns about electricity theft, which can now manifest as IT-related problems. To address these concerns, contemporary approaches that use data analytics, machine learning (ML), and predictive techniques are required. The rise in sophisticated statistical methods, particularly those involving machine learning, has led to an interest in developing models and algorithms, e.g., for smart homes, that can interpret smart meter data to quickly identify signs of tampering.

Smart home appliances, which are becoming increasingly integral to modern households, are also vulnerable to electricity theft. This can have a significant impact on both utility providers and consumers, as these appliances often connect to the internet and transmit usage data to utility providers for billing and

[1]Doctoral Dissertation, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 14, 2024.

monitoring. This data transmission can be intercepted or manipulated, leading to false readings and unauthorized electricity use. The ability to remotely control appliances also presents a risk of unauthorized access, which can lead to the misuse or manipulation of consumption data. To address these vulnerabilities, emerging technologies such as advanced ML algorithms and enhanced encryption methods are being developed to detect anomalies in usage patterns and secure data transmission. These technologies can help reduce the vulnerability of smart home appliances to electricity theft and improve the overall security of the electricity grid.

This dissertation presents novel approaches to electricity theft detection (ETD) by introducing various classification algorithms to detect anomalies in non-intrusive appliance load monitoring (NIALM) or disaggregated smart meter networks. Our proposed framework utilizes ML knowledge-based synthetic attack data (KBSAD) to train an attack classifier. These data consist of benign attack patterns that serve as the foundation for generating synthetic and simulated attacks that closely mirror real-world scenarios. The framework was validated using the Almanac of Minutely Power dataset version 2 (AMPds2), which contains fine-grained time-series data from a smart home. We preprocessed the data for binary classification to evaluate our synthetic attack model using real attack data. The Extreme Gradient Boosting algorithm performed best with an average area under curve (AUC) score of 98.74% and 98.69% for detecting and classifying anomalies in real and simulated attacks, respectively. These methods outperformed legacy unsupervised methods (LUM). By integrating the KBSAD, our approach eliminates the need for extensive data collection for real attacks and seamlessly combines synthetic attacks with genuine consumption readings, representing a significant advancement in the field of smart-home electricity theft detection.

Keywords:

Electricity Theft Detection, Smart homes, Data analytics, Cybersecurity, Machine Learning, Synthetic Attack Data, Real Attack Data, Non-Intrusive Appliance Load Monitoring (NIALM), Encryption methods, Anomaly Detection.

Contents

1 Introduction **1**

1.1 Problem Statement 3

1.2 Research Objectives and Contributions 5

1.2.1 Research Objectives 5

1.2.2 Research Contributions 6

1.3 Research Scope and Limitation 7

1.4 Dissertation Layout 9

2 Literature Review **10**

2.1 Preliminaries 10

2.2 Smart Grid Fundamentals 10

2.2.1 Overview of AMI 12

2.2.2 Architectural Elements of AMI 13

2.2.3 Smart Metering Devices 14

2.2.4 Networking Framework for Advanced Metering Infrastructure 16

2.2.5 Operations and Advantages of AMI 17

2.2.6 Security Concerns and Challenges in AMI 18

2.2.7 Energy Theft: A Multi-Billion Dollar Concern 19

2.3 Electricity Losses in Distribution Systems 20

2.3.1 Technical Losses (TLs) 20

2.3.2 Non-Technical Losses (NTLs) 20

2.3.3 Addressing Energy Theft 21

2.4 Classification of Attacks in Conventional and Smart Grid Systems 23

2.4.1 Physical Attacks 23

2.4.2 Cyber Attacks 24

2.4.3 Data Attacks . . . 25
2.5 Strategies for Energy Theft . . . 25
2.6 Approaches to Energy Theft in High Voltage Meters . . . 27
2.7 NTL Detection Methods . . . 28
2.7.1 State-based Detection: . . . 28
2.7.2 Game-Theoretic Approaches in NTL Detection . . . 29
2.7.3 Approaches in Classification-Based NTL Detection . . . 30
2.8 Advances in Non-Intrusive Load Monitoring . . . 31
2.9 Evaluating Machine Learning . . . 33
2.9.1 Model Performance Metrics . . . 33
2.10 Chapter Summary and Overview . . . 34

3 Unauthorized Power Usage Detection (UPUD) System 36
3.1 Introduction . . . 36
3.2 Related Work . . . 37
3.2.1 AMPds2 Dataset . . . 38
3.2.2 Dataset Preliminary Preprocessing . . . 38
3.3 Modeling Attack Scenarios . . . 42
3.3.1 False Appliance Injection Attack Scenarios . . . 42
3.4 GB-based UPUD model on smart meter disaggretated data . . . 43
3.4.1 Proposed Gradient Boosting Classifier Algorithm . . . 45
3.4.2 Classifier Training Process . . . 46
3.4.3 Proposed Machine Learning Approach for Training NIALM 46
3.5 Experiment results and Performance Evaluation . . . 47
3.5.1 Discussion . . . 49
3.6 Chapter Summary . . . 51

4 Electricity Theft Detection for Smart Homes with Knowledge-Based Synthetic Attack Data (KBSAD) Framework 52
4.1 Introduction . . . 52
4.2 Related Work . . . 53
4.3 Electricity Theft Attacks in Smart Homes . . . 54
4.3.1 Attack Model . . . 56
4.3.2 Attack Scenarios . . . 56

4.4 Electricity Theft Detection with Synthetic Attack Data 58
4.4.1 Multiclass classification approach 58
4.5 Dataset for Electricity Theft Detection 59
4.5.1 Overview 59
4.5.2 Data Profiles 61
4.5.3 Attack Impact 62
4.6 Evaluation 63
4.6.1 Experiment Settings 63
4.6.2 Performance Overview 66
4.6.3 Performance by Attack Class 68
4.7 Discussion 69
4.8 Chapter Summary 69

5 Electricity Theft Detection for Smart Homes: Harnessing the Power of Machine Learning with Real and Synthetic Attacks 71
5.1 Introduction 71
5.2 Related Work 71
5.3 Attacks Model Beyond The Distribution Board 72
5.4 Knowledge-Based Attack Simulation Framework 73
5.4.1 Attack Data Generation 74
5.4.2 Data Labeling and Preprocessing 74
5.4.3 Attack Classification 75
5.5 Dataset for electricity theft detection 76
5.5.1 Data Collection 76
5.5.2 Overview of the Feature Selection 78
5.5.3 Dataset preprocessing for binary classification 81
5.6 Data preparation 84
5.6.1 Data anonymization 84
5.6.2 Normalization 84
5.6.3 Framework for binary class from multi-class attack scenarios 85
5.6.4 Data Augmentation 86
5.6.5 Circular Shifting 86
5.6.6 Feature Vector (X) and Label (Y) 87
5.6.7 UTokyo Data - Real Attack Data 90

5.7 Synthetic Binary Discriminator Model (SYNBDM) 92
5.7.1 Proposed models Characteristics overview 93
5.7.2 XGBoost (XGB) 93
5.7.3 Random Forest (RF) 94
5.7.4 Multi-Layer Perceptron (MLP) 94
5.7.5 Maximum Likelihood Estimation (MLE) 95
5.8 Legacy unsupervised model (LUM) 96
5.8.1 ETA Based Autoencoder Detection Algorithm 96
5.8.2 Training Phase 98
5.8.3 Threshold Determination 99
5.8.4 Testing Phase 99
5.9 Experiments Results and Performance Evaluation 103
5.9.1 Model Performance by ROC and Confusion Matrices . . . 103
5.9.2 Model Training and test error 107
5.9.3 Model selection 109
5.9.4 Trade-off between training and test errors 111
5.10 Privacy Concerns and Model Development 112
5.10.1 Privacy Concerns in Smart Home Environments 112
5.11 Appliance Authentication Methods 114
5.11.1 Authentication of Appliance 114
5.11.2 Appliance Signature Analysis 114
5.11.3 Smart Meter Data Utilization 114
5.11.4 Integration with Home Automation Systems 115
5.11.5 Real-Time Monitoring and Authentication Checks 115
5.11.6 Machine Learning for Anomaly Detection 115
5.11.7 User Interaction and Feedback 115
5.11.8 Security and Privacy Considerations 116
5.11.9 Physical Fingerprint Appliances Authentication. 116
5.11.10 Appliance Authentication Using Physically Unclonable Functions 117
5.11.11 Advantages of Using PUFs with SYNBDM 118
5.11.12 Authentication without Biometrics 118
5.11.13 Privacy Concern Comparison of Recent Studies on ETD . 119

5.12 Model Updates . . . 121
5.12.1 Implementation of the SYNBDM Algorithm . . . 121
5.13 ETD Model Implementation in AMI Using Anonymized Aggregated Appliance Consumption Data . . . 125
5.13.1 Data Anonymization . . . 125
5.13.2 Feature Engineering and Anomaly Detection . . . 125
5.13.3 Train Models . . . 125
5.13.4 Model Evaluation . . . 125
5.13.5 Visualizations of the Anonymized Dataset . . . 126
5.14 Model Performance Comparison . . . 130
5.14.1 Comparison with Benchmark Models . . . 130
5.14.2 Model Performance on Synthetic Attack Data . . . 132
5.14.3 Model Performance on Real Attack Data . . . 134
5.15 Differences Between Consumption and Prevention in Smart Home Electricity Management . . . 136
5.15.1 The Consumption Pilot Approach . . . 137
5.16 Analysis of Economic Implications . . . 137
5.16.1 Economic Implication of SYNBDM Deployment . . . 137
5.16.2 For Utility Companies and Electricity Providers . . . 137
5.16.3 For Homeowners and Consumers . . . 138
5.16.4 For Technology Providers . . . 138
5.16.5 For the Broader Economy . . . 138
5.16.6 Economic Equity . . . 139
5.17 Chapter Summary . . . 139

6 Discussion and Future Work **141**
6.1 Discussion . . . 141
6.1.1 Summary of our findings . . . 142
6.1.2 Model Development and Evaluation . . . 143
6.1.3 Machine Learning Algorithms . . . 143
6.1.4 Performance Metrics . . . 143
6.1.5 Challenges in Model Training . . . 143
6.1.6 Data Augmentation Techniques . . . 143
6.1.7 Synthetic Attack Data Utilization . . . 144

6.1.8 Legacy vs. Smart Attacks 144
6.1.9 Future Directions 144
6.1.10 Limitations of the Proposed Model 144

7 Conclusion **146**

Acknowledgements **148**

References **150**

Publication List **161**

List of Figures

1.1 Smart Home ETD Research Scope. 7

2.1 Overview of Electrical and Data Flow in the Smart Grid. 11
2.2 Architecture of AMI in Smart Grid[93]. 13
2.3 Models of conventional power and smart grids[93]. 15
2.4 Direct tapping to the power line[93]. 26
2.5 Breaking control wire[93]. 27
2.6 NILM event capturing [89] 33

3.1 Home Appliances Disaggregated Metering Diagram with load value 38
3.2 Base power consumption data pattern 40
3.3 False Injection of CDE - Consumption pattern changes 41
3.4 False Injection of HPE - Consumption pattern changes 41
3.5 Distribution of imbalance dataset 43
3.6 Block diagram of our proposed UPUD system 44
3.7 Smart meter NIALM at benign model diagram 44
3.8 Smart meter NIALM attack model diagram 45
3.9 Evaluation Report of imbalance dataset 49
3.10 Classification Model Report after data augmentation 50
3.11 Accuracy of the balanced dataset 50

4.1 The Power Distribution board and ETA Scenarios 57
4.2 The framework for electricity theft detection with synthetic attack data. 58
4.3 Electricity usage of Home A with synthetic attack data on different days. 62

4.4 2D projections of attack contained electricity usage with the benign case by UMAP 62
4.5 Confusion matrices of the trained models. The number (No. #) indicates the rank of the overall accuracy. 67

5.1 ETD framework for processing real power consumption data to detect anomalies. 74
5.2 Flow of ETD model 76
5.3 Maximum power values for each appliance 78
5.4 Features selection with correlation comparison 82
5.5 Features selection with mutual information Scores 83
5.6 Multi-class attack scenarios with unspecified attacks preprocessed to binary attack class for anomaly detection. 86
5.7 Features Vector and Labels 88
5.8 Framework of Data Augmentation 89
5.9 Electricity power consumption pattern of Home A for original class and augmented data class for a day 90
5.10 Electricity power consumption pattern of all homes for original class and augmented data class for a day 91
5.11 Flow of evaluation with Supervised Binary Descriminator 92
5.12 Flow of evaluation with Unsupervised Autoencoder 98
5.13 Threshold determination for Home A training dataset 102
5.14 Sampled ROC curves for comparison performance evaluation of the proposed SYNBDM of some selected homes. 107
5.15 Performance comparison of ROC curves for LUM across homes. . 107
5.16 Confusion matrices of the SYNBDM and LUM in order of overall performance. 108
5.17 Training and test error comparison for our proposed model 110
5.18 SYNBDM implementation process 116
5.19 Histograms of selected feature distributions 126
5.20 Correlation heatmap of selected features 127
5.21 Selected sample of first five features for time series plot 128
5.22 Selected sample of Box Plots for Anomaly Visualization 128
5.23 SYNBDM implementation in AMI 129

5.24 Performance metrics comparison analysis 131
5.25 Perfomance across homes for the SYNBDM 133
5.26 Perfomance across homes for the LUM 134

List of Tables

3.1 Appliances and units in AMPds 39
3.2 NIALM-Hacking Dataset - Benign and Attack Instances 42
3.3 GB Multiclass Classification Report 47
3.4 Experimental Results of different Algorithms 48

4.1 The Configuration of Synthetic Attack Data For Supervised Learning in the experiment. 60
4.2 Appliances and units in AMPds 60
4.3 The Profile of the dataset generated with synthetic attack data, and corresponding attack impacts (AI). 63
4.4 Models parameter configuration 64
4.5 Experimental results of different algorithms. 65

5.1 Home configurations based on appliance data points 77
5.2 Distribution of simulated and real binary dataset 91
5.3 Parameters for the binary supervised discriminator 96
5.4 Flow of the evaluation with Synthetic Binary Discriminator 97
5.5 Parameters for legacy unsupervised models 100
5.6 Legacy unsupervised AE Threshold values for Anomaly detection 103
5.7 Evaluation of the ETD model with AUC and accuracy Scores . . 104
5.8 Models training and testing error report for all homes. 109
5.9 Comparison of models based on accuracy and AUC scores 130
5.10 Performance metrics of different multiclass classification algorithms. 132
5.11 Proposed model performance metric comparison with binary class. benchmark. 132

1. Introduction

The advent of smart meters has led to the provision of high-resolution electricity data at the residential level [17], which has significantly enhanced our understanding of energy usage and how it benefits both electricity providers and consumers. In Europe, the shift to renewable energies has been accompanied by the adoption of intelligent power meters[98]. These devices enable quick adaptation to fluctuating energy demands and the variability of renewable energy sources like wind and photovoltaic energy. This evolution has transformed the power grid into an intelligent, decentralized network that not only transmits energy but also data, thereby functioning as a communication channel[71]. The smart grid, a next-generation electrical network, features bi-directional communication among various sensors and actuators across several networks, enabling the collection and analysis of detailed information[27].

Smart grid (SG), equipped with Advanced Metering Infrastructure (AMI), smart meters, and connected smart home appliances, are engineered for resilience against disruptions[56]. This resilience is tested by the integration of decentralized, fluctuating renewable energy sources. Within this complex network, the AMI plays a pivotal role in facilitating two-way communication between consumers and utility providers, enhancing the grid's adaptability and response to changing energy demands. Smart meters, a critical component of the AMI, provide real-time, high-resolution data crucial for accurate energy demand prediction. This data underpins essential tasks such as consumer profile forecasting, dynamic energy pricing, and the monitoring of individual smart home appliances. These appliances, ranging from smart thermostats, refrigerators, dishwashers, washing machines and dryers, ovens and stoves, security cameras, smart locks, smart doorbells, home hubs, and controllers, plugs, and switches to energy-efficient lighting systems, contribute to a comprehensive demand response strat-

egy, adjusting consumption patterns to optimize energy usage[26].

The security of this interconnected smart grid system through the widespread of Internet of Things (IoT) and its prominent application is paramount. It must be robustly safeguarded against cyber threats, which encompass not only direct attacks on the infrastructure but also fraudulent activities and potential software failures. The integrity and reliability of the smart grid, particularly the data transmitted and processed by AMI and smart meters, are essential for maintaining both the efficiency of the grid and the trust of consumers using these smart appliances.

Historical research in energy demand, spanning Non-Intrusive Appliance Load Monitoring (NIALM), energy forecasting, residential energy demand modeling, and Typical Load Classification (TLC) [34], has been a cornerstone in shaping the evolution of smart grid research. Originating in the early 80s, these studies have provided valuable insights into energy consumption patterns, which are now critical in the realm of Electricity Theft Detection (ETD) at the level of smart home appliances within the Home Area Network (HAN). In a modern HAN, various devices such as smart meters, intelligent thermostats, connected lighting systems, and smart home security devices are interconnected. These components collectively monitor and manage energy usage within the smart home ecosystem. By employing NIALM techniques, each appliance's energy consumption can be analyzed non-intrusively, enabling the detection of abnormal usage patterns that could indicate electricity theft. This approach transforms the way residential energy demand is monitored, moving from aggregate to appliance-specific analyses, thus enhancing the precision of ETD.

Furthermore, this granularity of monitoring brings to the fore the vulnerability of appliance consumption patterns within the broader cybersecurity ecosystem of smart homes. The interconnected nature of HAN devices makes them susceptible to cyber threats, which can manifest as unauthorized access to energy usage data or manipulation of smart meter readings. The insights gained from historical research on energy demand, particularly in the development of TLC methods, are now instrumental in identifying and mitigating such vulnerabilities. By understanding typical consumption patterns, anomalies - potentially indicative of cybersecurity breaches or electricity theft - can be more readily identified and

addressed. This integration of traditional energy demand studies with contemporary cybersecurity measures reflects the evolving challenges and technological advancements in safeguarding smart home environments against electricity theft.

1.1 Problem Statement

The emergence of smart grids, with their intricate webs of interconnected smart meters and home appliances, represents a significant leap forward in energy management and efficiency. However, with the sophistication of the smart grid comes an increased vulnerability to cyber threats that can compromise the integrity of the system. The high-resolution data provided by smart meters, essential for dynamic energy pricing, load forecasting, and monitoring of smart appliances, now stands at risk of unauthorized access and manipulation, which can lead to electricity theft and other fraudulent activities.

The AMI at the heart of the smart grid facilitates crucial two-way communication but also opens the door to potential security breaches. These breaches can distort the real-time data that is foundational for the grid's efficiency and the reliability of consumer demand response strategies. Moreover, the integration of volatile renewable energy sources, although outside the scope of this dissertation, adds another layer of complexity to the smart grid, necessitating even more robust and resilient security measures to maintain stability.

In the HAN, every connected device, from intelligent thermostats to smart lighting systems, not only enhances user convenience but also adds to the potential entry points for cyber attacks. The shift from aggregate energy monitoring to appliance-specific analysis through NIALM techniques improves ETD capabilities but simultaneously highlights the security challenges at the appliance level.

Anomalies in energy consumption patterns, once detectable only at a macro level, can now be traced to individual devices, suggesting potential electricity theft or cybersecurity issues within the HAN. The challenge lies in leveraging historical energy demand studies and Typical Load TLC methods to fortify the smart grid against such threats. This integration of traditional research with the latest cybersecurity protocols is critical for identifying, mitigating, and ultimately preventing the unauthorized use and manipulation of energy data within smart homes.

The scarcity of labeled data for energy theft incidents poses significant challenges in training robust detection models, necessitating advanced data augmentation and unsupervised learning techniques to enhance the detection accuracy of legacy and modern smart home systems.

As the smart grid evolves into an increasingly decentralized and user-driven network, the need for a sophisticated approach to secure the vast amount of data it generates and processes becomes paramount. This thesis addresses the pressing problem of securing smart home appliance data against electricity theft and cyberattacks, ensuring the reliability and efficiency of the smart grid, and safeguarding the trust and safety of the consumers it serves within the smart home area network. To identify fraudulent customers within the smart home network, we trained our model on fine-grained power consumption appliance data for unauthorized power usage detection(UPUD) leveraging smart meter data disaggregation or Non-Intrusive Appliance Load Monitoring (NIALM).

1.2 Research Objectives and Contributions

In this research, our primary objectives and contribution are as follows:

1.2.1 Research Objectives

In this research, we aim to design and develop an effective Electricity Theft Detection (ETD) framework for smart homes with knowledge-based synthetic attack data (KBSAD) by simulating real-life attack scenarios with validation attack data and applying different Machine Learning (ML) algorithm approaches.

1. RO1.1: Design and develop an effective ML-based UPUD in AMI networks.
2. RO1.2: Transfer and adapt RO1.1 into Smart Home networks

- RO1.1: Design and develop an effective ML-based UPUD in AMI networks.
  - To develop an effective framework that allows the training of attack classifiers only from legitimate power consumption data.
  - To optimize the Machine Learning models to achieve the best detection performance.
  - To investigate different machine learning algorithms based on performance, accuracy, and effectiveness.

  RO1.1 - the objective is to optimize machine learning (ML) models to achieve reasonable detection accuracy and reduce false positive and false negative rates. We thoroughly investigate how to optimize machine learning models. Our experiment results demonstrate that the right hyper-parameter values selection is essential to developing a robust ETD system. Additionally, transfer and adapt the above objective into smart home networks. The justification is presented in the later chapters.

  As we affirm, several electricity theft cyber-attacks can transfer into the disaggregated appliance consumption patterns (NIALM) in smart home networks; hence, we intend to transfer our solution into the NIALM system.

- RO1.2: Design and develop an effective ETD with Knowledge-Based Synthetic Attack Data (KBSAD) for smart homes.

- To provide an effective pre-processing method for developing an effective supervised classification model.
- To simulate real-world attack scenarios in apartments/office rooms and optimize the Machine Learning models to achieve the best detection performance.
- To deploy and validate KBSAD with real attack data from a building in the detection of electricity theft (ETD) in the smart home network.

RO1.2 - the research's principal aim is to generate Knowledge-Based Synthetic attack datasets (KBSAD) for five samples of real-world attack scenarios in smart homes by arithmetically adding stolen power as power consumption and adding onto real power consumptions for training an attack classifiers and propose an effective pre-processing method to develop a robust ETD for smart home network system. Finally, we investigate different machine learning models to determine the most suitable algorithms for developing the ETD smart home network attack detection. Although we study datasets from an open source for a house in Canada (AMPds2), we believe our ETD will be effective in any apartment, congested building complex, and/or office rooms where power line cables are not easily traceable. The aforementioned proposal details are presented in Chapters 5, 6 & 7.

1.2.2 Research Contributions

1. **Cost-efficiency:** Training ETD models with synthetic attack data bypasses the need to gather real-life attack patterns and saves time and resources.
    - We thoroughly investigate how to optimize the machine learning approaches to attain higher detection rates.
2. **Real-World Evaluation:** Enables the testing of ETD models against synthetic attacks and real data, ensuring cost-effective and practical assessment.
3. **Simplified Calculation:** Streamlines the process of incorporating attack scenarios into meter readings by adding synthetic data for legitimate usage.

4. **Previously Unattainable:** Offers a solution to the challenge in computer network communications, where node behavior analysis depends on actual attack occurrences.

5. **Enhanced flexibility:** The adaptability of synthetic data, informed by a wide knowledge base, can significantly improve ETD model training.

6. **Improved Training and Precision:** Leverages the diversity of synthetic attacks to train more robust ETD models, leading to accurate attack detection and classification in smart homes and by extension to SG through AMI.

1.3 Research Scope and Limitation

The scope of this research includes the design and development of an electricity theft detection (ETD) system that can be utilized in a smart home network with refined aggregated appliance consumption patterns. The first part includes the NIALM-based Unauthorized Power Usage Detection (UPUD) which optimizes machine learning (ML) models to achieve reasonable detection accuracy and reduce false positive and false negative rates. It also introduces six different models to determine the best-performing algorithm in UPUD of smart home electricity theft attack (ETA).

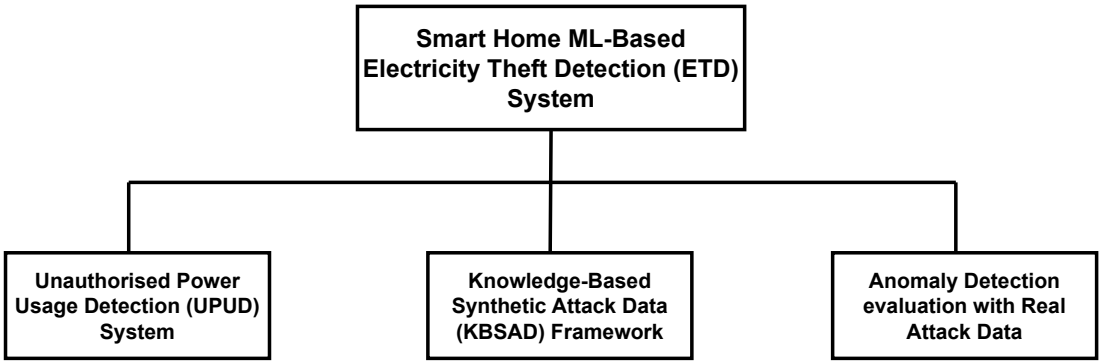


Figure 1.1: Smart Home ETD Research Scope.

In the second part, we include the introduction of the Knowledge-Based Synthetic Attack Data (KBSAD) framework for Electricity Theft Detection (ETD)

in the smart home network. This is due to one of the challenges with using ML classifiers in ETD, which is data imbalance, i.e., the numbers of normal and abnormal samples are not in the same range. Benign samples are easily available using historical data but attack/theft samples on the other hand, rarely or do not exist for a given customer. Besides, in many cases, samples of attack classes cannot be obtained from historic data due to zero-day attacks. However, the problem of imbalanced data and zero-day attacks was addressed by generating a synthetic attack dataset, benefitting from the fact that theft patterns are predictable. The final part focuses on anomaly detection of ETA in smart home appliance consumption patterns. We evaluated our synthetic data with binary classified real attack data measured from a power distribution board in a building to test our model performance. The simulation results include a comparison of supervised and unsupervised learning frameworks with the synthetic and real attacks in three different smart homes (Home A, Home B, and Home C).

1.4 Dissertation Layout

The dissertation outline is as follows: In Chapter 2, we discuss the preliminary studies of the Smart Grid and the Smart Home Area Network (HAN), Non-Intrusive Appliance Load Monitoring (NIALM) - Disaggregated Smart Meter. We also discuss the machine learning models' performance evaluation matrix. Chapter 3, presents our studies about Unauthorised Power Usage Detection (UPUD) for Machine Learning-based ETD for smart home networks. In Chapter 4, we explain our proposed knowledge-based synthetic Attack Data (KBSAD) ETDS regarding some real-world simulated attack scenarios in a smart home/office complex environment. We provide details on how to optimize the model and improve the detection accuracy of the ETD. In Chapter 5, we discuss the detailed development of the binary classification of Electricity Theft Attack Detection (ETA-DD) in smart homes and also detail the anomaly detection algorithm for ETD in smart home networks. We also discuss the performance evaluation with different models and model applications. In Chapter 6, we present discussion and future work.Chapter 7 concludes the dissertation.

2. Literature Review

2.1 Preliminaries

This chapter discusses an overview of the Smart Grid, the Advanced Metering Infrastructure (AMI) and different types of attacks. We will discuss Non-Intrusive Appliance Load Monitoring (NIALM) - Disaggregated Smart Meter and consumption patterns of some appliances and also mention the machine learning models' performance evaluation matrix.

2.2 Smart Grid Fundamentals

Advanced Metering Infrastructure (AMI), serving as a foundation for Smart Grids (SGs), is a cornerstone of the modern electrical infrastructure, progressively replacing antiquated power systems in both residential and commercial sectors. SGs represent an evolutionary leap in power grids, incorporating bidirectional information and communication technology (ICT) alongside pervasive computing to enhance energy management and distribution. This sophisticated integration fosters enhanced control, heightened efficiency, augmented reliability, and increased safety in energy distribution [92]. The United States Department of Energy's modern grid program designates Smart Grids (SGs) as a combination of control methods, integrated communication systems, and advanced sensing technologies within the existing electrical power infrastructure. This innovation allows both consumers and utility providers to better monitor, manage, and predict energy usage patterns. As depicted in Figure 2.1, the process starts with Bulk Generation, involving large-scale electricity production from various sources [47]. This electricity is transmitted at high voltage from generation plants to substations [75], and then distributed at lower voltages to end-users [10]. Substations convert

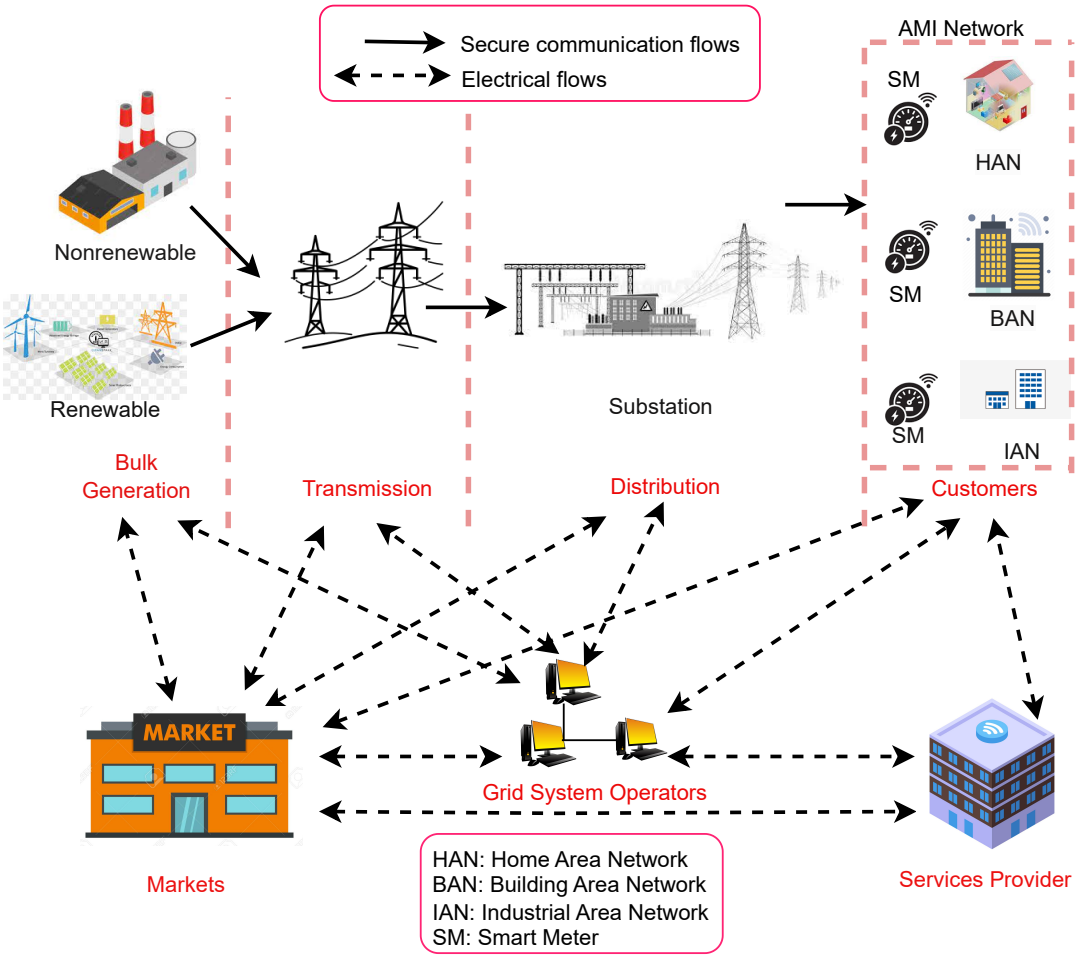


Figure 2.1: Overview of Electrical and Data Flow in the Smart Grid.

transmission to distribution voltage [94]. The Advanced Metering Infrastructure (AMI) Network facilitates two-way data communication between customers and utilities [24]. Smart Meters (SM) record and report energy usage [97]. Home Area Network (HAN), Building Area Network (BAN), and Industrial Area Network (IAN) manage local energy usage and generation [30]. Markets enable the buying and selling of electricity [13], while Grid System Operators ensure the power system's operation [12]. Finally, Service Providers offer diverse services from energy supply to efficiency solutions [79].

SGs facilitate the integration of microgrids and various distributed energy resources (DER), such as solar and wind power, as well as energy storage systems.

This combination aims to tackle prevailing energy management challenges, resulting in a robust, self-healing grid system. Moreover, SGs employ demand-side management (DSM) strategies to encourage consumers to consciously adjust their energy consumption. This incentivizes consumers to shift their usage to off-peak hours, thereby reducing consumption rate fluctuations and balancing peak to average energy demand. This shift is essential in achieving various objectives such as reducing greenhouse gas emissions, combating global warming, and striving for national energy independence [57]. The following section, 2.2.1, delves into the intricacies of Advanced Metering Infrastructure (AMI), which is crucial for collecting information and data from consumers and loads, underpinning the operation of Smart Grids.

2.2.1 Overview of AMI

The integration of Advanced Metering Infrastructure (AMI) represents a pivotal advancement in the evolution of electrical grids. Central to the smart grid (SG) paradigm is the active engagement and empowerment of consumers—a vision articulated by the U.S. Department of Energy in 2008 [85]. AMI is instrumental in providing a dynamic metering ecosystem that equips consumers with critical insights, fostering informed energy usage and enabling proactive participation. Concurrently, utility providers (UPs) benefit from enhanced customer engagement and optimized operational efficiency through sophisticated asset management facilitated by granular metering analytics. AMI serves as a conduit for real-time energy consumption data and grid event notifications to both consumers and UPs, fostering collaborative energy management. This synergy enhances billing accuracy, aids in demand response initiatives, and supports judicious decision-making within SGs. The role of AMI extends beyond mere data relay; it constitutes a critical nexus connecting UPs, consumers, and the broader spectrum of generation and storage assets, underpinned by a suite of integrated technologies encompassing advanced communication systems, intelligent metering devices, Home Area Networks (HAN), Building Area Network(BAN), Industrial Area Network (IAN) and robust interfaces for utility operations and data management. This thesis delves into the appliance consumption patterns in smart homes and the implications of smart meters (one of the components of AMI) within the

electrical power distribution landscape.

2.2.2 Architectural Elements of AMI

The AMI comprises a complex network of intelligent devices, including smart meters (SMs), data aggregation units, and Internet of Things (IoT) devices, as well as a robust data management infrastructure, which includes systems such as Meter Data Management Systems (MDMS) and central communication hubs. These components are interconnected through a sophisticated network infrastructure that seamlessly integrates data into various software applications and physical interfaces. AMI's distinct feature is its bidirectional flow of both communication and electricity, enabling a comprehensive overlay of informational and power exchanges, as noted by Rashed Mohassel et al [63]. The integrated network design and AMI's role within the smart grid are illustrated in Figure 2.2.

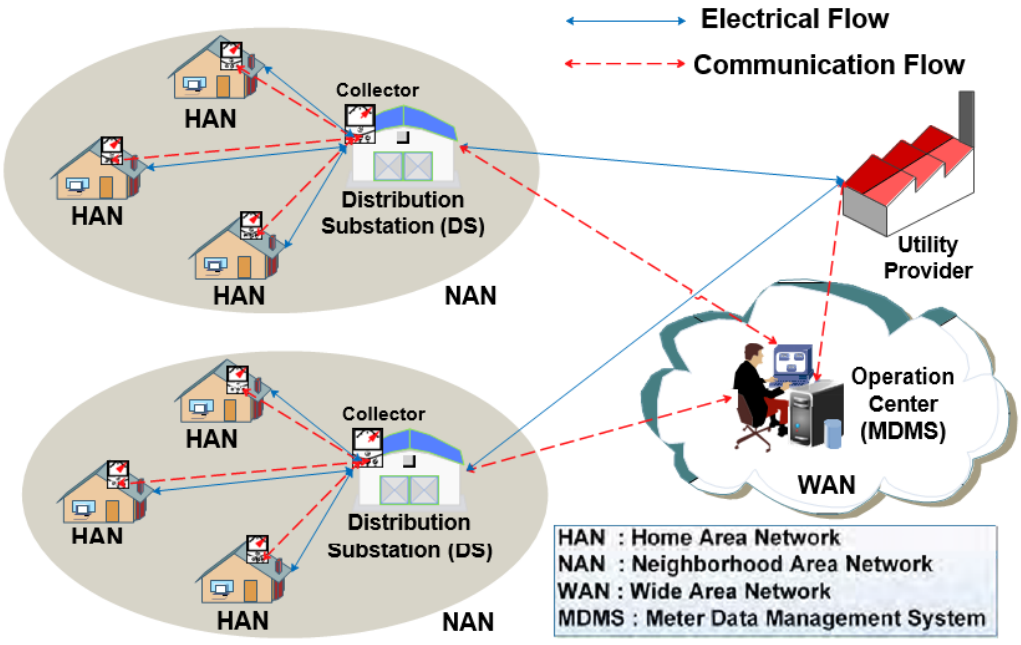


Figure 2.2: Architecture of AMI in Smart Grid[93].

2.2.3 Smart Metering Devices

Smart metering devices represent the cutting edge of technological innovation, equipped with sophisticated hardware and software designed for precise data collection and measurement at predetermined intervals. These devices are configured and programmed by system administrators to transmit data to relevant entities at regular intervals. As part of the bidirectional communication framework of the AMI, these Internet of Things (IoT) devices not only receive but also transmit signals to perform designated actions. Moreover, information on utility pricing from UPs enables these devices to adjust energy consumption according to user-defined preferences and operational parameters.

Traditionally, conventional meters were exclusively used for billing consumer energy usage. However, with the emergence of SGs, contemporary SMs have become increasingly prevalent, enhancing the resilience and operational efficiency of power systems that incorporate DER and distributed demand response initiatives, as identified by W. Wang and Lu [90]. SMs are essential components within AMI ecosystems, serving as advanced energy meters that record consumer energy usage in regular, predetermined intervals. Unlike traditional meters, SMs possess the capability to capture detailed energy usage metrics, such as voltage, current, frequency, and phase angle. These meters collaborate with central data aggregators and communication gateways to ensure power quality and secure data transmission in real-time. Furthermore, SMs possess the ability to both execute and receive control directives, whether locally or remotely. In addition to their role in consumption regulation and system monitoring, SMs also gather diagnostic data on domestic appliances, the distribution grid, and power-related events. They can remotely regulate energy supply to consumers, thus controlling the maximum energy consumption. The data collected by SMs includes consumption figures, timestamps, and unique meter identifiers, as discussed by Depuru, Wang, & Devabhaktuni [23]. Figure 2.3 illustrates the metering architectures of both conventional power grids and SGs.

The essential features of SMs encompass a wide array of functionalities, such as:

- The capacity for real-time data collection and evaluation

- The ability to upgrade and operate remotely
- Notification of system failures and malfunctions
- Implementation of dynamic pricing structures based on time
- Monitoring of power quality
- Energy trading through net metering
- Load scheduling for Demand Side Management (DSM)
- Load limiting to encourage demand response
- Detection of energy theft and meter tampering using alert systems and sensors
- Advancement of energy conservation for the sake of environmental sustainability.

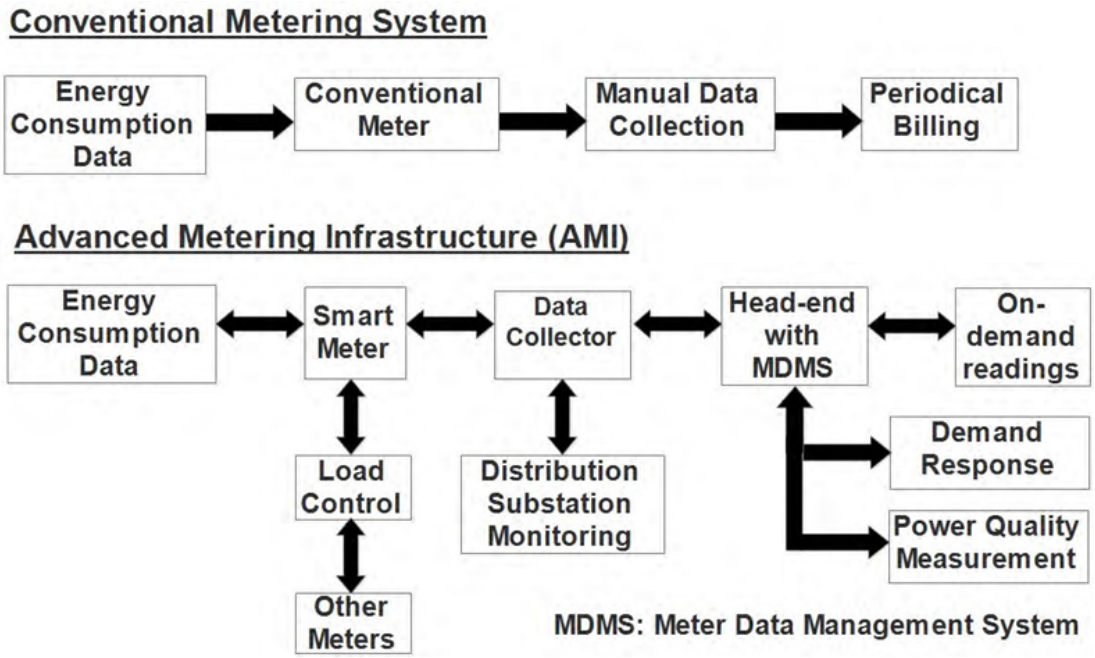


Figure 2.3: Models of conventional power and smart grids[93].

2.2.4 Networking Framework for Advanced Metering Infrastructure

The Advanced Metering Infrastructure (AMI), as depicted in Figure 2.2, is supported by a tripartite networking system consisting of the Home Area Network (HAN), Neighborhood Area Network (NAN), and Wide Area Network (WAN), each possessing unique characteristics and functions:

- **HAN:** Within the domestic realm, HAN connects smart meters (SMs), intelligent appliances, local generation units, energy storage systems, and plug-in hybrid electric vehicles (PHEVs), along with their respective control units. To facilitate high bandwidth and low-energy signal transmission, wireless modalities are the most appropriate communication methods for HAN. Protocols such as ZigBee, WiFi operating at 2.4 GHz, HomePlug standards, and IEEE 802.11 are commonly utilized to fulfill these requirements, as indicated by the U.S. Department of Energy [86].

- **NAN:** Comprising data aggregator units and SMs situated within consumer properties, the NAN facilitates the exchange of information within a localized community. The choice of communication technology within the NAN is tailored to suit various operational contexts and may include Power Line Communication (PLC), RS485 serial connections, GPRS/3G mobile networks, or ZigBee interfaces.

- **WAN:** Serving as the communication backbone, the WAN provides connectivity between the data aggregation points and the central operations hub. The selection of communication channels within the WAN is predicated upon regional requirements and may encompass PLC, Ethernet, or GPRS/3G networks.

Subsequent to data acquisition, the collated information is conveyed to an operations center. This center houses both a Meter Data Management System (MDMS) and a head-end system. The head-end system orchestrates the communication protocols, aggregates, and preserves metering information, and ensures compatibility with internet protocols while interfacing with various devices. In parallel, the MDMS, as the core component of the AMI's management architecture, oversees the monitoring of the distribution network and undertakes the

analysis, maintenance, and operational management tasks integral to the functioning of Smart Grids.

2.2.5 Operations and Advantages of AMI

AMI presents numerous benefits and implications that impact consumers, Utility Providers (UPs), and society at large, as acknowledged by the U.S. Department of Energy (2008).

For Consumers:

- Improved Energy Management: AMI offers extensive insights into energy consumption and grid status, empowering consumers to make informed decisions about their energy use. As a result, consumers can expect enhanced power quality, increased reliability, and precise billing, ultimately leading to reduced utility expenses.

For UPs:

- Operational Efficiency and Billing Process Improvement: AMI enables UPs to gain advantages in areas such as operational efficiency and billing processes.

- Cost Savings: The implementation of Smart Meters (SMs) facilitates automated data transmission and remote firmware updates, eliminating the need for physical meter readings and maintenance. Consequently, operational and labor costs are reduced.

- Demand Response and Load Management: AMI equips UPs with the ability to monitor grid load continuously, facilitating dynamic pricing and direct load control. This not only reduces peak demand but also mitigates the need for new-generation facilities and eases transmission congestion.

- Grid Health Monitoring: Real-time monitoring allows UPs to analyze line losses and optimize the electrical power infrastructure.

- Enhanced Outage Management: Smart metering aids UPs in promptly detecting and locating outages, enabling more efficient repair crew dispatch.
- Non-Technical Loss Reduction: SMs and data collectors in AMI play a crucial role in identifying and mitigating energy theft, thereby reducing NTLs.

For Society:

Environmental Sustainability: AMI contributes to a greener environment by enhancing energy delivery and usage efficiency. It promotes the adoption of Distributed Generation (DG) and DER, potentially reducing carbon dioxide emissions significantly, as suggested by Siddiqui [76].

2.2.6 Security Concerns and Challenges in AMI

The widespread adoption of SMs across homes and businesses in the United States and initiatives like Tenaga Nasional Berhad (TNB) [83] pilot project in Malaysia highlight the growing implementation of AMI. However, this expansion brings about significant security challenges. This thesis will delve into these security concerns, emerging from the new power infrastructure, and explore potential mitigation strategies.

Privacy Concerns for Consumers

In the SG ecosystem, while consumers collaborate with UPs for efficient energy management, privacy concerns arise due to the necessity of sharing detailed consumption data. Third parties can utilize fine-grained SM data for load profiling, which reveals not only appliance usage but also personal lifestyle patterns. The ability to deduce such detailed information poses privacy risks, ranging from burglary risks to unsolicited marketing and competitive disadvantages for industrial consumers. Instances like the Dutch Parliament's rejection of SG implementation in 2009 due to privacy concerns [20] underscore the need for robust privacy protection measures. Establishing a regulatory framework that defines data collection, sharing, and usage policies is crucial to maintaining consumer trust in

SG adoption. Moreover, the security and reliability of smart devices must be rigorously assessed by academia, the government, and the energy industry to ensure comprehensive protection against privacy invasions.

2.2.7 Energy Theft: A Multi-Billion Dollar Concern

Energy theft within electrical power distribution systems has long been an international concern. Northeast Group [29] estimates the global annual cost of energy theft to be approximately $96 billion in 2017, with a notable portion occurring in emerging economies like India, Brazil, and Russia. Historically, methods of energy theft involved physically tampering with meters to disrupt accurate measurements. However, with the introduction of Smart Meters (SMs), the nature of energy theft is evolving towards more complex cyber-related attacks, such as remote device hacking and data manipulation. These advanced attacks can lead to minor alterations in energy readings or even large-scale infrastructure threats, as detailed by McDaniel & McLaughlin [57].

Smart Meters, integral to the Advanced Metering Infrastructure (AMI), are built using common software and hardware, exposing them to typical vulnerabilities found in networked and communication systems. This includes risks such as unauthorized access, distributed denial-of-service (DDoS) attacks, and various forms of malware, making SMs prime targets for cybercriminals [57]. The exploitation of these vulnerabilities can lead to significant financial implications, not only in terms of energy theft but also in the costs of replacing compromised meters. Misuse of SMs could also pose serious risks to the power infrastructure, potentially leading to misguided operational decisions by UPs and hiding imminent threats.

The future development of SG is highly dependent on the regulatory frameworks established by governments and UPs. Addressing the security issues introduced by AMI is crucial for transitioning to a more efficient and cost-effective power grid.

2.3 Electricity Losses in Distribution Systems

Energy loss in electrical distribution systems, as defined by Nagi et al. [66], refers to the discrepancy between the energy supplied and that reported by consumers. Assessing these losses is key to evaluating system performance. Generally, UPs face two types of losses: technical losses (TLs), inherent to energy distribution and transformation, and non-technical losses (NTLs), which are related to energy fraud and metering errors. Although the supplied and recorded energy should ideally match, discrepancies due to systemic losses are inevitable [28].

2.3.1 Technical Losses (TLs)

TLs in electrical distribution systems result from energy dissipation in electrical components during transmission and distribution, contributing to increased costs and carbon emissions [33]. Influencing factors include the physical properties of the electrical equipment, voltage levels, and grid design. Common sources of TLs encompass transformer winding losses and resistive losses in feeders and networks [5].

According to Congres International des Reseaux Electriques de Distribution [19], TLs primarily arise from load losses (variable resistive and reactance losses), no-load losses (fixed losses), and losses due to network services. Advancements in Information and Communication Technology (ICT) and data acquisition have enabled more precise computation and verification of TLs, as demonstrated by predictive models using transformer and SM data [53].

2.3.2 Non-Technical Losses (NTLs)

NTLs, which occur independently of TLs, are often more challenging to measure and attribute to external factors such as energy theft and meter irregularities [74]. These losses have significant implications for political stability and the financial well-being of UPs. High NTLs are commonly linked to governance issues like political instability and corruption [22]. Addressing NTLs is crucial for enhancing the efficiency of electrical distribution systems and reducing costs.

Key causes of NTLs include:

- Non-payment by consumers: Leading to systemic financial issues for UPs.
- Meter irregularities: Resulting from incorrect meter readings and equipment malfunctions.
- Energy theft: Entailing illegal connections and meter tampering.

Non-Payment by Consumers

Failure to pay utility bills can lead to a cascade of financial and systemic problems [14]. Strategies like service disconnection and prepayment meters are implemented to mitigate these issues [43].

Meter Irregularities

Meter irregularities, contributing significantly to NTLs, occur when meters inaccurately record energy consumption. Common causes include incorrect readings and equipment malfunctions (Tenaga Nasional Berhad, 2018). With AMI, UPs can effectively monitor SMs in real-time, aiding in the detection and resolution of these issues. This comprehensive monitoring enables UPs to identify discrepancies in energy reporting and manage the distribution system more efficiently.

Addressing NTLs is vital for the effective operation of electrical distribution systems, enhancing network efficiency, reducing costs, and improving reliability. Effective strategies involve ensuring accurate meter readings, proper installation and maintenance, and implementing systems to detect and prevent energy theft.

2.3.3 Addressing Energy Theft

Meter tampering and energy theft are primary contributors to Non-Technical Losses (NTLs) in Malaysia, as identified by Tenaga Nasional Berhad (2006). NTLs have persistently troubled Utility Providers (UPs) globally since the inception of energy billing. Current data suggests that worldwide energy theft amounts to an astonishing $96 billion annually (Northeast Group, 2017). In the United States, energy theft results in annual losses of about $6 billion (Karaim, 2015), while UPs in India face yearly losses of around $4.5 billion due to energy fraud (Ahmad et al., 2018). In Canada, British Columbia Hydro incurs yearly

losses of $100 million, largely attributed to illegal marijuana cultivation operations (Meuse, 2016). The increasing sophistication of energy theft techniques poses a significant challenge in detecting NTLs, leading to inflated costs for consumers and substantial government subsidies. This thesis, therefore, focuses on identifying and combating energy theft and meter irregularities in Smart Grids (SGs).

Various methods are employed to under-report energy usage:

- At the consumer level, common tactics include tampering with meters or siphoning energy from unoccupied premises.
- At the grid level, fraudulent practices often involve bypassing meters by directly connecting high-load appliances or the entire electrical system to the feeder with an unauthorized distribution transformer (DT).
- At the utility level, inaccuracies in billing, whether due to unintentional meter errors or intentional manipulation by corrupt personnel, can lead to profit losses.

To counteract energy theft, TNB Malaysia has formed specialized teams for physical meter inspections (Tenaga Nasional Berhad, 2006). These teams, equipped with additional technicians and resources, conduct investigations into suspected cases of energy fraud.

The implementation of Advanced Metering Infrastructure (AMI) and Smart Meters (SMs) is a strategic response to energy theft. SMs are engineered to detect and report tampering, mitigating certain vulnerabilities present in traditional analog meters. TNB's pilot project in 2015 involved installing 1,000 SMs in Malacca and Putrajaya (Tenaga Nasional Berhad, 2017), aiming to offer enhanced digital services, including failure alarms and tamper detection for analyzing NTLs.

However, the integration of smart metering infrastructure introduces new vulnerabilities. For instance, SMs are not immune to tampering (McLaughlin et al., 2010). An energy thief could gain unauthorized access to a SM, disrupting its communication and neutralizing automated alarms. Additionally, the high rate of false positives in meter alarms complicates the UPs' task of differentiating between fraudulent and legitimate consumers.

Literature review (McLaughlin et al., 2010, 2013; Jiang et al., 2014; Accenture, 2011; Y. Liu et al., 2018; Tellbach & Li, 2018) categorizes energy theft techniques into three main types:

1. Physical attacks
2. Cyberattacks
3. Data attacks

Data attacks can result from both cyber and physical breaches, compromising consumer usage data at various stages: recording, transmission, or storage (Xiao et al., 2013). These techniques are summarized in Section 2.4 below, which will be referenced in Section 4.3 to construct an attack model encompassing various known energy theft methods. Subsequently, energy theft and meter irregularity scenarios will be simulated by altering benign SM readings to assess the proposed anomaly detection frameworks in Chapters 5 & 6.

2.4 Classification of Attacks in Conventional and Smart Grid Systems

2.4.1 Physical Attacks

These attacks involve direct interference with the electrical metering and distribution equipment.

1. Meter Swapping: Replacement of meters with units from unoccupied or low-usage locations.
2. Power Diversion: Redirecting power supply within a neighborhood.
3. Meter Tampering: Includes meter removal/disconnection, reversing the meter's position, using magnets to affect readings, obstructing or damaging the rotating coil, introducing substances to impair meter function, unauthorized access to Smart Meters (SMs), and modifying the current transformer (CT) ratio.

4. Meter Bypass: Direct wiring of high-load appliances to the grid, bypassing the meter completely.

5. Illicit Grid Connection: Unauthorized connections to the primary voltage grid or distribution feeder.

6. Corruption: Bribery of utility personnel for altered billing.

7. Meter Calibration and Regulation Issues: Inappropriate calibration and regulation of meters.

2.4.2 Cyber Attacks

Cyber attacks target the digital and networked components of the grid.

1. Credential Theft: Unauthorized access to meters using stolen login credentials.

2. Firmware Hacking: Remotely hacking into Smart Meter firmware.

3. Data Tampering: Altering meter storage data including total energy consumption, audit logs, and encryption keys.

4. Network Exploitation: Compromising meter readings through network vulnerabilities and intercepting meter communications.

5. Bandwidth Flooding: Overloading the Neighborhood Area Network (NAN) bandwidth.

6. Resource Exhaustion: Depleting meter memory or central processing unit (CPU) capacity.

7. Event Logging Interference: Deleting or interrupting logged events.

8. Communication Disruption: Disrupting radio frequency (RF) communications.

9. Value Injection: Injecting forged values into communications between Utility Providers (UPs) and Smart Meters.

10. Traffic Modification: Altering traffic between UPs and Smart Meters.
11. Meter Spoofing and Jamming: Impersonation of meters and RF signal jamming.
12. Malware: Designing and injecting malware into Smart Meters.
13. Pricing Manipulation: Altering predictive pricing algorithms.

2.4.3 Data Attacks

These attacks involve manipulation of consumption data.

1. Zero/Negative Reporting: Falsely reporting zero or negative energy consumption.
2. Consumption Report Alteration: Stopping or altering energy usage reports.
3. Measurement Exclusion: Removing high-consumption appliances from measurements.
4. Under-reporting: Reporting less energy consumption than actual.
5. Load Profile Modification: Changing appliance load profiles to hide larger loads.

2.5 Strategies for Energy Theft

From the outset of energy billing, fraudulent consumers have employed numerous methods to alter meters and their inputs for energy theft. These methods fall primarily into two categories: meter tampering, involving manipulation of the meter's internal structure, and line tampering, which includes bypassing the meter connection. This thesis focuses on identifying such attacks under the assumption that the integrity of meter consumption readings has been compromised, and does not address methods of manipulating communication signals.

Low Voltage Meters (230V Single Phase):

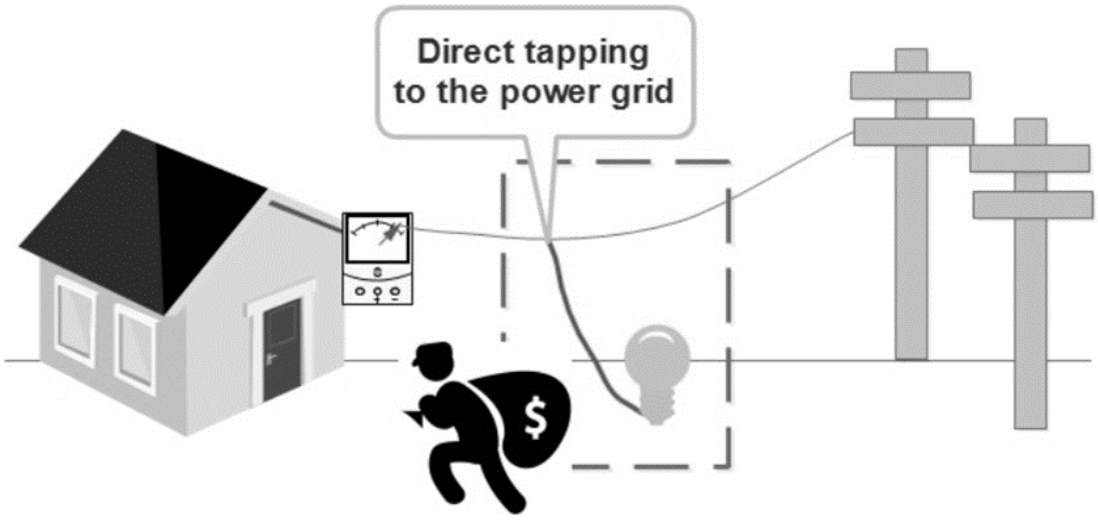


Figure 2.4: Direct tapping to the power line[93].

1. **Direct Connection to Power Grid (Bypassing the Meter):** Consumers, particularly in domestic or SME sectors, often directly tap into the LV 230V single-phase/415V three-phase power lines. This method is more straightforward and considered safer compared to tampering with HV lines.

2. **Meter Tampering:**
   - Breaking the meter's enclosure seal for internal access and tampering.
   - Reversing number dials in analog meters or modifying CT turns.
   - Using magnets to slow down the rotor disk, affecting consumption measurement.
   - Installing remote control switches to regulate energy consumption recording.

3. **Circuit Bypass/Hidden Switch:** Utilizing jumper wires or hidden switches to bypass the metering circuit, avoiding energy consumption registration. This method is becoming increasingly sophisticated with the use of advanced gadgets and wiring.

These strategies illustrate the evolving nature of energy theft, ranging from simple mechanical tampering to sophisticated electronic interference. The chal-

lenge for utility providers lies in detecting and mitigating these diverse methods of energy theft, ensuring accurate billing and fair energy distribution.

2.6 Approaches to Energy Theft in High Voltage Meters

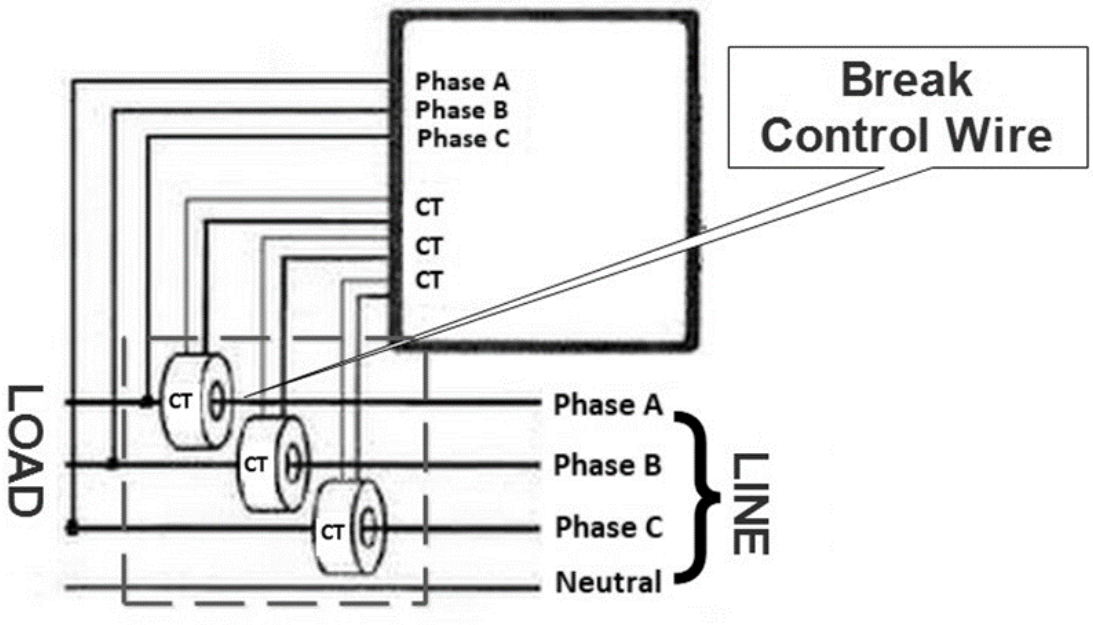


Figure 2.5: Breaking control wire[93].

High Voltage (HV) meters, particularly three-phase watt-hour meters, are commonly installed at industrial sites with substantial energy demands. These meters typically employ a "two-watt-hour meters" connection in a three-phase energy meter for measurement. Energy theft in these settings, particularly prevalent in commercial and industrial sectors, contributes significantly to Non-Technical Losses (NTLs) in countries like Malaysia.

High Voltage Meters (12kV or 24kV three-phase, three or four-wire primary):

1. Direct Power Line Connection: Direct tapping into HV power lines is more challenging, with heightened risks due to the high voltage.

2. Meter Tampering: Similar methods as those employed in LV meters are applied, including physical alteration of the meter's components.

3. Control Wire Tampering: Energy theft executed by manipulating the control wires of the CT, resulting in reduced current readings.

4. Terminal/Meter Seal Tampering: Tampering with terminal seals located below the meter to alter voltage phase readings.

5. Voltage Tap Interference: Disrupting voltage readings by shorting voltage taps to the ground.

2.7 NTL Detection Methods

Several recent studies have proposed various NTL detection methods, such as Support Vector Machine (SVM), load profiling, neural networks, and decision trees (Viegas et al., 2017). This thesis categorizes these detection schemes into three groups: state-based, game theory-based, and classification-based (Jiang et al., 2014), focusing primarily on data-oriented schemes that rely on consumer data like energy consumption measurements.

2.7.1 State-based Detection:

State-based detection involves monitoring the power system's state using specific equipment to identify energy fraud.

Xiao et al. (2013) proposed three inspection schemes for identifying anomalous Smart Meters (SMs), which, while effective, increase cost due to the additional metering equipment required. McLaughlin et al. (2013) designed an Advanced Metering Infrastructure Intrusion Detection System (AMIDS) using various data sources, though this raises privacy concerns due to non-intrusive load monitoring (NILM). Selvapriya (2014) suggested using control units to compare individual and aggregated consumption against feeder input levels, with alerts sent via GSM to investigate suspected energy theft. Khoo and Cheng (2011) integrated RFID technology for better meter inventory management and energy theft reduction. Huang et al. (2013) and Sahoo et al. (2015) used variance analysis and temperature-dependent predictive models, respectively, to identify abnormal energy consumption patterns.

Various recent studies have proposed NTL detection methods, focusing primarily on data-oriented schemes that rely on consumer data like energy consumption measurements.

2.7.2 Game-Theoretic Approaches in NTL Detection

Game-theoretic approaches to NTL detection conceptualize the interaction between fraudulent consumers and Utility Providers (UPs) as a strategic game (Cardenas et al., 2012; Amin et al., 2015). In this framework, energy thieves aim to under-report electricity consumption while minimizing detection risk. Conversely, UPs focus on enhancing the likelihood of identifying theft while also aiming to reduce the operational costs associated with these detection mechanisms. This perspective offers a novel way to analyze and mitigate NTLs.

In the game-theoretic framework proposed by Amin et al. (2015), two scenarios are considered: perfect competition and unregulated monopoly. The study develops an elaborate game-theoretic model to assess various statistical techniques for detecting energy theft. However, this model necessitates certain impractical assumptions about the nature of fraudulent activities. The research yields precise estimates of detection capabilities under these specific assumptions.

Cardenas et al. (2012) also developed a game-theoretic model, formulating a strategic interaction between energy thieves and UPs. They determined the Nash equilibrium of this game as the probability density function that both defenders and attackers should adopt when sending Advanced Metering Infrastructure (AMI) measurements. Furthermore, the study includes a preliminary analysis of the optimal sampling interval for Smart Meters (SMs), balancing consumer privacy concerns with the retention of demand response program benefits.

However, the development of utility functions for all involved parties, including energy thieves and UPs, and the formulation of their potential strategies remain complex challenges within this approach.

Game-theoretic approaches to NTL detection conceptualize the interaction between fraudulent consumers and Utility Providers (UPs) as a strategic game. In this framework, energy thieves aim to under-report electricity consumption while minimizing detection risk, whereas UPs focus on enhancing the likelihood of identifying theft and reducing operational costs.

2.7.3 Approaches in Classification-Based NTL Detection

Classification-based NTL detection utilizes a comparative approach between forecasted and actual energy consumption values, where a significant discrepancy often indicates potential fraud [59]. Both ARIMA and ARMA models, renowned for time series forecasting, play a pivotal role in this process. Badrinath et al., [9] found that ARIMA tends to be more effective for residential consumers. Expanding on this, Krishna et al., [46], investigated the application of Kullback-Leibler Divergence (KLD) for detecting sophisticated energy theft, where KLD is used to assess the disparity between a set of measurements and a historical baseline. This method is particularly adept at uncovering smart attacks that blend anomalous usage into normal patterns using legitimate ARIMA models. Jindal et al. [40] explored decision trees and SVM-based classifiers for in-depth analysis of energy consumption data to pinpoint fraud. Their approach integrates two levels of data processing, with the decision tree output serving as input for the SVM classifier. Villar-Rodriguez et al.,[87] developed a novel algorithm for outlier detection in SGs, which accommodates irregularities in consumer consumption habits by focusing on consumption trends rather than temporal attributes.

However, protecting consumer privacy while detecting energy theft is a critical aspect of SGs, as highlighted by Mirzaee et al. [61] and Llaria et al. [51]. They proposed a Lower-Upper Decomposition (LUD) algorithm to calculate consumer honesty coefficients in a privacy-preserving manner. The LUD approach is unique in that it gauges the likelihood of having multiple energy thieves in a community based on the honesty coefficient vector. Despite its innovative approach, the LUD algorithm has limitations, such as not accounting for technical losses and being constrained by the dimensionality of consumption data, which might necessitate adjustments in data granularity.

Moreover, some classification-based methods are susceptible to contamination attacks, where energy thieves subtly alter data to mislead learning algorithms. The dependency of machine learning-based methods on extensive long-term monitoring often results in a delay in theft detection due to the requirement of large sample sizes [41]. Additionally, most NTL detection schemes overlook technical losses, which could impede their practical application.

To overcome these limitations, this thesis proposes an electricity theft detec-

tion (ETD) framework for smart homes with knowledge-based synthetic attack classifiers and evaluated with real attack data. These frameworks are capable of identifying electricity theft and meter irregularities without being constrained by the dimensions or granularity of power consumption data. They are designed to detect anomalies in the appliance consumption patterns, an NTL, occurring minutely or intermittently throughout the day. The proposed frameworks also consider the impact of electricity theft attacks (ETA) at three different homes assumed to have different appliance configurations and different ETA scenarios, aiming for a high detection rate with minimal false positives. A range of NTL attack scenarios is evaluated to validate the reliability of these frameworks in real-world AMI fraud and metering defect scenarios. Importantly, these frameworks can be implemented without additional hardware costs, offering a cost-effective solution for UPs and ultimately benefiting consumers through reduced operational costs.

2.8 Advances in Non-Intrusive Load Monitoring

In the realm of Smart Grids, AMI plays a pivotal role in generating data at the household level concerning power consumption. However, appliance-level data, offering more granularity, are increasingly sought after by homeowners and building managers for a detailed analysis of each appliance's contribution to total energy usage and cost. The conventional method of collecting such data is known as intrusive load monitoring, involving the physical installation of sensors on each appliance. While this approach yields precise usage data for each appliance, it is limited by its lack of scalability, efficiency, and cost-effectiveness.

Contrastingly, Non-Intrusive Load Monitoring (NILM) employs software algorithms to disaggregate whole-house data, effectively deducing appliance-level power consumption. This innovative technique was first introduced in the literature [99], where it focuses on identifying changes in voltage and current to ascertain when different appliances are turned on and off. An illustrative example Figure 2.6 in Hart et al., [45] demonstrates how NILM algorithms discern individual appliance events based on their impact on overall household energy consumption.

NILM algorithms can be categorized into three distinct types based on their

objectives:

- **On/Off Classification:** This model simplifies the task of determining whether an appliance is on or off, disregarding variations during operation. Its strengths lie in high accuracy and simplicity. Due to the binary states of each appliance (on or off), these states are easier to detect, reducing the complexity of the models. However, this approach might lead to inaccurate estimations of energy consumption by treating all active states uniformly.

- **Multiple States Recognition:** Often associated with Markov Chain models as seen in Liu et al.[50], and Azaza et al. [8], this approach, akin to a Finite State Machine (FSM), differentiates various operational states of an appliance. It aims to identify the specific state an appliance is in during operation, thus capturing more nuanced activities. However, the challenge lies in manually defining the power consumption ranges for each state.

- **Wave Reconstruction:** This method aspires to reconstruct the power wave of each appliance. Perfect reconstruction is challenging due to the inherent variability in appliance operation and circuit-induced noise. Furthermore, the high correlation among appliance operations can lead to compounded errors from any noise or inaccuracies.

In practical applications, the first NILM product was introduced by Enetics, Inc., a certified meter Data service provider, with the launch of their SPEED software in 1996 (www.enetics.com). Since then, numerous energy service companies have offered NILM services globally, catering to both residential and commercial sectors. These services encompass various applications such as fault detection in appliances, scheduling, and consumer education, providing disaggregated data in real-time, or at various intervals like hourly or minutely. For instance, the 'Trickl' mobile application by London Hydro Inc., a Canadian energy provider, shares hourly NILM data with its customers. According to a study by Home Energy Analysis, Inc., NILM technologies have contributed to an average energy consumption reduction of 12.8% (www.nilm.eu).

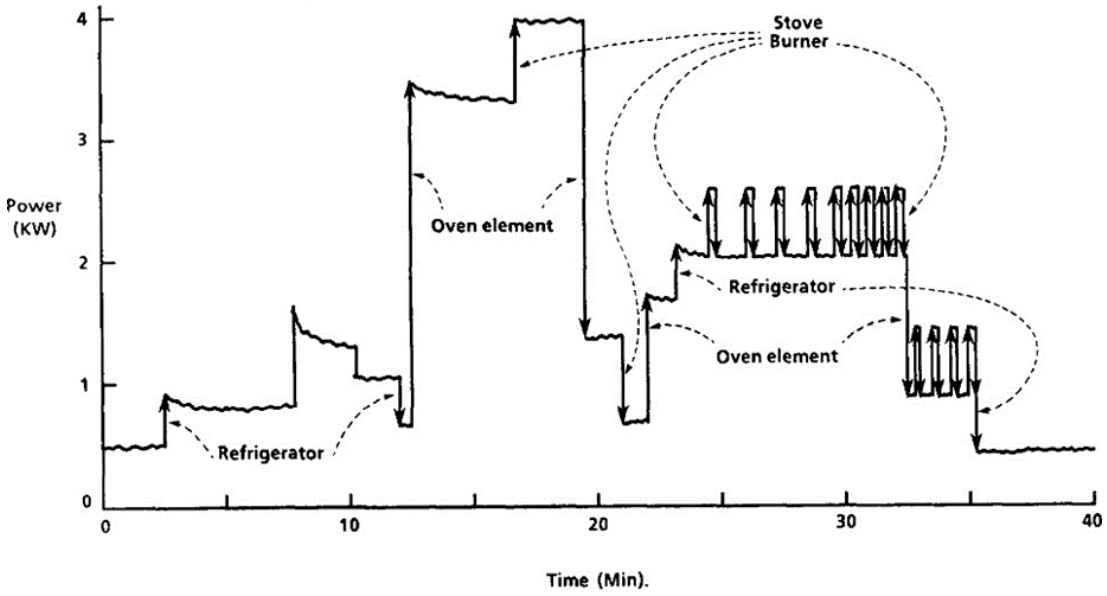


Figure 2.6: NILM event capturing [89]

2.9 Evaluating Machine Learning

2.9.1 Model Performance Metrics

In machine learning, especially in the context of network attack detection, it is critical to measure performance accurately. Performance is commonly evaluated using metrics such as the Area Under The Curve (AUC) of Receiver Operating Characteristics (ROC) curves and the F1 score [3]. The F1 score is particularly crucial in scenarios where datasets are imbalanced. Relying solely on accuracy for performance evaluation in such cases is insufficient. The F1 score, nearing 1.0, signifies a robust model, representing a harmonized average of precision and recall. Additionally, the assessment includes the False Positive Rate (FPR) and False Negative Rate (FNR), which are vital for gauging the efficacy of detection systems.

The formulas for calculating these rates are as follows:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{2.1}$$

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}} \tag{2.2}$$

Accuracy, denoted in Equation 3, is the proportion of correctly classified attack instances out of the total number of observations [33].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{2.3}$$

Where TP is True Positive, FP is False Positive, TN is True Negative, and FN is False Negative.

Recall, as indicated in Equation 4, is the fraction of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.4}$$

The F1 score, as described in Equation 5, is derived from precision and recall [61, 73].

$$\text{F1-Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \tag{2.5}$$

2.10 Chapter Summary and Overview

Utility Providers (UPs) globally strive to reduce operational expenses and enhance revenue, often confronting challenges like technical losses (TLs) and non-technical losses (NTLs), with energy theft being a primary factor in NTLs. This type of loss not only inflates operational costs but also indirectly burdens honest consumers through higher energy prices. The advent of Smart Grid (SG) technologies has ushered in more effective strategies for detecting and analyzing potential energy fraud. The fine-grained data provided by Smart Meters (SMs) and other intelligent devices can be utilized through data analytics and software applications to pinpoint instances of energy fraud and metering defects with precision. The integration of these smart devices into revenue protection systems offers UPs substantial returns on their SG investments.

The current paradigm in combating energy theft involves leveraging consumers' detailed energy usage data coupled with sophisticated data analytics. This approach is akin to fraud detection mechanisms used in other industries, such as

banking and credit card transactions, and is highly recommended for adoption by UPs.

This chapter has delved into the fundamentals of SGs and various aspects of Advanced Metering Infrastructure (AMI). It provided a comprehensive review of the literature concerning electricity losses in electrical distribution systems, encompassing both TL and NTL activities. It also discussed different methods of energy theft, including direct connections, meter tampering, and methods to slow down the meter's rotating disk in both low voltage (LV) and high voltage (HV) energy meters. Finally, the chapter presented a detailed examination of existing NTL detection schemes, categorizing them into state-based, game theory-based, and classification-based detection methods. Additionally, the chapter covered the advances in Non-Intrusive Load Monitoring (NILM) as a significant development in AMI, highlighting its role in enhancing the granularity of energy consumption analysis and its importance in modern SGs.

3. Unauthorized Power Usage Detection (UPUD) System

3.1 Introduction

This research aims to enhance the detection of unauthorized power usage (UPUD) in Advanced Metering Infrastructure (AMI) within the Smart Grid (SG) context. The focus is on identifying irregularities in electricity consumption patterns of users, leveraging historical data to analyze user behavior. This study employs a supervised machine learning (ML) approach to UPUD, specifically targeting abnormal or fraudulent usage in SG meter data. Prior research has explored various ML techniques for detecting unauthorized power usage, including artificial neural networks, autoregressive integrated moving averages, time series methods, and support vector machines, as indicated in the literature [41]. Recent advancements, such as those highlighted in [52], demonstrate the effectiveness of Gradient Boosting Classifier (GBC) in non-technical loss (NTL) detection.

This paper aims to conduct an exhaustive comparative analysis of six distinct ML algorithms: Logistic Regression (LG), Support Vector Machine, Decision Tree Classifier (DTC), Random Forest (RF), Ridge Regression Classifier (RRC), and Gradient Boosting Classifier (GBC). These analyses are based on the feature engineering-based preprocessing module of the GBC, which enhances detection rates, reduces the false positive rate (FPR), and optimizes time complexity. The GBC's preprocessing module includes a stochastic feature generation function that augments FPR and detection rate (DR) by using daily electricity consumption patterns as features. Additionally, it features a weighted feature importance (WFI) extraction mechanism that minimizes training time complexity by eliminating irrelevant data, thereby also reducing storage needs for customer data in

SGs.

For each SG customer, historical real-usage data are readily available. However, UPUD samples may be rare or absent. To address this, the study simulates UPUD cases by modifying real usage data based on mathematical formulas, reflecting the concept that theft typically involves reporting lower consumption than actual usage or shifting high usage to low-tariff periods, as proposed in [1]. Unlike [2], which focuses on detecting primarily unintentional fraud, this paper concentrates exclusively on intentional unauthorized power usage detection at the appliance level (NIALM). The dataset used here contains 50% NTL samples, generated based on a previously studied dataset [1], as opposed to the 5.38 to 8.37% NTL samples in [2]. This approach enables a fair and reliable comparison of classifier performance in terms of detection efficacy.

In this chapter, our ML algorithms are based on load disaggregation otherwise known as Non-Intrusive Load Appliance Monitoring (NIALM). This technique generates appliance-level power consumption data based on a single smart meter reading to improve detection performance as well as time complexity.

3.2 Related Work

This section summarizes pivotal research in detecting power theft within smart meter networks. Jiang et al. [39] and Jokar et al. [84] focused on classification methodologies using fuzzy clustering and SVM for abnormal consumption pattern recognition. Salinas et al. [60] and Huang et al. [36] developed distributed techniques and neural network models, respectively, to identify dishonest or malicious users in the system.

Challenges in using SVM for electricity theft detection were explored by Depuru et al. [25], while Ren et al. [72] incorporated ML algorithms for False Data Injection Attacks (FDIA) in smart meters. Dayaratne et al. [21] demonstrated cost-effective FDIA using a Demand Response (DR) scheme. Park et al. [68] and Wang et al. [89] proposed anomaly detection approaches and deep learning models for Non-Intrusive Appliance Load Monitoring (NIALM), respectively.

Despite these studies, limited research has been conducted on applying ML to design NIALM smart meter consumption patterns, which has implications for billing accuracy. Our work aims to bridge this gap by using various ML classifiers

based on historical data. Security aspects of these models, crucial in the context of household energy usage prediction [78], and appliance-level power theft detection in ML models [67], have been relatively underexplored and represent a key focus of this research.

3.2.1 AMPds2 Dataset

In this research, we consider the Almanac of Minutely Power Dataset version 2 [16] dataset to detect unauthorized power usage attacks. Measurements are made available, and it is composed of two years of measurements recorded for the whole house and 20 appliances Table 4.2 inside the house. [55] Figure 2 describes the disaggregation of each appliance from the main meter into sub-meters with their loads. WHE represents the monitoring point of the whole house, while HPE represents the monitoring point of the heat pump, CWE represents the washer and DWE represents the Dishwasher. All these monitoring points are characterized by high voltage (240V) power-consuming appliances sub-meters. We discussed the UPUD attack scenarios in Section 2.1.

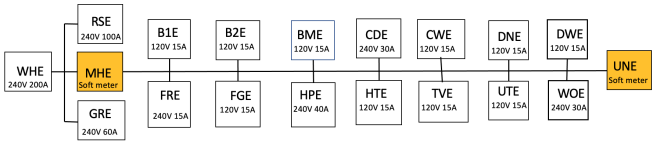


Figure 3.1: Home Appliances Disaggregated Metering Diagram with load value

3.2.2 Dataset Preliminary Preprocessing

We extracted our data from the AMPds2 public dataset, Table II, which lists 20 disaggregated appliances Figure 3.1 and their associated attributes in chronological sequence, with their unit label. This dataset has 730 rows and 1441 columns. Each row represents a day. The first 1440 columns are the power consumption [W] of a minute from 00:00 every day. For example,

- The first column is the power consumption of 00:00 of the day.
- The second column is the power consumption of 00:01 of the day.

Table 3.1: Appliances and units in AMPds

Appliance ID	Description	Appliance ID	Description
WHE	Whole House Meter	FRE	HVAC/Furnace
B1E	North Bedroom	GRE	Garage
B2E	Master/South BR	HPE	Heat Pump
BME	Basement Plugs/Lights	HTE	Instant Hot Water Unit
CDE	Clothes Dryer	OFE	Home Office
CWE	Clothes Washer	OUE	Outside Plug
DNE	Dinning Room Plugs	TVE	Entertainment TV/PVR/AMP
DWE	Dishwasher	UTE	Utility Room Plug
EBE	Electronics Workbench	WOE	Wall Oven
EQE	Security/Network	UNE	Unmetered Loads
FGE	Kitchen Fridge		.

- The 1440th column is the power consumption of 23:59 of the day.

The last column (the 1441th) column is the label.

The meaning of the label is as follows.

- Label 0 (BASE): Base power consumption of the house, which is the aggregated power of ['B1E', 'B2E', 'BME', 'CWE', 'DNE', 'DWE', 'EBE', 'EQE', 'FRE', 'OFE', 'OUE', 'TVE', 'UTE', 'WOE'] as shown in Fig 8 visualization of base minutely consumption pattern for three consecutive days.

- Label 1 (False Injection by 'CDE'): 'CDE' is added to the base power consumption only if 'CDE' has experienced any power consumption on the day.

- Label 2 (False injection by 'HPE'): 'HPE' is added to the base power consumption only if 'HPE' has experienced any power consumption on the day.

- Label 3 (False injection by 'HTE'): 'HTE' is added to the base power consumption only if 'FGE' has experienced any power consumption on the day.

- Label 4 (False injection by 'FGE'): 'FGE' is added to the base power consumption only if 'FGE' has experienced any power consumption on the day.
- Label 5 (False injection by 'WOE'): 'WOE' is added to the base power consumption only if 'WOE' has experienced any power consumption on the day.

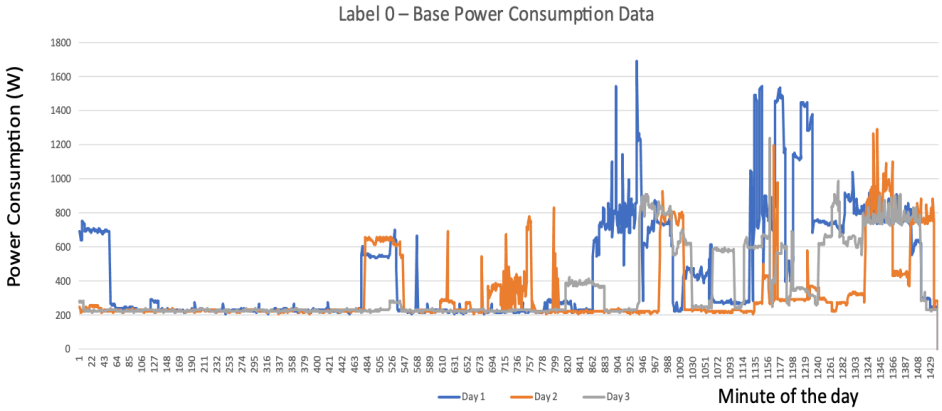


Figure 3.2: Base power consumption data pattern

The purpose of these load additions is to detect "unauthorized power usage" from the main power line.

Example of Label 0 (Base power consumption), Figure 3.2, Benign Data. This data has been pre-cleaned to provide consistent and comparable accuracy results among researchers and machine learning algorithms.

Dataset Preprocessing for Augmentation

Our labeled dataset has 6 categorical features in a numerical representation which are assigned numbers for each category. In this investigation we discover dataset imbalance, Figure 3.5. This led to overfitting of our training set as the model could not generalize well because the random fluctuations in the training data are picked up and learned as concepts by the model.

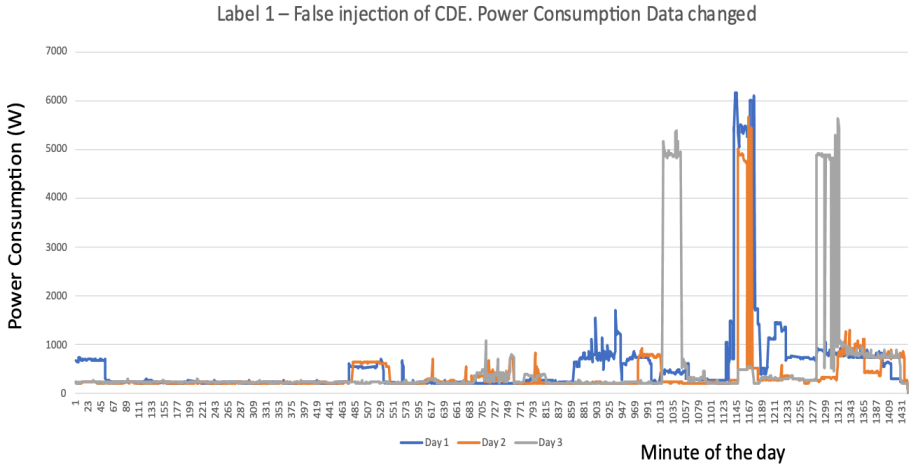


Figure 3.3: False Injection of CDE - Consumption pattern changes

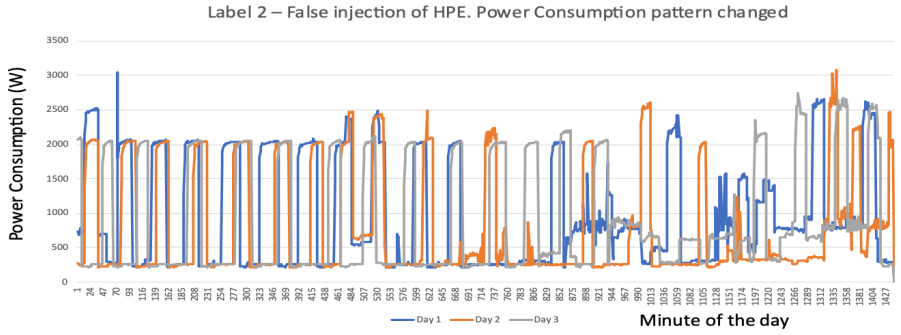


Figure 3.4: False Injection of HPE - Consumption pattern changes

We deployed a data augmentation algorithm by oversampling minority class using the Synthetic Minority Over Sampling Technique (SMOTE) to make up for the imbalance multiclass classification label. In this process, we use the K-Nearest Neighbour algorithm and try to produce a synthetic sample from our sample.

Table 3.2: NIALM-Hacking Dataset - Benign and Attack Instances

Type of Attacks	**Attack Description**	**Attack Messages**	**Number of Instances**
0 (Benign)	Base power	Appliances consumption pattern	730
1 (CDE)	Cloth dryer injection	pattern changes	283
2 (HPE)	Heat pump injection	pattern changes	730
3 (HTE)	Instant hot water injection	pattern changes	729
4 (FGE)	Kitchen Fridge injection	pattern changes	730
5 (WOE)	Wall Oven injection	pattern changes	91

3.3 Modeling Attack Scenarios

This study primarily focuses on experimenting with five types of attacks by injecting high voltage-consuming home appliances into the system. These appliances include the Clothes Dryer (CDE), Heat Pump (HPE), Instant Hot unit (HTE), Kitchen Fridge (FGE), and Wall Oven (WOE), as detailed in Table 3.2. For each type of attack, a corresponding message is defined to represent the unauthorized usage. In the context of our training model, 'Benign' refers to the base consumption pattern encompassing all appliances.

3.3.1 False Appliance Injection Attack Scenarios

Assumption:

- We assume that false injection constitutes unauthorized power usage through the illegitimate addition of appliances to the network.

In constructing our threat model, understanding the adversaries' knowledge is essential. Therefore, based on our preprocessed dataset, we have crafted six distinct attack labels. These attacks are categorized into various scenarios, each defined by different levels of energy consumption from the injected appliances.

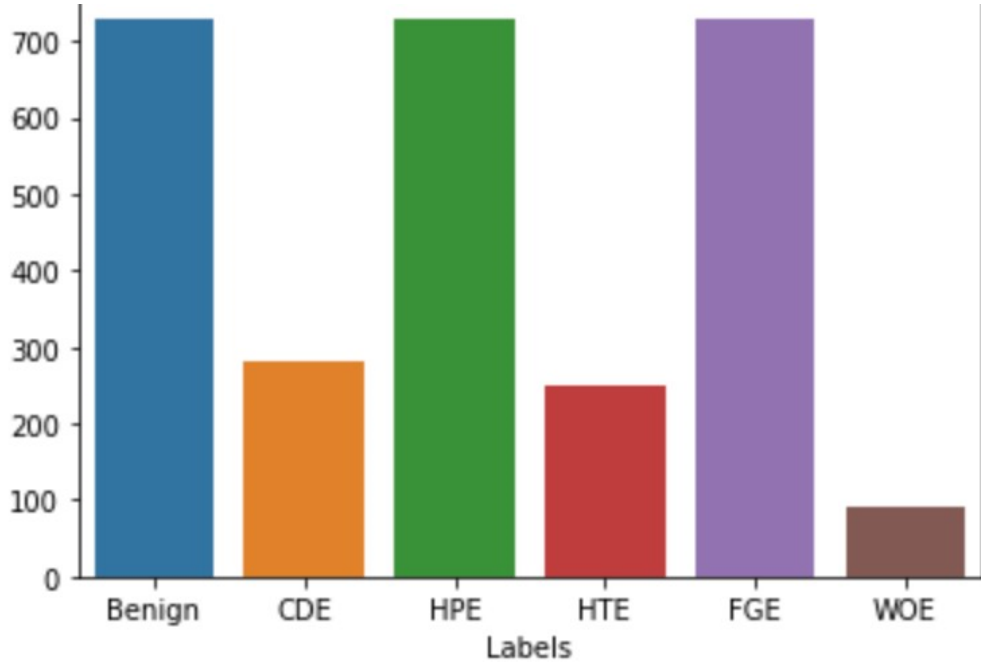


Figure 3.5: Distribution of imbalance dataset

The scenario of unauthorized power usage is a classic case of an imbalanced dataset, where the frequency of one class is significantly less than that of other classes. This imbalance is depicted in the multiclass distribution shown in Figure 3.5.

3.4 GB-based UPUD model on smart meter disaggretated data

To effectively train the Non-Intrusive Appliance Load Monitoring (NIALM) classifier, it is imperative to have access to detailed power consumption data for each appliance over a prolonged duration. For this purpose, the AMPds2 (Almanac of Minutely Power Dataset version 2) dataset was utilized. This dataset encompasses two years of comprehensive power measurements for both the entire household and 20 individual appliances, as detailed in Table 4.2, [55]. Figure 3.1 illustrates the disaggregation process, showing how each appliance's consumption is separated from the main meter into individual sub-meters, each representing a

different high voltage (240V) consuming appliance.

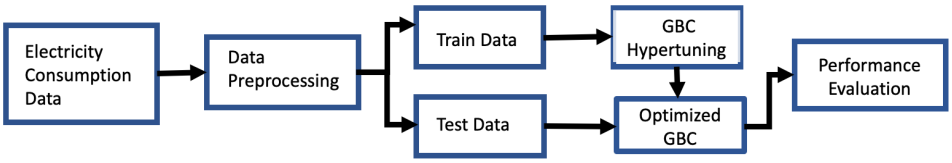


Figure 3.6: Block diagram of our proposed UPUD system

The UPUD (Unauthorized Power Usage Detection) system's workflow is depicted in Figure 3.6. Figure 3.7 further highlights the importance of sensing and decision-making in a home prediction model. Data from smart meter aggregated appliances are collected following conventional patterns and then fed into the NIALM classifier. This classifier disaggregates each appliance's consumption pattern for training the model through a sophisticated communication process where decision-making occurs.

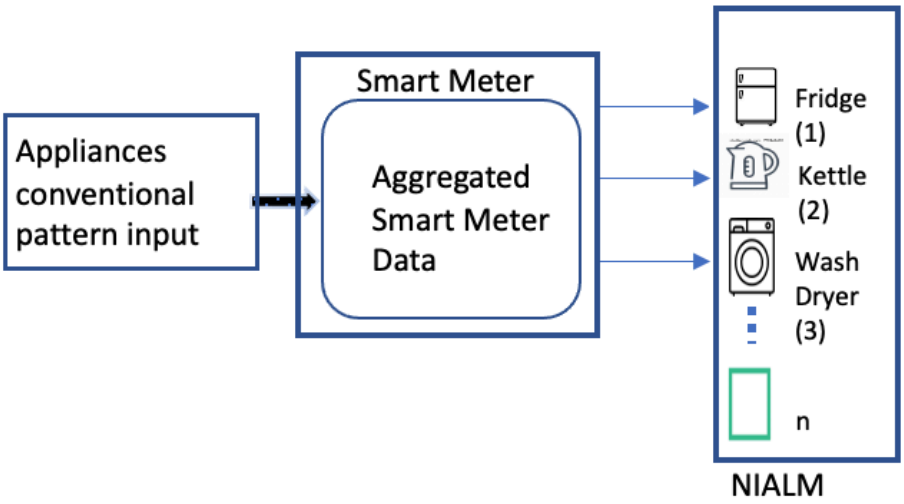


Figure 3.7: Smart meter NIALM at benign model diagram

An intelligent attacker, however, poses a significant threat by executing what is known as a false appliance injection attack or unauthorized power usage, as shown in Figure 3.8. This type of attack, indicated by the dotted rectangular

box in the NIALM model, can manifest in two ways. Firstly, it can occur during the transmission of information, akin to a man-in-the-middle or spoofing attack, where the transmitted data is susceptible to alteration. Secondly, the attack can be integrated into the machine learning model as a form of false data injection. Our proposed model is designed within this context, addressing these specific attack paradigms and ensuring robust unauthorized power usage detection.

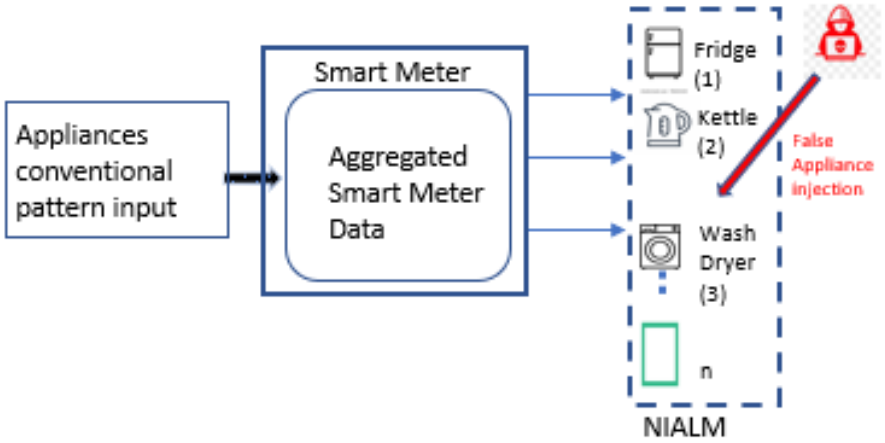


Figure 3.8: Smart meter NIALM attack model diagram

3.4.1 Proposed Gradient Boosting Classifier Algorithm

In this study, we employed machine learning (ML)-based algorithms to identify unauthorized power usage patterns in customers' electricity consumption data. The approach centered on training a classifier model in a supervised manner, utilizing a dataset that encompassed daily power usage profiles of both regular and fraudulent users.

A crucial initial step in developing this model involved preprocessing the data with a specialized data preparation technique. This step was particularly important, as the classification process might be biased towards certain classes due to data imbalances, as highlighted in Figure 3.5. To enhance the model's performance and mitigate class disparity issues, we implemented data augmentation

techniques during the preprocessing phase.

Following the preparation of the data, the model underwent hyper-tuning to refine its parameters. The optimized version of the model was then subjected to testing using a separate set of test data to evaluate its effectiveness. The entire methodology, from data preparation to model testing, is concisely illustrated in Figure 3.6, providing a clear overview of the process involved in developing the proposed Gradient Boosting Classifier Algorithm.

3.4.2 Classifier Training Process

The Machine Learning (ML) module in our model training process adopts a method akin to a forward (input) function, which ultimately yields the output. At its core, simple machine learning involves taking the input, applying weights and biases, processing it through multiple hidden layers, and then producing the output. In essence, ML serves as a technique to approximate an unknown function based on historical data or observations from a specific domain.

For our model, we have utilized a classification algorithm, a form of supervised learning, to identify new observations. This could include detecting a new appliance injection based on the patterns learned from the training data. Alternatively, regression algorithms are employed when there is a discernible relationship between input and output variables. These are typically used for predicting continuous variables, such as in weather forecasting or market trend analysis. Supervised learning models like ours are instrumental in solving various real-world challenges, including fraud detection and spam filtering. Our model, focusing on fraud detection, employs a supervised classification approach.

3.4.3 Proposed Machine Learning Approach for Training NIALM

One of the key aspects of training Non-Intrusive Appliance Load Monitoring (NIALM) is the extraction of detailed features from appliances for accurate state identification. Machine Learning facilitates this automatic feature learning process.

In the context of NIALM, numerous potential features could be extracted. For instance, the washing machine's ramping patterns, including the subtle move-

Table 3.3: GB Multiclass Classification Report

Attack labels	Accuracy	TPR	FPR	Recall	AUC	F-1 score
Benign	72.34%	0.94	0.81	0.94	0.97	0.82
CDE	100.00%	0.75	0.86	0.70	0.88	0.83
HPE	91.42%	0.89	0.90	0.86	0.92	0.88
HTE	91.04%	0.64	0.75	0.76	0.95	0.85
FGE	94.44%	0.93	0.93	0.95	0.91	0.96
WOE	67.55%	0.22	0.33	0.33	0.92	0.50

ments of the drum during its operation, are critical features. The ML algorithm takes a vector input where each input is assigned a weight. It multiplies each input by its corresponding weight and sums up the results, incorporating a non-linearity to process these inputs effectively.

3.5 Experiment results and Performance Evaluation

For our investigation, we used python PyCharm IDE 2022.5.30 and Keras with TensorFlow as the backend. We did our experiment with CPU Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz 2.50 GHz, 16GB RAM. Windows 10 (64-bit) processor.

In this study, we rigorously evaluate the effectiveness of our proposed Machine Learning (ML) model in accurately identifying instances of false appliance injections within a baseline smart meter dataset. Central to this assessment is the performance analysis of the Gradient Boosting Classifier (GBC) in the Unauthorized Power Usage Detection (UPUD) context. We focus on ensuring the reliability and robustness of our experimental analysis using various performance metrics.

Table 3.3 showcases the performance of our method in detecting six types of attacks across different training rates. The table includes metrics such as Accuracy, True Positive Rate (TPR), False Positive Rate (FPR), Recall, Area Under Curve (AUC), and F-1 Score. The test results highlight the model's exceptional capability in identifying all types of attacks.

Table 3.4: Experimental Results of different Algorithms

Models	Accuracy	TPR	FPR	FNR	Recall
Baseline	25.56%	0.04	0.17	0.07	0.00
Support Vector Machine (SVM)	68.89%	0.65	0.645	0.50	0.41
Decision Tree Classifier (DTC)	70.65%	0.61	0.62	0.62	0.70
Logistic Regression (LR)	44.86%	0.41	0.34	0.35	0.81
Random Forest (RF)	81.85%	0.74	0.68	0.70	0.41
Ridge Regression Classifier (RRC)	44.99%	0.41	0.34	0.35	0.41
Gradient Boosting (GB)	87.99%	0.93	0.76	0.80	0.87

For this experiment, we reprocessed the public AMPds2 dataset to create specific attack labels. We then conducted various attack scenarios using our multi-class classification dataset with six different ML algorithms, excluding the baseline for comparison. Notably, the Gradient Boosting Classifier emerged as the top performer, achieving an accuracy of 87.99%.

Figure 3.11 illustrates the relationship between the number of iterations and the accuracy of the augmented dataset. This graph demonstrates that higher training and testing accuracy significantly bolsters the strength of our model.

Moreover, we conducted a multi-class classification metric evaluation to determine the most accurate model in comparison to others, including the baseline, in detecting UPUD, as detailed in Table 3.4. The confusion matrix, depicted in Figure 3.9, visually represents the performance of our classification models against a given set of test data. It effectively contrasts the actual values with the model's predictions, particularly for the imbalanced dataset.

After augmenting our dataset to address imbalances, we observed an enhanced performance. Our approach involved using the Synthetic Minority Over Sampling Technique (SMOTE) for over-sampling minority classes, as detailed in the classification model report shown in Figure 3.10. This technique played a pivotal role in improving the model's detection capabilities in our investigation.

The evaluation of our balance dataset is shown in Figures 3.10 and 3.11: Accuracy and deviance graphs.

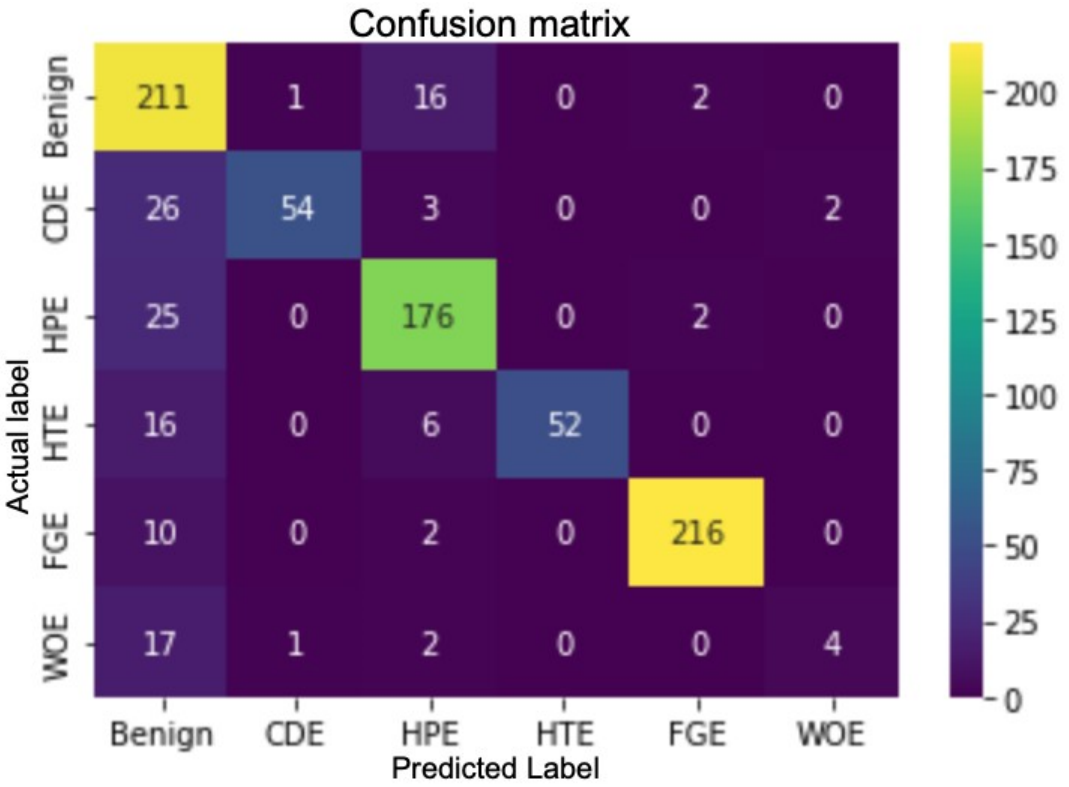


Figure 3.9: Evaluation Report of imbalance dataset

3.5.1 Discussion

The execution of this experiment presented several challenges, particularly in generating attack label datasets for Non-Intrusive Appliance Load Monitoring (NIALM).

Notwithstanding the effectiveness of the proposed method in this context, there are notable limitations when considering its application in real-world scenarios. Firstly, our method's current focus is solely on electricity consumption data. This approach may yield limited insights, as it overlooks other critical factors. Future enhancements should integrate additional data sources, such as climatic conditions (like temperature), regional characteristics, and electrical parameters (current and voltage), to provide a more comprehensive analysis.

Secondly, the scarcity of training data in practical settings poses a significant challenge. Electric companies often refrain from sharing customer data for re-

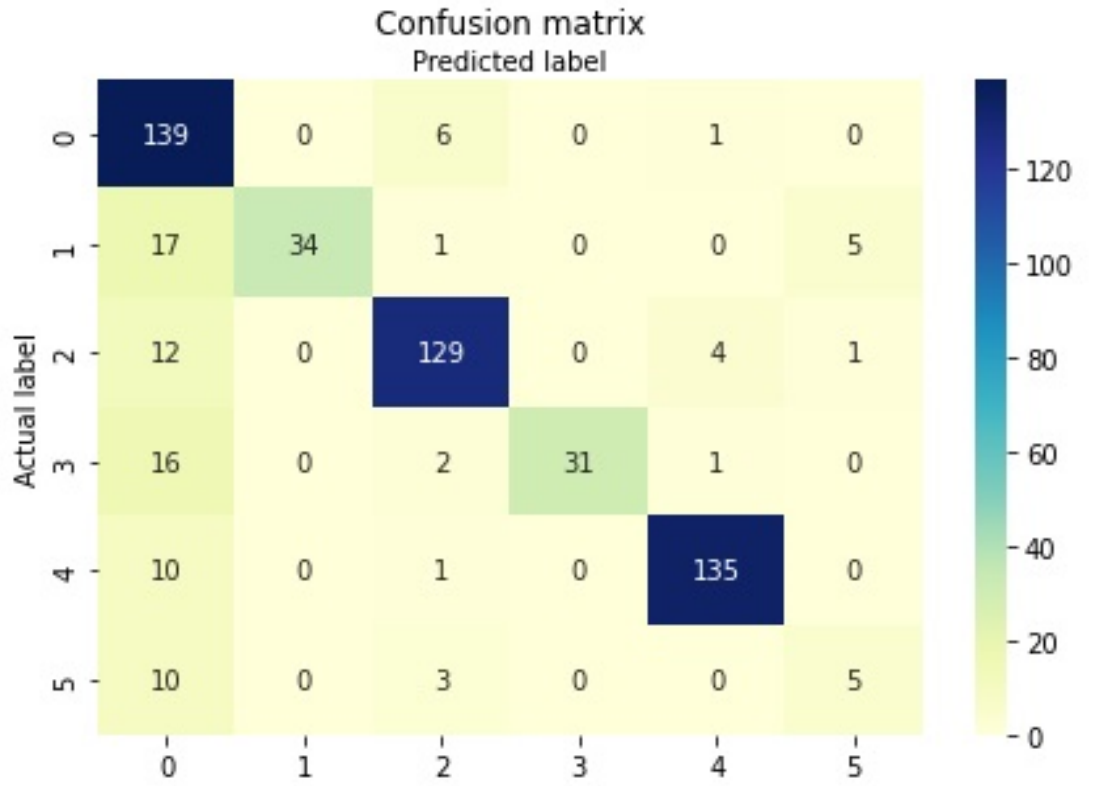


Figure 3.10: Classification Model Report after data augmentation

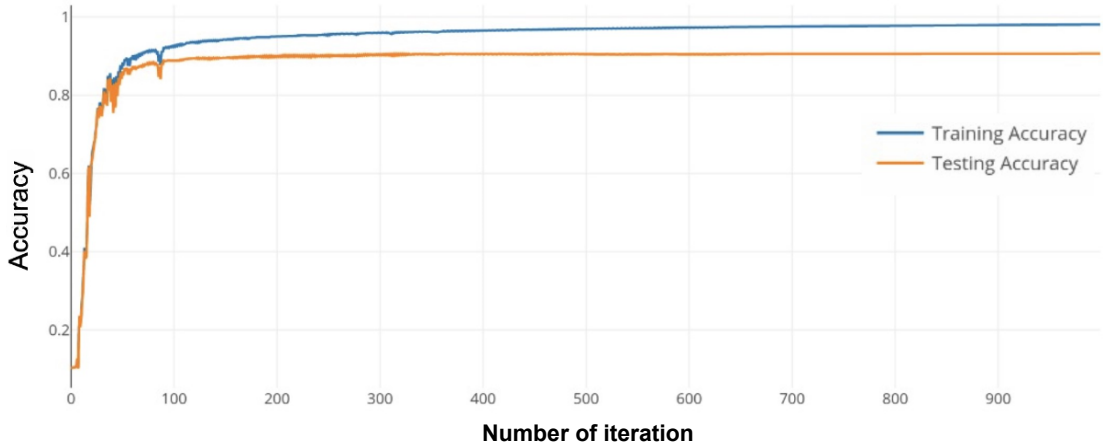


Figure 3.11: Accuracy of the balanced dataset

search purposes, primarily due to privacy concerns. Furthermore, a common issue in UPUD within NIALM is the imbalance in data sets, where malevolent or attack samples outnumber benign ones. Addressing this imbalance in a real-world setting requires the application of multiple techniques, as there is no universally applicable solution to this challenge.

Looking ahead, our future endeavors will focus on harnessing the capabilities of Deep Neural Networks and incorporating a more extensive set of training data. This approach aims to refine and enhance the outcomes of our previous experiments, further advancing the effectiveness of our methods in detecting unauthorized power usage in NIALM systems.

3.6 Chapter Summary

This study introduces a method for detecting unauthorized power usage (UPUD) in disaggregated smart meter networks, utilizing a Gradient Boosting Classifier (GBC). The GBC-based UPUD approach is designed to iteratively refine the model by adding new trees that address errors from prior iterations. The model is trained on a mix of benign and malicious samples, using the AMPd2 power preprocessed public dataset, which includes six types of attacks. The experimental results demonstrate that the GBC-based UPUD model surpasses various AI-based models such as SVM, DTC, RRC, LRC, and RF classifiers in terms of performance. Notably, the application of the Synthetic Minority Over-sampling Technique (SMOTE) effectively addresses the imbalance in the training set, contributing to the model's high performance across various parameter ranges Table 3.3. Furthermore, GBC enhances detection rates and reduces False Positive Rates (FPR) while also minimizing storage requirements and processing time for customer data. Comparative analysis using different algorithms, including Logistic Regression (LG), Support Vector Machine (SVM), Decision Tree Classifier (DTC), Random Forest (RF), Gradient Boosting (GB), and Ridge Regression Classifier (RRC), reveals that GB achieves the highest accuracy of 87.99% compared to others Table 3.4.

4. Electricity Theft Detection for Smart Homes with Knowledge-Based Synthetic Attack Data (KBSAD) Framework

4.1 Introduction

Electricity, though intangible, holds significant material value, making Electricity Theft Attacks (ETAs) a global concern, particularly prevalent in developing countries. For instance, India loses over a fifth of its electricity production to theft [70]. Traditional detection methods, such as meter inspections and user reports [7], are human-reliant and struggle with the complexities of power systems.

In smart meters, electricity consumption is aggregated, blending legitimate and malicious loads, complicating theft detection through simple analysis [65]. However, machine learning's advancement offers new avenues for detecting electricity theft from these aggregated patterns.

This paper introduces a framework for detecting electricity theft in smart homes using knowledge-based synthetic attack data. We explore five attack scenarios, focusing on detailed time-series data to extract electricity load features and predict attack types.

A major challenge in Machine Learning (ML) for Electricity Theft Detection (ETD) is data imbalance, where benign samples outnumber theft samples. Additionally, historical data often lacks examples of attack classes, particularly for zero-day attacks. We address this by generating synthetic attack datasets [41], which enhances detection capabilities across various attack types.

Training ETD models with synthetic data [69] is promising due to the predictability of theft patterns and the cost-effectiveness of not waiting for real-life attack data. Synthetic data allows for incorporating a wide range of knowledge,

aiding in effective model training.

Unlike previous studies [69] that used simple statistical models for synthetic attacks, we propose more realistic scenarios like Baseload, Weakload, Peakhour, Midnight, and Evil-Twin attacks. These scenarios, grounded in real-world intentions, offer more realistic dimensions than purely statistical approaches.

ETD research often focuses on protecting utility companies [21, 78, 69], utilizing smart meter data for demand control and dynamic pricing. However, the fine-grained data available within smart meters, though private and not reportable to utilities, can be instrumental in detecting theft at the household level for smart home applications.

It's important to differentiate between protecting utility companies and individual homes. If an attacker steals power by tampering with a meter, it reduces the home's electricity bills. Conversely, if power is stolen from one home by another, it results in increased bills for the victim, necessitating household-level theft detection.

In our study, we compare nine ML algorithms, demonstrating the superior performance of Gradient-boosting algorithms, with Random Forest as a secondary option. We also assess how these algorithms perform in classifying various attack types, noting that Weakload and Peakhour attacks present classification challenges in these models, while other algorithms struggle with different attack types.

4.2 Related Work

Aldegheishem et al. [4] have outlined five prevalent methods of electricity theft from a utility company's perspective, including bypassing meters, hacking meters, direct hooking, tapping, and meter tampering. Additionally, False Data Injection Attacks (FDIA) have been recognized as a significant cyber threat to cyber-physical systems [21, 78]. Jokar et al. [41] have also considered FDIA along with physical attacks like bypassing or tampering with meters, as these actions ultimately reflect in the meter readings as reduced electricity consumption [3]. This has led to a focus on detecting electricity theft by analyzing meter readings.

Various machine learning methods have been explored for pattern classification and electricity theft detection, including Support Vector Machines (SVM) [25,

41], Decision Trees [40], and more recently, Gradient-Boosting [69, 37]. These approaches primarily rely on supervised learning, utilizing both benign and attack data samples; however, obtaining accurately labeled data in real-life scenarios remains challenging.

Anomaly detection is another strategy employed for electricity theft detection, especially in cases where data labels are not readily available. This includes approaches like Anomaly Pattern Detection based on Hypothesis Testing (APD-HT) [68], Hierarchical Self-Organizing Maps (SOM) [84], and Stacked Sparse Denoising Autoencoders [36]. One limitation of anomaly detection is its tendency to flag all abnormal patterns, including non-malicious irregular appliance use.

Our research adopts a practical approach, acknowledging the absence of real-life labels, and integrates knowledge of potential attack scenarios with synthetic attack data to train a supervised model from unlabeled datasets. We focus on safeguarding homes against electricity theft, considering scenarios where an attacker steals electricity not from the grid but from other households, a concern often overlooked by utility companies.

Punmiya et al.'s work [69] is similar in using Gradient Boosting and synthetic theft patterns, but it is aimed at protecting utility companies rather than individual homes. Their approach involves designing theft patterns to reduce billing, while our work considers theft scenarios that increase billing. We simulate more realistic electricity theft patterns and train a multi-classification model to identify specific attack types, a novel and more complex approach than previous studies. Our research utilizes fine-grained time-series data in a smart home environment, advancing beyond current smart grid capabilities.

4.3 Electricity Theft Attacks in Smart Homes

Electricity theft in smart homes, particularly through manipulation of appliance consumption patterns, can be executed in several ways. These theft methods exploit the smart home's electrical system or the data communication network of smart meters and appliances. Below are various types of electricity theft attacks:

1. **Meter Tampering:** Physically manipulating the smart meter so it under-reports consumption. This can involve tampering with the meter's hardware

to slow down or alter its recording of electricity usage.

2. **Bypassing the Meter:** Illegally connecting electrical lines before the meter or bypassing the meter entirely. This allows electricity to be used without being measured by the meter.

3. **Remote Hacking:** Gaining unauthorized access to the smart meter's software or firmware through hacking. This can be used to alter the reported consumption data or to interfere with the meter's normal operation.

4. **Data Fabrication/Injection:** Injecting false data into the system to under-report actual consumption. This can be done by manipulating the data transmitted from the smart meter to the utility provider.

5. **Appliance Misreporting:** Hacking into smart appliances to report lower consumption than is actually occurring. Since many smart homes rely on the Internet of Things (IoT) devices, these appliances can be targets for hackers to alter reported energy use.

6. **Energy Resale or Redistribution:** Illegally siphoning off electricity to sell to others or redistributing to different locations. This involves using one connection to supply multiple users, with only one user (or none) being billed.

7. **Cloning Smart Meters:** Copying the identity of a legitimate smart meter and using it for another connection. This allows a user to consume electricity under another user's account.

8. **Signal Jamming or Interference:** Using devices to jam or interfere with the communication signals of smart meters. This can disrupt the transmission of accurate consumption data to the utility company.

9. **Algorithmic Predictive Manipulation:** Sophisticated theft methods could involve using machine learning or algorithmic techniques to predict when audits or inspections are likely to happen and temporarily reduce theft activities to avoid detection.

10. **Power Reselling/Redistribution Via Smart Inverters:** In homes with renewable energy systems like solar panels, smart inverters can be manipulated to disguise stolen grid power as self-generated power.

These methods represent a mix of old-fashioned physical tampering and modern digital hacking, reflecting the challenges posed by the digital transformation of energy systems. Utilities and smart home providers are continuously working to strengthen security measures to prevent such thefts. This includes improved physical security for meters, advanced encryption for data transmissions, regular firmware updates, and anomaly detection algorithms to identify unusual consumption patterns indicative of theft.

4.3.1 Attack Model

In this dissertation, the hypothesis is that the adversaries, represented by dishonest consumers, manipulate their smart meters (SMs) to falsify energy consumption data, thereby lowering their bills. Their objective is to report lower energy usage, resulting in financial gain at the utility providers' (UPs) expense. Section 2.3.2 outlines various established methods for illicit energy extraction from power grids. It's important to remember that these methods of energy theft, common in both traditional and smart grids (SGs), fall into three main categories: physical, cyber, and data attacks, as detailed by McLaughlin et al. (2013). Data attacks, in particular, can stem from vulnerabilities in both the cyber and physical domains.

4.3.2 Attack Scenarios

A house has an electric meter for accounting purposes. The household has to pay the bill based on the accumulated power counted by the meter. An attacker may physically access the power distribution board under the electric meter and steal power from the house to make the household pay the bills (Fig. 4.1) on their behalf. An attacker may also steal power from an outlet if it is physically accessible outside.

This kind of scenario may happen in apartments, congested complex buildings, and/or office rooms where power line cables are not easily traceable. For example, some houses might be leased with pre-deployed lines for electricity theft.

In this paper, we have identified the following five attack classes for theft consumption patterns.

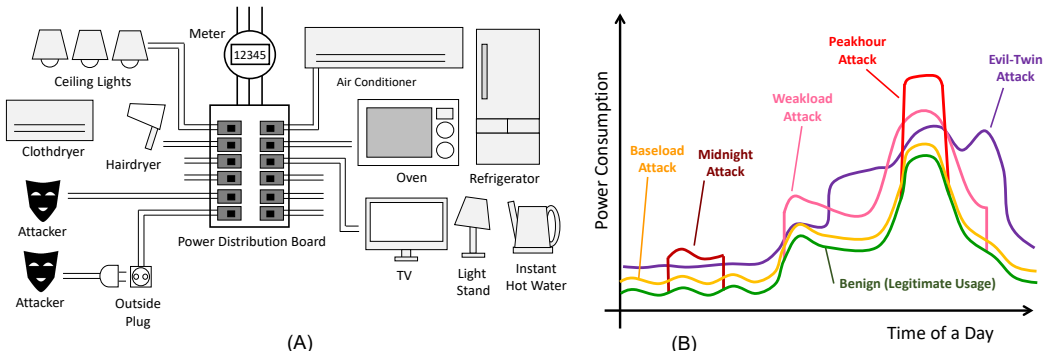


Figure 4.1: The Power Distribution board and ETA Scenarios

1. **Baseload Attack:** In the Baseload attack, the attacker continuously and constantly steals the power from the power line just as it is leaked. The instant power is not large, but accumulated theft power becomes huge. The owner of the power will not notice that the power is theft.

2. **Weakload Attack:** In the Weakload attack, the attacker has a sensor and steals the power weakly only when the power consumption of the home is high – in other words, when some appliances, such as TVs, washing machines, and refrigerator, are consuming the power. The owner of the house will simply misunderstand that the power consumption increases because they use their appliances.

3. **Peakhour Attack:** In the Peakhour attack, the attacker has a sensor and steals the power largely only when the power consumption of the home is very high – in other words, when some appliances, such as oven, heat pump, and hairdryer are consuming the power. The owner of the house will simply misunderstand like in the Weakload attack case. The attacker consumes large power during the short peak periods.

4. **Midnight Attack:** In the Midnight attack, the attacker steals the power at midnight when the households are sleeping. This attack can happen when a power outlet is available outside the house. They physically connect the

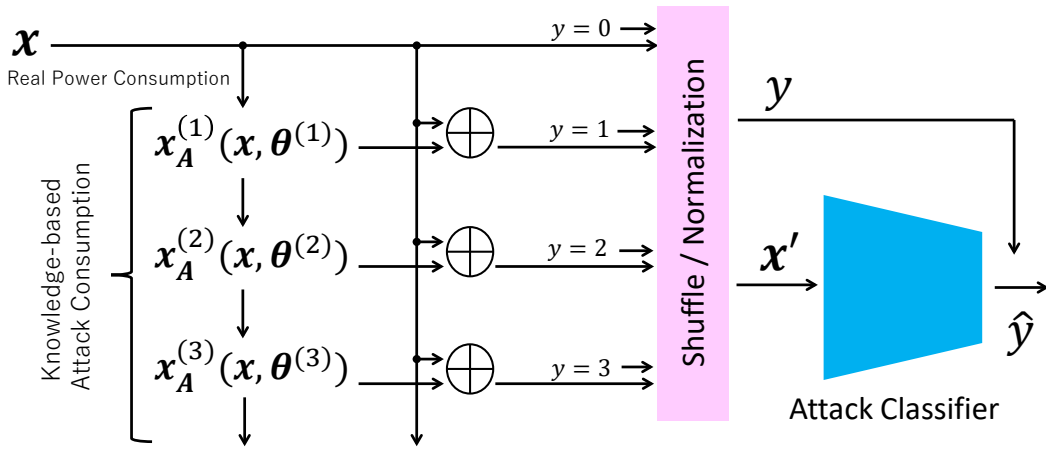


Figure 4.2: The framework for electricity theft detection with synthetic attack data.

cable to the power outlet at midnight and steal to charge their battery including electric vehicles or to boil hot water (i.e., thermal storage).

5. **Evil-Twin Attack:** In the Evil-Twin attack, the attacker steals and uses the power at their house. The pattern of power consumption will look like the aggregated power consumption of two houses. In this scenario, the same kinds of home appliances will co-exist in the consumption domain. For example, two refrigerators, two washing machines, and two TVs will appear in the power consumption patterns.

4.4 Electricity Theft Detection with Synthetic Attack Data

4.4.1 Multiclass classification approach

To detect such electricity theft attacks, we take the approach of monitoring power consumption with enough granularity in both time and power domains at the power aggregation point. In this paper, we explore the possibility of synthetic attack data deployed to validate machine learning applications for identifying the attack types only from legitimate electricity consumption patterns.

Figure 4.2 shows the framework of our synthetic attack data learning for electricity theft detection.

Let us consider $\mathbf{x}$ – a vector of power consumption. This vector contains the power consumption of each timeslot in the time order. For example, $\mathbf{x}$ may represent a power consumption of a certain day, and the i-th element x_i corresponds to the power usage at i-th minute from the beginning of the day. In this case, $\mathbf{x}$ has 1440 elements: i.e., $60 \times 24 = 1440$.

In supervised learning, we assume that each $\mathbf{x}$ has a corresponding label y for training a classification model for power consumption patterns. In our case, we can assume that label $y = 0$ as a benign case, $y = 1$ as a Baseload attack, $y = 2$ as a Weakload attack, and so on.

In the real, practical scenario, we will only get a collection of $\mathbf{x}$ from a house as a result of long-term monitoring, and we will not get real attack-enabled cases with labels. However, many power-stealing cases can be simulated just by arithmetically adding stolen power as a power consumption.

Let $\mathbf{x}_A$ be a vector of stolen power by an attacker. As we assume the attacker changes the stealing power based on the consumption of the house, $\mathbf{x}_A$ is a function of $\mathbf{x}$ and attacking parameters θ: e.g., $\mathbf{x}_A(\mathbf{x}, \theta)$.

Depending on the attack cases, i.e., depending on the label $y(\neq 0)$, we can consider different stolen power vectors: $\mathbf{x}_A^{(y)}(\mathbf{x}, \theta^{(y)})$.

Finally, we can get the labeled dataset as follows.

$$(\mathbf{x}', y) = \begin{cases} (\mathbf{x}, 0) & y = 0 \\ (\mathbf{x} + \mathbf{x}_A^{(y)}(\mathbf{x}, \theta^{(y)}), y) & y \neq 0 \end{cases} \tag{4.1}$$

Supervised machine learning models can be applied to the collection of $(\mathbf{x}', y)$.

4.5 Dataset for Electricity Theft Detection

4.5.1 Overview

We have used AMPds2 (Almanac of Minutely Power Dataset version 2) [55] as a benchmark for the two-year power consumption of homes. AMPds2 consists of minute-level measured powers at the outputs of a power distribution board. They have put the names of their monitoring points as Table 4.2.

Depending on the configuration of the houses, they may not have some appliances such as Clothes Dryers, Wall Ovens, or Dishwashers. So, by removing the

Table 4.1: The Configuration of Synthetic Attack Data For Supervised Learning in the experiment.

Attack Class	Configuration
Baseload	Theft = 100W
Weakload	Threshold=500W Theft=500W
Peakhour	Threshold=1500W Theft=2000W
Midnight	Starting from 1 AM - 2 AM (at random) Duration = 120 minutes Theft=1000W
Evil-Twin	One day was randomly chosen in the dataset.

Table 4.2: Appliances and units in AMPds

Appliance ID	**Description**	**Appliance ID**	**Description**
WHE	Whole House Meter	FRE	HVAC/Furnace
B1E	North Bedroom	GRE	Garage
B2E	Master/South BR	HPE	Heat Pump
BME	Basement Plugs/Lights	HTE	Instant Hot Water Unit
CDE	Clothes Dryer	OFE	Home Office
CWE	Clothes Washer	OUE	Outside Plug
DNE	Dinning Room Plugs	TVE	Entertainment TV/PVR/AMP
DWE	Dishwasher	UTE	Utility Room Plug
EBE	Electronics Workbench	WOE	Wall Oven
EQE	Security/Network	UNE	Unmetered Loads
FGE	Kitchen Fridge		.

power consumption of some optional appliances, we have virtually set up three types of homes, as follows.

- **Home A** is composed of a full set of appliances, which is the aggregated power consumptions of B1E, B2E, BME, CWE, DNE, HTE, EBE, EQE, FRE, OFE, OUE, TVE, UTE, CDE, HPE, DWE, FGE, and WOE. Some of

those appliances such as those associated with HPE and WOE make peak power consumption, which may allow the peak-hour attacker to steal power more efficiently.

- **Home B** excludes CDE and WOE from Home A, assuming that the existence of a cloth dryer and wall oven may influence the accuracy of attacker detection.

- **Home C** excludes DWE, HPE, and TVE from Home A, assuming that the existence of a dishwasher, heat pump, and small appliances may influence the accuracy of attacker detection.

As the electric power consumption is time-dependent data, we have picked up the first 80% of data (i.e., 584 days) for the training data, applying our synthetic attack data method, and the last 20% of data (i.e., 146 days) for the test data. The test data also contains the simulated attacks in this study. Please also note that the test data does not cover the whole year because we wanted to use a larger amount of data for training.

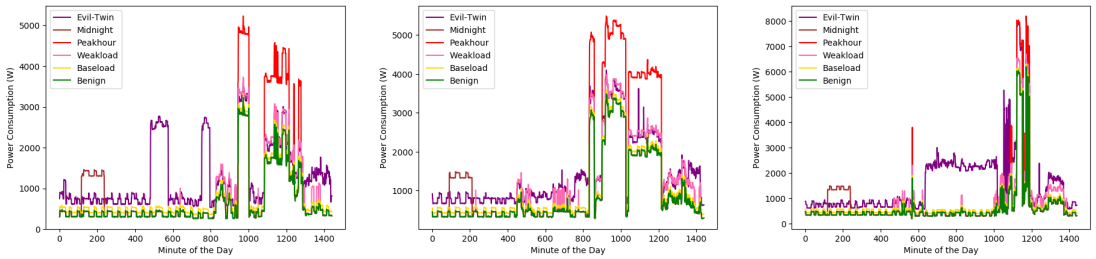
4.5.2 Data Profiles

Table 4.3 shows the profiles of the data for our ETA detection study. We have 584 benign records for training for each home from the original monitoring data. We simulated and added Baseload, Weakload, Peakhour, Midnight, and Evil-Twin attacks based on the definitions in Section III, and the method of Section IV. For more details, Table 4.1 shows the parameters of each attack. In some conditions, if the base power consumption of the home does not reach the threshold, Weakload and Peakhour attacks are not triggered. Such attack samples are not included in the dataset.

Figure 4.3 shows examples of Benign and Attack data samples of different three days of Home A. From these figures, we can observe that the electric consumption pattern has a huge change by the day, and we cannot easily recognize the attack class only from a single power consumption data.

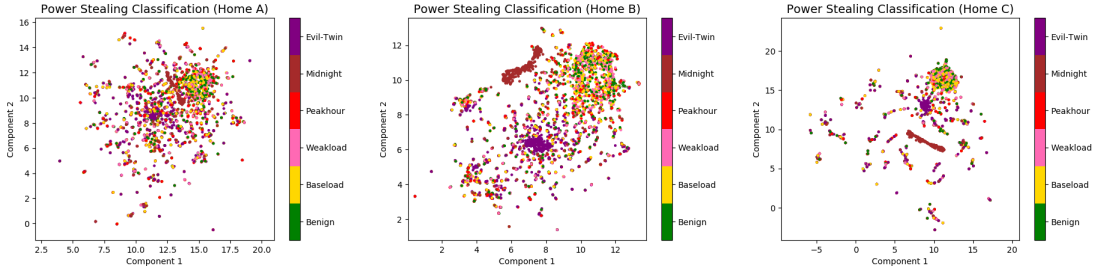
Figure 4.4 shows the Uniform manifold approximation and projection (UMAP) [58] of the data samples for Homes A, B, and C. Each plot corresponds to a day.

We can observe that different homes have different characteristics in data. For example, in Home A, benign data is scattered having overlaps with all the types of attack samples, whereas in Home C, we can observe more clusters that might be easier for classification.



(a) Day 1 (b) Day 2 (c) Day 3

Figure 4.3: Electricity usage of Home A with synthetic attack data on different days.



(a) Home A (b) Home B (c) Home C

Figure 4.4: 2D projections of attack contained electricity usage with the benign case by UMAP

4.5.3 Attack Impact

We define attack impact (AI) as a score to measure electricity theft. This attack impact is the amount of stolen energy in the bill.

Let w_i be the weighted price of the power of x_i on the day at each index i. Let us denote the vector of w_i by $\mathbf{w}$.

The attack impact of electricity theft of the day is :

$$I = \mathbf{w} \cdot \mathbf{x_A} = \sum_i w_i x_{Ai} \tag{4.2}$$

Here, x_{Ai} represents the power stolen at index i of $\mathbf{x_A}$.

For calculating AI on our dataset, we have assumed 0.20 USD/kWh constantly for the unit price. The daily bill mounted by the attacker is around 1 USD on average (see, Table 4.3). It will be about 30 USD monthly, and 360 USD in a year. If the base power consumption of the home is larger, such as in the case of an office, shop, restaurant, or factory, the attacker will be able to steal much more power.

Table 4.3: The Profile of the dataset generated with synthetic attack data, and corresponding attack impacts (AI).

Home	Category	Benign	Baseload	Weakload	Peakhour	Midnight	Evil-Twin	Averaged AI	AI Ratio
A	Train	584	584	580	490	584	584	1.15 USD	1.77
	Test	146	146	146	140	146	146	1.17 USD	1.72
B	Train	584	584	580	437	584	584	1.08 USD	1.80
	Test	146	146	146	130	146	146	1.10 USD	1.77
C	Train	584	584	576	306	584	584	0.78 USD	1.66
	Test	146	146	146	84	146	146	0.76 USD	1.68

4.6 Evaluation

4.6.1 Experiment Settings

In this study, we have taken the approach of surveying a wide range of machine learning algorithms with the synthetic framework in order to find the ranking of the algorithms for our problem.

Table 4.4 shows the configuration of the experiment in our evaluation. The details of these configurations are given below.

For Gradient Boosting (GB) classifier, we have set up (1) the learning rate which shrinks the contribution of each tree to default value = 0.1, (2) the number of boosting stages to perform n_estimator to default value = 100.

For Extreme Gradient Boosting (XGB), we have used an open-source library 'xgboost' [1] that provides an efficient and effective implementation of the gradient boosting algorithm. We used the python library to run Extreme Gradient Boosting to predict power output using the default extreme gradient boost classifier model: objective="multi:softprob", random_state=0.

For Random Forest (RF) estimator, we used the default hyperparameter values with n_estimator = 100.

Table 4.4: Models parameter configuration

Model	Library	Configuration
GB	sklearn	learning rate = 0.1 n_estimator=100
XGB	xgboost	objective="multi:softprob" random_state=0
RF	sklearn	n_estimator=100
DTC	sklearn	random_state=12
MLP-4	sklearn	hidden_layer_1=720 hidden_layer_2=200 epoch=500
MLP-3	sklearn	hidden_layer_1=720 epoch=500
LR	sklearn	copy=True with_mean=True with_std=True
RRC	sklearn	copy=True with_mean=True with_std=True
SVM	sklearn	c=1000 gamma=0.01 kernel='rbf'

Table 4.5: Experimental results of different algorithms.

Ranking	**Models**	**Home**	**Acc.**	**P**	**R**	**F1**
1	Gradient Boosting(GB)	A	92.98%	0.94	0.93	0.93
		B	94.53%	0.95	0.95	0.95
		C	93.61%	0.94	0.94	0.93
2	Extreme Gradient Boosting(XGB)	A	92.53%	0.94	0.93	0.92
		B	95.00%	0.96	0.95	0.95
		C	92.75%	0.94	0.93	0.92
3	Random Forest (RF)	A	86.78%	0.91	0.87	0.87
		B	89.19%	0.92	0.89	0.90
		C	88.94%	0.93	0.89	0.88
4	Decision Tree Classifier(DTC)	A	70.58%	0.75	0.71	0.71
		B	77.33%	0.81	0.77	0.78
		C	80.47%	0.82	0.80	0.80
5	Multilayer Perceptron 4 (MLP-4)	A	66.44%	0.70	0.66	0.67
		B	71.28%	0.76	0.71	0.72
		C	77.52%	0.78	0.78	0.78
6	Multilayer Perceptron 3 (MLP-3)	A	65.75%	0.69	0.66	0.66
		B	69.88%	0.73	0.70	0.71
		C	71.01%	0.76	0.71	0.72
7	Logistic Regression (LR)	A	54.83%	0.54	0.55	0.54
		B	56.86%	0.56	0.57	0.55
		C	70.76%	0.73	0.71	0.71
8	Ridge Regression Classifier(RRC)	A	52.30%	0.53	0.52	0.52
		B	54.88%	0.56	0.55	0.54
		C	62.04%	0.67	0.62	0.62
9	Support Vector Machine(SVM)	A	52.29%	0.51	0.52	0.51
		B	56.16%	0.56	0.56	0.55
		C	67.69%	0.69	0.68	0.67

Acc., P, R, F1 refer to the Accuracy, Precision, Recall and F1-score.

For Multilayer Perceptrons (MLPs), we have tested with 3-layer MLP (MLP-

3) and 4-layer MLP (MLP-4). For MLP-3, we have set 720 features for the sole hidden layer. For MLP-4, we have set 720 features for the first hidden layer and 200 features for the second hidden layer. We have run 500 epochs for training. We used the default values of sklearn for other hyperparameters.For Decision Tree Classifier (DTC) we used random_state=12 using the Sklearn library.

For Logistic Regression (LR) and Ridge Regression Classifier (RRC), we have used sklearn.preprocessing.StandardScaler (*, copy=True, with_mean=True, with_std=True) default parameters to avoid the dataset behaving badly if the individual feature does not more or less look like standard normally distributed data to machine learning estimators.

For SVM, we used the default hyperparameter, error regularization, c = 1000, preventing overfitting with gamma = 0.01 and defining the function to transform the dataset, kernel= 'rbf'.

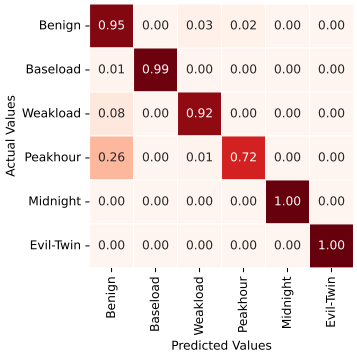
For carrying out these experiments, we have used Google Colaboratory.

4.6.2 Performance Overview

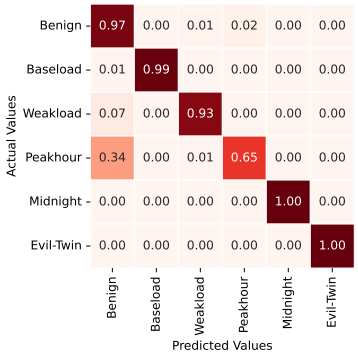
Table 4.5 shows the ranking of the algorithms based on the accuracy of Home A. GB and XGB performed the best in every home scenario. RF could also achieve higher accuracy than the others. Compared to them, MLPs did not perform well. We found that SVM, LR, and RRC do not fit our electricity theft dataset well.

In many algorithms except RF, GB, and XGB, we have observed the performances of homes as Home C > Home B > Home A. There were about 5% to 15% gaps between their accuracies. This is probably because Home A contains a lot of legitimate consumption peaks, whereas Homes B and C do not. This indicates that the detection of electricity theft in Home A is potentially more difficult than in Homes B and C.

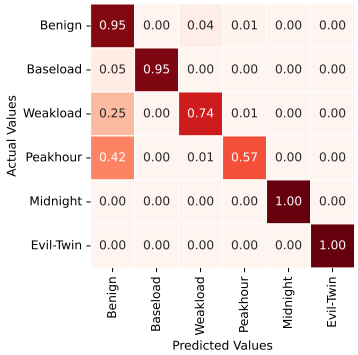
However, RF, GB, and XGB could successfully achieve high accuracy in Home A although it is still lower than the other homes. These results indicate that RF, GB, and XGB, especially GB and XGB are promising algorithms for detecting electricity theft with smart meters.



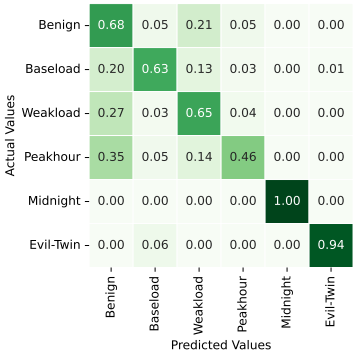
(No.1) Gradient Boosting (GB)



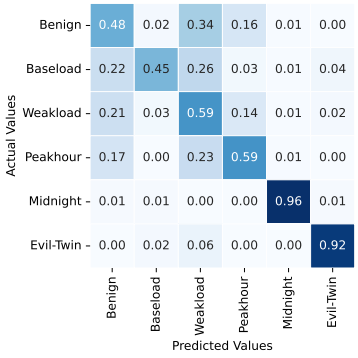
(No.2) XG Boosting (XGB)



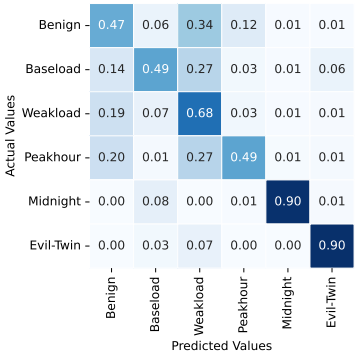
(No.3) Random Forest (RF)



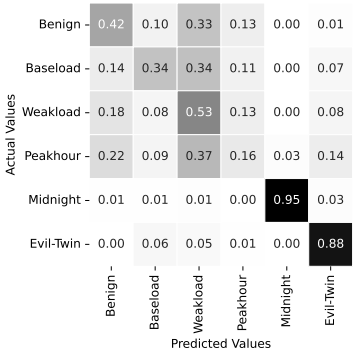
(No.4) Decision Tree (DTC)



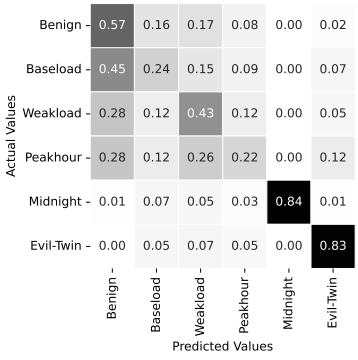
(No.5) Multilayer Perceptron 4 (MLP-4)



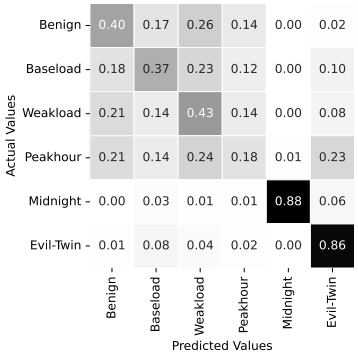
(No.6) Multilayer Perceptron 3 (MLP-3)



(No.7) Logistic Regression



(No.8) Ridge Regression Classifier (RRC)



(No.9) Support Vector Machine (SVM)

Figure 4.5: Confusion matrices of the trained models. The number (No. #) indicates the rank of the overall accuracy.

4.6.3 Performance by Attack Class

Figure 4.5 shows the confusion matrices of the trained models for Home A. The colors of this figure are:

- Red if the overall accuracy is more than 80%.
- Green if the overall accuracy is between 70% and 80%.
- Blue if the overall accuracy is between 60% and 70%.
- Gray if the overall accuracy is less than 60%.

The numbers (i.e., No. #) indicate the rank, which corresponds to Table 4.5.

The colors indicate the group of accuracy. Midnight and Evil-Twin attacks were relatively easily detected by all the models, especially achieving 100% accuracy with GB, XGB, Baseload, Weakload, and Peakhour attacks were difficult to classify but GB and XGB could perform well. The false positive rates of GB, XGB, and RF were about 5% which is much better than other models.

We can observe that Midnight and Evil-Twin attacks were well classified in all the classifiers. Especially, with GB, XGB, and RF – i.e., the best three algorithms, have achieved 100% accuracy for these attacks. Compared to these attacks, Baseload, Weakload, and Peakhour attacks were difficult to classify. For example, we can observe confusion in classifying these attacks with Benign samples in DTC, MLP, LR, RRC, and SVM. Especially, the false positives of these algorithms were not good. For example, in LR, 33% of the Benign cases were predicted as Weakload attacks. In MLPs, 34% of the Benign cases were predicted as Peakhour attacks. These false positives should be critical in real operational scenarios. GB and XGB also have false positives in Weakload and Peakhour attacks. However, the ratio has decreased to less than 3%. They have false negatives in Peakhour and Weakload attacks with 26-34% and 7-8% respectively, which can be eventually detected if they continuously perform the attacks every day onwards.

4.7 Discussion

This research presents a machine learning approach for identifying and categorizing electricity theft in smart home environments, employing synthetic attack data informed by specific attack scenarios.

Our investigation revealed that the Gradient Boosting model demonstrates superior performance when compared to a range of machine learning algorithms. This finding suggests its effectiveness in this specific context of electricity theft detection.

Looking ahead, our research can be extended in several directions. Firstly, there is scope for expanding the range of attack types beyond the five currently defined in our study. Additionally, exploring diverse household profiles and varying patterns of electricity consumption, particularly in the context of sophisticated attacks like the Evil-Twin, could further refine the model's accuracy.

Another critical aspect to consider is the impact of the attack on the detection capability. The relationship between the magnitude of an attack's impact and its detectability needs thorough exploration. High-impact attacks are more likely to be detected, whereas low-impact ones might elude detection despite their potential cumulative significance over time.

Moreover, adapting to short monitoring durations and seasonal variations, such as increased heat pump usage, could enhance model accuracy. The uniqueness of electricity usage, influenced by factors like family composition, home size, location, and personal habits, underscores the need for personalized models. One potential avenue to address these privacy-sensitive issues is through the adoption of federated learning, which allows for the personalization of a general model without compromising data privacy. Future research could explore this domain, as suggested by Tan et al. in their recent work [81].

4.8 Chapter Summary

This research focuses on the detection of electricity theft in smart homes, utilizing synthetic attack data informed by specific knowledge. We identified five distinct attack types that illicitly siphon electricity from other homes. Our analysis, based on the minute-level, two-year monitoring data from the AMPds2 dataset, revealed

that Gradient Boosting algorithms outperformed others, achieving an impressive 93% accuracy, with Random Forest as a close alternative at 87%.

In our classification, Baseload, Midnight, and Evil-Twin attacks—referred to as legacy attacks—were almost perfectly identified with near 100% accuracy. On the other hand, the more sophisticated Weakload and Peakhour attacks posed a greater challenge and were not as effectively detected. While further investigation into additional scenarios is warranted, the preliminary results suggest that our methodology holds significant promise for practical application in identifying and categorizing electricity theft in smart home settings.

5. Electricity Theft Detection for Smart Homes: Harnessing the Power of Machine Learning with Real and Synthetic Attacks

5.1 Introduction

In the dynamic field of Electricity Theft Detection (ETD), the ability to discern anomalous consumption patterns within smart home environments is critical. Given the inherent complexity and variability in electricity usage data, enhancing the diversity and volume of training data is crucial for the development of robust and reliable anomaly detection models. This study introduces a novel approach to data augmentation, an essential component in addressing the challenge of imbalanced datasets, a common occurrence in ETD scenarios.

5.2 Related Work

Although there is extensive research on appliance consumption patterns [77], [91],[95], there is a lack of specific focus on using these patterns to train a classifier for electricity theft detection. Existing literature primarily discusses load disaggregation, human activity recognition, and energy consumption forecasting based on appliance power consumption patterns. However, the direct application of these patterns to train a classifier for electricity theft detection has not been explicitly addressed.

The literature provides insights into the challenges associated with appliance consumption patterns, such as the complexity of identifying appliance-specific consumption patterns and overlapping operation of appliances, which makes event detection difficult [88].

Additionally, some studies discuss the application of appliance power consumption patterns for simulating human living activities [95] and improving residential load disaggregation [64]. Furthermore, some studies have emphasized the accuracy of identifying appliance usage patterns using the proposed models [77], [91].

However, the specific task of using appliance consumption patterns to train a classifier for electricity theft detection in smart homes remains underexplored. The references did not directly address the development of a classifier for detecting electricity theft based on appliance consumption patterns. Therefore, there is a clear gap in existing knowledge regarding this specific application.

Although the literature provides valuable insights into appliance consumption patterns and their applications, there is a notable gap in knowledge concerning the direct utilization of these patterns to train a classifier for electricity theft detection in smart homes.

Our work operates within the context of real-life scenarios where labeled data are scarce. However, we introduced a novel approach by incorporating knowledge of potential attack scenarios and synthetic attack data to train a supervised model using a non-labeled real-world dataset. Furthermore, our work capitalizes on fine-grained time-series data within a smart home environment, a resource that is currently unavailable in today's smart grid landscape.

In our previous work [2], we introduced nine algorithms for detecting five real-world simulated attack classes in smart homes based on appliance consumption patterns. The present work is an extension of [2]. This paper is focused on electricity theft detection in smart homes and improvements relative to [2] including making the algorithm robust against unclassified attacks, application of synthetic binary discriminator, and legacy unsupervised techniques to enhance classification accuracy, employment of real building appliance consumption dataset for performance evaluations and model comparison with other existing models.

5.3 Attacks Model Beyond The Distribution Board

In the context of a Synthetic Binary Discriminator Model (SYNBDM) used for detecting electricity theft and ensuring appliance usage authentication in smart homes, various attack scenarios can threaten the integrity, availability, and confidentiality of the system. These threats are not limited to physical interactions

with the distribution board, as in Figure 4.1, but can also involve digital intrusions and manipulative tactics. Here are some potential attack scenarios:

Physical Attacks:

1. **Meter Swapping:** Swapping meters with those from vacant or low-consumption premises.
2. **Power Diversion:** Rerouting the power supply within a community.
3. **Meter Tampering:** This encompasses removing or disconnecting meters, inverting meters, employing magnets to disrupt readings, and unauthorized Smart Meter (SM) access.

Cyber Attacks:

1. **Credential Theft:** Gaining unauthorized meter access via stolen login details.
2. **Firmware Hacking:** Compromising Smart Meter firmware remotely.
3. **Data Tampering:** Modifying stored meter data, including total energy use, audit trails, and cryptographic keys.

Data Attacks:

1. **Zero/Negative Reporting:** Incorrectly reporting no or negative energy use.
2. **Consumption Report Alteration:** Halting or modifying energy consumption reports.
3. **Measurement Exclusion:** Excluding high-usage appliances from records.

5.4 Knowledge-Based Attack Simulation Framework

The framework proposed in Figure 5.1, employs a knowledge-based approach to generate synthetic attack scenarios on power consumption data, encapsulating

domain expertise within its operational paradigm. The core of this methodology is the utilization of the actual power consumption profiles, denoted as x, as the foundational dataset from which attack patterns are derived. This framework benefits from recognizing both specific attacks and unclassified anomalies, integrating the strengths of both approaches for a robust security posture.

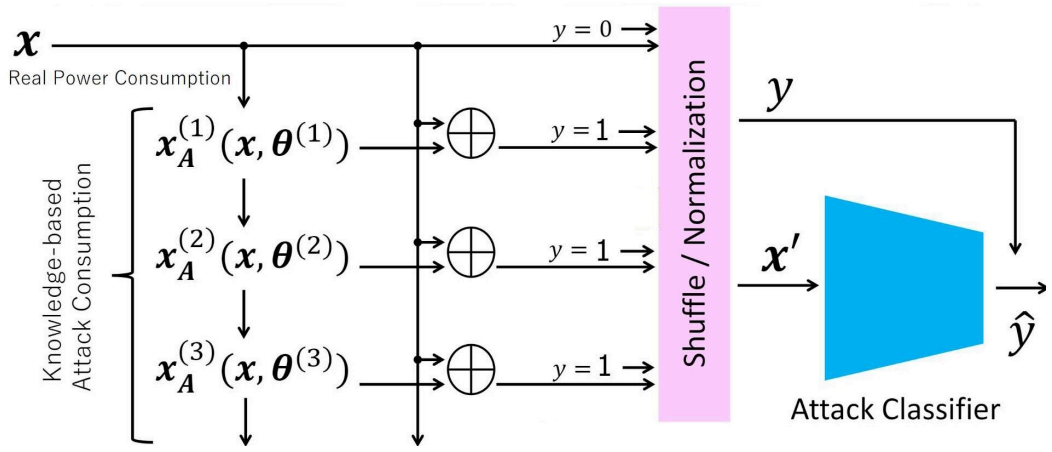


Figure 5.1: ETD framework for processing real power consumption data to detect anomalies.

5.4.1 Attack Data Generation

Distinct attack scenarios are simulated through a series of transformations applied to real consumption data, parameterized by θ. These transformations—$x_A(x, \theta^{(1)})$, $x_A(x, \theta^{(2)})$, and $x_A(x, \theta^{(3)})$—are crafted based on expert insights into the modus operandi of various attack vectors, with each θ iteration representing a unique attack typology.

5.4.2 Data Labeling and Preprocessing

Further cementing its knowledge-driven architecture, the framework classifies consumption data into normal ($y = 0$) and anomalous ($y = 1$) states, employing preestablished criteria that delineate normalcy from theft-related anomalies. Before classification, data undergo a shuffling and normalization process, for eliminating potential classifier bias attributable to sequential order or feature scale disparities.

5.4.3 Attack Classification

The culmination of the framework is the attack classifier, a predictive model trained on a rich historical corpus comprising known instances of consumption patterns, both benign and malignant. This classifier is not merely a data-driven algorithm but a knowledge-infused system tuned to recognize and react to the subtle intricacies of electricity theft within smart grid environments.

The framework's reliance on domain-specific knowledge for the generation and processing of data points designates it as knowledge-based. This is exemplified by the methodical application of expert understanding to the identification of theft signatures, which is paramount for effective discrimination between legitimate and fraudulent electricity usage patterns.

Our knowledge-based framework sets a new benchmark for electricity theft detection systems, marrying the depth of domain knowledge with the rigor of machine learning classification. This synergy promises a robust and discerning methodology, poised to advance the state-of-the-art in smart grid security.

Our research is propelled by the crucial need to equip home operators with the ability to effectively detect and categorize cyber threats. Hence the following benefits are manifold:

- **Rapid response**: Our system enables operators to recognize threats immediately by employing binary classification techniques, prompting a quick response to cyber incidents.
- **In-depth analysis and prevention**: Through multiclass classification, we provide comprehensive insights into the nature of attacks, which is essential for effectively strategizing prevention and allocating defensive resources.
- **Strategic planning & resilience**: Our approach to specific attack classifications underpins the development of customized incident response strategies, thereby fortifying system resilience.
- **Compliance and confidence**: Our method ensures compliance with cybersecurity regulations, thereby reinforcing customer trust in protective measures.

- **Operational integrity**: The efficiency of our detection and classification system is paramount for sustaining operational continuity and mitigating the impacts of cyber threats.

Figure 5.2 depicts the architecture of the proposed ETD mode to consolidate appliance consumption data from smart homes, ensuring privacy through anonymization and consistency through normalization. Simulated theft scenarios enhance the dataset, with each instance labeled as normal or fraudulent as described in Section V.

A training set derived from this data trains algorithms to detect consumption patterns, whereas a testing set comprising simulated and real data evaluates accuracy. The model employs both traditional machine learning and neural network classifiers, benchmarked against metrics such as accuracy, area under curve (AUC), and F1-score. In post-validation, the model was deployed, with ongoing retraining to refine its detection capabilities.

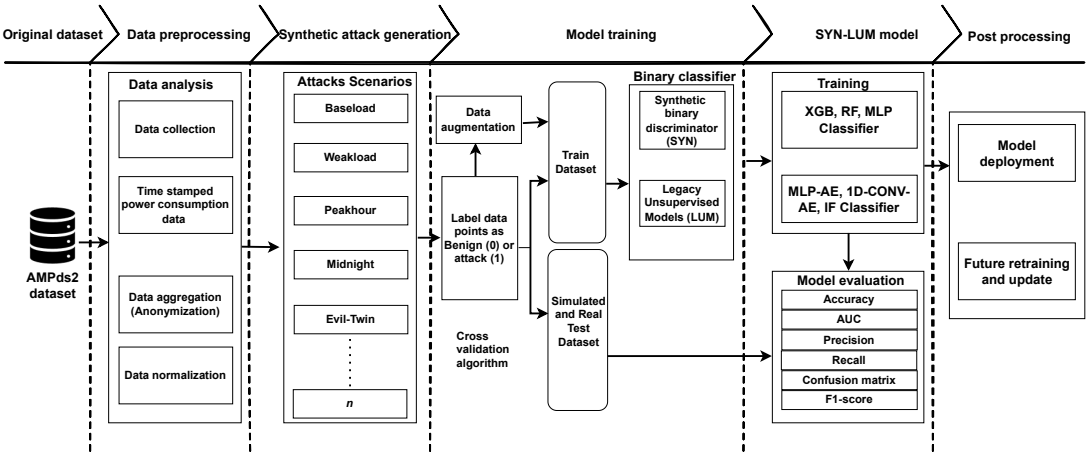


Figure 5.2: Flow of ETD model

5.5 Dataset for electricity theft detection

5.5.1 Data Collection

For our study, we utilized the AMPds2 dataset (Almanac of Minutely Power Dataset version 2)[55] as a benchmark, representing two years' worth of home

power consumption data. AMPds2 includes minute-by-minute power measurements recorded at the outputs of power distribution boards as shown in Figure 4.1A in the previous chapter. The monitoring points within this dataset are detailed in Table 4.2. Given the varying configurations of homes, it is possible that certain appliances, such as Clothes Dryers, Wall Ovens, or Dishwashers, may not be present in some households. To address this variability, we simulated three distinct home types by excluding the power consumption of these optional appliances as listed in Table 5.1.

In configuring each home, we assumed the following:

- **Home A**, some appliances associated with HPE and WOE have peak power consumption, (Figure 4.3), which may allow the peak-hour attacker to steal power more efficiently.
- **Home B** that the existence of a cloth dryer and wall oven may influence the accuracy of the attacker detection.
- **Home C** that the existence of a dishwasher, heat pump, and small appliances may influence the accuracy of attacker detection.

Table 5.1: Home configurations based on appliance data points

Home	**Aggregated Appliances**	**Excluded Appliances**
A	B1E, B2E, BME, CWE, DNE, HTE, EBE, EQE, FRE, OFE, OUE, TVE, UTE, CDE, HPE, DWE, FGE, WOE	None (full set)
B	B1E, B2E, BME, CWE, DNE, HTE, EBE, EQE, FRE, OFE, OUE, TVE, UTE, HPE, DWE, FGE	CDE, WOE
C	B1E, B2E, BME, CWE, DNE, HTE, EBE, EQE, FRE, OFE, OUE, UTE, CDE, FGE, WOE	DWE, HPE, TVE

Recognizing that electric power consumption data are inherently time-dependent, we adopted a data segmentation strategy. Specifically, we selected the initial 80% of the data, equivalent to 584 days, for our training dataset. During this phase,

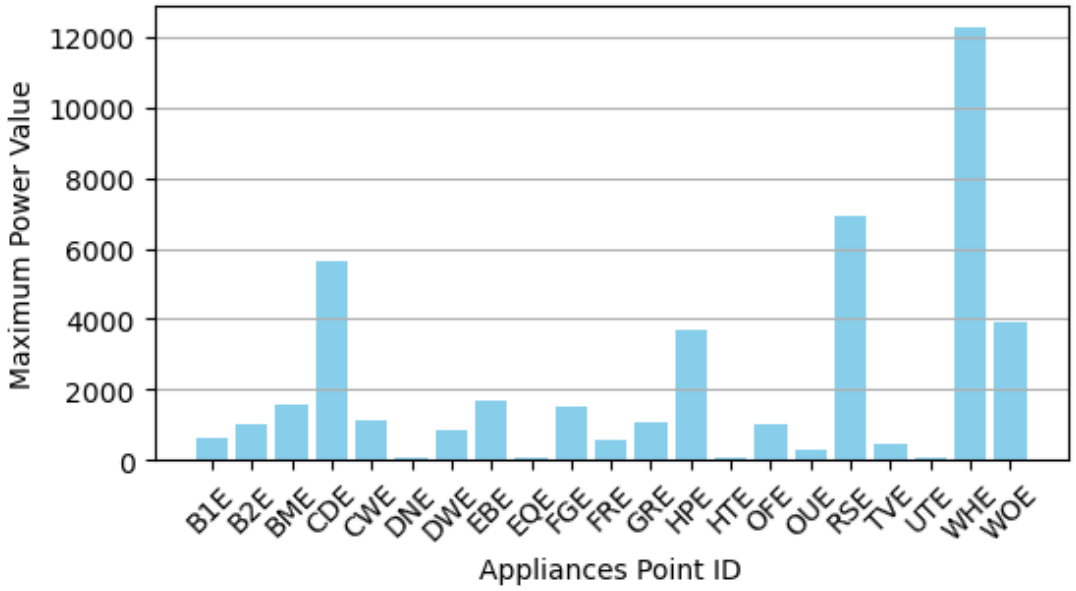


Figure 5.3: Maximum power values for each appliance

we applied our synthetic attack data methodology to enrich the dataset. The remaining 20% of the data, spanning 146 days, was reserved for our test dataset in our benchmark experiment as listed in Table 4.3. Importantly, the test dataset included simulated attacks generated as part of this study.

5.5.2 Overview of the Feature Selection

1. **Aggregated Power Consumption Patterns (APCP)** The feature for each minute m on day d is represented as the power consumption at that minute, denoted by $P_{d,m}$, where $d = 1, 2, \ldots, 730$ (for 730 days) and $m = 1, 2, \ldots, 1440$ (for 1440 minutes in a day). The aggregated power consumption for a day d is given by:

$$\text{APCP}_d = \sum_{m=1}^{1440} P_{d,m} \tag{5.1}$$

2. **Deviation from Typical Consumption (DTC)**

   This feature measures the deviation of actual consumption from the ex-

pected (baseline) consumption. Let $B_{d,m}$ represent the baseline power consumption for minute m on day d. The deviation for that minute is:

$$\text{DTC}_{d,m} = P_{d,m} - B_{d,m} \tag{5.2}$$

The total deviation for a day d can be aggregated as:

$$\text{DTC}_d = \sum_{m=1}^{1440} |\text{DTC}_{d,m}| \tag{5.3}$$

3. **Temporal Features (TF)**

   Temporal features could include binary indicators for peak and off-peak hours. We define $\text{PeakHour}(m)$ as a function that returns 1 if minute m is within peak hours, and 0 otherwise. The temporal feature for a day d is:

$$\text{TF}_d = \sum_{m=1}^{1440} \text{PeakHour}(m) \times P_{d,m} \tag{5.4}$$

4. **Combined Feature Vector**

   For a machine learning model, these features collectively form the input vector for each day d, represented as:

$$\text{FeatureVector}_d = [\text{APCP}_d, \text{DTC}_d, \text{TF}_d] \tag{5.5}$$

To determine the best feature for measuring appliance consumption patterns, we consider how each feature relates to energy consumption and how well it might differentiate between different consumption patterns [**?**]. Let us now review each feature:-

1. ***V*** **(voltage)**: While voltage levels can affect power consumption, in most residential and commercial settings, the voltage is relatively constant. It is not a direct measure of consumption but can be relevant in some analyses.

2. ***I*** **(Current)**: The current is directly related to the power consumption ($P = VI$, where V is the voltage and I is the current). Fluctuations in current can indicate changes in appliance consumption patterns.

3. ***f* (Frequency)**: The frequency is stable in most power systems. Variations are usually an indication of grid instability rather than appliance consumption patterns.

4. ***DPF* (Displacement Power Factor)**: This measures the efficiency of power usage but does not directly indicate consumption levels. This is more about the quality of consumption than the quantity.

5. ***APF* (Apparent Power Factor)**: Similar to *DPF*, it indicates the efficiency of power usage and is more about power quality.

6. ***P* (Power)**: Power is a direct measure of energy consumption at any given moment. This is one of the most direct measures of appliance consumption.

7. ***Pt* (Total Power)**: If this is cumulative power over time, it is an excellent measure of total consumption but less useful for instantaneous consumption patterns.

8. ***Q* (Reactive Power)**: This is related to the energy stored in the load and returned to the source and is more about the type of load than the quantity of consumption.

9. ***Qt* (Total Reactive Power)**: Similar to Q, but cumulative. It is more relevant to assessing load type over time than consumption patterns.

10. ***S* (Apparent Power)**: This is a combination of reactive power and real power and provides a total power figure but doesn't directly measure consumption efficiency.

11. ***St* (Total Apparent Power)**: Cumulative apparent power over time. Like S, it encompasses active and reactive power but does not directly indicate consumption patterns.

Based on this analysis, when we apply the correlation coefficient for feature selection, for most relevant features, P and Q are relative to S. Figure 5.4 shows that feature P has a higher correlation to feature S (the orange bar) and a slightly lower, yet still high correlation with feature 2 Q (the blue bar). Also Figure 5.5

shows that **(*P*) (Power)** is the best feature for measuring appliance consumption patterns using a mutual information algorithm [**?**] because it directly reflects the amount of electric power being used at any given moment. In the case of power consumption patterns, the machine automatically learns the key features from the raw sequence of power consumption data without providing any statistically processed data as explicit features.

The power_base dataset has the following features: • Time Components: • Day: The day number relative to the baseline (day_offset). • Minute: The minute of the day. • Aggregated Power Readings: • We extracted the active power P for its processing (as seen in power=float(row[6])), the aggregated data in power_base consists of the sum of active power readings from selected meters (base_names) for each minute of each day. Therefore, each entry in power_base represents the total active power consumption (in watts) from a subset of meters, Figure 5.3, for each minute of each day over 730 days. This aggregated dataset forms a baseline for normal power consumption patterns against which deviations (such as potential electricity theft) can be compared.

This dataset is pivotal for our analysis and predictive modeling. The structure of `power_base` is organized as a two-dimensional array, where each entry in `power_base[day][min]` denotes the aggregated power consumption for a specific category, as listed in Table 5.1 for a given day and minute. The data spanned a temporal resolution of one minute, totaling to 1440 min (24 h) per day. This granularity allows for a detailed analysis and forecasting of power consumption patterns.

5.5.3 Dataset preprocessing for binary classification

The electricity theft attack detection dataset (ETA-DD) consists of:

1. two training sub-datasets and
2. two testing sub-datasets for homes A, B, and C

This ETA-DD is assumed for binary classification problems (*Benign* or *Attack*). This is because it is intended for applying unsupervised learning for attack detection and evaluating detection accuracies with real building data. The real building

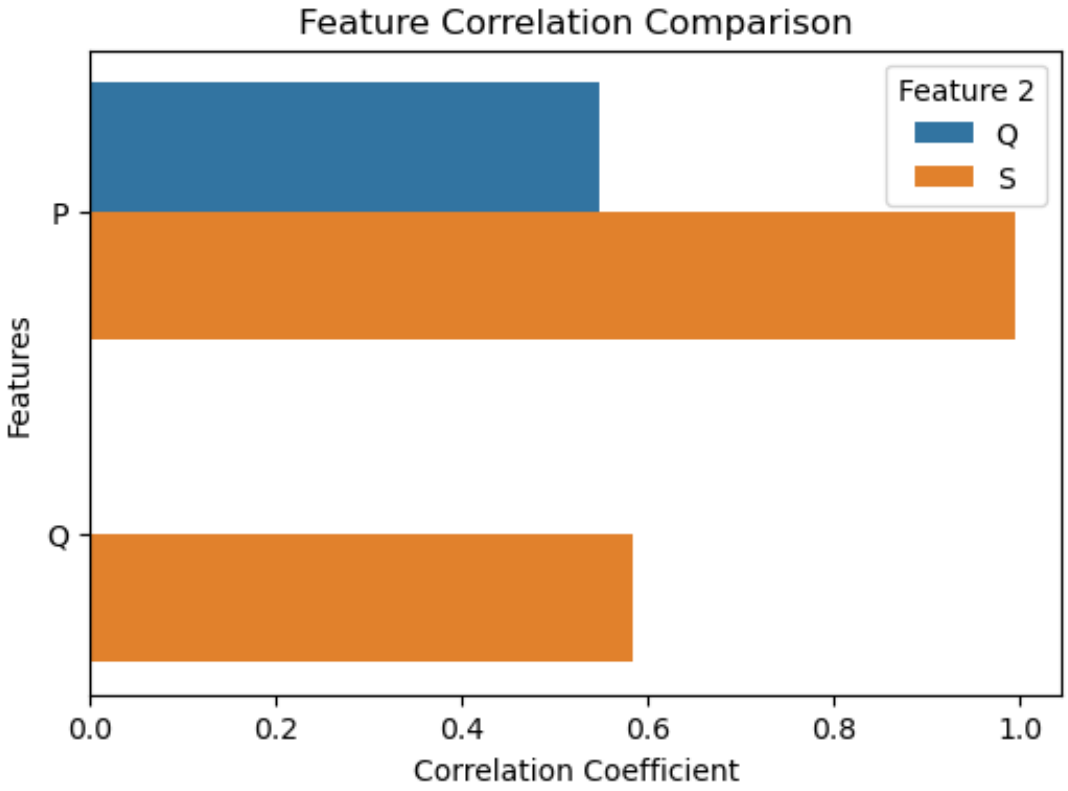


Figure 5.4: Features selection with correlation comparison

power consumption cannot be easily labeled for the predefined attack cases (as mentioned in subsection A above). These are the reasons why this study focuses on attack detection, rather than classification. Even though it is not intended for attack classification, the evaluation scope has drastically widened. This paper expands the limitations of the model to detect attacks that are not classified hence we preprocessed our dataset to accommodate the 5 simulated attack scenarios to include unknown or new attacks, n, as *1* and the benign or non-attack data as *0* as shown in Figure 5.6.

If we denote the original class labels by C where $C \in \{\text{Benign}, \text{Baseload}, \text{Weakload}, \text{Peakhour}, \text{Midnight}, \text{Evil-twin}, \ldots, n\}$, with their attack parameters as shown in

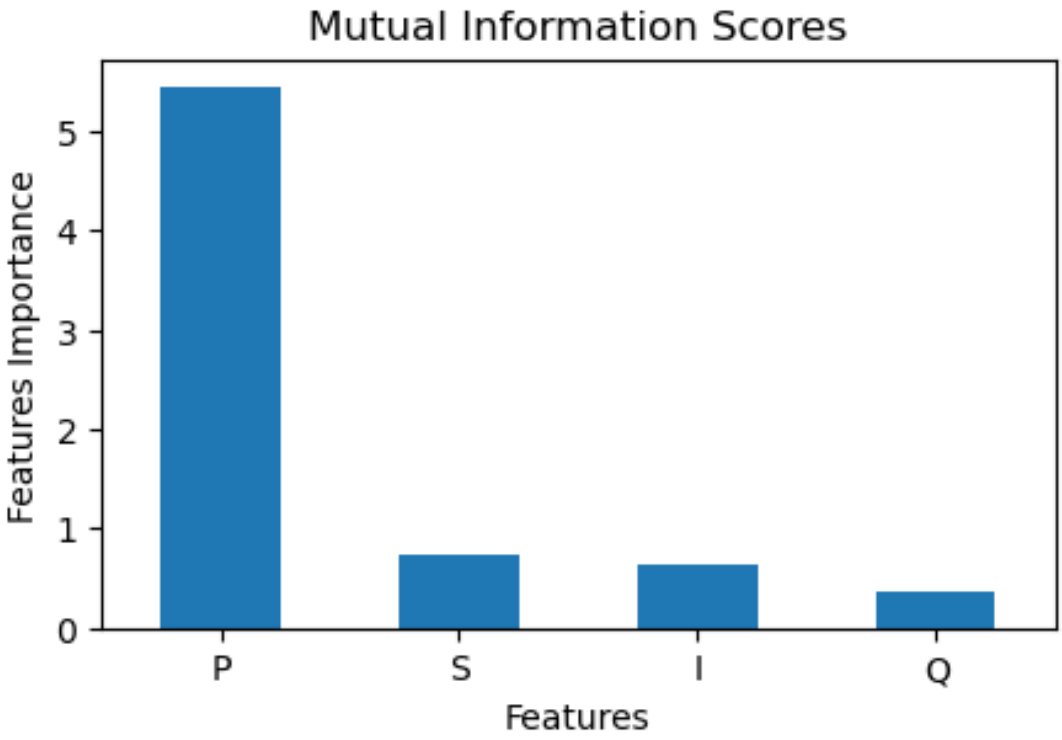


Figure 5.5: Features selection with mutual information Scores

Table 4.1, the binary classification function $f(C)$ can be defined as

$$f(C) = \begin{cases} 0 & \text{if } C = \text{Benign} \\ 1 & \text{otherwise} \end{cases} \tag{5.6}$$

Here, label *0* corresponds to the normal (benign) class, and label *1* corresponds to any kind of attack scenario. This binary labeling strategy is a common approach in anomaly or intrusion detection systems [**?**] where the focus is on differentiating between normal and abnormal behaviors, regardless of the specific type of abnormal activity.

Our real test attack data which have the same features (aggregated power_base consumption patterns), further solidify the evaluation of our model's performance.

We extended our research by transitioning from multiclass[2] to binary classification, incorporating data from three distinct homes (Home A, Home B, and Home C as depicted in Table 5.1.

5.6 Data preparation

5.6.1 Data anonymization

Data from various appliances are combined to create a comprehensive view of a home's power usage, with measures to protect user privacy.

The main feature of our dataset is the aggregation of the power consumption of each appliance in each home. The values are only numerical readings without any direct personal identifiers.

The aggregated baseline power consumption for each home (Table 5.1), $P_{\text{total base}}$ for n appliances is given by the sum of individual baseline power consumptions $P_{\text{base},i}$:

$$P_{\text{total base}} = \sum_{i=1}^{n} P_{\text{base},i} \tag{5.7}$$

where $P_{\text{base},i}$ is the baseline power consumption of the i-th appliance.

The total energy consumption E_{total}, considering the duration of usage D_i for each appliance, is calculated as:

$$E_{\text{total}} = \sum_{i=1}^{n} (P_{\text{base},i} \times D_i) \tag{5.8}$$

where D_i is the duration of usage for the i-th appliance in hours, and $P_{\text{base},i}$ is as defined earlier.

5.6.2 Normalization

In our ETD for smart homes, we deployed both StandardScaler and MinMax to normalize the feature vectors before synthetic binary discriminators (SYNBDM) and legacy unsupervised models (LUM) experiments respectively.

- For SYN-supervised models, we used **StandardScaler** to normalize the feature vectors by removing the mean and scaling to unit variance which helped improve the performance and stability of our models.

  The scaler is first fitted on the training data:

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} X_{train_{ij}} \tag{5.9}$$

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{train_{ij}} - \mu_j)^2} \tag{5.10}$$

where μ_j and σ_j are the mean and standard deviation for each feature j, and n is the number of training samples.

The training data are then transformed using these parameters:

$$X_{train\_scaled_{ij}} = \frac{(X_{train_{ij}} - \mu_j)}{\sigma_j} \tag{5.11}$$

The same transformation is applied to the test data:

$$X_{test\_scaled_{ij}} = \frac{(X_{test_{ij}} - \mu_j)}{\sigma_j} \tag{5.12}$$

This ensures that both training and test data are on the same scale.

- **MinMaxScaler** was deployed to ensure that the input features contribute equally to the model training, enhancing the learning process of anomaly detection.

  The MinMax Scaler linearly transforms each feature to a common scale, typically between 0 and 1. The transformation is defined as:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5.13}$$

  where $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the feature in the training dataset, respectively, and X represents the original feature value.

5.6.3 Framework for binary class from multi-class attack scenarios

From the previous chapter, in which we categorized smart home attacks into five categories, this study expands the limitations of the models to detect attacks that are not classified; hence, we preprocessed our dataset to accommodate the five attack scenarios generated from the knowledge-based attack consumption in Figure 4.2. For example, $\mathbf{x}_A^{(1)}(\mathbf{x}, \theta^{(1)})$, $\mathbf{x}_A^{(2)}(\mathbf{x}, \theta^{(2)})$, corresponds to baseload attack with attack parameter $\theta^{(1)}$ of 100W increase in power consumption during the

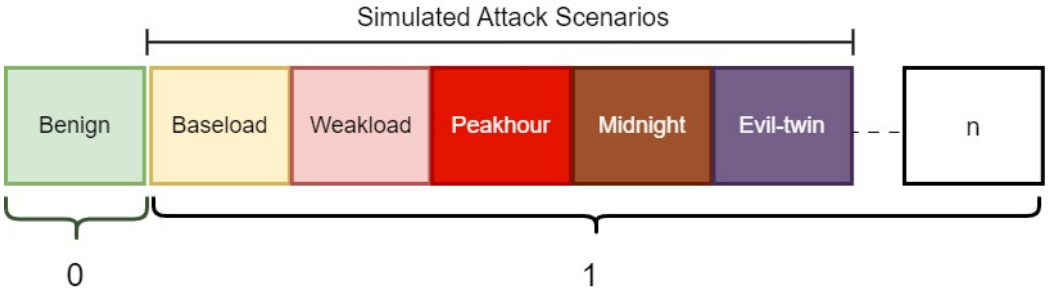


Figure 5.6: Multi-class attack scenarios with unspecified attacks preprocessed to binary attack class for anomaly detection.

attack period and weakload attack with attack parameter $\theta^{(2)}$ of 500W stolen power weakly only when the power consumption of the home is high, etc., including $\mathbf{x}_A^{(n)}(\mathbf{x}, \theta^{(n)})$, corresponding to unclassified attack with attack parameter $\theta^{(n)}$ of anomaly pattern classified as *'1'* as shown in Figure 5.6 and the benign or non-attack data as *'0'* for our proposed ETD framework to detect known attacks and classify unknown as anomalies.

Our methodology, as illustrated in Figure 5.8, focuses on the strategic manipulation of time offsets to generate varied yet realistic benign consumption patterns. This approach is grounded in the core principle of our data augmentation framework, which involves the application of circular shifting techniques to the original dataset.

5.6.4 Data Augmentation

5.6.5 Circular Shifting

Circular shifting is a data augmentation technique that's particularly useful in ML for image and time-series data[18]. It involves rotating or shifting the data points in a dataset circularly, meaning the data wraps around. This method can be visualized as moving the last data point to the first position, and shifting all other data points one position forward.

For time-series data, such as electricity consumption patterns, circular shifting involves shifting time points. For instance, earlier data points might move to the end of the series and vice versa. This is useful for creating variations in datasets

where the sequence is important but the exact starting point is arbitrary.

Overall, circular shifting as a data augmentation method enhances the diversity of training data, thereby improving the model's ability to generalize and reducing the risk of overfitting to a limited set of data patterns.

This dataset comprises pairs of feature vectors (X) and corresponding labels (Y), each vector encapsulating 1380 data points, representing minute-by-minute power consumption over 23 hours.

The augmentation process introduces random time offsets, ranging from 0 to 59 minutes, to the original feature vectors. This temporal shifting, executed using circular shifting methods as discussed by Chen et al. [15], effectively simulates variations in daily power usage while maintaining the integrity of the original data patterns. Importantly, the label for each augmented sample remains constant, ensuring that the classification of 'benign' or 'attack' is consistently applied across all variations.

5.6.6 Feature Vector (X) and Label (Y)

Each file starts from a header row, and then consumption records follow. Each consumption record is organized as Figure 5.7.

Let $\mathbf{X}$ be the original feature vector of length $n = 1380$, representing power consumption readings for every minute of 23 hours:

$$\mathbf{X} = [x_1, x_2, \ldots, x_n] \tag{5.14}$$

Let k be the time offset for circular shifting, where $k \in \{0, 1, \ldots, 59\}$.

The augmented feature vector $\mathbf{X}'$ after applying a circular shift of k positions is defined as:

$$\mathbf{X}' = [x_{(i-k) \mod n}, x_{(i-k+1) \mod n}, \ldots, x_{(i-1) \mod n}, x_i, \ldots, x_{(i-k-1) \mod n}] \tag{5.15}$$

where $i = 1, 2, \ldots, n$ and x_i is the power consumption at the i-th minute. The modulo operation $\mod$ ensures that the index wraps around the vector when the shift exceeds the vector's start.

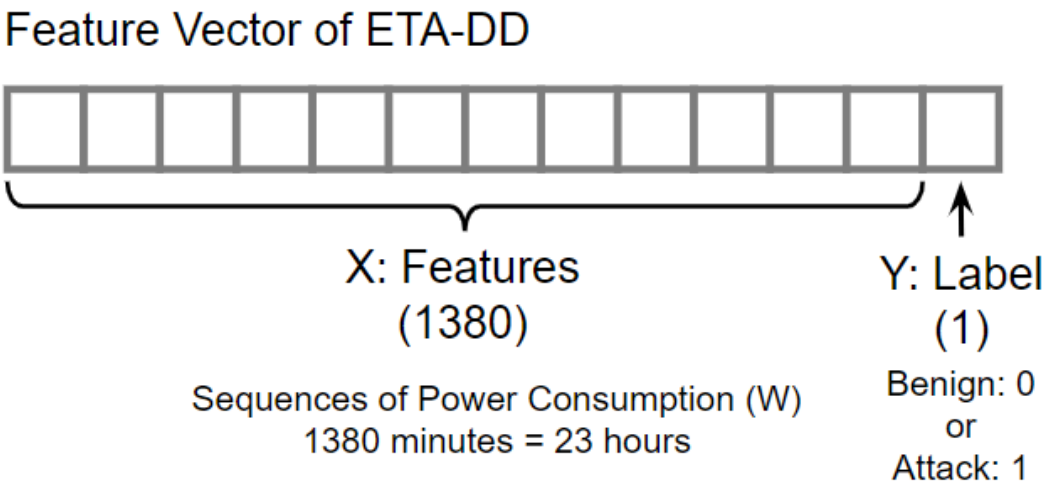


Figure 5.7: Features Vector and Labels

The label Y remains unchanged for all augmented records. If the original record is labeled as benign ($Y = 0$) or attack ($Y = 1$), then all augmented versions of the record retain this label:

$$Y' = Y \tag{5.16}$$

For each record in the dataset, this augmentation process is repeated 60 times, corresponding to each possible offset k, effectively increasing the number of benign records.

The length of the augmented feature vector $\mathbf{X}'$ is 1380 instead of the full 1440 which represents the total number of minutes in 24 hours. The reasoning for this is twofold:

1. A circular shift by an offset k ranging from 0 to 59 minutes implies that we need to have a buffer at the end of the vector to accommodate the maximum possible shift without wrapping into the data of the next day.

2. By limiting the feature vector to 1380 minutes, we ensure that for the largest shift of 59 minutes, the augmented data still represents a continuous 23-hour

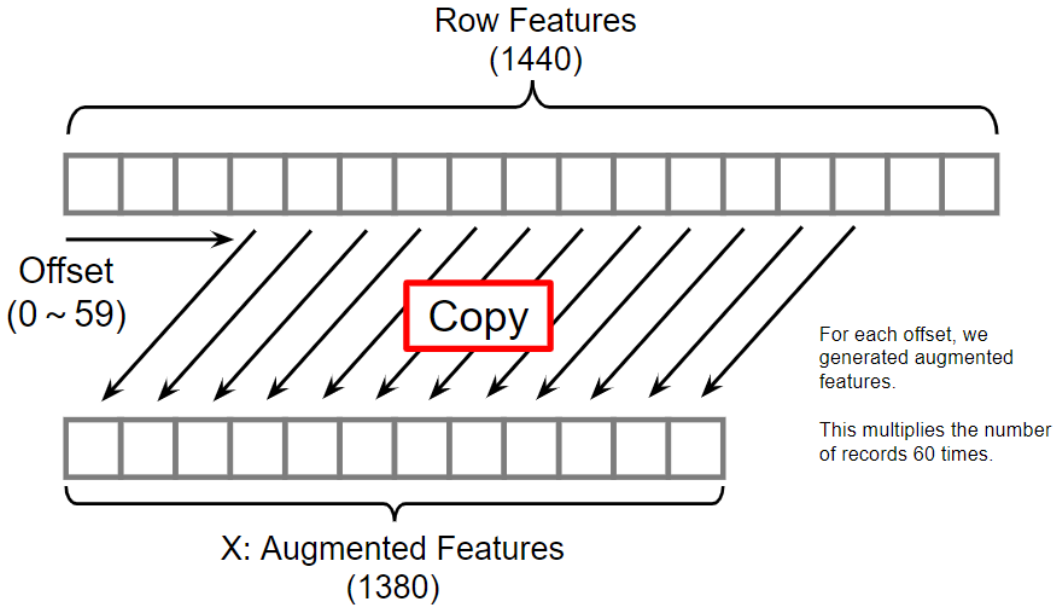


Figure 5.8: Framework of Data Augmentation

window from the same day. This is crucial for maintaining the integrity of daily patterns in power consumption without mixing data from two different days.

The label Y, which represents whether the original 1440-minute vector corresponds to a benign or attack pattern, remains associated with the corresponding 1380-minute augmented vector $\mathbf{X}'$. This ensures that the model learns to detect anomalies based on the most representative and complete daily consumption patterns possible within the constraints of the data augmentation process.

To illustrate the effectiveness of this augmentation technique, Figures 5.9 and 5.10 provide visualizations of the augmented power consumption patterns for a single home (Home A) and multiple homes (Homes A, B, and C), respectively. This strategy not only enriches the dataset but also reflects the everyday fluctuations in appliance operation times in residential settings, thereby creating a more comprehensive and realistic training environment for machine learning models employed in anomaly detection for ETD.

Through this innovative data augmentation strategy, our research aims to sig-

nificantly enhance the performance of binary classification models in detecting electricity theft. This methodology promises to advance the accuracy and reliability of ETD systems in smart home settings, contributing to the evolving landscape of smart grid security and management.

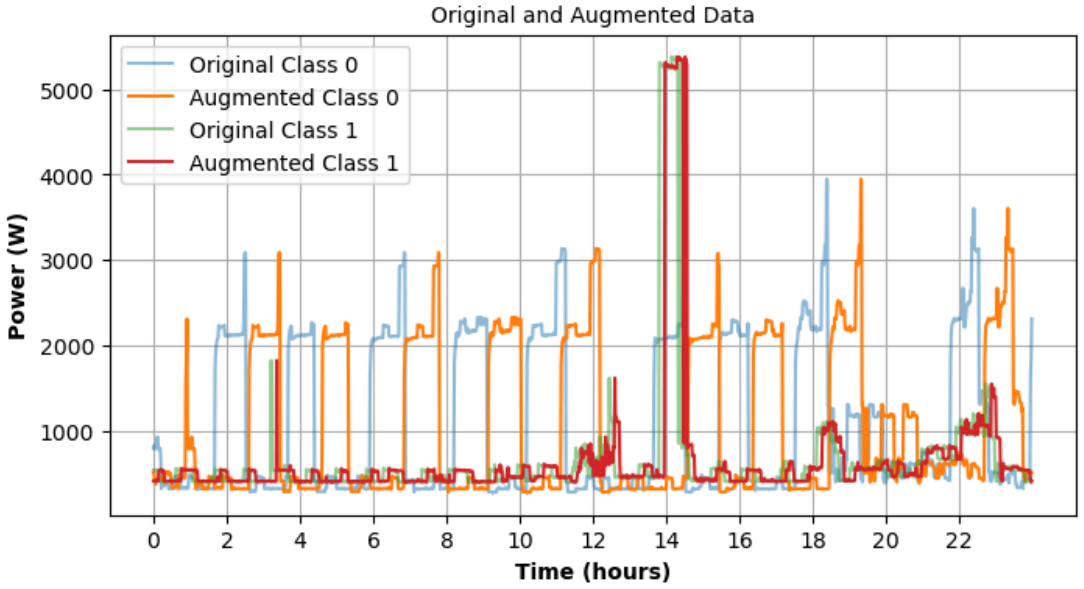


Figure 5.9: Electricity power consumption pattern of Home A for original class and augmented data class for a day

In our daily life, the operation of home appliances may be shifted for about one hour. Based on this idea, we have shifted the original data over the time axis up to 60 minutes and extracted them also as a benign record.

5.6.7 UTokyo Data - Real Attack Data

To test the performance of ETD in both synthetic and unsupervised approaches, we consider including the consumption pattern of real rooms of the University of Tokyo as attack data for the model tests. We measured the power consumption at the power distribution boards of the I-REF building (6th-floor building) from 2012 with a sampling frequency of 1 min. We selected the daily consumption patterns in which the attack impact (AI) corresponded to the attack on the simulation test dataset.

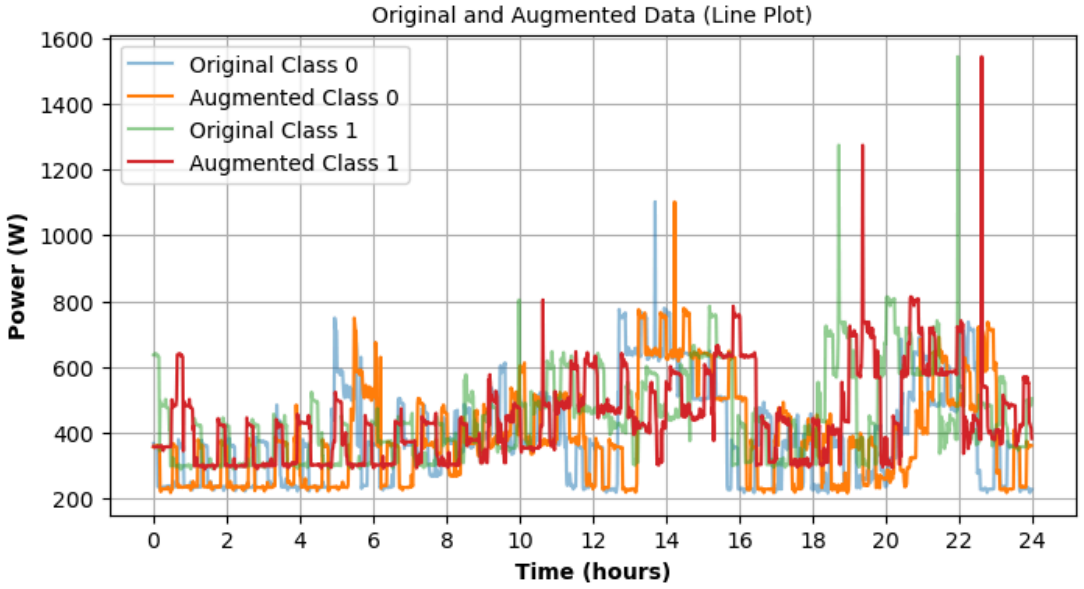


Figure 5.10: Electricity power consumption pattern of all homes for original class and augmented data class for a day

Table 5.2: Distribution of simulated and real binary dataset

Dataset	Home	Benign samples	Attack samples
SYN Train	A	35040	33818
	B	35040	33242
	C	35040	31533
UN Train	A	35040	0
	B	35040	0
	C	35040	0
Sim Test	A	146	144
	B	146	142
	C	146	133
Real Test	A	8760	8760
	B	8760	8760
	C	8760	8760

Note: SYN = Synthetic, UN = Unsupervised

5.7 Synthetic Binary Discriminator Model (SYNBDM)

Figure 5.11 shows the evaluation flow of our SYNBDM. We examine the performance of supervised XGB, RF, and MLP classifiers in different homes. Suppose X represents the input features, from the synthetic train dataset, and Y is the output prediction for binary classification,

$$f(X) = Y \tag{5.17}$$

where f represents the learning function of the Binary Discriminator.

During the testing phase, the trained model, f is evaluated using two different datasets:

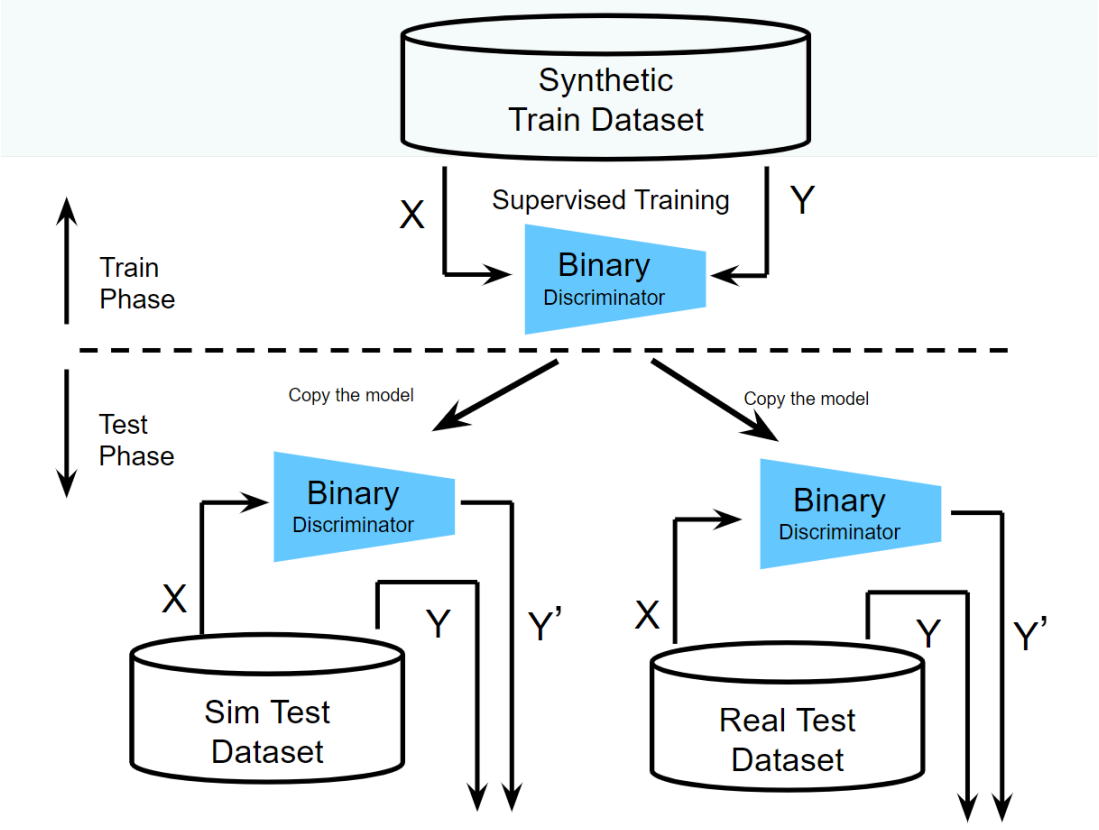


Figure 5.11: Flow of evaluation with Supervised Binary Descriminator

For the Simulated Test Dataset:

$$f(X_{\text{sim}}) = Y'_{\text{sim}} \tag{5.18}$$

where X_{sim} are the input features and Y'_{sim} is the output predicted by the model.

For the Real Test Dataset:

$$f(X_{\text{real}}) = Y'_{\text{real}} \tag{5.19}$$

where X_{real} is the input feature and Y'_{real} is the output predicted by the model.

The performance of the model was assessed based on the accuracy of the predictions Y'_{sim} and Y'_{real} in comparison to the true labels.

5.7.1 Proposed models Characteristics overview

5.7.2 XGBoost (XGB)

XGBoost is a prominent ensemble learning method that, primarily utilizes decision tree structures. It employs gradient boosting, a technique that iteratively refines models by integrating multiple weak learners to formulate a robust predictive framework.

- Regularization: A distinctive feature of XGBoost is the incorporation of a regularization term into its objective function [16]. This term is instrumental in mitigating the risk of overfitting, thereby enhancing the model generalization.

- Objective Function:

$$\text{Objective}_{XGB}(\theta) = \sum_{i=1}^{n} \text{loss}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5.20}$$

  where θ denotes the model parameters, n is the number of observations, $\text{loss}(y_i, \hat{y}_i)$ is the loss function, and $\sum_{k=1}^{K} \Omega(f_k)$ is the regularization component.

5.7.3 Random Forest (RF)

Random Forest is an ensemble learning technique based on decision tree algorithms. It constructs a multitude of decision trees during training, and their collective output obtained through averaging or majority voting, constitutes the final model prediction.

- Overfitting Reduction: The algorithm introduces randomness in tree generation, effectively reducing overfitting compared to individual decision trees [31].
- Objective Function:

$$\text{Objective}_{RF}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \text{loss}(y_i, \hat{y}_i) \tag{5.21}$$

  where θ represents the model parameters, n is the number of data points, and $\text{loss}(y_i, \hat{y}_i)$ are the loss functions.

5.7.4 Multi-Layer Perceptron (MLP)

Multilayer Perceptron (MLP) is a class of feedforward artificial neural networks, characterized by multiple layers of nodes. Each layer is interconnected through weights and biases, enabling MLPs to capture complex, non-linear relationships in the data [6].

- Backpropagation: MLPs rely on backpropagation for training, which is an algorithm that iteratively adjusts weights and biases to minimize the error between the actual and predicted outcomes.
- Objective Function:

$$\text{Objective}_{MLP}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \text{loss}(y_i, \hat{y}_i) + \alpha \sum_{i=1}^{L-1} ||W_i||^2 \tag{5.22}$$

  where θ denotes the model parameters, n is the number of observations, $\text{loss}(y_i, \hat{y}_i)$ the loss function, α is the regularization parameter, and L the number of network layers.

5.7.5 Maximum Likelihood Estimation (MLE)

For each model, parameter estimation can often be described using Maximum Likelihood Estimation (MLE), which for classification problems, involves maximizing the log-likelihood function:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log P(y_i|X_i;\theta) \tag{5.23}$$

where θ represents the parameters, $P(y_i|X_i;\theta)$ is the probability of the target y_i given the input X_i, and $\hat{\theta}$ is the set of parameters that maximizes the likelihood, that is, involves finding the values of hyperparameters, for example, learning rate (eta), max-depth or the number of trees (n_estimators) that maximize the likelihood of observing the actual data.

We performed a grid search with cross-validation techniques, where the objective function, which is a combination of the loss function and regularization was minimized during training, and selection was made based on the hyperparameter tuning of each model.

We normalized our dataset with a StandardScaler to improve performance and, trained our SYNBDM classifier with benign and synthetic attack samples, as shown in Table 5.2 for Home A. We evaluated the degree to which the classifier is capable of detecting attack instances with real test data. We repeated the same experiment with Homes B and C, on both the simulated and real attack datasets. In the experiment, we used Jupyter Notebook, an open-source web application, written in Python; hence, it was easy to use with TensorFlow. Table 5.3 lists the parameter values used in the binary classification experiment. We selected the best hyperparameter values by experimenting with grid search. We performed validation through the *fit()* function using validation data, that is, simulated and real data. After training and testing each home, we calculated the accuracy and loss based on the number of correctly classified instances.

Our framework integrates model selection and feature selection to optimize the machine learning pipeline for ETD. It uses synthetic training data for the initial model training and hyperparameter tuning. The model is then validated on synthetic and real test datasets to ensure that it generalizes well to unseen, real-world data. This is achieved by performing a test each time the new appli-

Table 5.3: Parameters for the binary supervised discriminator

Model	Parameters	Values
XGB	learning rate	0.1
	n_estimators	100
	reg_alpha	0.01
	reg_lambda	1.0
RF	n_estimators	50
	max_depth	10
	min_samples_leaf	1
	min_samples_split	2
MLP	hidden_layer_sizes	50, 100
	activation	tanh
	batch_size	50
	learning_rate_init	0.001

ances are connected. Each new appliance was preprocessed and converted into a proper format consistent with the training set. The proposed XGB is applied to a new sample format to determine whether it belongs to the benign or attack class. The framework incorporates MLE to optimize the objective function, ensuring that the models are well-calibrated, and providing probabilistic outputs that can be interpreted as risk scores for ETD. Table 5.4 presents the results of our experiments for all the homes.

5.8 Legacy unsupervised model (LUM)

The process depicted in Figure 5.12 involves an unsupervised learning model, specifically an autoencoder, which is trained to detect anomalies based on the reconstruction error.

5.8.1 ETA Based Autoencoder Detection Algorithm

In our model, we deployed a reconstruction error threshold to classify data points as normal (benign) or anomalous (attack) and also metrics like area under the re-

Table 5.4: Flow of the evaluation with Synthetic Binary Discriminator

Home	**DataSet**	**Model**	**Acc**	**Recall**	**F1-score**	**Prec**	**AUC**
A	Sim	XGB	0.9521	0.9169	0.9492	0.9839	0.9876
		RF	0.9197	0.8452	0.9113	0.9886	0.9647
		MLP	0.9733	0.9633	0.9723	0.9816	0.9656
	Real	XGB	0.9555	0.9209	0.9530	0.9875	0.9891
		RF	0.9223	0.8505	0.9146	0.9894	0.9702
		MLP	0.9243	0.9048	0.9212	0.9383	0.9668
B	Sim	XGB	0.9615	0.9281	0.9590	0.9921	0.9878
		RF	0.9359	0.8718	0.9296	0.9957	0.9756
		MLP	0.9514	0.9237	0.9486	0.8718	0.9724
	Real	XGB	0.9624	0.9293	0.9599	0.9927	0.9878
		RF	0.9327	0.8662	0.9259	0.9942	0.9764
		MLP	0.9441	0.9250	0.9413	0.9583	0.9795
C	Sim	XGB	0.9526	0.9116	0.9480	0.9837	0.9853
		RF	0.9269	0.8539	0.9168	0.9906	0.9625
		MLP	0.9168	0.8791	0.9168	0.9406	0.9537
	Real	XGB	0.9570	0.9209	0.9540	0.9894	0.9853
		RF	0.9291	0.8612	0.9216	0.9911	0.9654
		MLP	0.9293	0.9026	0.9251	0.9487	0.9649

ceiver operating characteristic curve (AUC-ROC), F1-score, precision, and recall, which are often more informative in such cases. High accuracy can be achieved by simply classifying everything as benign, which does not help detect attacks.

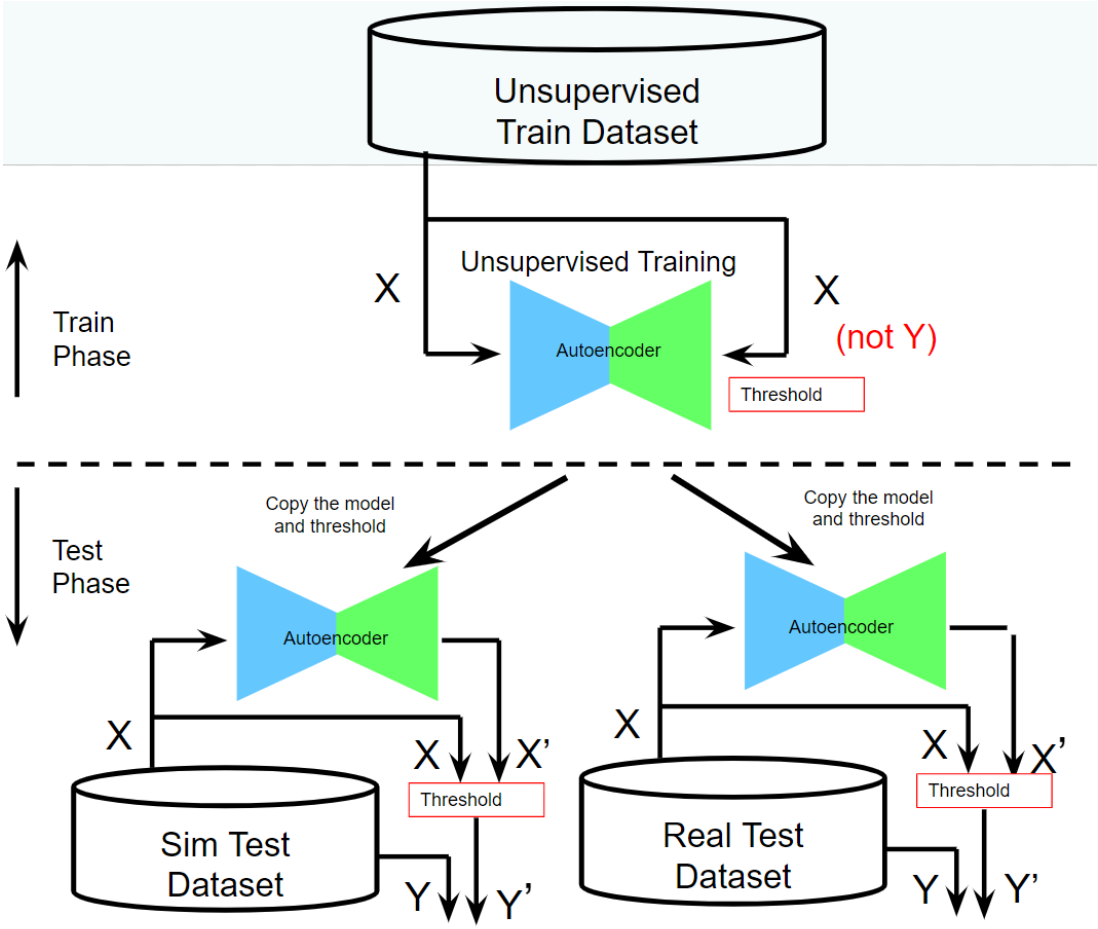


Figure 5.12: Flow of evaluation with Unsupervised Autoencoder

5.8.2 Training Phase

During the training phase, the autoencoder learns parameter θ by minimizing the reconstruction error of the input historical electricity consumption data (benign only) X as follows:

$$\min_{\theta} \ \|X - \hat{X}(\theta)\|^2 \tag{5.24}$$

where $\hat{X}(\theta)$ denotes the reconstructed output of the autoencoder, and θ denotes its parameters.

5.8.3 Threshold Determination

A threshold τ was established based on the reconstruction error distribution during the training phase as shown in Table 5.6. These thresholds were used to classify the data points as either normal or anomalous.

5.8.4 Testing Phase

In the testing phase, the autoencoder reconstructs new data from both Simulated and Real Test Datasets:-

$$\hat{X}' = \text{Autoencoder}(X') \tag{5.25}$$

Subsequently, the reconstruction error E for each data point is computed:

$$E = \|X' - \hat{X}'\| \tag{5.26}$$

An anomaly is flagged if the reconstruction error E exceeds the predetermined threshold τ:

$$Y' = \begin{cases} \text{Anomaly} & \text{if } E > \tau \\ \text{Normal} & \text{if } E \leq \tau \end{cases} \tag{5.27}$$

The performance of the autoencoder in anomaly detection is contingent on the accuracy of the threshold τ and its capability to accurately learn the representation of normal data during training.

The binary classification output Y' indicates whether a data point is normal or anomalous, based on the reconstruction error relative to the threshold.

Our experimental parameter settings in Table 5.5 reference attack detection based on unsupervised binary classification models [35]: Multi-Layer Perceptron Autoencoder (MLP_AE) employed a fixed learning rate of 0.001. The autoencoder has a single hidden layer consisting of 32 neurons. A batch size of 32 was used

for the training. The model was trained for 100 epochs, and the Adam optimizer was utilized. A validation split of 10% was employed during the training.

1D Convolutional Autoencoder (1D-CONV_AE): A fixed learning rate of 0.001 was used. The latent space dimension was set to 64. A batch size of 32 was used during the training. The model was trained for 100 iterations. The Adam optimizer was utilized and a validation split of 10% was employed during training. We used 100 trees in the Isolation Forest (IF) algorithm. The random state was set to 42 to ensure reproducibility. The contamination parameter was set to 0.05, which represented the assumed proportion of outliers in the dataset.

Table 5.5: Parameters for legacy unsupervised models

Model	**Parameters**	**Values**
MLP_AE	learning rate	0.001
	hidden_layer_sizes	32
	batch_size	32
	epoch	100
	optimizer	adam
	Validation_split	0.1
1D-CONV_AE	learning rate	0.001
	latent_dim	64
	batch_size	32
	epoch	100
	optimizer	adam
	Validation_split	0.1
IF	n_estimators	100
	random state	42
	contamination	0.05

Similar to the supervised binary discriminator, we first normalized the data using MinMaxScaler and trained the autoencoder with historical electricity consumption data X.

Equations (30), (31), (32), and (33) indicate that TP, TN, FP, and FN are true positive, true negative, false positive, and false negatives respectively. A TP refers

to a sample that is malicious and is detected as malicious. TN indicates a benign sample that was detected as benign. FP indicates that the sample is benign but is detected as malicious. An FN represents a malicious sample detected as benign [80].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.28}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.29}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.30}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5.31}$$

For our electricity consumption data from various homes and datasets, normalization facilitates the scaling of input features. This scaling ensures that the features, such as aggregate power consumption, have uniform scales across different homes and datasets; therefore, the original feature values X are transformed into scaled values X_{scaled} within the *[0, 1]* range. This standardized scaling process is essential for the autoencoder to accurately learn and detect anomalies in electricity consumption patterns.

During training, the autoencoder leveraged these scaled features to reconstruct benign data, and a threshold was determined using the statistical method Median Absolute Deviation (MAD), (Equations 34 and 35), and reconstruction errors to identify anomalies. For example, Figure 5.13 shows the threshold determination from the training set only for home A while Table 5.6 depicts the threshold values used in training MLP-AE and 1D-CONV-AE for all homes.

The robust Z-score method uses the Median Absolute Deviation (MAD) [73] instead of the standard deviation, and is not significantly affected by outliers.

The mathematical representation of the Modified Z-score is:

$$\text{MAD} = \text{median}\left(|x_i - \tilde{x}|\right) \tag{5.32}$$

$$\text{Modified Z-score} = 0.6745 \times \frac{x_i - \tilde{x}}{\text{MAD}} \tag{5.33}$$

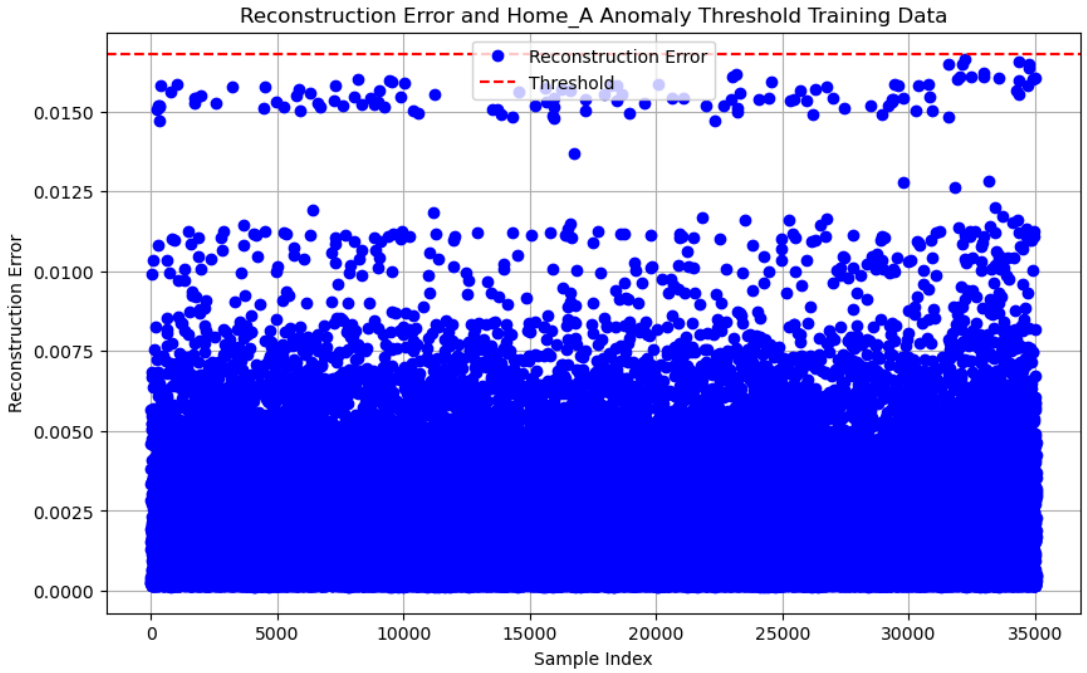


Figure 5.13: Threshold determination for Home A training dataset

0.6745 is the 0.75th quartile of the standard normal distribution, to which the MAD converges.

- $\tilde{x}$ which is just the median of the sample
- MAD, is calculated by taking the absolute difference between each point and the median, and then calculating the median of those differences.

This feature scaling contributes to the robustness and accuracy of anomaly detection in the context of protecting homeowners from energy theft by identifying unusual electricity consumption patterns, as indicated in the experimental data from different homes in Table 5.2.

Table 5.7 shows the results for all the algorithms in different homes. We deployed the receiver operating characteristic curve (AUC-ROC), F1-score, precision, and recall metrics, which are often more informative in such cases. High accuracy can be achieved by simply classifying everything as benign, which does not help in detecting attacks.

In Figure 5.15 we plotted the ROC curves for the MLP-AE, 1D-CONV-AE, and IF for both simulated and real attack scenarios in Home A. The MLP-AE detector outperformed both 1D-CONV-AE and IF with an AUC score of 0.76 for simulated and 0.59 for real attacks, respectively while 1D-CONV-AE had an ROC value of 0.67 and 0.61; IF has AUC values 0.64 and 0.54 respectively.

Table 5.6: Legacy unsupervised AE Threshold values for Anomaly detection

Home	Dataset	Model	Threshold
A	Sim & Real	UN-MLP-AE	0.017565
		UN-1D-CONV-AE	0.000020
B	Sim & Real	UN-MLP-AE	0.018085
		UN-1D-CONV-AE	0.000055
C	Sim & Real	UN-MLP-AE	0.019676
		UN-1D-CONV-AE	0.000055

In many real-world scenarios, the attack patterns vary and evolve constantly. Rare attack patterns can be vastly exceeded by benign data. Consequently, the autoencoder may not have sufficient examples of these rare attacks to learn effective representations, making it difficult to detect new attacks.

5.9 Experiments Results and Performance Evaluation

We present a comprehensive performance evaluation of various machine learning models for ETD, utilizing both real-world and synthetic attack datasets. We primarily focus on the AUC metric from Table 5.7 as the primary evaluation criterion and complement it with additional metrics from Table 5.4, including F1-score, accuracy, precision, and recall.

5.9.1 Model Performance by ROC and Confusion Matrices

Figures 5.14 and 5.15 further generate ROC curves for the remaining supervised benchmark detectors to facilitate a comparative analysis. The ROC curve provides a general view of the model's performance across all thresholds and, provides a sense of discrimination ability. In contrast, the confusion matrix provides de-

Table 5.7: Evaluation of the ETD model with AUC and accuracy Scores

Home	DataSet	Model	AUC	ACC
A	Sim	SYN-XGB	**98.76%**	95.21%
		SYN-RF	96.47%	91.97%
		SYN-MLP	96.56%	**97.33%**
		UN-MLP-AE	74.61%	58.28%
		UN-1D-CONV-AE	67.31%	54.83%
		UN-IF	63.55%	63.79%
	Real	SYN-XGB	**98.91%**	**95.55%**
		SYN-RF	97.02%	92.23%
		SYN-MLP	96.68%	92.43%
		UN-MLP-AE	58.96%	51.26%
		UN-1D-CONV-AE	61.12%	50.51%
		UN-IF	53.66%	53.66%
B	Sim	SYN-XGB	**98.78%**	**96.15%**
		SYN-RF	97.56%	93.59%
		SYN-MLP	97.24%	95.14%
		UN-MLP-AE	78.90%	64.58%
		UN-1D-CONV-AE	77.04%	65.63%
		UN-IF	61.64%	62.15%
	Real	SYN-XGB	**98.78%**	**96.24%**
		SYN-RF	97.64%	93.27%
		SYN-MLP	97.95%	94.41%
		UN-MLP-AE	60.61%	52.58%
		UN-1D-CONV-AE	71.77%	53.58%
		UN-IF	55.01%	55.01%
C	Sim	SYN-XGB	**98.53%**	**95.26%**
		SYN-RF	96.25%	92.69%
		SYN-MLP	95.37%	91.68%
		UN-MLP-AE	75.04%	55.20%
		UN-1D-CONV-AE	64.95%	60.22%
		UN-IF	67.66%	68.82%
	Real	SYN-XGB	**98.53%**	**95.70%**
		SYN-RF	96.54%	92.91%
		SYN-MLP	96.49%	92.93%
		UN-MLP-AE	65.64%	50.26%
		UN-1D-CONV-AE	54.95%	51.10%
		UN-IF	74.35%	74.35%

tailed information regarding the performance of our model at a specific threshold level.

In the ROC curve, the AUC for each model (XGBoost, Random Forest, MLP) provides a single measure of performance across all possible classification thresholds, summarizing the trade-off between TPR and FPR.

The confusion matrices in Figure 5.16 provide a more granular view. For in-

stance:

- The MLP for Home A simulated (Home A_sim) confusion matrix indicates that the model correctly identifies 95% of benign cases (TN), the exact value is 6655, and 89% of attack cases (TP), 6065, at a specific threshold.
- The Random Forest confusion matrix showed a high TN rate of 99%, but a lower TP rate of 84%.
- The XGBoost confusion matrix showed a similarly high TN rate of 99% and a better TP rate of 91%.

The ROC curve does not show the actual values of TP, FP, TN, and FN; rather, it shows the rate at which these values change with the different thresholds as shown in Figure 5.14 (a),(b), and (c), samples - taken from home A (sim and real), and home C(real attack) respectively. A high AUC reflects a model with a high TPR and low FPR across different thresholds, which generally corresponds to high values of TP and TN and low values of FP and FN in the confusion matrices at a particular operating threshold.

The same principles were applied to legacy unsupervised models (LUM). For instance, consider the confusion matrices for the simulated and real data from Home B using the 1D-CONV-AE model in Figure 5.15(b):

- The AUC of 0.78 and 0.72 for simulated and real data respectively on the ROC curve suggests that the model's ability to distinguish between the classes is reasonably good for simulated data and less so for real data.
- For the corresponding confusion matrices in Figure 5.16, we see high TN rates (0.99 for simulated, 1.00 for real) but varying TP rates (0.78 for simulated, 0.98 for real). This suggests that, while the model is quite good at identifying negative cases (benign), its performance on positive cases (attacks) is inconsistent between the simulated and real data.
- In the confusion matrices, we observed the specific number of instances that are correctly and incorrectly classified, which was reflected in the ROC curve by the closeness of the curve to the top left corner.

From the ROC curve for the isolated forest model in Figure 5.15(c):

- Home A Simulated has an AUC of 0.64, meaning that the model has a 64% chance of correctly distinguishing between a benign and an attack instance for the simulated environment of Home A.
- Home A Real had a lower AUC of 0.54, suggesting that the model was less effective in distinguishing between benign and attack instances in the real-world data of Home A.

Similarly, Home B Simulated and Home B Real have AUCs of 0.62 and 0.55, respectively, and Home C Simulated and Home C Real have AUCs of 0.68 and 0.74. The higher the AUC, the better the model is at distinguishing between positive (attacks) and negative (benign) classes. For example, the model performed best on real data for Home C, with an AUC of 0.74.

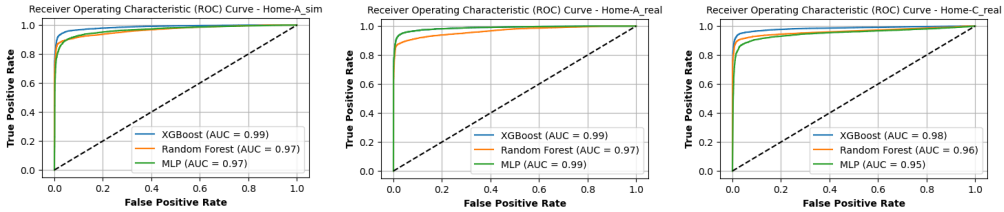
From the confusion matrix for Home C real data:

- True Positive (TP): 4737 - The model correctly identified 4737 attack instances.
- True Negatives (TN): 8289 - The model correctly identified 8289 instances as benign.
- False positive (FP): 471 - The model incorrectly identified 471 benign instances as attacks.
- False Negatives (FN): 4023 - The model failed to identify 4023 attacks, mistakenly classified as benign.

Relating the Confusion Matrix to the ROC Curve:

- The specific values in the confusion matrix correspond to a single point on the ROC curve for Home C's real data. The point is determined by the sensitivity (TPR) and FPR.
- These values would indicate a corresponding point on the ROC curve, but the exact point was not marked in the ROC curve. However, we know that point exists, and if the threshold is adjusted, this point moves along the curve, resulting in different values in the confusion matrix.

Generally, the ROC curve indicates how well the model can separate the two classes, and provides a holistic view of the model's performance across all thresholds. By contrast, the confusion matrix indicates exactly where the model makes mistakes at a specific threshold.

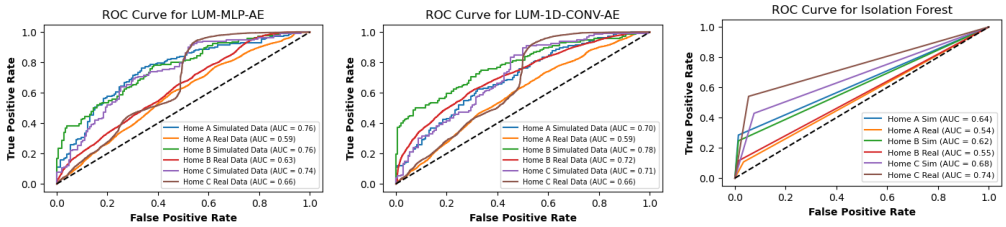


(a) ROC curves for three model-based simulated attacks at Home A

(b) ROC curves for the three models based on real attacks in Home A

(c) ROC curves for the three models based on a real attack at Home C

Figure 5.14: Sampled ROC curves for comparison performance evaluation of the proposed SYNBDM of some selected homes.



(a) MLP-AE: sim vs. real attack

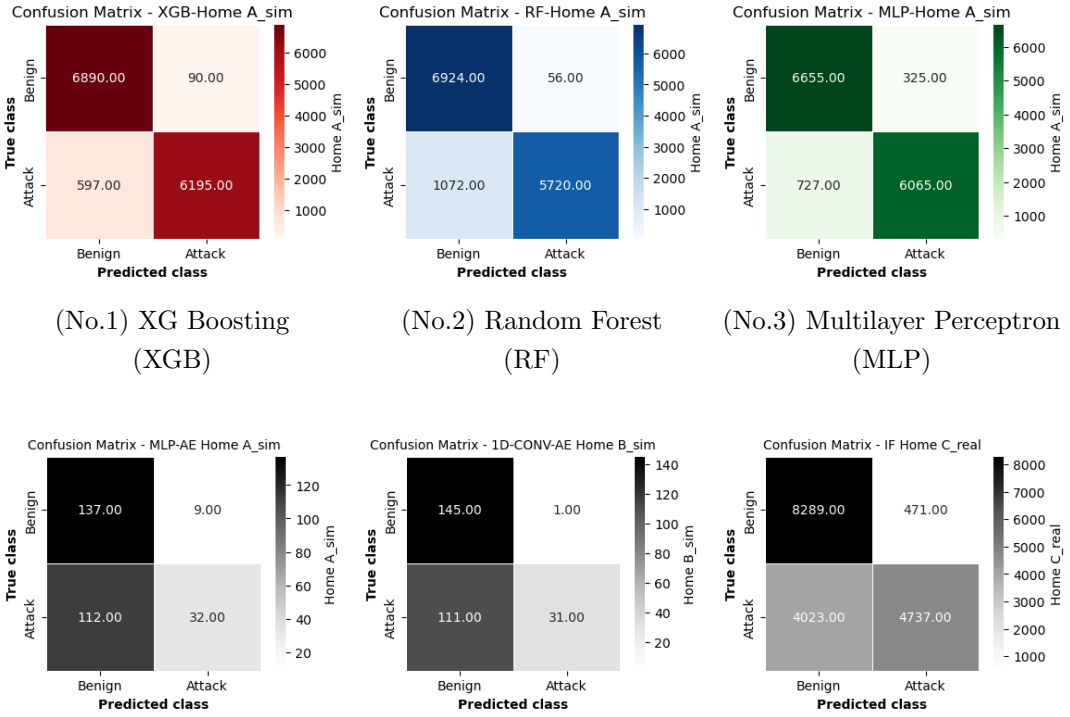
(b) 1D-CONV-AE: sim vs. real attack

(c) IF: sim vs. real attack

Figure 5.15: Performance comparison of ROC curves for LUM across homes.

5.9.2 Model Training and test error

To ensure balanced optimization between the training and test errors, we employed GridSearchCV for meticulous hyperparameter tuning. A five-fold cross-validation was incorporated to enhance the robustness of our model evaluations.



(No.1) XG Boosting (XGB)

(No.2) Random Forest (RF)

(No.3) Multilayer Perceptron (MLP)

(No.4) Multilayer Perceptron Autoencoder

(No.5) 1D-Convolutional Autoencoder

(No.6) Isolation Forest Outliers (IF)

Figure 5.16: Confusion matrices of the SYNBDM and LUM in order of overall performance.

For the XGBoost model, the regularization parameters L1 (`reg_alpha`) and L2 (`reg_lambda`) were utilized. Conversely, in the context of the Random Forest model, the parameters `min_samples_split` and `min_samples_leaf` serve a regulative function by constraining the complexity of the decision trees. The MLP model employed an L2 penalty term (`alpha`) and a maximum number of iterations (`max_iter`), combined with an early stopping criterion to prevent overfitting.

The subsequent results, as shown in Table 5.8, were obtained through the post-application of the aforementioned hyperparameter tuning and regularization strategies, as detailed for each model in Table 5.3.

The bar charts in Figure 5.17 show the training and test error rates for various models, based on simulated and real data aggregate performance outputs in Table 5.8. The error rates were calculated as one minus the AUC and ac-

Table 5.8: Models training and testing error report for all homes.

Home	DataSet	Model	AUC	ACC	Tr_Err	Ts_Err
A+B+C	Sim_av	SYN-XGB	**0.9869**	**0.9554**	**0.0446**	none
		SYN-RF	0.9676	0.9275	0.0725	”
		SYN-MLP	0.9632	0.9472	0.0528	”
		UN-MLP-AE	0.7618	0.5935	0.4065	”
		UN-1D-CONV-AE	0.6977	0.6023	0.3877	”
		UN-IF	0.6428	0.6492	0.3508	”
	Real_av	SYN-XGB	**0.9874**	**0.9583**	none	**0.0417**
		SYN-RF	0.9707	0.9280	”	0.0720
		SYN-MLP	0.9704	0.9326	”	0.0674
		UN-MLP-AE	0.6174	0.5137	”	0.4863
		UN-1D-CONV-AE	0.6261	0.5170	”	0.4830
		UN-IF	0.6428	0.6101	”	0.3899

Note: Tr_Err = Train Error, Ts_Err = Test Error

curacy (ACC) values for each model. From the charts, we can observe that for the detection of electricity theft in smart homes utilizing aggregated appliance consumption patterns, the comparative performance analysis of various models is pivotal. Our investigation encompassed the following: both supervised and unsupervised learning paradigms, with the supervised models demonstrating superior efficacy.

5.9.3 Model selection

From our performance analysis with other ETD models for smart homes through aggregated appliance consumption patterns, the XGBoost model (SYN-XGB) emerged as a standout performer, hence it was selected as the best model for our proposed ETD for the following reasons:

It achieved the highest Area Under the Curve (AUC) of 98.69% and accuracy (ACC) of 95.54% among the evaluated models. The XGBoost (SYN-XGB) model also exhibited the lowest error rates across simulated and real datasets, indicating its robustness and high accuracy in discerning normal consumption from theft-

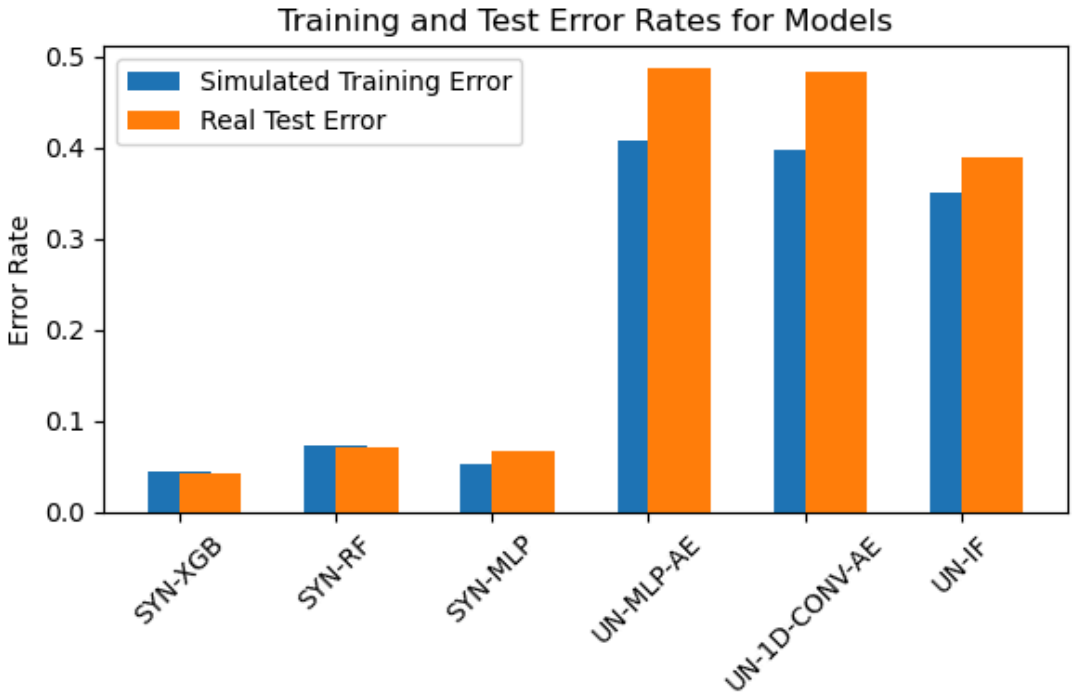


Figure 5.17: Training and test error comparison for our proposed model

related anomalies. The robustness of XGBoost is further enhanced by its ensemble learning framework that employs boosting [96], which has proven effective in managing anomalies and noise prevalent within electricity consumption data. Incorporating regularization within the objective function of the model serves as a bulwark against overfitting, thus facilitating better generalization to unseen data. The model's adeptness in binary classification makes it particularly well-suited for anomaly detection tasks, such as identifying instances of electricity theft in smart home environments. Given its exemplary performance across key metrics, XGBoost is highly recommended for deployment in ETD systems, where the accurate and reliable identification of theft-related irregularities is paramount.

The synthetic random forest (SYN-RF) and synthetic multilayer perceptron (SYN-MLP) also perform well, although they exhibit marginally higher error rates in comparison to SYN-XGB, suggesting room for optimization.

Conversely, unsupervised models, which include the MLP Autoencoder (UN-MLP-AE), the 1D Convolutional Autoencoder (UN-1D-CONV-AE), and the IF

(UN-IF), exhibit significantly higher error rates. This is particularly notable in scenarios simulating training conditions, which may point to challenges these models face in capturing complex patterns inherent to electricity theft without labeled training data. Notwithstanding, the UN-IF model demonstrates a lesser increase in error rate transitioning from simulated to real datasets, hinting at a certain level of stability in model performance despite lower overall accuracy.

The findings suggest that, in the context of smart home electricity theft detection, supervised models adeptly leverage the nuanced patterns within aggregated appliance consumption data, thus providing a strong foundation for the development of reliable theft detection systems.

5.9.4 Trade-off between training and test errors

In the dedicated exploration of Electricity Theft Detection (ETD) within smart homes, a critical aspect of our experimental design was ensuring equilibrium between the training and test error rates. This balance, a trade-off between training and test errors, is imperative to avert the model's overfitting to training data, which could compromise its generalization capabilities on new, unseen data. This phenomenon could skew the detection of electricity theft.

Our methodology encompasses the strategic application of GridSearchCV to perform exhaustive hyperparameter tuning, [82], a practice that aids in identifying optimal model parameters setting recorded in Table 5.3. To further bolster the reliability of our findings, we used a five-fold cross-validation scheme, which provides a more rigorous validation of the model's predictive probability.

In the domain of unsupervised anomaly detection, particularly when employing autoencoders, the goal is to minimize the reconstruction error across both training and test datasets. However, too low a training error (overfitting) may result in poor generalization of the test data. To achieve a good trade-off, we deploy early stopping techniques, and validation data splits, which form the cornerstone of our strategy to fine-tune the model's complexity. Additionally, the nuanced adjustment of hyperparameters, Table 5.5, including the dimensionality of the encoding and hidden layers, as well as the learning rate, was instrumental in achieving a judicious balance between underfitting and overfitting.

5.10 Privacy Concerns and Model Development

5.10.1 Privacy Concerns in Smart Home Environments

Developing a model to protect smart home users' privacy using an appliance consumption patterns dataset involves a comprehensive approach aimed at ensuring the model's capability to analyze or predict patterns without exposing sensitive personal information. The methodology encompasses several key steps outlined below:

1. Understanding the Dataset

- **Data Inspection:** This involves understanding the dataset comprehensively, including the types of appliances, usage patterns, timestamps, and any user-identifiable information present.
- **Privacy Concerns Identification:** Identifying elements within the data that might pose privacy risks, such as usage patterns that could infer when a household is occupied.

2. Data Anonymization

- **Anonymizing Data:** Techniques like pseudonymization (replacing private identifiers with fictitious ones) and aggregation (summarizing data to obscure individual details) are applied.
- **Noise Addition:** Introducing randomness into the data helps further mask individual patterns.

3. Differential Privacy (Optional)

Implementing differential privacy techniques is recommended for highly sensitive datasets. This ensures that the dataset's analysis remains unaffected by the alteration of any single record, thereby safeguarding individual privacy.

4. Feature Engineering

- **Select Relevant Features:** Features that are pertinent to the analysis yet less likely to compromise privacy are chosen.
- **Create Derived Features:** Derived features can sometimes offer valuable insights without disclosing sensitive information.

5. Model Development

- **Choose a Model:** A machine learning model suitable for the objective, such as predicting energy usage or identifying abnormal consumption patterns, is selected.
- **Train the Model:** The model is trained using the anonymized and engineered dataset.

6. Model Validation and Testing

- **Cross-Validation:** Techniques like k-fold cross-validation are employed to ensure the model's validity.
- **Performance Metrics:** The model's performance is evaluated using metrics such as accuracy and F1-score, ensuring that privacy-preserving measures do not significantly impair effectiveness.

7. Deployment and Monitoring

- **Deploy the Model:** The model is integrated into the smart home system.
- **Monitor and Update:** Continuous performance monitoring and necessary updates are carried out to accommodate new patterns or privacy concerns.

8. Compliance and Ethical Considerations

Ensuring compliance with data protection laws (like GDPR, CCPA) and regularly reviewing ethical considerations, particularly regarding data usage and the implications of model predictions, is crucial.

Tools and Technologies

Programming languages such as Python and R are used for data analysis and model development, with libraries like Pandas, NumPy, Scikit-learn, TensorFlow, or PyTorch supporting various stages of the process. For implementing differential privacy, libraries such as TensorFlow Privacy, PySyft, or Opacus can be utilized.

Security Measures

Ensuring secure data storage and transmission, alongside regular audits for data and model security, is paramount.

In developing such a model, maintaining a balance between utility and privacy is crucial to ensure the model's effectiveness without compromising user privacy.

5.11 Appliance Authentication Methods

5.11.1 Authentication of Appliance

Authentication strategies for electrical appliances focus on ensuring that only registered and verified devices operate within a given network, enhancing security and enabling efficient electricity use monitoring.

5.11.2 Appliance Signature Analysis

- Each electrical appliance exhibits a unique power consumption signature (load profile), distinguishable through variations in voltage, current, and frequency during different operational states.
- Algorithms are implemented to learn and recognize these power signatures, employing machine learning techniques to differentiate between authenticated appliances within a home environment.

5.11.3 Smart Meter Data Utilization

- Smart meters provide fine-grained electricity usage data, enabling the identification of operational patterns specific to authenticated appliances.

- This data aids in detecting anomalies by highlighting deviations from established consumption patterns, potentially indicating unauthorized usage or tampering.

5.11.4 Integration with Home Automation Systems

- The proposed SYNBDM model integrates with home automation systems, which catalog information on registered appliances, facilitating cross-reference checks between power usage data and authenticated appliance lists.
- Anomalies are flagged when discrepancies arise between actual power consumption and expected usage patterns of registered devices.

5.11.5 Real-Time Monitoring and Authentication Checks

- Real-time monitoring of electricity consumption is complemented by periodic authentication checks (updating), ensuring alignment between power usage and registered appliance profiles as depicted in Figure 5.18.
- Unauthorized appliance usage or abrupt changes in consumption patterns trigger immediate alerts for investigation.

5.11.6 Machine Learning for Anomaly Detection

- Both supervised and unsupervised machine learning algorithms are utilized to identify a typical usage patterns that deviate from recognized appliance profiles.
- The model is trained on datasets encompassing normal operations and various theft scenarios to enhance detection accuracy.

5.11.7 User Interaction and Feedback

- The SYNBDM system allows users to provide feedback on generated alerts, enabling confirmation of whether an anomaly signifies actual theft or a legitimate alteration in usage.

- This feedback loop plays a crucial role in refining the model's accuracy and reliability.

5.11.8 Security and Privacy Considerations

- Data collection and processing protocols are designed to comply with privacy regulations, ensuring user information is handled securely.
- Secure communication protocols are implemented to safeguard data transmitted between smart meters, appliances, and the SYNBDM system from unauthorized access.

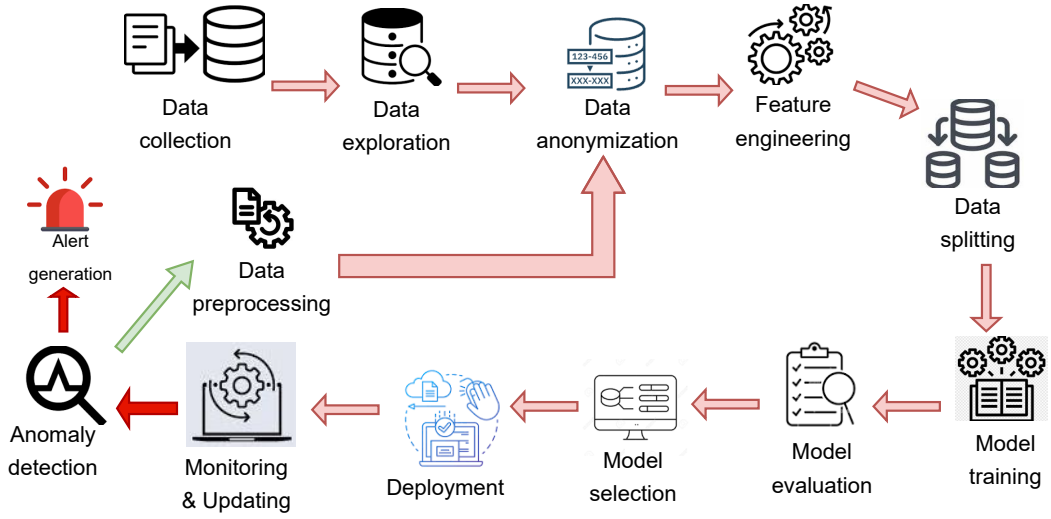


Figure 5.18: SYNBDM implementation process

5.11.9 Physical Fingerprint Appliances Authentication.

Some appliances have unique, measurable characteristics in their electricity usage that can be observed without internal data access, akin to a "fingerprint" [11]. In training the SYNBDM, these physical fingerprints—patterns in power consumption—are crucial. They serve as the dataset's features, enabling the model

to learn and distinguish between normal operation and anomalies, such as electricity theft. This approach leverages the appliance's consumption behavior as a distinctive signature to improve detection and authentication processes within the smart home environment.

To combine appliance physical fingerprints with consumption patterns in training the SYNBDM model, you would first collect detailed usage data for each appliance, including power draw, usage cycles, and any unique electrical signatures. This data forms the physical fingerprint. Next, integrate this with consumption patterns—how and when the appliance is typically used. The model is trained on these combined features to recognize normal operating conditions. By learning the nuanced differences between appliances and their typical usage, the model can more accurately detect anomalies indicative of electricity theft, enhancing its predictive capabilities.

5.11.10 Appliance Authentication Using Physically Unclonable Functions

To authenticate appliances with physical fingerprints within the context of a Synthetic Binary Discriminator Model (SYNBDM), we leverage the concept of Physically Unclonable Functions (PUFs). PUFs are unique, unclonable identifiers derived from the inherent physical variations of hardware devices. These variations, which occur naturally during the manufacturing process, can serve as a secure and tamper-evident fingerprint for each appliance.

Relevance to Previous Research

The integration of PUFs into electricity theft detection models like SYNBDM builds upon the foundation laid by previous research in both cybersecurity and smart grid management. Previous studies have highlighted the effectiveness of PUFs in various applications, from secure key generation to hardware authentication [54, 32], underscoring their potential for enhancing security in smart home environments.

5.11.11 Advantages of Using PUFs with SYNBDM

- **Enhanced Security:** PUFs provide a high level of security due to their unclonability and inherent resistance to physical tampering. This makes the SYNBDM more resilient against attempts to bypass or deceive the system through hardware manipulation.
- **Unique Identification:** Each appliance's PUF serves as a unique identifier, enabling precise authentication and monitoring of individual devices. This specificity aids SYNBDM in accurately detecting anomalous behavior indicative of electricity theft.
- **Reduced False Positives:** By ensuring that only authenticated appliances are monitored, PUFs can help minimize false positives in theft detection, improving the overall accuracy of the SYNBDM.
- **Low Overhead:** Incorporating PUFs into the SYNBDM does not significantly increase computational overhead, as the authentication process leverages inherent physical properties rather than complex cryptographic procedures.

5.11.12 Authentication without Biometrics

The focus on non-biometric user authentication aims to link appliance usage to specific users through interaction patterns, offering an alternative to direct biometric methods.

Non-Biometric Authentication Approaches

- Behavioral patterns and device interaction profiles serve as the basis for user authentication, leveraging smart home interfaces and IoT devices to monitor and authenticate user interactions with appliances.
- Techniques such as user-specific PIN codes, Radio Frequency Identification (RFID) tags, and smart device interaction monitoring enable the system to recognize and authenticate user activities without relying on biometric data.

This approach not only enhances the system's ability to monitor and control appliance usage but also significantly improves the detection capabilities for unauthorized use and electricity theft, thereby ensuring more secure and efficient energy consumption.

5.11.13 Privacy Concern Comparison of Recent Studies on ETD

The comparison of the methodologies used in these papers as related to our model is shown in the table below:

Paper Name	Methods Used	Contributions	Limitations	Practical Implications
Improving Home Appliance-Based Electricity Theft Detection: Insights from Real and Synthetic Attack Scenarios *(This thesis)*	SYNBDM-XGB, RF, MLP, LUM-MLP-AE, 1D-CONV-AE, Isolation Forest	High accuracy in theft detection with AUC scores up to 98.74%. Expands model to detect unknown attacks.	Dependence on data quality, overfitting, and privacy concerns.	Enhances energy security and management in smart homes.
Electricity Theft Detection in AMI Using Customers' Consumption Patterns [41]	SVM, CPBETD algorithm, Kernel function, Cluster analysis	Novel algorithm for detecting energy theft in AMI. Robust against contamination attacks.	False alarm rate, need for on-site inspection, privacy risks with high sampling rate.	High-performance solution for energy theft detection in smart grids.
Fully Homomorphic Encryption with Table Lookup for Privacy-Preserving Smart Grid [49]	FHE, Table Lookup with FHE, Integer encoding	Efficient protocol for function evaluation with FHE. More practical than previous methods.	Limited to addition and multiplication on encrypted data. Lookup table scalability.	Protects user data in smart grids while allowing function evaluation.
Privacy-Preserving Data Falsification Detection in Smart Grids using Elliptic Curve Cryptography & Homomorphic Encryption [42]	ECC-based homomorphic encryption	18x faster execution than CKKS scheme. Ensures data privacy with smaller memory space.	Long execution time for HE schemes. Privacy concerns with customer data analysis.	Fast and privacy-preserving detection of data falsification in smart grids.
Look-Up Table based FHE System for Privacy Preserving Anomaly Detection in Smart Grids [48]	Homomorphic LUT-based FHE, Private information retrieval with FHE	First implementation of flexible control over detection accuracy and time. Detects various attack types.	Differential privacy limits accuracy. SMC has high communication costs.	Enables privacy-preserving anomaly detection with flexible accuracy and time control.
Towards Privacy-preserving Anomaly-based Attack Detection against Data Falsification in Smart Grid [38]	Optimization of encryption for resource-constrained devices, Homomorphic encryption	Framework detects energy theft and data integrity attacks. 40x faster encryption for low-power devices.	Balancing privacy and security. Future work on quantification of information leakage.	Optimized for automated billing and load monitoring in smart grids, preserving user privacy.

5.12 Model Updates

5.12.1 Implementation of the SYNBDM Algorithm

Application of the SYNBDM algorithm proceeds as follows:

1. In each home, one or more transformers, Figure 5.23, gauge the total electricity delivered to the homeowner within the smart home network. This measured value is then compared with the aggregate amount of consumption as reported by the associated distribution transformer.

2. Every new sample undergoes preprocessing and is transformed into a format that aligns with the training set.

3. XGB is utilized on a new sample to ascertain whether it falls into the benign or attack class.

4. If step 1 fails to identify an anomaly and the XGB classifies the new sample as benign, this sample is then incorporated into the benign dataset. Concurrently, the relevant attack patterns are created and appended to the attack dataset.

5. If NTL is identified in step 1 and the classifier detects an attack, the appliance's behavior in the smart home is marked as suspicious. Electricity theft is confirmed when this suspicious appliance behavior in a smart home occurs n times within a specified timeframe. During this period, new samples are collected in a temporary database. Upon confirmation of electricity theft, an appropriate response, such as a physical inspection, is initiated. Homes in areas with higher NTL, as determined in step 1, are given priority for inspection. If theft is confirmed, the samples from the temporary database are moved to the attack dataset. If no theft is found, these samples are instead added to the benign dataset, along with their associated attack patterns.

6. Another possibility arises when step 1 fails to detect an NTL, but XGB identifies an anomaly. This scenario can be attributed to three potential causes. It might be a result of XGB misclassification or an error in the NTL

calculation in step 1. Such cases are not expected to recur frequently on consecutive days. Alternatively, the condition might stem from changes in consumption habits, such as alterations in residents or appliances, leading to significant shifts in usage patterns. In this instance, the condition will persist. Consequently, when XGB detects an anomaly while step 1 shows no NTL, the new sample is placed in a temporary database. Should this condition repeat often in the following days, the current dataset will be substituted by a new dataset created from the samples in the temporary database. Upon reaching a sufficient size, the classifier is retrained. Each smart home appliance is assigned a credibility factor, $\mathbf{cf}_i$, which is a binary variable initially set to one. In cases where a non-malicious anomaly is detected as described, $\mathbf{cf}_i$ is reduced to zero, and it is reset to one once the issue is resolved. In instances of detected electricity theft, smart homes with a $\mathbf{cf}_i$ of 1 are prioritized for further action. This aspect of the algorithm enhances SYNBDM's resilience against non-malicious alterations in consumption patterns.

7. If NTL is identified in step 1, but the (XGB) model does not detect any anomalies and the situation persists, this may indicate an ongoing attack that XGB is unable to recognize. In such instances, the safe dataset of the appliances is examined for indications of a data contamination attack. This attack type deceives the learning machine into misinterpreting an aberrant pattern as normal by gradually altering the data, thereby corrupting the dataset. The long-term usage pattern of the appliance is analyzed, with a downward trend in the long-term consumption graph potentially signaling contamination. Should historic data analysis not reveal a contamination attack, SYNBDM will sound an alarm (alert generation), Figure 5.18, suggesting that an attack may be occurring, although it remains undetected by the algorithm, which then continues its standard operation for new samples. This situation is uncommon, such as when a new, high-consumption load is directly connected to a feeder. This phase in the algorithm's process enhances SYNBDM's resilience against contamination attacks.

The SYNBDM algorithm uses a more mathematical-style notation: Let:

- S represents the new sample.
- $E(S)$ be the total electricity measured by the transformer for sample S.
- $R(S)$ be the reported consumption by the distribution transformer for sample S.
- $P(S)$ be the preprocessing and format conversion of sample S.
- $C(S)$ be the classification of sample S by XGB (returns 'benign' or 'attack').
- DB_{benign} and DB_{attack} represent the benign and attack datasets, respectively.
- DB_{temp} be the temporary database.
- CF_{i} represents the credibility factors.

$$E_{\text{total}} = \text{TransformerMeasurement}(S) \tag{5.34}$$

$$E_{\text{reported}} = \text{DistributionTransformerReport}(S) \tag{5.35}$$

$$\text{NTL} = \begin{cases} 1 & \text{if } E_{\text{total}} \neq E_{\text{reported}} \\ 0 & \text{otherwise} \end{cases} \tag{5.36}$$

$$S_{\text{processed}} = \text{Preprocess}(S) \tag{5.37}$$

$$\text{class} = \text{XGBClassify}(S_{\text{processed}}) \tag{5.38}$$

$$\text{DatasetUpdate}(S_{\text{processed}}, \text{class}, \text{NTL}) \tag{5.39}$$

$$\text{TheftHandling}(\text{NTL}, \text{class}, S_{\text{processed}}) \tag{5.40}$$

$$\text{AnomalyHandling}(\text{NTL}, \text{class}, S_{\text{processed}}) \tag{5.41}$$

$$\text{ContaminationCheck}(\text{NTL}, \text{class}) \tag{5.42}$$

Algorithm 1 SYNBDM Algorithm for Electricity Theft Detection in Smart Home Networks

Input: New sample S

Output: Updated benign_dataset DB_{benign}, attack_dataset DB_{attack}, temp_database DB_{temp}, credibility_factors CF_{i}

Initialize: benign_dataset DB_{benign}, attack_dataset DB_{attack}, temp_database DB_{temp}, credibility_factors CF_{i}

Step 1: Measure Total Electricity and Detect NTL $E(S) \leftarrow$ Transformer_Measurement(S) $R(S) \leftarrow$ Distribution_Transformer_Report(S) NTL(S) $\leftarrow E(S) \neq R(S)$

Step 2: Preprocess Sample $S' \leftarrow P(S)$

Step 3: Classify Sample Class(S') $\leftarrow C(S')$

Step 4: Update Datasets if $\neg NTL(S)$ *and Class*(S') = *'benign'* **then**

    DB_{benign}.add(S') Generate_And_Add_Attack_Patterns(DB_{attack})

Step 5: Detect And Handle Theft if $NTL(S)$ *and Class*(S') = *'attack'* **then**

    Report_Suspicious_Behavior() **if** *Suspicious_Behavior_Repeated(m, period)* **then**

        Take_Appropriate_Action() **if** *Theft_Verified()* **then**

            DB_{attack}.add(DB_{temp})

        else

            DB_{benign}.add(DB_{temp})

Step 6: Handle Anomalies if $\neg NTL(S)$ *and Class*(S') = *'anomaly'* **then**

    DB_{temp}.add(S') **if** *Anomaly_Persists(days)* **then**

        Discard_Old_And_Create_New_Dataset(DB_{temp}) Retrain_Classifier() Adjust_Credibility_Factors(CF_{i}, appliance)

Step 7: Detect Contamination Attack if $NTL(S)$ *and Class*(S') $\neq$ *'anomaly' and Condition_Persists()* **then**

    if *Check_For_Contamination_Attack(appliance_data)* **then**

        Raise_Alert()

    else

        Continue_Normal_Operations()

return DB_{benign}, DB_{attack}, DB_{temp}, CF_i

5.13 ETD Model Implementation in AMI Using Anonymized Aggregated Appliance Consumption Data

5.13.1 Data Anonymization

The dataset utilized in this study has been pre-anonymized, containing only numerical readings devoid of direct personal identifiers. It is imperative to scrutinize the dataset thoroughly for columns harboring potentially identifiable information, albeit not evident in the initial rows. Such columns necessitate anonymization or elimination to uphold privacy standards.

5.13.2 Feature Engineering and Anomaly Detection

A pivotal aspect of our methodology involves identifying salient features that significantly contribute to the detection of abnormal consumption patterns. This process may entail the creation of novel features, such as the aggregation of readings to encapsulate total consumption over specified intervals, thereby enhancing the model's predictive prowess.

5.13.3 Train Models

For the model training phase, the smaller subset designated for training is bifurcated into distinct training and validation sets. Subsequently, models including XGBoost, Random Forest, and Multilayer Perceptron (MLP) are meticulously trained on these datasets.

5.13.4 Model Evaluation

Model performance is rigorously evaluated using a dedicated test dataset, denoted as `psa_journal_home_A_sim_test.csv`. Key performance metrics such as Accuracy, Recall, Precision, F1-Score, and the Area Under the Curve (AUC) are computed to ascertain the efficacy of each model.

5.13.5 Visualizations of the Anonymized Dataset

To illustrate the dataset's anonymized state while showcasing its utility for privacy-conscious insights, several visualization techniques are employed:

1. **Histograms of Feature Distributions:** These histograms elucidate the distribution of readings across the first 10 features, as depicted in Figure 5.19, offering insights into data spread and prevalent value ranges, thereby ensuring privacy through aggregation.

2. **Correlation Heatmap:** A heatmap, Figure 5.20, depicting the correlation among selected features is generated to unveil patterns and relationships without compromising individual privacy.

3. **Time Series Plots:** For features representing time-series data, plots are crafted to display general consumption trends over time, highlighting overall patterns rather than individual behaviors as shown in Figure 5.21.

4. **Box Plots for Anomaly Visualization:** Utilization of box plots, Figure 5.22, facilitates the identification of outliers or anomalies within the data, crucial for detecting abnormal consumption patterns.

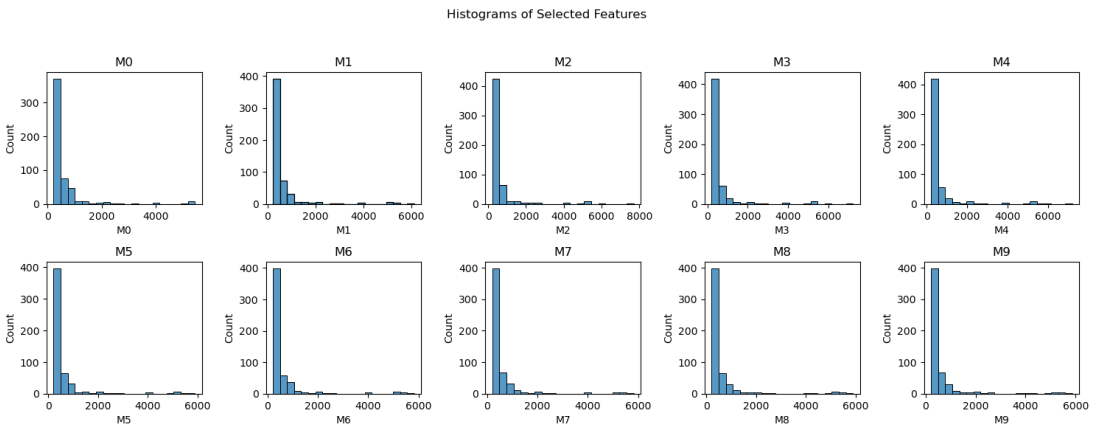


Figure 5.19: Histograms of selected feature distributions

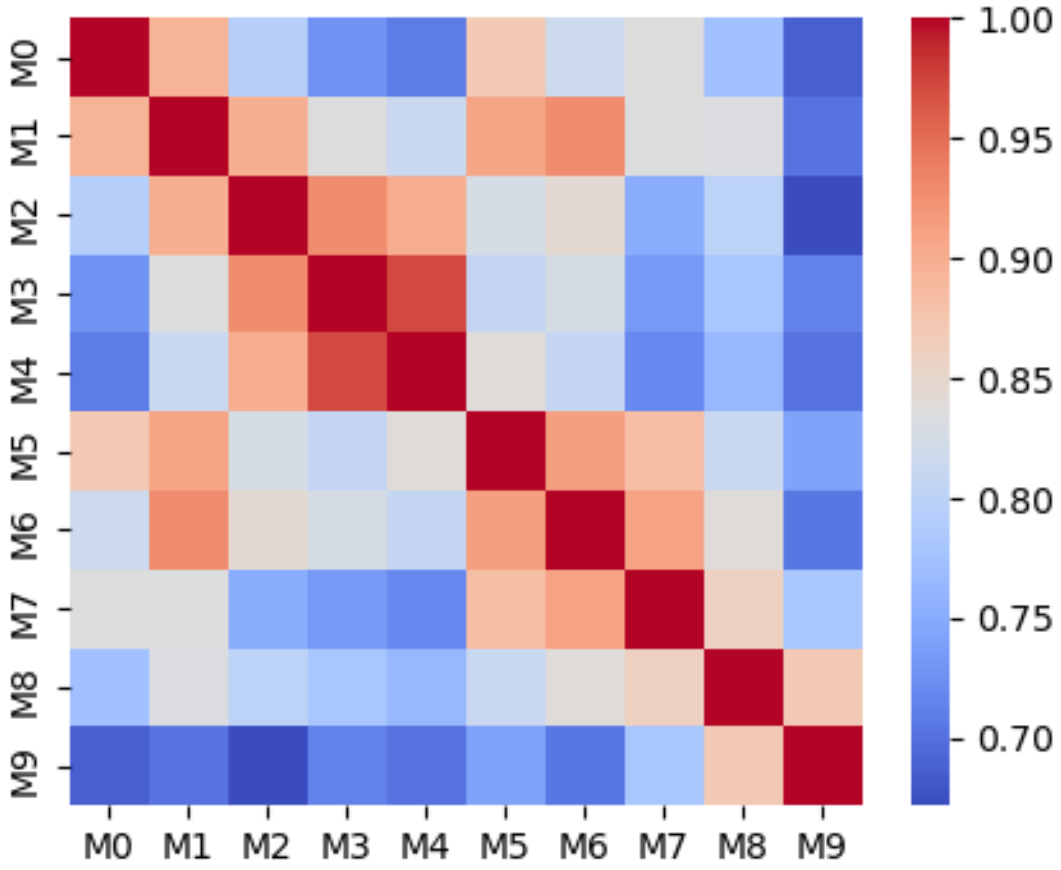


Figure 5.20: Correlation heatmap of selected features

These visualizations underscore the feasibility of deriving meaningful insights from data while strictly adhering to privacy considerations. The deliberate absence of direct or indirect personal identifiers, such as names or specific timestamps, ensures the anonymity of the dataset.

The Figure 5.23 provides a visual representation of the implementation of the ETD (Electricity Theft Detection) model, known as SYNBDM, within an Advanced Metering Infrastructure (AMI) system. This implementation leverages anonymized aggregated appliance consumption pattern data to enhance energy management and efficiency.

In this context, the AMI system plays a crucial role by collecting granular energy consumption data from smart meters deployed in distinct areas, namely

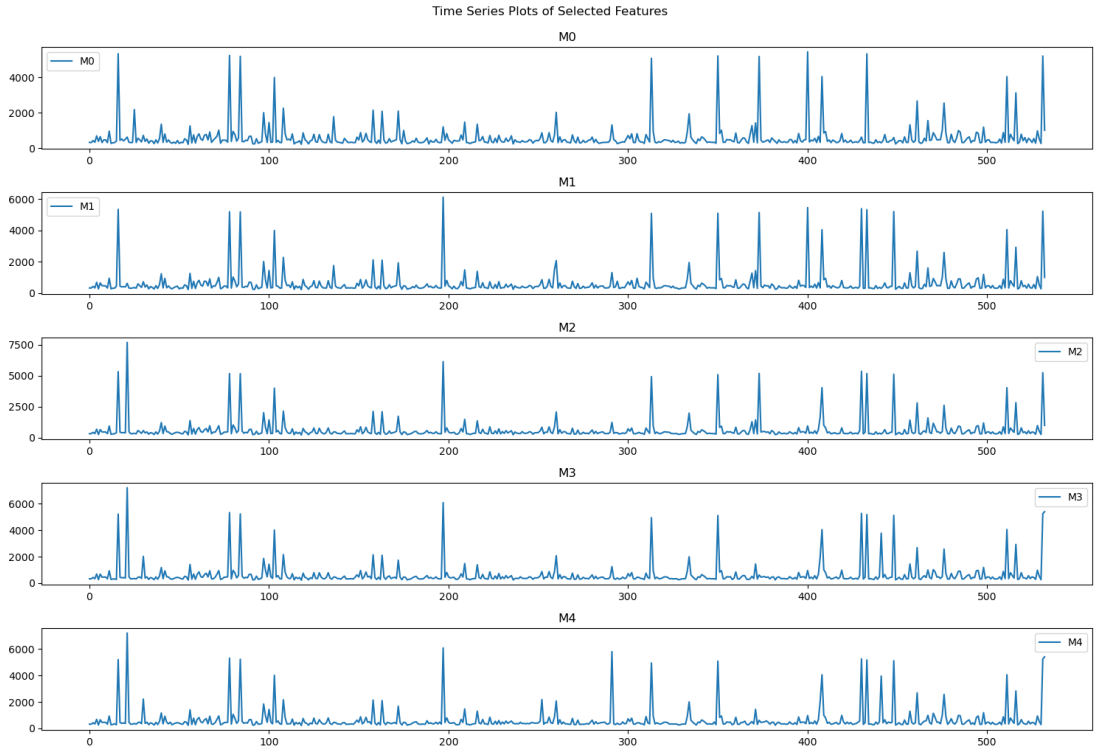


Figure 5.21: Selected sample of first five features for time series plot

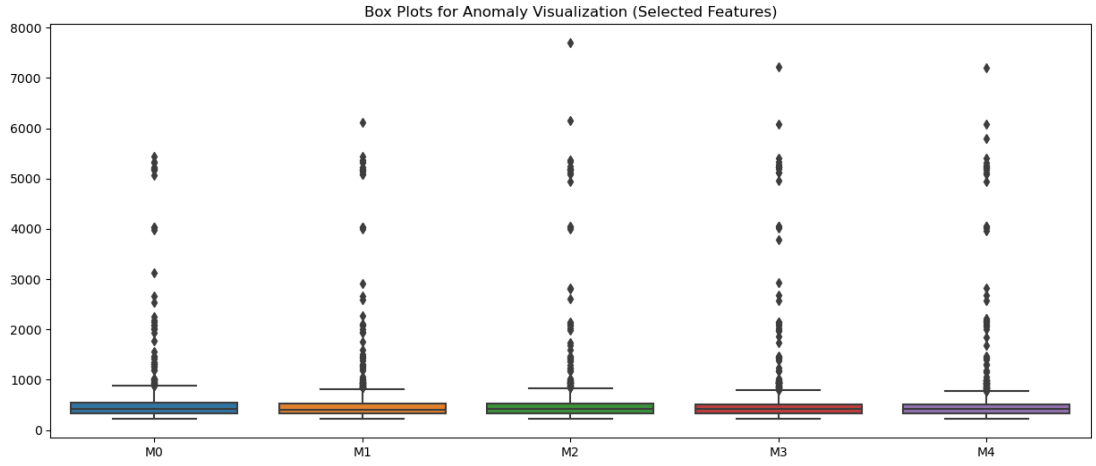


Figure 5.22: Selected sample of Box Plots for Anomaly Visualization

Area A for Smart Home Users and Area B for Commercial Users. This data serves as the foundation for analyzing appliance consumption patterns.

The flow of energy consumption data follows a path from the smart meters to a central data concentrator and then to the Data Management System. This streamlined data flow ensures centralized data processing and analysis, facilitating meaningful insights.

The SYNBDM model is employed to make sense of the aggregated data, focusing on tasks such as optimizing energy distribution, identifying usage trends, and improving overall energy efficiency. Notably, the implementation prioritizes data privacy and security through the anonymization of collected data, adhering to data protection regulations while enabling effective analysis of consumption patterns.

The overarching objective of this implementation is to enhance energy management practices, improve efficiency, and potentially offer insights for future infrastructure development and policy-making decisions related to energy distribution.

This implementation exemplifies a sophisticated approach to energy management, capitalizing on modern AMI systems and advanced data analysis models while maintaining the utmost consideration for user privacy and data security.

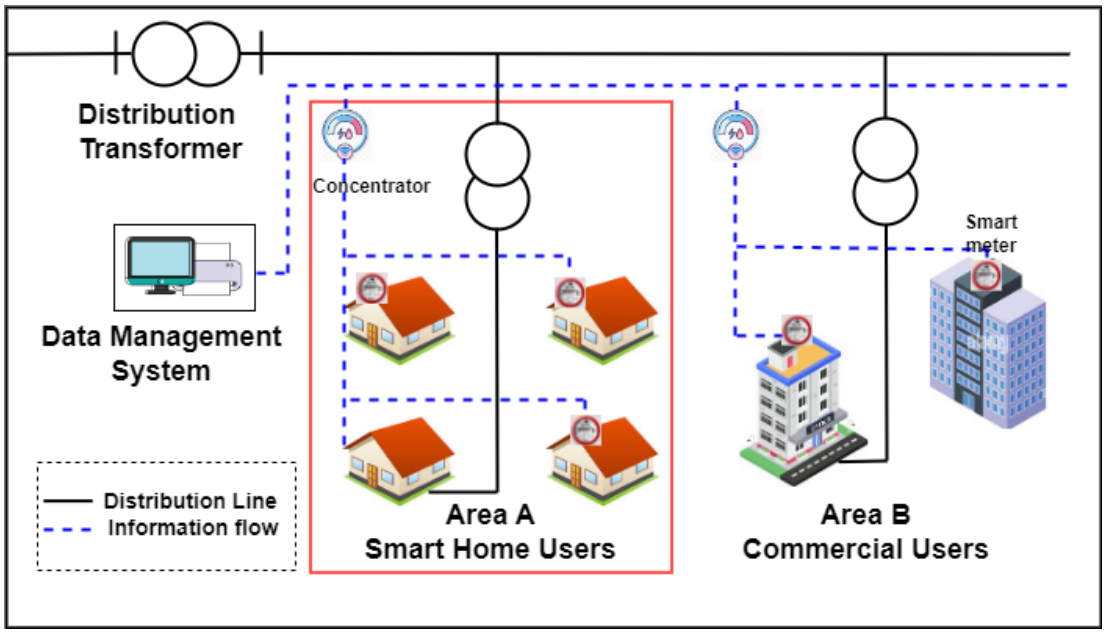


Figure 5.23: SYNBDM implementation in AMI

5.14 Model Performance Comparison

5.14.1 Comparison with Benchmark Models

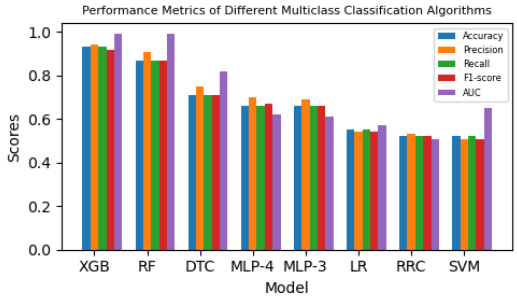
The comparison of Table 5.9 with the accuracy and AUC scores and with other classifiers in Table 5.10 provides a clearer picture of our model performance with benchmark algorithms such as SVM and LR, each model in multiclass and binary classification tasks.

Table 5.9: Comparison of models based on accuracy and AUC scores

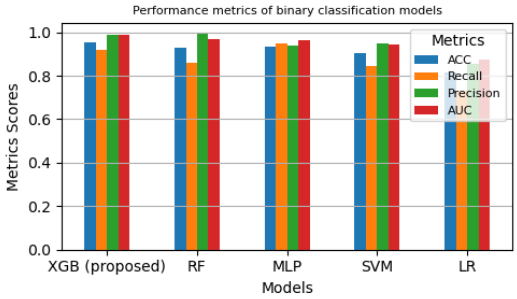
Model	Multiclass Accuracy	Binary Class Accuracy	Multiclass AUC	Binary Class AUC
XGB	0.9330	0.9569	0.9851	0.9872
RF	0.8652	0.9276	0.8896	0.9692
MLP	0.6610	0.9399	0.6189	0.9668
LR	0.5537	0.8148	0.5743	0.8748
SVM	0.5220	0.9035	0.6478	0.9430

XGB had the highest scores across the board, achieving 93.30% accuracy and 98.51% AUC in multiclass, and 95.69% accuracy and 98.72% AUC in binary classification. RF is a strong contender with 86.52% multiclass accuracy and 88.96% AUC, (Figure 5.24(a)), along with 92.76% binary accuracy and 96.92% AUC.Figure 5.24(b). MLP performs moderately with 66.10% multiclass accuracy and 61.89% AUC, improving binary classification with 93.99% accuracy and 96.68% AUC. LR and SVM, while viable, offer lower accuracy and AUC, suggesting that more advanced methods may be preferable for complex classification challenges.

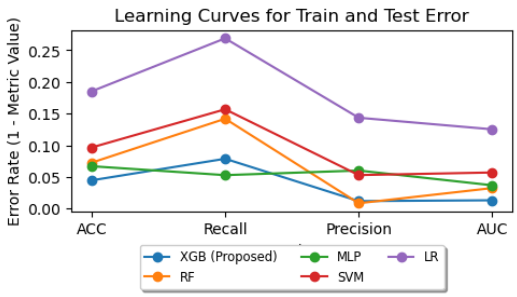
The line plot presented in Figure 5.24(c) illustrates the error rates for various evaluation metrics across five different machine learning models: XGBoost (proposed), Random Forest (RF), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Logistic Regression (LR) derived from experimental result Table 5.11. The error rate for each model is computed as a $1 - metric\ value$, where the metrics include accuracy (ACC), Recall, Precision, and Area Under the Curve (AUC). Figure 5.24(d) shows the aggregate AUC and ACC binary



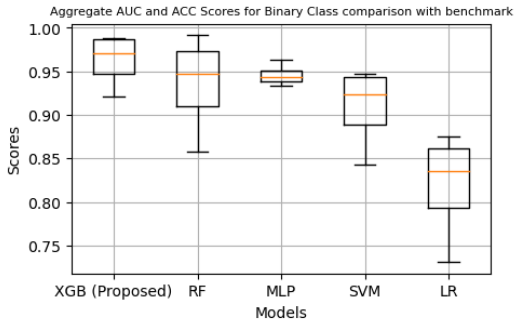
(a) Comparison of multiclass classification models



(b) Comparison of binary classification models



(c) Comparison of training and test error with benchmark model



(d) Model AUC and ACC scores comparison with benchmark

Figure 5.24: Performance metrics comparison analysis

class performance comparison of the proposed models with the benchmark, existing literature, SVM, and LR models.

- The **XGBoost model (proposed)** shows the most favorable error rates across all metrics, signifying its superior performance relative to the other models.
- The **Random Forest (RF)** and **Multilayer Perceptron (MLP)** models displayed competitive performance, with error rates marginally higher than those of the XGBoost model.
- The **Support Vector Machine (SVM)** and **Logistic Regression (LR)** models exhibit higher error rates, indicating that their performance is not as robust as that of the models above for the tasks evaluated.

Table 5.10: Performance metrics of different multiclass classification algorithms.

Models	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC**
XGB	0.9330	0.9417	0.9291	0.9174	0.9851
RF	0.8652	0.9082	0.8694	0.8701	0.8896
DTC	0.7142	0.7477	0.7068	0.7129	0.8208
MLP-4	0.6610	0.7044	0.6552	0.6812	0.6189
MLP-3	0.6582	0.6877	0.6571	0.6601	0.6097
LR	0.5537	0.5417	0.5543	0.5383	0.5743
RRC	0.5148	0.5280	0.5194	0.5178	0.5060
SVM	0.5220	0.5109	0.5189	0.5085	0.6478

Table 5.11: Proposed model performance metric comparison with binary class. benchmark.

Metric	XGB (Proposed)	RF	MLP	SVM	LR
ACC	0.9554	0.9275	0.9331	0.9035	0.8148
Recall	0.9213	0.8581	0.9472	0.8432	0.7311
Precision	0.9882	0.9916	0.9399	0.9471	0.8565
AUC	0.9869	0.9676	0.9632	0.9430	0.8748

The graph underscores the effectiveness of the XGBoost model in minimizing error rates, which correlates with the high predictive accuracy and model reliability for our ETD in smart homes.

5.14.2 Model Performance on Synthetic Attack Data

For Home A, when evaluated on simulated synthetic attack data, for example, the SYN-XGB model stands out with an impressive AUC score of 98.76%. It achieves a high F1-score of 94.92%, indicating robust performance in capturing fraudulent electricity consumption patterns as illustrated in Table 5.4. SYN-RF and SYN-MLP also exhibit strong AUC scores of 96.47% and 96.56% respectively, indicating their efficacy in identifying anomalies. It is noteworthy that SYN-MLP

has a slightly higher AUC score metric performance detection rate of 0.06% than SYN-RF and also has a higher F1-score of 97.23% which is 2.31% and 6.10% performance rate better than SYN-XGB and SYN-RF respectively in Home A simulated attack only, suggesting a trade-off between precision and recall.

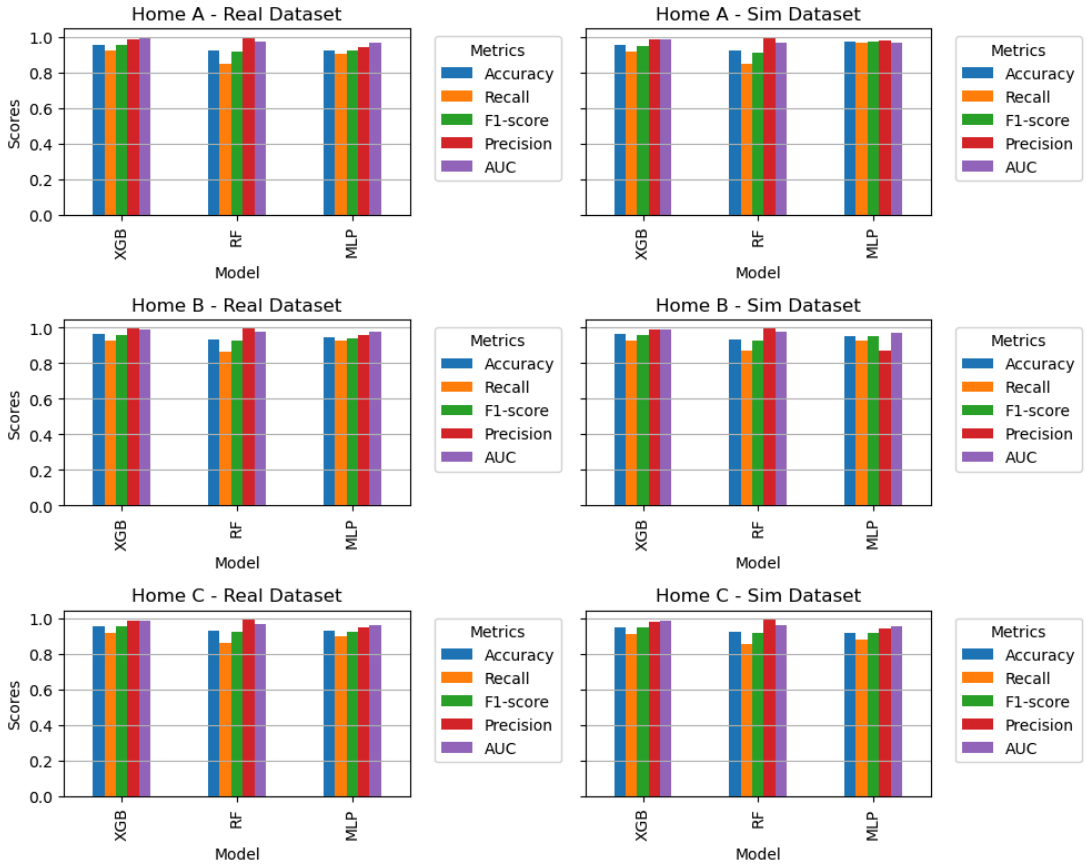


Figure 5.25: Perfomance across homes for the SYNBDM

In contrast, the legacy unsupervised models(LUM), UN-MLP-AE, UN-1D-CONV-AE, and UN-IF, struggle to match the performance of supervised models on synthetic data. UN-MLP-AE achieves an aggregated AUC score of 76.18% and 61.74%, while UN-1D-CONV-AE (69.77% and 62.61%), and UN-IF (64.28%), for simulated and real attacks for all homes, lag further behind, respectively as depicted in Table

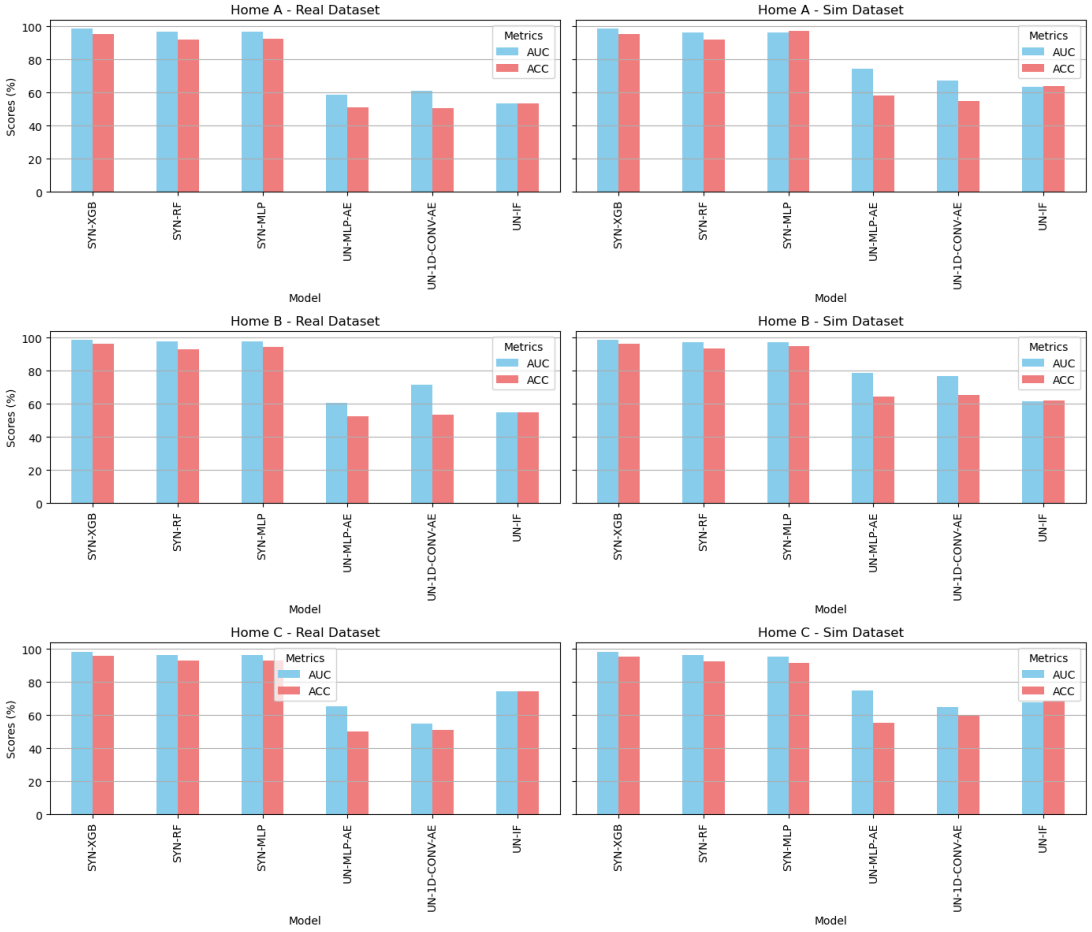


Figure 5.26: Perfomance across homes for the LUM

5.14.3 Model Performance on Real Attack Data

When assessing the same models on real attack data, for example in Home A, SYN-XGB continues to excel with an AUC score of 98.91%, reaffirming its capability to detect electricity theft accurately. Similarly, SYN-RF and SYN-MLP maintain strong AUC scores of 97.02% and 96.68%, respectively. These models also exhibit competitive F1-scores, indicating their reliability in real-world scenarios.

In contrast, the LUM's performance drops significantly when faced with a real attack scenario. UN-MLP-AE, UN-1D-CONV-AE, and UN-IF struggle to achieve AUC scores above 60%, signaling their limitations in detecting electricity theft

in practical settings.

This Chapter centered on developing and evaluating machine learning models for Electricity Theft Detection (ETD), focusing on binary classification via a Binary Discriminator model. The model's effectiveness was assessed in two phases: training and testing, using both simulated and real datasets. This approach ensured the model's generalizability to real-world scenarios.

The chapter delved into the mathematical framework underpinning the primary models employed in this study: XGBoost, Random Forest (RF), and Multilayer Perceptron (MLP). Each model's objective function and prediction mechanisms were outlined, employing Maximum Likelihood Estimation (MLE) for optimizing their respective objective functions.

Comprehensive performance evaluations were conducted, showcasing the models' effectiveness in ETD. This was achieved through detailed analyses using Receiver Operating Characteristic (ROC) curves and confusion matrices, which provided insights into each model's discriminative ability and detailed classification performance.

Key findings from the evaluations included:

- The XGBoost model's high accuracy and low false positive rate, making it a standout performer.
- The Random Forest model also demonstrated high True Negative rates, although with a slightly lower True Positive rate compared to XGBoost.
- The MLP model showed high accuracy in benign case identification, with slightly lower accuracy in attack case detection.

Furthermore, the chapter addressed the challenge of training and test errors, emphasizing the importance of balancing these errors to avoid overfitting or underfitting. Techniques like GridSearchCV for hyperparameter tuning and cross-validation were employed to refine the models further.

In conclusion, the chapter reaffirmed the potential of machine learning models in effectively detecting and classifying electricity theft, highlighting their role in enhancing smart home security against such fraudulent activities. The exploration of different models and their performance metrics offered a thorough

understanding of their capabilities and limitations, setting a foundation for future advancements in ETD technologies.

5.15 Differences Between Consumption and Prevention in Smart Home Electricity Management

In the context of electricity theft detection within smart homes, the terms "consumption" and "prevention" hold distinct significances that are pivotal to the management and protection of electrical resources:

- **Consumption:** Refers to the electrical energy usage by various appliances and systems within the smart home. Monitoring and analyzing consumption patterns serve multiple objectives, such as optimizing efficiency, calculating costs, and identifying anomalous behaviors. Devices like smart meters and home energy management systems are essential for tracking the electricity usage across different appliances, their consumption rates, and timing. This information is integral to establishing baseline consumption patterns within a household.
- **Prevention (Electricity Theft):** Entails adopting measures and strategies aimed at curtailing unauthorized access or theft of electricity. Various methodologies for electricity theft include meter tampering, meter bypassing, and infiltrating smart meter systems. Prevention strategies encompass:
  - Technical Measures: Implementation of advanced security features within smart meters, including tamper detection sensors, encrypted data transmission, and robust authentication protocols.
  - Data Analysis and Anomaly Detection: Utilizing consumption data to identify unusual or suspicious patterns indicative of potential theft, such as unaccountable surges in electricity use or consumption patterns that deviate from the household's typical profile.
  - Regular Inspections and Audits: Conducting physical checks on the meters and electrical installations to uncover any tampering or unauthorized modifications.

5.15.1 The Consumption Pilot Approach

In the realm of smart homes and smart grids, the interplay between electricity consumption and theft prevention is paramount. By meticulously analyzing electricity consumption data, we can not only streamline energy management but also significantly enhance our ability to thwart electricity theft. It is through the meticulous examination of normal consumption patterns that anomalies indicative of theft emerge more conspicuously, enabling timely investigation and intervention.

The concept of a "consumption pilot approach" is pivotal in this discourse. It denotes a preliminary study or trial endeavor designed to gain insights into, monitor, and refine electricity usage. Such approaches are instrumental in evaluating the efficacy of novel technologies, methodologies, or behavioral strategies prior to their widespread implementation. This methodology is integral to our Electricity Theft Detection (ETD) model, which leverages data-driven insights to pinpoint and mitigate unauthorized electricity usage effectively.

5.16 Analysis of Economic Implications

5.16.1 Economic Implication of SYNBDM Deployment

The deployment of a Synthetic Binary Discriminator Model (SYNBDM) in smart homes for electricity theft detection and appliance usage authentication can have several economic implications, both for the stakeholders directly involved (such as utility companies, homeowners, and technology providers) and for the broader economy. These implications can vary widely depending on the scale of deployment, the effectiveness of the technology, and the specific context in which it is implemented. Here are some key economic implications to consider:

5.16.2 For Utility Companies and Electricity Providers

- **Reduction in Electricity Theft:** By accurately identifying and mitigating instances of electricity theft, utility companies can significantly reduce losses. This improvement in revenue protection can translate to better financial health for these companies.

- **Investment Costs:** The initial outlay for developing, testing, and deploying SYNBDM systems can be substantial. However, if the system effectively reduces theft, the long-term savings could outweigh these costs.
- **Operational Efficiency:** Implementing advanced detection systems can streamline operations, reduce the need for manual inspections, and decrease the incidence of false theft accusations, leading to cost savings.

5.16.3 For Homeowners and Consumers

- **Increased Energy Costs:** The cost of implementing and maintaining such a system may be passed on to consumers in the form of higher energy rates or service charges. However, if the system effectively deters theft, it could contribute to a more equitable distribution of costs and potentially lower rates over the long term.
- **Privacy and Security Concerns:** There could be additional costs related to ensuring data privacy and security, especially given the sensitive nature of biometric data like fingerprints.

5.16.4 For Technology Providers

- **Market Opportunities:** The development and deployment of SYNBDM systems open new markets for technology providers specializing in smart home devices, biometric authentication, and energy management systems. This can lead to increased revenues and opportunities for innovation.
- **Research and Development Costs:** Providers will incur costs in researching, developing, and continuously improving SYNBDM technologies. These costs are investments in the competitive advantage of their solutions.

5.16.5 For the Broader Economy

- **Innovation and Job Creation:** The push towards advanced energy management solutions like SYNBDM can stimulate innovation, leading to the creation of new jobs in technology, engineering, and data analysis sectors.

- **Energy Efficiency and Sustainability:** By reducing electricity theft and promoting efficient appliance use, SYNBDM deployment can contribute to broader goals of energy efficiency and sustainability. This can have long-term positive effects on the economy by conserving resources and reducing environmental impact.

- **Regulatory and Legal Framework:** Implementing such systems may require new regulations and legal frameworks to address issues like privacy, data security, and consumer protection, potentially leading to costs associated with compliance and enforcement.

5.16.6 Economic Equity

- **Access and Affordability:** There's a risk that the costs associated with advanced systems like SYNBDM could exacerbate economic disparities, making smart home technologies less accessible to lower-income households.

The economic implications of deploying SYNBDM in smart homes are multifaceted, involving trade-offs between upfront costs and long-term benefits, as well as considerations of privacy, equity, and sustainability. The balance of these implications would depend on the effectiveness of the technology, regulatory environment, and how the costs and benefits are distributed among stakeholders.

5.17 Chapter Summary

The chapter presents an in-depth exploration of methodologies for detecting electricity theft in smart homes, significantly expanding the scope of previous research by incorporating a diverse range of attack scenarios. This expansion allows for more comprehensive detection capabilities, including both classified and unclassified attacks.

Central to the chapter is the sophisticated approach to data preprocessing, tailored to accommodate various attack scenarios such as baseload, weakload, and other unclassified attacks. This preprocessing is pivotal for preparing the dataset for effective model training and ensuring accurate attack detection.

A key innovation in the chapter is the implementation of data augmentation using circular shifting. This technique addresses the limitation of insufficient benign records, thereby enhancing the volume and diversity of the training dataset. As a result, the robustness of the Electricity Theft Detection (ETD) framework is significantly improved.

The chapter also delves into the integration of model selection and feature selection techniques, crucial for optimizing the machine learning pipeline for ETD. Employing Maximum Likelihood Estimation for parameter optimization ensures that the models are not only well-calibrated but also highly effective in practical scenarios.

Furthermore, the chapter provides insights into the specific features of the appliance consumption dataset. These features are instrumental in training the models to accurately differentiate between normal and anomalous consumption patterns, thus enhancing the accuracy and reliability of the ETD system.

Also, the methodologies and approaches detailed in this chapter provide a comprehensive and advanced framework for electricity theft detection in smart homes. Through the integration of diverse attack scenarios, innovative data augmentation techniques, and careful model and feature selection, the chapter lays out a robust and efficient system capable of effectively safeguarding smart homes against electricity theft.

6. Discussion and Future Work

6.1 Discussion

In this study, we explored the complexities inherent in electricity theft detection (ETD) within smart home environments, focusing on the use of aggregated power consumption patterns of appliances. A critical challenge in ETD is the variability of appliance consumption patterns, which can be influenced by a range of non-attack factors, such as temporary electrical spikes, periodic variations, or even permanent changes in usage habits. Such variations pose a risk of false positives in theft detection systems, where benign changes in power usage might be mistakenly identified as malicious activities.

Our synthetic binary discriminator model (SYNBDM) is designed to address these challenges effectively. It incorporates mechanisms to differentiate between short-term unusual behaviors and actual theft incidents. For instance, a transient spike in power usage, which may occur due to atypical yet benign activities, is not immediately flagged as an anomaly. This approach significantly reduces the likelihood of false alarms triggered by such short-term changes. The utility of aggregated power base consumption patterns in this context cannot be overstated, as it plays a pivotal role in reducing false positives. By only reporting suspicious behavior when both the smart meters and the XGB model concurrently detect an anomaly, the system ensures a higher degree of accuracy. Consequently, a single appliance's unusual yet non-malicious behavior does not trigger a false alarm unless another appliance is simultaneously compromised, indicating a potential theft scenario.

Furthermore, to refine the system's accuracy, we emphasize the importance of calculating the false positive rate (FPR) and adjusting the sensitivity parameter, denoted as'm'. This adjustment is crucial, as it allows the system to avoid

overreacting to sporadic or isolated incidents of unusual power usage. By configuring the system to flag theft only upon the recurrence of suspicious behavior, we significantly enhance the reliability of our ETD model.

The application of uniform manifold approximation and projection (UMAP) for clustering adds another layer of sophistication to our approach. This technique enables the algorithm to discern and adapt to various distribution patterns in the dataset, allowing for the training of separate classifiers tailored to specific usage patterns. Such adaptation is particularly beneficial in accounting for the differences in appliance usage between weekdays and weekends or across different seasons. If time-dependent patterns are observed within these clusters, the corresponding classifiers are labeled accordingly, ensuring that new instances are evaluated using the most relevant classifier for that specific time frame.

Lastly, our model is adept at identifying and adjusting to permanent changes in consumption patterns, such as those resulting from new appliances or shifts in weather conditions. This adaptability is key to maintaining the long-term effectiveness of the SYNBDM, ensuring that it remains reliable despite evolving household dynamics.

Our study not only addresses the immediate challenges of detecting electricity theft in smart homes but also lays the foundation for future advancements in this field. By considering a wide array of factors that influence power consumption and employing advanced analytical techniques, our approach demonstrates a comprehensive and robust strategy for ETD.

6.1.1 Summary of our findings

This comprehensive study was dedicated to advancing the field of Electricity Theft Detection (ETD) in smart home environments, utilizing sophisticated machine learning models. The primary focus was on devising and validating models capable of discerning irregularities in electricity usage patterns, a key indicator of potential theft.

6.1.2 Model Development and Evaluation

At the core of our research, we developed a Binary Discriminator model for binary classification tasks. This model was rigorously trained and tested using both simulated and real datasets to ensure its efficacy and adaptability to real-world scenarios.

6.1.3 Machine Learning Algorithms

The investigation employed various machine learning algorithms, including XGBoost, Random Forest (RF), and Multilayer Perceptron (MLP). Each model was meticulously configured, and its performance was optimized using Maximum Likelihood Estimation (MLE) techniques.

6.1.4 Performance Metrics

Our evaluation process heavily relied on Receiver Operating Characteristic (ROC) curves and confusion matrices. These tools provided a detailed analysis of each model's capability to differentiate between normal and anomalous electricity usage patterns. Particularly, the XGBoost model demonstrated exceptional accuracy and a low false positive rate, marking it as a notably effective tool in ETD.

6.1.5 Challenges in Model Training

One of the significant challenges we addressed was the balancing of training and test errors to prevent overfitting or underfitting. This was achieved through methods such as GridSearchCV for hyperparameter tuning and the inclusion of cross-validation techniques.

6.1.6 Data Augmentation Techniques

To combat the issue of limited benign records, data augmentation strategies were employed. This involved creating time-offset variations of the input features, enriching the dataset, and enhancing the model's ability to generalize across different scenarios.

6.1.7 Synthetic Attack Data Utilization

A novel approach in our study was the use of synthetic attack data to train the models. This methodology allowed for a more comprehensive understanding of possible electricity theft scenarios and aided in the development of models that are more attuned to real-world attack patterns.

6.1.8 Legacy vs. Smart Attacks

The research distinguished between legacy attacks (e.g., Baseload, Midnight, EvilTwin) and smart attacks (Weakload, Peakhour). While legacy attacks were detected with high accuracy, smart attacks posed more challenges, indicating the need for further refinements in model training and data analysis techniques.

6.1.9 Future Directions

Looking ahead, the study suggests exploring additional attack types and including varied electricity consumption patterns to enhance the models' robustness. Furthermore, considering factors such as seasonal effects and personalization through federated learning could significantly advance the effectiveness of ETD systems.

6.1.10 Limitations of the Proposed Model

Although our models mark significant advancements in detecting electricity theft in smart homes, it is imperative to acknowledge certain limitations that accompany our current methodology.

Dependence on Data Quality and Granularity: The effectiveness of our models is closely tied to the quality and granularity of aggregated appliance consumption data. Any inadequacies in data resolution or representativeness could potentially affect the predictive accuracy of the models.

Assumption of Consistent Consumption Patterns: Our models operate under the assumption that consumption patterns within a household remain relatively stable over time. Significant behavioral changes or the introduction of new appliances could alter these patterns, potentially affecting the model's performance until retraining occurs.

Overfitting Risks and Model Complexity: Despite the implementation of regularization techniques, there remains the risk of overfitting, particularly if the model complexity is not finely calibrated.

Privacy and Ethical Considerations: Although our dataset was anonymized, the utilization of detailed electricity consumption data raises privacy concerns. This can be resolved using federated learning or secure multiparty computation in smart home energy consumption monitoring[62], [44]. The potential for re-identification or misuse of these data, even in an anonymized form, cannot be entirely ruled out. Ensuring ongoing compliance with privacy regulations and maintaining ethical standards for data usage are paramount.

Computational Demands and Resource Constraints: The computational complexity associated with our models, especially in terms of hyperparameter tuning and processing, presents limitations in terms of resource allocation.

In future work, we plan to address these limitations knowing that embarking on further research in these areas will contribute to the ongoing development of robust and effective ETD systems for smart homes. As the field continues to evolve, these challenges provide exciting avenues for future exploration and innovation in the quest for more secure and reliable smart grids.

7. Conclusion

Electricity theft poses a significant and widespread global challenge, leading to elevated utility expenses, additional financial burdens on compliant consumers, and various safety concerns. Recent advancements, including smart metering and Internet-based software in the smart grid, have increased its vulnerability to electricity theft. However, despite these developments, utility providers continue to face challenges in identifying and addressing complex attacks targeting the metering infrastructure. Consequently, the imperative arises for the development of an anomaly detection framework that can discern irregular electricity consumption patterns indicative of non-technical loss (NTL) activities. This framework becomes essential in combating the illicit diversion of electricity through SM.

In this concluding chapter, we encapsulate the pivotal findings gleaned from the preceding chapters. Additionally, we offer insights into intriguing avenues for future research, recognizing the persistent need to address the multifaceted challenges associated with electricity theft within the context of evolving smart grid technologies.

In this research, we introduced the SYNBDM and LUM algorithms for electricity theft detection in smart homes, utilizing fine-grained appliance consumption data to distinguish between normal and malicious usage. By employing uniform manifold approximation and projection (UMAP) to identify varied data distributions across different homes, our algorithm effectively leverages aggregated power consumption patterns for robust anomaly detection. Our tests on a real building appliance dataset demonstrated high performance, even with anonymized data, highlighting the algorithm's capability to balance effective theft detection with customer privacy. Though highly effective, we noted that unsupervised learning models require further refinement to better handle the complexities of real-world attack data. Our findings underline the potential of machine learning in enhanc-

ing energy security and stress the importance of incorporating appliance consumption patterns in electricity theft detection. This approach offers significant benefits to both consumers and energy providers, aiming for more efficient and secure energy management in smart homes.

Acknowledgements

I would like to express my heartfelt gratitude and give all glory to the Almighty God for His blessings and for being alive to complete my Ph.D. journey.

I am sincerely thankful to my supervisor, Professor Dr. Youki Kadobayashi, for his invaluable guidance, encouragement, and support, particularly in facilitating my collaboration with The University of Tokyo and attending the IEEE PES annual general meetings. These experiences significantly shaped my doctoral research journey. My appreciation extends to Associate Professor Hideya Ochiai from The University of Tokyo. His support, advice, and teachings were instrumental in my research. The discussions during our research meetings were not only enlightening for my work but also beneficial personally.

I am also grateful to Prof. Keiichi Yasumoto, Prof. Yuichi Hayashi, Associate Prof. Dr. Yuzo Taenaka, and Assistant Prof. Dr. Delwar Hossain for their valuable advice and insights. My gratitude goes to the secretaries and fellow laboratory members for their cooperation and support in both academic and daily life matters.

I am indebted to the administration of Nara Institute of Science and Technology for selecting me for the tuition fee exemption program, which greatly aided in achieving my doctoral ambitions. I also thank the staff of the International Student Affairs Section and the Graduate School of Information Science Office for their assistance with essential documentation and support.

In the pages of this dissertation, the silent yet profound support of my family—my wife and children—resonates deeply. Your unwavering support and ceaseless prayers have been the cornerstone of my strength and perseverance throughout my research. Your sacrifices have not gone unnoticed, and I am forever indebted to you for your love and encouragement.

I extend my heartfelt gratitude to Mr. Fati Odunayo Boluwaji, the commercial

manager at Eko Electricity Distribution Company Plc. His enduring encouragement and invaluable practical insights into the workings of electricity distribution significantly shaped the initial phase of this project. His guidance was instrumental in bridging the gap between theory and practical application, for which I am profoundly thankful.

To my family and friends, especially the FUTA'95 set, your camaraderie and support have been sources of joy and motivation. To my pastors and spiritual father, whose prayers have ceaselessly ascended on behalf of this endeavor, I am deeply moved by your faith and dedication. Your spiritual guidance has been a beacon of light on this journey.

I also express my sincere appreciation to the African students at NAIST, and my brethren across Japan and beyond, for their companionship and encouragement. Your solidarity has been a source of comfort and strength.

To all who have offered advice, guidance, and support, including those whose names I may not have mentioned but whose contributions have touched this work in countless ways, I am eternally grateful. Your collective wisdom and encouragement have been pivotal in the completion of this journey. Thank you for being part of this significant chapter of my life.

I extend my heartfelt gratitude to SHIMADZU CORPORATION for bestowing upon me the prestigious award for my presentation at the Doctoral Career Messe Kyoto 2023. This accolade is a testament to my dedication to the progression of scientific inquiry and underscores the value of synergy between academic scholarship and industrial expertise. I am profoundly motivated by this honor to persist in my endeavors for scholarly distinction in my domain. Additionally, my appreciation goes out to the NAIST Career Services Office for their impeccable coordination of the event, which played a pivotal role in this achievement.

Finally, I am profoundly grateful to the Industrial Cyber Security Center of Excellence (ICS-CoE) for the research assistantship that supported me throughout my Ph.D. research.

Bibliography

[1] XGBoost Documentation, 2022. https://xgboost.readthedocs.io/en/stable/ (accessed: 2022.12.23).

[2] O. A. Abraham, H. Ochiai, M. D. Hossain, Y. Taenaka, and Y. Kadobayashi. Electricity theft detection for smart homes with knowledge-based synthetic attack data. In *2023 IEEE 19th International Conference on Factory Communication Systems (WFCS)*, pages 1–8. IEEE, 2023.

[3] R. K. Ahir and B. Chakraborty. Pattern-based and context-aware electricity theft detection in smart grid. *Sustainable Energy, Grids and Networks*, 32:100833, 2022.

[4] A. ALDEGHEISHEM, M. ANWAR, N. JAVAID, N. ALRAJEH, M. SHAFIQ, and H. AHMED. Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks. *IEEE Access*, 2021.

[5] F. Aminifar, M. Abedini, T. Amraee, P. Jafarian, M. H. Samimi, and M. Shahidehpour. A review of power system protection and asset management with machine learning techniques. *Energy Systems*, pages 1–38, 2021.

[6] A. Arif, N. Javaid, A. Aldegheishem, and N. Alrajeh. Big data analytics for identifying electricity theft using machine learning approaches in microgrids for smart communities. *Concurrency and Computation: Practice and Experience*, 33(17):e6316, 2021.

[7] N. F. Avila, G. Figueroa, and C.-C. Chu. Ntl detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and

random undersampling boosting. *IEEE Transactions on Power Systems*, 33(6):7171–7180, 2018.

[8] M. Azaza and F. Wallin. Evaluation of classification methodologies and features selection from smart meter data. *Energy Procedia*, 142:2250–2256, 2017.

[9] V. Badrinath Krishna, R. K. Iyer, and W. H. Sanders. Arima-based modeling and validation of consumption readings in power grids. In *Critical Information Infrastructures Security: 10th International Conference, CRITIS 2015, Berlin, Germany, October 5-7, 2015, Revised Selected Papers 10*, pages 199–210. Springer, 2016.

[10] R. Bayindir, I. Colak, G. Fulli, and K. Demirtas. Smart grid technologies and applications. *Renewable and sustainable energy reviews*, 66:499–516, 2016.

[11] B. Bezawada, M. Bachani, J. Peterson, H. Shirazi, I. Ray, and I. Ray. Behavioral fingerprinting of iot devices. In *Proceedings of the 2018 workshop on attacks and solutions in hardware security*, pages 41–50, 2018.

[12] E. Bompard, A. Mazza, and L. Toma. Classical grid control: Frequency and voltage stability. *Converter-Based Dynamics and Control of Modern Power Systems*, pages 31–65, 2021.

[13] S. Borenstein and J. B. Bushnell. Do two electricity pricing wrongs make a right? cost recovery, externalities. Technical report, and efficiency. Working Paper 24756, National Bureau of Economic Research, 2018.

[14] J. W. Busby, K. Baker, M. D. Bazilian, A. Q. Gilbert, E. Grubert, V. Rai, J. D. Rhodes, S. Shidore, C. A. Smith, and M. E. Webber. Cascading risks: Understanding the 2021 winter blackout in texas. *Energy Research & Social Science*, 77:102106, 2021.

[15] S. Chen, E. Dobriban, and J. H. Lee. A group-theoretic framework for data augmentation. *The Journal of Machine Learning Research*, 21(1):9885–9955, 2020.

[16] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[17] Z. Chen, A. M. Amani, X. Yu, and M. Jalili. Control and optimisation of power grids using smart meter data: A review. *Sensors*, 23(4):2118, 2023.

[18] T.-H. Cheung and D.-Y. Yeung. Modals: Modality-agnostic automated data augmentation in the latent space. In *International Conference on Learning Representations*, 2020.

[19] Congres International des Reseaux Electriques de Distribution. Reduction of technical and non-technical losses in distribution networks. Online, 2017.

[20] C. Cuijpers and B.-J. Koops. Smart metering and privacy in europe: Lessons from the dutch case. *European data protection: Coming of age*, pages 269–293, 2013.

[21] T. Dayaratne, C. Rudolph, A. Liebman, and M. Salehi. We can pay less: Coordinated false data injection attack against residential demand response in smart grids. In *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, pages 41–52, 2021.

[22] F. de Souza Savian, J. C. M. Siluk, T. B. Garlet, F. M. do Nascimento, J. R. Pinheiro, and Z. Vale. Non-technical losses: A systematic contemporary article review. *Renewable and Sustainable Energy Reviews*, 147:111205, 2021.

[23] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy policy*, 39(2):1007–1015, 2011.

[24] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi. Smart meters for power grid—challenges, issues, advantages and status. In *2011 IEEE/PES Power Systems Conference and Exposition*, pages 1–7. IEEE, 2011.

[25] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and P. Nelapati. A hybrid neural network model and encoding technique for enhanced classification of

energy consumption data. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–8. IEEE, 2011.

[26] A. C. Duman, Ö. Gönül, H. S. Erden, and Ö. Güler. Survey-and simulation-based analysis of residential demand response: Appliance use behavior, electricity tariffs, home energy management systems. *Sustainable Cities and Society*, 96:104628, 2023.

[27] E. Esenogho, K. Djouani, and A. M. Kurien. Integrating artificial intelligence internet of things and 5g for next-generation smartgrid: A survey of trends challenges and prospect. *IEEE Access*, 10:4794–4831, 2022.

[28] E. R. Frederiks, K. Stenner, and E. V. Hobman. Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews*, 41:1385–1394, 2015.

[29] N. Group. $96 billion is lost every year to electricity theft. Online, May 2017.

[30] V. Ç. Güngör and G. P. Hancke. *Industrial wireless sensor networks: Applications, protocols, and standards*. Crc Press, 2013.

[31] S. K. Gunturi and D. Sarkar. Ensemble machine learning models for the detection of energy theft. *Electric Power Systems Research*, 192:106904, 2021.

[32] C. Herder, M.-D. Yu, F. Koushanfar, and S. Devadas. Physical unclonable functions and applications: A tutorial. *Proceedings of the IEEE*, 102(8):1126–1141, 2014.

[33] E. G. Hertwich, S. Ali, L. Ciacci, T. Fishman, N. Heeren, E. Masanet, F. N. Asghari, E. Olivetti, S. Pauliuk, Q. Tu, et al. Material efficiency strategies to reducing greenhouse gas emissions associated with buildings, vehicles, and electronics—a review. *Environmental Research Letters*, 14(4):043004, 2019.

[34] D. Hock. *Detecting Energy Theft and Anomalous Power Usage in Smart Meter Data*. PhD thesis, University of Plymouth, 2020.

[35] M. D. Hossain, H. Inoue, H. Ochiai, D. Fall, and Y. Kadobayashi. Lstm-based intrusion detection system for in-vehicle can bus communications. *IEEE Access*, 8:185489–185502, 2020.

[36] Y. Huang and Q. Xu. Electricity theft detection based on stacked sparse denoising autoencoder. *International Journal of Electrical Power & Energy Systems*, 125:106448, 2021.

[37] S. Hussain, M. W. Mustafa, T. A. Jumani, S. K. Baloch, H. Alotaibi, I. Khan, and A. Khan. A novel feature engineered-catboost-based supervised machine learning framework for electricity theft detection. *Energy Reports*, 7:4425–4436, 2021.

[38] Y. Ishimaki, S. Bhattacharjee, H. Yamana, and S. K. Das. Towards privacy-preserving anomaly-based attack detection against data falsification in smart grid. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6. IEEE, 2020.

[39] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen. Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology*, 19(2):105–120, 2014.

[40] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra. Decision tree and svm-based data analytics for theft detection in smart grid. *IEEE Transactions on Industrial Informatics*, 12(3):1005–1016, 2016.

[41] P. Jokar, N. Arianpoo, and V. C. Leung. Electricity theft detection in ami using customers' consumption patterns. *IEEE Transactions on Smart Grid*, 7(1):216–226, 2015.

[42] S. Joshi, R. Li, S. Bhattacharjee, S. K. Das, and H. Yamana. Privacy-preserving data falsification detection in smart grids using elliptic curve cryptography and homomorphic encryption. In *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 229–234. IEEE, 2022.

[43] N. Kambule, K. Yessoufou, and N. Nwulu. Formulating best practice recommendations for prepaid electricity meter deployment in soweto, south africa–capitalising on the developed-world's experiences. *Journal of Public Affairs*, 22(4):e2646, 2022.

[44] H. M. Khan, A. Khan, F. Jabeen, A. Anjum, and G. Jeon. Fog-enabled secure multiparty computation based aggregation scheme in smart grid. *Computers & Electrical Engineering*, 94:107358, 2021.

[45] N. A. Kipreos. *Nonintrusive load identification & monitoring: techniques and applications for smart meters*. Pontificia Universidad Catolica de Chile (Chile), 2011.

[46] V. B. Krishna, K. Lee, G. A. Weaver, R. K. Iyer, and W. H. Sanders. F-deta: A framework for detecting electricity theft attacks in smart grids. In *2016 46th Annual IEEE/IFIP international conference on dependable systems and networks (DSN)*, pages 407–418. IEEE, 2016.

[47] A. Kusiak. Smart manufacturing must embrace big data. *Nature*, 544(7648):23–25, 2017.

[48] R. Li, S. Bhattacharjee, S. K. Das, and H. Yamana. Look-up table based fhe system for privacy preserving anomaly detection in smart grids. In *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 108–115. IEEE, 2022.

[49] R. Li, Y. Ishimaki, and H. Yamana. Fully homomorphic encryption with table lookup for privacy-preserving smart grid. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 19–24. IEEE, 2019.

[50] L. Liu, Y. Peng, S. Wang, M. Liu, and Z. Huang. Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. *Information Sciences*, 340:41–57, 2016.

[51] A. Llaria, J. Dos Santos, G. Terrasson, Z. Boussaada, C. Merlo, and O. Curea. Intelligent buildings in smart grids: A survey on security and privacy issues related to energy management. *Energies*, 14(9):2733, 2021.

[52] J. Lloret, J. Tomas, A. Canovas, and L. Parra. An integrated iot architecture for smart metering. *IEEE Communications Magazine*, 54(12):50–57, 2016.

[53] A. L'Heureux, K. Grolinger, and M. A. Capretz. Transformer-based model for electrical load forecasting. *Energies*, 15(14):4993, 2022.

[54] R. Maes, V. van der Leest, E. van der Sluis, and F. Willems. Secure key generation from biased pufs: extended version. *Journal of Cryptographic Engineering*, 6:121–137, 2016.

[55] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich. Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014. *Scientific data*, 3(1):1–12, 2016.

[56] J. F. Martins, A. G. Pronto, V. Delgado-Gomes, and M. Sanduleac. Smart meters and advanced metering infrastructure. In *Pathways to a smarter power system*, pages 89–114. Elsevier, 2019.

[57] P. McDaniel and S. McLaughlin. Security and privacy challenges in the smart grid. *IEEE security & privacy*, 7(3):75–77, 2009.

[58] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[59] G. M. Messinis and N. D. Hatziargyriou. Unsupervised classification for non-technical loss detection. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2018.

[60] G. Micheli, E. Soda, M. T. Vespucci, M. Gobbi, and A. Bertani. Big data analytics: an aid to detection of non-technical losses in power utilities. *Computational Management Science*, 16(1):329–343, 2019.

[61] P. H. Mirzaee, M. Shojafar, H. Cruickshank, and R. Tafazolli. Smart grid security and privacy: From conventional to machine learning issues (threats and countermeasures). *IEEE access*, 10:52922–52954, 2022.

[62] P. H. Mirzaee, M. Shojafar, Z. Pooranian, P. Asef, H. Cruickshank, and R. Tafazolli. FIDS : A Federated Intrusion Detection System for 5G Smart Metering Network.

[63] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar. A survey on advanced metering infrastructure. *International Journal of Electrical Power & Energy Systems*, 63:473–484, 2014.

[64] A. Moradzadeh, O. Sadeghian, K. Pourhossein, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam. Improving residential load disaggregation for sustainable development of energy via principal component analysis. *Sustainability*, 12:3158, 2020.

[65] M. Nabil, M. Ismail, M. Mahmoud, M. Shahin, K. Qaraqe, and E. Serpedin. Deep learning-based detection of electricity theft cyber-attacks in smart grid ami networks. In *Deep learning applications for cyber security*, pages 73–102. Springer, 2019.

[66] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE transactions on Power Delivery*, 25(2):1162–1171, 2009.

[67] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami. Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, pages 506–519, 2017.

[68] C. H. Park and T. Kim. Energy theft detection in advanced metering infrastructure based on anomaly pattern detection. *Energies*, 13(15):3832, 2020.

[69] R. Punmiya and S. Choe. Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Transactions on Smart Grid*, 10(2):2326–2329, 2019.

[70] R. Razavi and M. Fleury. Socio-economic predictors of electricity theft in developing countries: An indian case study. *Energy for Sustainable Development*, 49:1–10, 2019.

[71] M. H. Rehmani, M. Reisslein, A. Rachedi, M. Erol-Kantarci, and M. Radenkovic. Integrating renewable energy resources into the smart grid: Recent developments in information and communication technologies. *IEEE Transactions on Industrial Informatics*, 14(7):2814–2825, 2018.

[72] K. Ren, T. Zheng, Z. Qin, and X. Liu. Adversarial attacks and defenses in deep learning. *Engineering*, 6(3):346–360, 2020.

[73] J. Rodrigues. Outliers make us go mad: Univariate outlier detection. http://tinyurl.com/2vnz7nvb, 2018. [Online; accessed 20-July-2023].

[74] M. S. Saeed, M. W. Mustafa, N. N. Hamadneh, N. A. Alshammari, U. U. Sheikh, T. A. Jumani, S. B. A. Khalid, and I. Khan. Detection of non-technical losses in power utilities—a comprehensive systematic review. *Energies*, 13(18):4727, 2020.

[75] P. Schavemaker and L. Van der Sluis. *Electrical power system essentials*. John Wiley & Sons, 2017.

[76] O. Siddiqui. The green grid: Energy savings and carbon emissions reductions enabled by a smart grid. Online, 2008.

[77] S. Singh and A. Yassine. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, 11:452, 2018.

[78] S. K. Singh, K. Khanna, R. Bose, B. K. Panigrahi, and A. Joshi. Joint-transformation-based detection of false data injection attacks in smart grid. *IEEE Transactions on Industrial Informatics*, 14(1):89–97, 2017.

[79] G. Strbac, N. Hatziargyriou, J. P. Lopes, C. Moreira, A. Dimeas, and D. Papadaskalopoulos. Microgrids: Enhancing the resilience of the european megagrid. *IEEE Power and Energy Magazine*, 13(3):35–43, 2015.

[80] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin. Variational auto-encoder-based detection of electricity stealth cyber-attacks in ami networks. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1590–1594. IEEE, 2021.

[81] A. Z. Tan, H. Yu, L. Cui, and Q. Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[82] H. Tatsat, S. Puri, and B. Lookabaugh. *Machine Learning and Data Science Blueprints for Finance*. O'Reilly Media, 2020.

[83] Tenaga Nasional Berhad. Tenaga nasional berhad annual report 2006. TNB, 2006.

[84] M. Toshpulatov and N. Zincir-Heywood. Anomaly detection on smart meters using hierarchical self organizing maps. In *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–6. IEEE, 2021.

[85] U.S. Department of Energy. The netl modern grid strategy powering our 21st-century economy: Advanced metering infrastructure. National Energy Technology Laboratory, February 2008.

[86] U.S. Department of Energy. Communications requirements of smart grid technologies. National Energy Technology Laboratory, October 2010.

[87] E. Villar-Rodriguez, J. Del Ser, I. Oregi, M. N. Bilbao, and S. Gil-Lopez. Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis. *Energy*, 137:118–128, 2017.

[88] H. Wang and W. Yang. An iterative load disaggregation approach based on appliance consumption pattern. *Applied Sciences*, 8:542, 2018.

[89] J. Wang. Deep learning on smart meter data: Non-intrusive load monitoring and stealthy black-box attacks. 2020.

[90] W. Wang and Z. Lu. Cyber security in the smart grid: Survey and challenges. *Computer networks*, 57(5):1344–1371, 2013.

[91] S. Wilhelm and J. Kasbauer. Exploiting smart meter power consumption measurements for human activity recognition (har) with a motif-detection-based non-intrusive load monitoring (nilm) approach. *Sensors*, 21:8036, 2021.

[92] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE communications surveys & tutorials*, 15(1):5–20, 2012.

[93] S. C. Yip. *Anomaly detection frameworks for identifying energy theft and meter irregularities in smart grids/Yip Sook Chin*. PhD thesis, Universiti Malaya, 2019.

[94] J. Zhang, X. Chen, M. Ni, T. Wang, J. Luo, et al. A security scheme for intelligent substation communications considering real-time performance. *Journal of Modern Power Systems and Clean Energy*, 7(4):948–961, 2019.

[95] X. Zhang, T. Kato, and T. Matsuyama. Learning a context-aware personal model of appliance usage patterns in smart home. *2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, 2014.

[96] C. Zhao, D. Wu, J. Huang, Y. Yuan, H.-T. Zhang, R. Peng, and Z. Shi. Boosttree and boostforest for ensemble learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[97] J. Zheng, D. W. Gao, and L. Lin. Smart meters in smart grid: An overview. In *2013 IEEE green technologies conference (GreenTech)*, pages 57–64. IEEE, 2013.

[98] S. Zhou and M. A. Brown. Smart meter deployment in europe: A comparative case study on the impacts of national policy schemes. *Journal of cleaner production*, 144:22–32, 2017.

[99] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012.

Publication List

Journals

1. Olufemi, Abiodun Abraham, Hideya Ochiai, Md Delwar Hossain, Yuzo Taenaka and Youki Kadobayashi. “Electricity Theft Detection for Smart Homes: Harnessing the Power of Machine Learning with Real and Synthetic Attacks.” IEEE Access, 12 (2024): 240204-000239.

International Conferences

1. Olufemi, Abiodun Abraham, Hideya Ochiai, Md Delwar Hossain, Yuzo Taenaka, Youki Kadobayashi, "Electricity Theft Detection for Smart Homes with Knowledge-Based Synthetic Attack Data", In 2023 IEEE PES General Meeting Student Poster Competition, July 2023.

2. Olufemi, A. Abraham, Hideya Ochiai, Md Delwar Hossain, Yuzo Taenaka, Youki Kadobayashi, "Electricity Theft Detection for Smart Homes with Knowledge-Based Synthetic Attack Data", In Proceedings of 19th IEEE International Conference on Factory Communication Systems (WFCS 2023), April 2023.

3. Olufemi, A. Abraham, Hideya Ochiai, Kabid Hassan Shibly, Md Delwar Hossain, Yuzo Taenaka, Youki Kadobayashi, "Unauthorized Power Usage Detection in Disaggregated Smart Meter Home Network", In IEEE Future Networks World Forum (FNWF'2022), pp. 688-693, October 2022.