

# Load Disaggregation: Towards Energy Efficient Systems

Attique Ur Rehman

A thesis submitted to Auckland University of Technology in fulfilment of the  
requirements for the degree of Doctor of Philosophy (PhD)

2020

School of Engineering, Computer, and Mathematical Sciences  
Auckland University of Technology  
New Zealand

## Abstract

Electricity is one of the valuables and widely used forms of energy. However, with the fast-paced technological development in the electrical and electronics market, the electricity demand is on a constant rise. To tackle energy and sustainability issues, two paths can be followed by the world community. Either new generation plants to be established with an expense of millions of dollars or to explore the existing system by integrating innovative techniques that can lead to energy efficiency and conservation. The latter one is a more viable solution, and the research community is extensively working to propose and develop innovative techniques towards energy efficiency and conservation. In this context, energy monitoring is one of the key techniques that play a significant role in the field of sustainable energy.

Load disaggregation is one of these promising energy monitoring techniques, where a non-intrusive<sup>1</sup> load disaggregation technique commonly referred to as non-intrusive load monitoring (NILM) is widely adopted to provide individual load profiles to the stakeholders. Appliance-level energy monitoring is not only beneficial for the consumers in terms of having valuable information regarding the operation status of their loads and corresponding consumption but also benefit the system operators, policymakers, and manufacturers in terms of analyzing the network's energy flow, creating policies/tariffs, and manufacturing of smart appliances, respectively.

This research work contributes to the existing research and development of non-invasive load disaggregation systems, by proposing and developing a robust event-based non-intrusive load disaggregation approach for low sampling data granularity. As a way forward, this research work contributes to different aspects of a NILM system. For NILM event detection, three new low complexity and computationally fast algorithms based on statistical parameters are proposed and validated on real-world datasets. In terms of electrical load features, a set of nine distinct load features based on statistical, geometrical, and power features is proposed. The extracted load features are further investigated in terms of significance using different feature selection methodologies and the extracted results are validated in the context of classification performance. For load classification, this research work investigated different supervised machine learning models<sup>2</sup> towards an optimal learning model for the given conditions. In addition to

---

<sup>1</sup> The terminologies of non-invasive and non-intrusive are used interchangeably in this thesis.

<sup>2</sup> In the context of machine learning, the terminologies of model, technique, classifier, and algorithm are used interchangeably in this thesis.

standalone machine learning models, this research also presents a combinatorial learning model, i.e., ensemble learning, for load classification in the context of the NILM. Further, a comprehensive comparative evaluation of these techniques is also part of this thesis.

The entire digital simulations and corresponding analysis presented in this research work are based on real-world electricity datasets, originating from different geographical regions, i.e., New Zealand and the United States of America. Based on the low data granularity of the employed databases, three different appliances/circuits, i.e., air conditioning unit, electric vehicle charging, and water heating are successfully disaggregated using the proposed non-intrusive load disaggregation approach. Moreover, a proof of concept in terms of real-world deployment, i.e., the application of the proposed non-intrusive load disaggregation, is also proposed and validated in this research work.

Due to low data granularity nature, this research work is more relevant for the existing metering infrastructure. Therefore, the proposed methodologies and corresponding simulation studies presented in this research work will significantly contribute to the existing state of the art on low sampling NILM systems particularly in terms of event detection, electrical load features, and learning model selection. The study presented in this thesis will also facilitate future research in terms of real-world deployment of NILM systems and its broader applications. Concisely, this research work based on a non-invasive load disaggregation approach is a way forward for energy efficient systems.

## Contents

List of Figures .....	7
List of Tables.....	9
Abbreviations and Symbols .....	11
Declaration .....	15
Acknowledgment .....	16
Chapter 1    Introduction .....	17
1.1    Rationale of the Study .....	17
1.2    Scope and Significance .....	18
1.3    Contributions .....	19
Chapter 2    Research Design.....	22
2.1    Literature Review .....	22
2.1.1    Energy Monitoring .....	22
2.1.2    Intrusive Load Monitoring and Smart Appliances .....	23
2.1.3    Non-Intrusive Load Monitoring .....	24
2.1.3.1    Data Acquisition.....	25
2.1.3.2    Load Features .....	25
2.1.3.3    Classification.....	28
2.2    Research Gap Analysis.....	29
2.2.1    Working Principle .....	30
2.2.2    Data Granularity .....	31
2.2.3    Research Objectives .....	33
2.3    Evaluation and Approval.....	33
2.4    Proposed Methodology.....	34
2.4.1    Load Databases .....	35
2.4.2    Performance Evaluation Metrics.....	37
2.5    Concluding Remarks .....	39
Chapter 3    Data Pre-processing and Event Detection.....	40
3.1    Data Pre-processing.....	40
3.2    Event Detection .....	42
3.2.1    Proposed Event Detection Algorithms.....	43
3.2.1.1    MSW Algorithm .....	44
3.2.1.2    VSW and MAD-SW Algorithms .....	45
3.2.2    Event Detection Simulations and Evaluation.....	46
3.2.2.1    Dataport.....	46

3.2.2.1.1	MSW Algorithm.....	47
3.2.2.1.2	VSW Algorithm .....	50
3.2.2.1.3	MAD-SW Algorithm.....	53
3.2.2.1.4	Inter-Algorithm Comparison.....	56
3.2.2.2	NZ GREEN Grid.....	58
3.3	Concluding Remarks .....	62
Chapter 4	Feature Engineering and Load Classification .....	63
4.1	Introduction .....	63
4.1.1	Feature Engineering .....	63
4.1.2	Load Classification.....	64
4.1.2.1	Support Vector Machine .....	65
4.1.2.2	Logistic Regression.....	66
4.1.2.3	Decision Trees.....	66
4.1.2.4	Random Forest .....	66
4.1.2.5	k-Nearest Neighbors.....	66
4.1.2.6	Naïve Bayes .....	66
4.1.2.7	Gaussian Process .....	66
4.1.2.8	Multi-Layer Perceptron.....	67
4.2	Research Methodologies .....	67
4.2.1	Feature Extraction and Reduction.....	68
4.2.2	Feature Selection.....	69
4.2.3	Learning Models .....	70
4.2.4	Learning Models Evaluation.....	71
4.3	Simulations and Results .....	72
4.3.1	Dataport.....	73
4.3.2	NZ GREEN Grid.....	78
4.3.2.1	Classifiers and Feature Space Evaluation .....	81
4.3.2.1.1	Reduced Feature Space Simulations .....	88
4.3.2.1.2	Comparative Analysis .....	91
4.3.2.2	Feature Selection Simulations.....	92
4.3.2.3	Ensemble Learning Approach.....	98
4.4	Concluding Remarks .....	101
Chapter 5	Non-Intrusive Load Disaggregation Applications .....	103
5.1	State of the Art on Applications .....	103
5.2	Proof of Concept .....	104
5.2.1	Demand Response Program .....	106

5.2.2	Non-Intrusive Load Disaggregation Assisted Demand Response .....	107
5.2.2.1	Rationale and Significance.....	107
5.2.2.2	Problem Formulation and Methodology .....	108
5.2.2.3	Case Study and Validation .....	110
5.3	Miscellaneous Applications .....	113
5.4	Concluding Remarks .....	113
Chapter 6	Conclusion and Future Work .....	115
6.1	Synopsis and Conclusion.....	115
6.1.1	Event Detection .....	115
6.1.2	Feature Engineering .....	115
6.1.3	Classification.....	116
6.1.4	Applications .....	117
6.2	Potential Challenges and Future Research Scope .....	118
References	.....	120
Appendix	.....	135

## List of Figures

Figure 1	Rationale of the Study .....	18
Figure 2	PhD Research Study Flow .....	22
Figure 3	Energy Monitoring Techniques .....	23
Figure 4	Traditional NILM Framework .....	25
Figure 5	Relationship between Data Acquisition and Features in NILM Domain .....	27
Figure 6	Machine Learning Models Framework .....	29
Figure 7	Proposed Event-Based Non-Invasive Load Disaggregation Approach .....	35
Figure 8	Geographic Regions of Employed Load Databases .....	35
Figure 9	Impulse and Ripple Phenomenon .....	41
Figure 10	Outcome of Median Filtering Technique .....	42
Figure 11	Graphical Depiction of Events .....	42
Figure 12	Algorithms' Working Principle Based on Sliding Window Concept .....	44
Figure 13	Working Principle of VSW and MAD-SW .....	45
Figure 14	Event Detection Simulation Flow .....	46
Figure 15	Event Detection Results by MSW Algorithm .....	47
Figure 16	MSW Algorithm Results Validation .....	48
Figure 17	MSW Algorithm Performance Trend for Different Window Width ' $\omega$ ' .....	49
Figure 18	VSW Algorithm Event Detection Results .....	50
Figure 19	VSW Performance Trend for Different Window Width ' $\omega$ ' .....	53
Figure 20	MAD-SW Algorithm Event Detection Results .....	54
Figure 21	MAD-SW Performance Trend at Different Window Width ' $\omega$ ' .....	56
Figure 22	Inter-Algorithm Comparison in terms of F-Score .....	57
Figure 23	MAD-SW Sensitivity Analysis in terms of $\omega$ for NZ GREEN Grid .....	59
Figure 24	Delay Tolerance Phenomenon .....	60
Figure 25	MAD-SW Sensitivity Analysis in terms of $\Delta\tau$ for NZ GREEN Grid .....	61
Figure 26	Feature Engineering Semantic Network .....	65
Figure 27	Working Principle of Cross-Validation Approaches .....	72
Figure 28	Generalized Employed Simulation Flow .....	72
Figure 29	Optimal k-value Selection for Testing Data ID. 26 .....	76
Figure 30	Optimal k-value Selection for Testing Data ID. 3036 .....	77
Figure 31	Comparison of Ground-truth and Predicted Load Element Status .....	79
Figure 32	Non-Intrusive Load Disaggregation Simulation Flow .....	81
Figure 33	Circuit-Level Inference Results for Household ID. rf_01 .....	84

Figure 34 Circuit-Level Inference Results for Household ID. rf_42 .....	85
Figure 35 Overall Classifiers' Performance in Combination with $\mathcal{F}$ .....	87
Figure 36 Overall Classifiers' Performance in Combination with $\mathcal{G}$ .....	91
Figure 37 Feature Space Dimensionality vs. Classification Performance .....	92
Figure 38 Feature Selection Based Non-Intrusive Load Inference Approach .....	93
Figure 39 Voting Classifier Framework .....	99
Figure 40 Intra-Ensemble Model Classification Performance .....	101
Figure 41 Categorization of NILM Applications .....	104
Figure 42 Demand Response Categories .....	106
Figure 43 Proposed Application Framework .....	109
Figure 44 Information Flow Between Consumer and Utility .....	110
Figure 45 Water Heating Circuit Operation Authentication .....	111



## List of Tables

Table 1	Comparison of Energy Disaggregation Datasets .....	26
Table 2	Recent Research on Features in NILM Domain .....	28
Table 3	Comparison between Event and Non-Event Based NILM Systems.....	31
Table 4	Confusion Matrix .....	38
Table 5	Kappa Index Range and Corresponding Labels .....	39
Table 6	MSW Algorithm Description .....	44
Table 7	Event Detection Simulation Attributes for Dataport .....	47
Table 8	MSW Algorithm Performance Results .....	48
Table 9	MSW Sensitivity Study Results.....	49
Table 10	Comparison of Detected Events by VSW and Ground-truth Events .....	51
Table 11	VSW Algorithm Performance Results.....	52
Table 12	VSW Sensitivity Study Results .....	52
Table 13	Comparison of Detected Events by MAD-SW and Ground-truth Events ...	55
Table 14	MAD-SW Algorithm Performance Results .....	55
Table 15	MAD-SW Sensitivity Study Results.....	56
Table 16	Inter-Algorithm Comparison .....	57
Table 17	NZ GREEN Grid Event Detection Simulation Attributes .....	58
Table 18	MAD-SW Sensitivity Analysis in terms of $\omega$ for NZ GREEN Grid .....	59
Table 19	MAD-SW Sensitivity Analysis in terms of $\Delta\tau$ for NZ GREEN Grid.....	61
Table 20	Classifiers Strengths and Weaknesses .....	67
Table 21	Event Detection Simulation Parameters .....	73
Table 22	Event Detection Results for Different Households of Dataport .....	73
Table 23	Extracted Load Features ' $\mathcal{F}$ ' for Dataport ID 26.....	74
Table 24	k-NN Evaluation Based on Cross-Validation Approach .....	75
Table 25	Appliance-Level Classification Results of Dataport .....	75
Table 26	Optimal k-value Based Inference Results Validation.....	77
Table 27	NZ GREEN Grid Household Attributes and Event Detection Results .....	81
Table 28	Circuit-level Classifiers' Performance in Combination with $\mathcal{F}$ .....	83
Table 29	Household-level Classifiers' Performance in Combination with $\mathcal{F}$ .....	86
Table 30	Circuit-level Classifiers' Performance in Combination with $\mathcal{F}$ .....	89
Table 31	Household-level Classifiers' Performance in Combination with $\mathcal{F}$ .....	90
Table 32	Extracted Load Features ' $\mathcal{F}_{ex\bar{T}}$ ' .....	94
Table 33	Features' Significance Score Based on $f\_classif$ Assessment .....	94

Table 34	Features' Significance Score Based on mutual_info_classif Assessment ...	94
Table 35	Features' Significance Score Based on Feature Importance Assessment....	95
Table 36	Load Features' Significance Ranking for all Testing Households .....	95
Table 37	Feature Selection Simulation Data .....	96
Table 38	Classifiers' Accuracy Performance Based on Feature Selection Methods..	97
Table 39	Ensemble Model Performance Comparison .....	100
Table 40	State of the art on Non-Intrusive Load Disaggregation Applications .....	105
Table 41	Water Heating Operation Inference Results .....	112
Table 42	Employed Learning Models' Parameters .....	135
Table 43	MSW Sensitivity Analysis in terms of $\Delta\tau$ for Dataport.....	136

## Abbreviations and Symbols

AC	Air Conditioning Unit
ADR	Automated Demand Response
AFHMM	Additive Factorial Hidden Markov Model
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
AMPds	Almanac of Minutely Power Datasets
ANOVA	Analysis of Variance
AUT	Auckland University of Technology
BLUED	Building Level Fully Labelled Dataset
COMBED	Commercial Building Energy Dataset
CT	Current Transformer
DOI	Digital Object Identifier
DR	Demand Response
DRED	Dutch Residential Energy Dataset
DRES	Distributed Renewable Energy Sources
DSM	Demand Side Management
DT	Decision Trees
ECO	Electricity Consumption & Occupancy
EMS	Energy Management System
EV	Electric Vehicle
FHMM	Factorial Hidden Markov Model
FN	False Negative
FP	False Positive
GP	Gaussian Process
GREEND	Greend Electrical Energy Dataset
HEMS	Home Energy Management System

HMM	Hidden Markov Model
HVAC	Heating Ventilation and Air Conditioning
I	Current
iAWE	Indian dataset for Ambient Water and Energy sensing
ICT	Information and Communication Technology
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
ILM	Intrusive Load Monitoring
ISGT	Innovative Smart Grid Technologies
k-NN	k-Nearest Neighbors
kW	Kilowatt
LR	Logistic Regression
LSTM-RNN	Long Short-Term Memory Recurrent Neural Network
MAD	Mean Absolute Deviation
MAD-SW	Mean Absolute Deviation Sliding Window
Misc.	Miscellaneous
Misc.ON	Miscellaneous Circuit Turn-on
Misc.OFF	Miscellaneous Circuit Turn-off
MLP	Multi-Layer Perceptron
MSW	Mean Sliding Window
NALM	Non-intrusive Appliance Load Monitoring
NB	Naïve Bayes
NIALM	Non-Intrusive Appliance Load Monitoring
NILM	Non-Intrusive Load Monitoring
NZ	New Zealand
P	Active Power
$P_{agg}(t)$	Aggregated Power Load at time $t$

PGR	Post Graduate Report
PhD	Doctor of Philosophy
$P_i(t)$	$i^{th}$ Appliance Power Load at time $t$
PLAID	Plug Load Appliance Identification Dataset
PRS	Personalized Recommendation System
PV	Photovoltaic
Q	Reactive Power
R&D	Research & Development
REDD	Reference Energy Disaggregation Dataset
REFIT	Personalised Retrofit Decision Support Tools for UK Homes Using Smart Home Technology
RF	Random Forest
RMS	Root Mean Square
SG	Smart Grid
sklearn	Scikit-learn
sqrt	Square Root
SustDataED	Sustainability Dataset – Energy Disaggregation
SVM	Support Vector Machine
TIM	Transactions on Instrumentation and Measurement
TP	True Positive
UDR	User-assisted Demand Response
UK-DALE	United Kingdom Domestic Appliance Level Electricity
USA	United States of America
V	Voltage
VSW	Variance Sliding Window
W	Watt
WH	Water Heating

$WH_{OFF}$	Water Heating Circuit Turn-off
$WH_{ON}$	Water Heating Circuit Turn-on
WT	Wind Turbine
$\delta$	Threshold Value
$\mu$	Mean
$\sigma$	Standard Deviation
$\omega$	Sliding Window Width
$\mathfrak{M}$	Median Absolute Deviation
$F_{\text{ext.}}$	Extended Feature set
$F_{FS}$	Feature Selection based Feature Set
$C_{Disp.}$	Coefficient of Dispersion
$C_{Var.}$	Coefficient of Variation
$P_{p2p}$	Peak to Peak Power
$\mathcal{S}_{\mathcal{E}}$	Event Slope
$\sigma^2$	Variance
$\tau_{width}$	Event Width
$\Delta\tau$	Delay Tolerance

## **Declaration**

I, Attique Ur Rehman, hereby declare that the contents of this research work are original, except where specific reference is made to the work of others. I also declare that to the best of my knowledge and belief, this work has not been submitted in whole or in part for consideration for any degree or qualification in this or any other university or institute of higher learning. This research thesis is my work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgment.

A handwritten signature in blue ink, appearing to read 'Attique Ur Rehman', with a stylized flourish at the end.

Attique Ur Rehman  
September 2020

## Acknowledgment

*In the name of ALLAH, the Most Merciful, the Most Beneficent.*

Looking back at the three years journey of my Doctor of Philosophy (PhD) studies, I am positive to conclude that during the entire time I was blessed with some of the fine people from whom I learn a lot and it would be unfair not to acknowledge their contributions.

Indeed, I would like to start with my supervisory team, by expressing my sincere gratitude to Prof. Tek Tjing Lie and Dr. Brice Vallès. I have no words that can comprehend their boundless contributions, because at every single aspect of my PhD life, whether its technical, social, financial, or personnel matters, Prof. Lie and Dr. Vallès are of great support and help to me. Hence, my utmost gratitude to both for their valuable time, support, guidance, and supervision. Further, I would also like to acknowledge and thank Dr. Shafiqur Rahman Tito, who always motivated me to aim higher and guided me with all his sincerity.

I also acknowledge the role of Callaghan Innovation and would like to express my gratitude for providing the research and development (R&D) student fellowship grant. The R&D fellowship by Callaghan Innovation not only facilitated me financially but also provided an excellent opportunity to work with one of the well-reputed energy companies of New Zealand, namely Genesis Energy Limited. I am also thankful to Genesis Energy Limited because collaborating and working there, not only provided me an insight into the corporate world but also facilitated me in my professional grooming by a dedicated mentorship program.

Last but not the least, I am thankful to my parents, siblings, and extended family for their sincere love, support, and prayers. Particularly, I am indebted to the unconditional love, support, and prayers of my most beloved wife, who is taking care of heaps of responsibilities back home for the last few years, while I am studying abroad. I acknowledge that this is the support and encouragement of my wife, which makes it possible for me to successfully pursue and complete my PhD studies abroad. And how can I forget my two dearest daughters, to whom I am thankful not only for their cute prayers but also for their care and love that acts as a driving force to complete my PhD within three years.



# Chapter 1 Introduction

This chapter presents a brief overview of the rationale of this research work along with the corresponding scope and significance. A detailed overview of the contribution is also presented in this chapter.

## 1.1 Rationale of the Study

The world has witnessed a tremendous transformation in the last few decades regarding every aspect of human life. Electricity which plays a key role in modern life has also gone through rapid changes in terms of technologies. Unlike a few decades ago, today's electricity flow is neither unidirectional nor centrally produced in large power plants. With the growing emergence of prosumers<sup>3</sup> and microgrids, the amount of electricity not produced by large traditional power plants is ever increasing. Moreover, the renewable generation, commonly known as distributed renewable energy sources (DRES), including but not limited to photovoltaic (PV), wind turbine (WT), geothermal, and biomass have also seen solid growth in the recent years. In this context, according to a recent report of the International Energy Agency, renewables will account for over 40% of total generation by the year 2040<sup>4</sup>. On the other side, some of the issues of renewables like PV and WT lie in their intermittent nature and being widely spread over large areas. It is then critical to be able to effectively manage the energy flow together with ensuring grid sustainability, i.e., to maintain an equilibrium between demand and supply. To address the challenges of DRES and maintain the said equilibrium, a concept of a smart grid (SG) system comes into existence. It is an updated electrical grid system based on digital or analog information and communication systems. Unlike the traditional grid system, a smart grid is a bi-directional energy flow system which collects and acts on the information regarding the behavior of the suppliers and the consumers in an automated way to increase the efficiency, reliability, and sustainability of the production and distribution of electricity [1].

One of the promising aspects of the smart grid is the transformation of consumers from passive<sup>5</sup> to active<sup>6</sup> consumers. Now the consumers can play a key role in improving the overall efficiency of the system [2], however, it is not realistic to expect consumers to play an effective role towards a sustainable system until and unless they are provided with the direct feedback; refers to real-time appliance-level consumption information [3]. In

---

<sup>3</sup> Customers who generate their own electricity, and can also utilise electric vehicle, battery storage etc.

<sup>4</sup> <https://www.iea.org/reports/world-energy-outlook-2018/renewables#abstract>

<sup>5</sup> Passive consumers do not actively participate in grid operations; just plug in their appliances and billed accordingly.

<sup>6</sup> Active consumers can monitor and control their electricity as per the market trends.

terms of direct feedback, consumers can not only effectively monitor their electricity usage but also significantly contribute to saving energy [3, 4]. In the said perspective regarding energy saving and the successful deployment of smart grid systems, effective energy monitoring is inevitable. Effective monitoring of energy usage not only contributes to grid stability but also leads to many promising applications in the context of energy efficiency and conservation. Figure 1 depicts the overall rationale of this study in the context of energy deficiency dilemma and its possible solutions (shaded by dark colors) and the factors (shaded by light colors) associated with it.

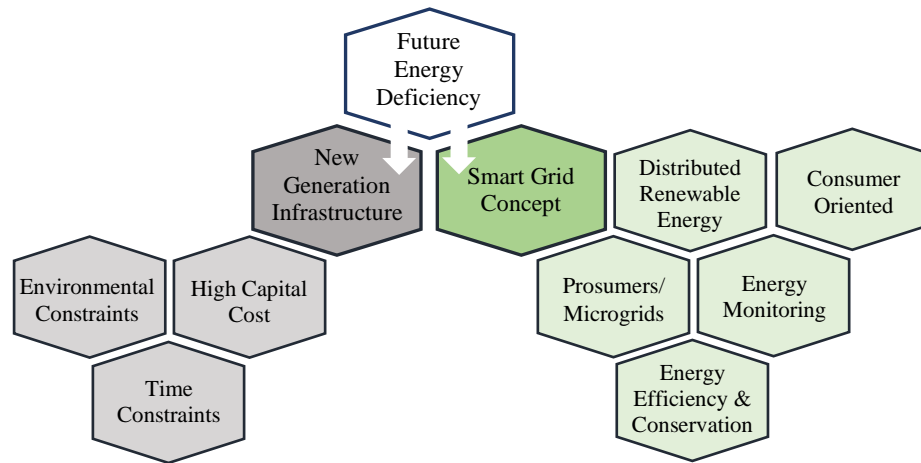


Figure 1 Rationale of the Study

As seen in Figure 1, establishing new powerplants requires high capital cost and substantial time to build, moreover, environmental constraints like carbon emission, in case of fossil-fuel based generation, are also associated with it. All these constraints limit the viability of establishing new generation plants for both developed and undeveloped societies. On the other side, adopting the smart grid concept towards sustainability is a more viable and attractive alternative, offering numerous promising solutions for both sides, i.e., generation and demand, as highlighted in Figure 1. On generation side, smart grid technologies effectively enable the integration of renewables, prosumers, and microgrids. Concurrently, on the demand side, the smart grid empowers the consumers by providing innovative tools that facilitate the consumers to interact and act more efficiently in terms of their energy usage. In the given context and from the larger perspective of sustainability, one of the key tools provided by the smart grid is the effective and interactive way of energy monitoring.

## 1.2 Scope and Significance

Based on the presented rationale, the scope of this research work is to investigate and develop a state-of-the-art approach that can facilitate effective energy monitoring. The

proposed approach must target the low data granularity to attain as much as possible compatibility with the existing metering infrastructure and further comprises attributes like robustness, low complexity, scalability, and cost-effectiveness. The developed approach will significantly facilitate the existing system in terms of energy efficiency and conservation. Consequently, avoid a large amount of investment in establishing new generation plants. This also leads to address the environmental concerns due to carbon emission in the case of establishing non-renewable generation plants.

Furthermore, the outcome of this research work in the form of low complexity and robust non-intrusive load disaggregation approach will not only be beneficial for the consumers in terms of having valuable information regarding the operation status of their loads and corresponding consumption but also benefit the system operators, policymakers, and manufacturers in terms of analyzing the network's energy flow, creating policies, and manufacturing of advanced appliances, respectively.

### **1.3 Contributions**

To advance the existing state of the art on non-invasive load disaggregation, this PhD research work has the following key attributes and contributions.

- This research work is based on low sampling data granularity making it more viable for the existing metering infrastructure.
- To realize real-world applications and deployment, the entire research work has been carried out on real-world electricity datasets.
  - For robustness, the proposed methodology has been employed on diverse load datasets based on different geographic regions, i.e., New Zealand and the United States of America.
- Three distinct low-complexity and computationally fast event detection algorithms are proposed and validated on both real-world datasets.
- A set of nine distinct electrical load features is proposed and evaluated.
  - The given load feature set is further evaluated in terms of feature space and individual feature significance using feature reduction and feature selection methodologies, respectively.
- Towards effective load classification, diverse supervised machine learning techniques are investigated towards an optimal classifier under given conditions.
  - In addition to standalone learning model configuration, the ensemble learning technique is also explored for load classification and a comprehensive comparative analysis is presented.

- Three different appliances/circuits having high potential towards real-world energy efficiency applications are successfully disaggregated using the proposed non-invasive load disaggregation approach.
- The proposed methodologies are extensively evaluated using well-known performance metrics as well as a newly introduced metric, i.e., the Kappa index, in the context of non-invasive load disaggregation.
- In terms of actionable feedback, a real-world application is proposed and validated using a case study.

Further, different research articles have been published, accepted, or submitted as part of this PhD research work. The details of the said articles are as follow.

### **Journal Articles**

- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low Complexity Statistical Features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 751-759, 2020.
  - DOI: [10.1109/TIM.2019.2904351](https://doi.org/10.1109/TIM.2019.2904351)
- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Intrusive Load Monitoring of Residential Water-Heating Circuit Using Ensemble Machine Learning Techniques," *Inventions*, vol. 5, no. 4, p. 57, 2020.
  - DOI: <https://doi.org/10.3390/inventions5040057>
- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Invasive Load-Shed Authentication Model for Demand Response Applications Assisted by Event-Based Non-Intrusive Load Monitoring", *Energy and AI*, p. 100055, 2021.
  - DOI: <https://doi.org/10.1016/j.egyai.2021.100055>

### **Conference Publications**

- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 2607-2612.
  - DOI: [10.1109/ISGT-Asia.2019.8881113](https://doi.org/10.1109/ISGT-Asia.2019.8881113)

- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Event Detection Algorithm for Non- Intrusive Load Monitoring Systems," in *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 746-751.
- DOI: [10.1109/ISGT-Asia.2018.8467919](https://doi.org/10.1109/ISGT-Asia.2018.8467919)

#### **Submitted Articles (Under Revision)**

- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Comparative Evaluation of Machine Learning Algorithms and Feature Space for Non-Invasive Load Disaggregation", *Journal of Modern Power Systems and Clean Energy*

## Chapter 2 Research Design

This chapter presents a comprehensive research design of this study. It is comprised of the literature review, research gap analysis, research objectives, proposed methodology, and performance evaluation criteria. A brief overview of the employed real-world load datasets is also presented in this chapter.

Figure 2 depicts the overall research flow of this PhD study, where each part is discussed in detail in the following sections.

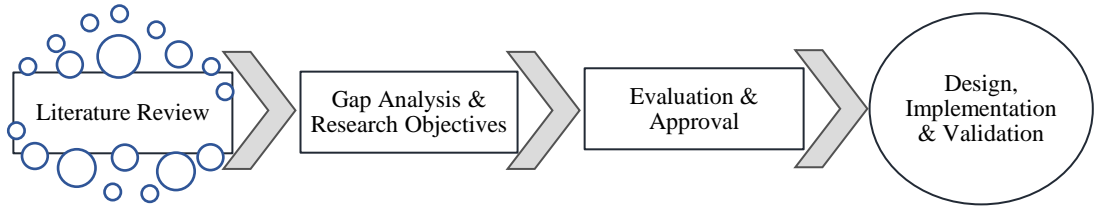


Figure 2 PhD Research Study Flow

### 2.1 Literature Review

This section presents a comprehensive review of the existing state of the art on energy monitoring techniques along with the corresponding advantages and disadvantages.

#### 2.1.1 Energy Monitoring

Due to the integration of information and communication technologies (ICT), today's power system is more advanced and smarter providing numerous promising solutions in terms of energy efficiency and conservation. But towards energy efficiency and conservation, effective management of energy is the key and for that energy monitoring is inevitable. Energy monitoring is a process that gathers consumers' consumption data, either by software or hardware means, analyses it, and then provides useful insight back to the consumers. Effective energy monitoring is one of the key solutions to energy efficient systems and can be performed either at the aggregated or segregated level. In terms of aggregated energy monitoring, advanced metering infrastructure (AMI) is one of the examples consisting of smart meters, ICT systems and database management systems, which plays a key role in terms of facilitating the concerned stakeholders to monitor the power consumption profiles and ultimately lead to efficient energy management [5].

In the context of segregated level energy monitoring, numerous approaches are available in the existing literature but a widely used approach is commonly referred to as load disaggregation, energy disaggregation, or power disaggregation [6]. It is a method where

the aggregated load profile is converted into the segregated load profiles of households' appliances. In the existing literature, a variety of load disaggregation methodologies is available that can be broadly classified as hardware and software-based methods [3, 7]. A broad categorization of energy monitoring techniques is depicted in Figure 3 [8]. The scope of this research work revolves around segregated energy monitoring particularly non-invasive (software) techniques as highlighted in Figure 3 (in blue). Therefore, the details regarding different categories of segregated energy monitoring are briefly discussed in the following sections.

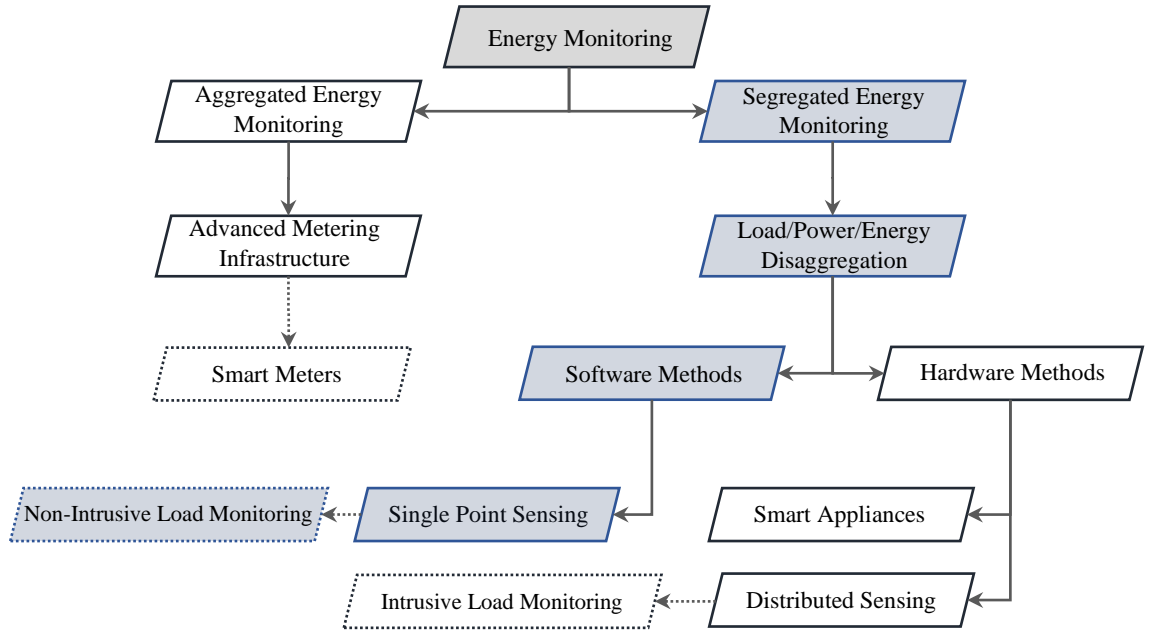


Figure 3 Energy Monitoring Techniques  
The portion in blue highlights the scope of this thesis.

### 2.1.2 Intrusive Load Monitoring and Smart Appliances

In the context of hardware-based methods, distributed sensing is a technique where discrete measurement sensors are used to monitor and report the energy consumption pattern of the individual appliances/circuits. Due to its invasive nature, it is also referred to as intrusive load monitoring (ILM). This technique is relatively simple to deploy, even the recent introduction of energy monitoring plug-in devices does not require any installation. These devices are ready-to-use and only need to be plugged in between an electric wall socket and the appliance's plug. But due to multiple numbers of sensors, reliability issues may exist [9] and in the context of the smart grid, a significant number of sensors need to be deployed which leads ILM to be cost-prohibitive in terms of capital and labor expenses [7].

Smart appliances also lie in the context of hardware-based load disaggregation. Smart appliances are those appliances that have integrated capabilities to monitor and report their power consumption [10] but these appliances are not widely in use due to their high market prices and interoperability issues [11].

### 2.1.3 Non-Intrusive Load Monitoring

Software methods provide an alternative and attractive solution to load disaggregation, where a frequently used approach is referred to as single point sensing technique, as shown in Figure 3. This technique uses only a single metering device, typically the main entry metering sensor, to monitor the aggregated load data and later different software methodologies are employed to disaggregate the monitored load data into individual appliances/circuit profiles. Due to its non-invasive nature, this technique is commonly referred to as non-intrusive load monitoring (NILM), non-intrusive appliance load monitoring (NALM) [12] or non-intrusive appliance load monitoring (NIALM) [13]. The concept of NILM was first introduced by Hart [14, 15] and refined over the years by the research community. NILM is one of the widely used techniques where the state of the operation of individual appliances/circuits is determined by the analysis of aggregated load data measured at a single metering point, typically the main power entry point of a building [16]. Consider a time-series power load curve monitored at a single metering point weighted as an algebraic sum of  $I$  number of appliances' power load, as shown in (1).

$$P_{agg}(t) = \sum_{i=1}^I P_i(t) + n(t) \quad (1)$$

$P_{agg}(t)$  and  $P_i(t)$  represent the total aggregated power and  $i^{th}$  appliance power at time  $t$ , respectively. Moreover,  $i=1, 2, 3, \dots, I$  and  $n(t)$  represents noise including measurement errors, baseloads, and loads not under consideration. The task of non-intrusive load disaggregation, i.e., NILM, is to estimate the state of the individual appliance power load,  $P_i(t)$ , with the only information of aggregated power load,  $P_{agg}(t)$ . A traditional NILM system consists of three main components, namely: data acquisition, feature extraction, and classification. Each component can be further categorized as depicted in Figure 4 [3] and further discussed in the following subsections.



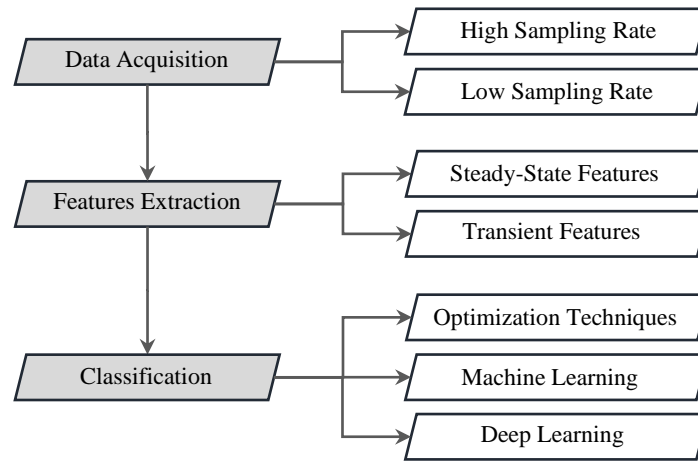


Figure 4 Traditional NILM Framework

### 2.1.3.1 Data Acquisition

Data acquisition is a pre-requisite stage of any NILM system where the aggregated load data is monitored and stored from a single metering point. This information can be acquired in the form of different variables such as power, current, voltage, and measured using different measurement devices at different sampling frequencies. Data acquisition granularity is a key factor that not only significantly influences the later stages of NILM in terms of selection of tools and analysis techniques but also determines the type and number of appliances to be precisely classified [4]. Based on the granularity of the monitored data, NILM algorithms in the available literature can be broadly categorized as high sampling rate (frequency  $\geq 50$  Hz) and low sampling rate (frequency  $\leq 1$  Hz) [17] as shown in Figure 4.

The research community has acquired electricity data at numerous sampling rates and publicly released these datasets allowing the researchers to evaluate their proposed NILM methodologies. Some of the well-known and widely used NILM datasets are Reference Energy Disaggregation Dataset (REDD) [18], the Almanac of Minutely Power datasets (AMPds) [19], Pecan Street Inc. Dataport [20], Electricity Consumption & Occupancy (ECO) [21], and GREEN Grid [22]. Table 1 presents a detailed comparative study of the available energy disaggregation datasets in terms of different attributes including data acquisition granularity.

### 2.1.3.2 Load Features

Each appliance is unique in terms of its consumption pattern. The explicit attribute of an appliance is known as its feature or also referred to as the signature of the appliance. The available literature on load features can be broadly categorized into steady-state features

Table 1 Comparison of Energy Disaggregation Datasets

Dataset	Resolution	Number of Buildings/Houses	Duration	Features	Location	Link
REDD [18]	1 Hz, 15 kHz	6	~2 weeks	$p$ , $i$ , and $v$	United States of America	<a href="http://redd.csail.mit.edu/">http://redd.csail.mit.edu/</a>
AMPds [19]	1 minute	1	2 years	$p$ , $q$ , $s$ , $i$ , and $v$	Canada	<a href="http://ampds.org/">http://ampds.org/</a>
UK-DALE [23]	6 sec, 16 kHz	5	3-51 months	$p$ , $i$ , and $v$	United Kingdom	<a href="https://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl/?WantComp=138">https://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl/?WantComp=138</a>
DRED [24]	1 Hz	1	2 months	$p$	Netherland	<a href="http://www.st.ewi.tudelft.nl/akshay/dred/">http://www.st.ewi.tudelft.nl/akshay/dred/</a>
Dataport [20]	1 minute	~1000	2011-ongoing	$p$	United States of America	<a href="https://www.pecanstreet.org/dataport/">https://www.pecanstreet.org/dataport/</a>
GREEND [25]	1 Hz	9	1 year	$p$	Italy & Austria	<a href="https://www.monergy-project.eu/?page_id=380">https://www.monergy-project.eu/?page_id=380</a>
ECO [21]	1 Hz	6	8 months	$p$ and $q$	Switzerland	<a href="https://www.vs.inf.ethz.ch/res/show.html?what=eco-data">https://www.vs.inf.ethz.ch/res/show.html?what=eco-data</a>
PLAID [26]	30 kHz	56	Summer-2013	$i$ and $v$	United States of America	<a href="http://plaidplug.com/">http://plaidplug.com/</a>
REFIT [27]	8 seconds	20	2013-2015	$p$	United Kingdom	<a href="https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned">https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned</a>
GREEN Grid [22]	1 minute	~45	2014-2018	$p$	New Zealand	<a href="http://reshare.ukdataservice.ac.uk/853334/">http://reshare.ukdataservice.ac.uk/853334/</a>
BLUED [28]	12 kHz	1	7 days	$i$ and $v$	United States of America	<a href="http://portoalegre.andrew.cmu.edu:88/BLUED/">http://portoalegre.andrew.cmu.edu:88/BLUED/</a>
SustDataED [29]	12.8 kHz	1	10 days	$i$ and $v$	Portugal	<a href="https://aveiro.m-iti.org/data/">https://aveiro.m-iti.org/data/</a>
iAWE [30]	1 Hz	1	73 days	$p$ , $f$ , $\phi$ , $i$ , and $v$	India	<a href="http://energy.iiitd.edu.in/Datasets.aspx">http://energy.iiitd.edu.in/Datasets.aspx</a>
COMBED [31]	30 seconds	-	1 month	$p$ , $i$ , and $e$	India	<a href="https://combed.github.io/">https://combed.github.io/</a>
Smart-Grid Smart-City Customer Trial Data [32]	30 minutes	-	2010-2014	-	Australia	<a href="https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details">https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details</a>
Smart <sup>7</sup> [33]	1 Hz	3	3 months	$p$ and $s$	United States of America	<a href="http://lass.cs.umass.edu/projects/smart/">http://lass.cs.umass.edu/projects/smart/</a>

**Note:**  $i$ ,  $v$ ,  $p$ ,  $q$ ,  $s$ ,  $f$ ,  $e$ , and  $\phi$ , and represents current, voltage, real power, reactive power, apparent power, frequency, energy, and phase respectively

<sup>7</sup> Smart refers to UMass Smart\* Home Data Set

and transient<sup>8</sup> features [3], as depicted in Figure 5. Both refer to identifying the changes in the operational characteristics of the appliance when it modulates from one operational state to another. But both features, i.e., steady-state and transient, differ in what data they are focusing on, referring to the granularity of data acquisition, the corresponding association is shown in Figure 5.

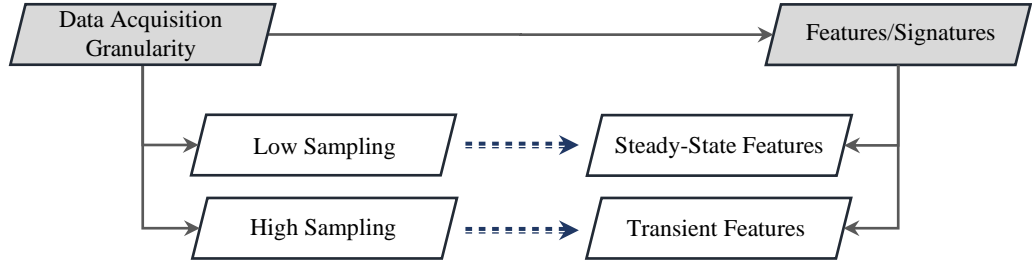


Figure 5 Relationship between Data Acquisition and Features in NILM Domain

Steady-state features like active and reactive power are more related to the appliances' power characteristics when it changes state. The extraction of these features does not require high-end metering devices and can be easily extracted from the RMS values of current and voltage [34]. On the other side, the transient features depict the appliances' features in terms of shape, duration, size, and harmonics at its transients [3]. To extract such features, high-end metering devices are required that lead to costly hardware to be installed at customers' premises because the existing smart meters are not capable of reaching the desired high sampling rates.

It is evident from the existing literature that a single feature does not perform well for all types of appliances' recognition hence use of multiple features is indeed promising. In this context, numerous features are proposed based on conventional parameters, i.e., voltage (V), current (I), and power (P), and non-conventional parameters, e.g., occupancy, dwelling, and weather information. The most widely employed features within the NILM domain are active (P) and reactive (Q) power [35-37]. Other commonly used features are based on waveforms [38], V-I trajectories<sup>9</sup> [39, 40], and harmonics [41]. Tabatabaei et al. [37] present a comprehensive review to identify the learning algorithms and features used in the NILM domain. Further discussion on features extraction in the NILM domain and a detailed graphical depiction in terms of a semantic network of the corresponding features can be found in [37]. Table 2 presents an overview of the recent research on feature extraction in the non-intrusive load disaggregation domain.

<sup>8</sup> The short-term momentary fluctuations associated with any state transition in signal before settling into a steady state value is known as transient.

<sup>9</sup> V-I trajectory is a graphical depiction of the instantaneous voltage and current in the voltage-current plane.

Table 2 Recent Research on Features in NILM Domain

Research Work	Load Features Categorization
Azaza and Wallin [42]	Active and reactive power variation, current waveform, voltage-current trajectory, harmonics, over/undershoot power amplitude, rise/fall time, settling time.
Sadeghianpourhamami et al. [43]	P-Q plane, wavelets, P-Q plane and macroscopic transient, P-Q plane and wavelets, P-Q plane and macroscopic transient and harmonics, shape features, P-Q plane and harmonics, raw waveforms, real power only, nonactive current, spectral envelop, low-frequency P, I, and V based features, combinational features.
Zeifmann and Roth [44]	Change of real power, change of real and reactive power, change of real and reactive power and the additional macroscopic features, harmonic and Fourier transform, beyond wavelet transform and geometrical shape.
Zoha et al. [3] Esa et al. [13]	Power change, time and frequency domain V-I features, V-I trajectory, Voltage noise, transient power, start-up current waveform, voltage noise.
Klemenjak and Goldsborough [45]	Power change, V-I features, V-I trajectory, harmonics, transient state features (shape, size, and duration of transient).

### 2.1.3.3 Classification

Load classification in non-intrusive load disaggregation comprises numerous pattern recognition algorithms that are used to identify an appliance specific state based on the extracted features as shown in Figure 4. The recent advancement in artificial intelligence (AI) and computational resources enable the researchers to perform load classification with more precision. Within AI, machine learning algorithms evolved significantly in the last two decades and established themselves as an imperative method in research as well as real-world application development [46]. These algorithms can be broadly categorized into supervised and unsupervised algorithms. Supervised learning requires a training phase where both the aggregated and ground-truth<sup>10</sup>, also referred to as label, data are required for training purposes [3]. Contrary, the unsupervised learning does not require appliances' ground-truth data for training purposes. Hence, the force that drives the selection of supervised or unsupervised algorithms is the availability of ground-truth data of individual load elements, which can be collected through the sub-metering of the individual load elements. Figure 6 presents a generalized framework of the aforesaid machine learning models.

<sup>10</sup> Ground-truth refers to the actual items, occur in reality.

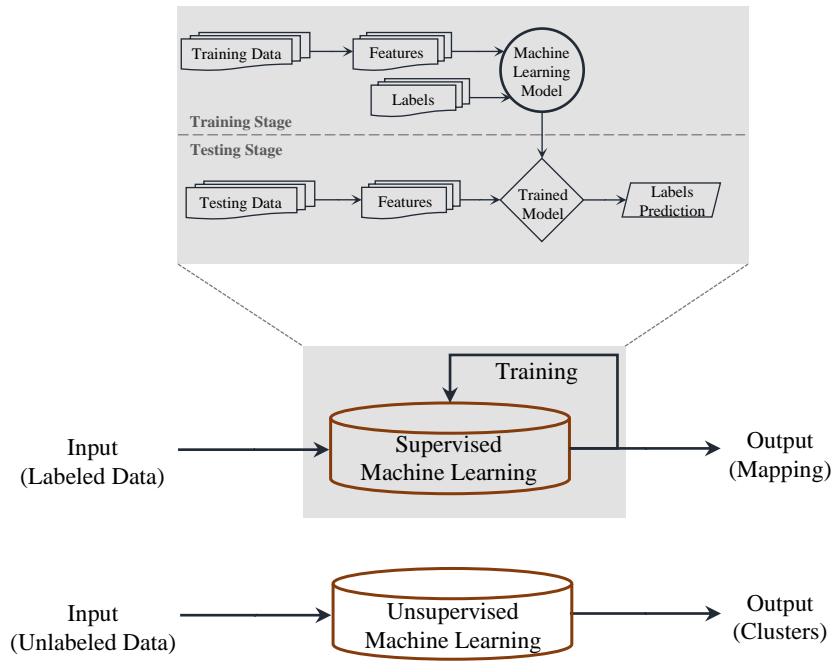


Figure 6 Machine Learning Models Framework

Recently, machine learning algorithms are extensively adopted in the domain of energy systems. In this context, Pérez-Ortizare et al. [47] present a comprehensive review of different classification problems and machine learning algorithms in renewable energy applications. Further, Lorena et al. [48] investigate different supervised machine learning algorithms towards an optimal classifier in potential distribution modeling. In the context of non-intrusive load disaggregation, supervised machine learning models [3, 49-51] are more frequently used as compared to any other AI-based methodologies, e.g., optimization [52-55], deep learning [56, 57]. A comprehensive comparative evaluation of different supervised machine learning algorithms in the context of low-sampling non-intrusive load disaggregation has been presented in [58].

## 2.2 Research Gap Analysis

With the worldwide deployment of smart meter infrastructure, intelligent power systems, and increasing awareness of energy efficiency and conservation, the research community has ignited renewed interest in the NILM domain. Consequently, extensive research within the NILM domain has been carried out, however, it is still an open research area having gaps that need to be addressed. To address the shortcomings of the existing research literature and contribute to the further advancement of non-intrusive load disaggregation, a comprehensive critical research gap analysis of the existing NILM literature has been carried out. The performed gap analysis aims at defining more viable and relevant research objectives for this study. It is worth noting that the following

research gap analysis does not intend to undermine the existing literature but acknowledges the existing work as a foundation towards further research developments.

Since the early concept of NILM, numerous techniques have been proposed and developed to enhance the performance of the NILM system. Comprehensive reviews of the proposed NILM in different contexts are presented in [3, 13, 34, 44, 45, 59-62]. Based on the existing literature trends, this research gap analysis is divided and conferred into two main categories in the context of the NILM system namely, the working principle and data granularity.

### **2.2.1 Working Principle**

Early NILM research was more focused on the disaggregation of high consumption loads [63-65] whilst currently, the trend is to disaggregate a higher number of appliances [4] with more precision. However, in terms of working principle, the existing literature regarding NILM systems can be broadly classified into two main categories namely, event-based and non-event based NILM systems [7, 66, 67]. The event-based NILM system relies on event<sup>11</sup> detection by using different edge<sup>12</sup> detection algorithms on the acquired aggregated load data. Later-on features are captured from the extracted events, that are further classified into respective appliance operation states using different pattern recognition techniques. Contrary, a non-event based NILM system does not rely on detecting appliance-state transitions (events) using edge detection algorithms before the classification stage. Rather, all the samples of the acquired aggregated load data are considered for inference.

One of the pioneering work in load disaggregation by Hart [15] in the early 90s was based on identifying appliances by their respective turning On/Off transitions by considering power consumption changes (events) both in the active and reactive power of the signal. To date, numerous event detection algorithms have been proposed and developed with diversity in terms of input variables, data granularity, and techniques. Most of the existing work is based on appliance consumed power [50, 68-73] as an input variable with some exception, e.g., the authors of [41, 74] employed current harmonics as an input feature to detect the events. Meziane et al. [67] also employed a current signal with the phenomenon of the sliding window. De Baets et al. [75] also performed event detection by taking active power as input feature but in the frequency domain instead of the time domain. Girmay

---

<sup>11</sup> In this context, event refers to appliance-state transition.

<sup>12</sup> An edge is defined as an abrupt change in terms of jumps, steps or shifts in the mean level of a time series or signal that is under consideration.

and Camarda [69] proposed a time-frequency based event detection using a goodness-of-fit Chi-squared test. Jin et al. [72] also proposed an event detector based on goodness-of-fit tests and performed a benchmark analysis with a conventional generalized likelihood ratio detector. Further, Wild et al. [41] proposed a new unsupervised event detection approach that is based on kernel Fisher discriminant analysis.

On the other side, work related to the hidden Markov model (HMM) [76] is an example of non-event based NILM systems. In this domain, Kolter and Jaakkola [77] considered an extension of HMM: additive factorial hidden Markov model (AFHMM), towards energy disaggregation. Kim et al. [78] also targeted the NILM using variants of the factorial hidden Markov model (FHMM) and concluded that the conditional FHMM model outperforms the other unsupervised disaggregation methods. These models are powerful, work offline, and can be supervised or unsupervised but they are not scalable and the complexity of these models increases exponentially with an increase in the number of appliances [42, 79]. In this context, Kim et al. [80] present a detailed discussion about the constraints associated with HMM and its variants and proposed an alternative in terms of advanced deep learning approach, i.e., long short-term memory recurrent neural network (LSTM-RNN).

In the context of the working principle of the NILM system, Table 3 summarizes both methodologies, i.e., event-based and non-event based NILM systems [7].

Table 3 Comparison between Event and Non-Event Based NILM Systems

<b>Event-Based NILM System</b>	<b>Non-Event Based NILM System</b>
Computationally more efficient as inference is carried out only on the detected events.	It does not rely on event detection hence inference is carried out on all samples, consequently, computational requirement increases.
False detection or misdetection of events may lead to errors.	Wrong estimation errors for a given sample can be corrected.

Based on the literature review, it is concluded that the event-based NILM is not only computationally more efficient compared to non-event based NILM but also more viable for the existing infrastructure in terms of computational cost and scalability. Therefore, this research work intends to contribute to the existing state of the art on event-based NILM systems by proposing computationally efficient event detection algorithms that will further facilitate the event-based non-intrusive load disaggregation approach.

### 2.2.2 Data Granularity

As discussed earlier for non-intrusive load disaggregation, the data granularity is of high significance because it drives the use of tools and analysis techniques. In this context, the

existing NILM literature can be categorized as, low sampling rate NILM ( $\leq 1$  Hz or  $\geq 1$  second) and high sampling rate NILM ( $\geq 50$  Hz or  $\leq 0.02$  seconds) [17] systems. In connection with the working principle, all high sampling rate data are generally analyzed using event-based techniques whereas low sampling data are analyzed using both event-based and non-event based techniques [68]. Low sampling rate data lead to macroscopic<sup>13</sup> (steady-state) features whereas high sampling rate data lead to microscopic<sup>14</sup> (transient) features [38], as graphically shown in Figure 5.

Many of the researchers agree that to attain a high accuracy of disaggregation, the microscopic features should be utilized [44]. Subsequently, most of the available literature is focusing on a high sampling rate NILM system [81]. For example, Anderson et al. [66] proposed a framework and evaluation metrics for event detection in the NILM domain using one week of aggregated voltage and current measurements sampled at a rate of 12 kHz. Meziane et al. [67] proposed a high accuracy event detector for the NILM system at a sampling frequency of 10 kHz. Furthermore, Gupta et al. [82] utilized a data acquisition system in the range of 36-500 kHz. The study claimed that the signature was both distinctive for a given appliance and variable enough to distinguish between similar devices in a household with a reported accuracy of up to 93.8% [82].

Besides the advantages and disadvantages associated with low and high sampling rate data acquisition, there is also a trade-off that exists. For example, a high sampling rate allows the extraction of transient features in addition to power features subsequently yielding not only better energy disaggregation [60] but also enabling more appliances recognition [4]. For high sampling data acquisition, the aforesaid improvements come at a price of high cost and complexity due to high-end, costly, and dedicated measurement devices that need to be installed at consumers' premises. Moreover, on social grounds, a high sampling rate also raises consumers' privacy concerns as their activities can be easily detected [83]. On the other side, the low sampling rate data acquisition fits well into the existing smart meter capabilities consequently avoiding the installation of high-end and costly metering devices. But this comes at the cost of a complicated disaggregation process because most of the information of the waveform is lost due to capturing the data at the low sampling frequency [17]. Further, at a low sampling rate, low consumption loads are difficult to detect. However, the high consumption load elements are detectable with reasonable precision. So far, a computationally efficient low sampling data

---

<sup>13</sup> Macroscopic features mostly refer to the power changes.

<sup>14</sup> Microscopic features mostly refer to the harmonics and signal waveforms.



acquisition based non-intrusive load disaggregation, i.e., NILM system, is a challenging and open research topic [3, 44].

As per detailed literature review and current research trends in the context of non-intrusive load disaggregation, it is noted that most of the NILM research is inclined towards high data granularity [52], subsequently, less research has been carried out in the domain of low sampling NILM system. To address the said research gap, this research work aims to exclusively focus on a non-intrusive load disaggregation approach based on low data granularity. Focussing on low data granularity based NILM system will not only contribute to the existing literature by addressing the research gap but also enable the deployment of the NILM system in real-world energy efficiency applications.

### **2.2.3 Research Objectives**

Based on the presented research gap analysis, two aspects of the non-intrusive load disaggregation literature have been identified, i.e., the working principle and data granularity, that need to be further addressed. For the identified research gaps, this research work aims to contribute as follows,

- *Working Principle:* To contribute to the existing state of the art on event-based NILM system by proposing and developing computationally efficient event detection algorithms that further facilitate the development of event-based non-intrusive load disaggregation methodologies.
- *Data Granularity:* To address the shortcoming of the existing literature regarding low sampling NILM research by proposing and developing a non-intrusive load disaggregation method that is solely based on low data granularity.
  - To realize the real-world potential, incorporate and validate the proposed NILM approach in the context of energy efficiency applications.

The aforesaid research objectives are well aligned with the rationale of this study presented in Chapter 1. As besides contributing to the existing state of the art on computationally efficient non-invasive load disaggregation, this research work and its broader real-world (prospective) applications will contribute significantly to the energy efficiency and conservation programs.

## **2.3 Evaluation and Approval**

Evaluation is a key component to be considered before plunging into designing and simulation studies. In this phase, a careful assessment of the extracted research gaps and

corresponding research objectives of this study are carried out in terms of their relevance, feasibility, and reliability in the broader context of energy efficiency and conservation.

It is important to evaluate whether the findings of the research gap analysis are relevant or not and whether it will solve any of the existing or future issues. Furthermore, it is also evaluated whether the methodology proposed to address the findings of the research gap analysis is viable enough for the existing system or not. Assessment of these conditions not only provides a base to opt/optimize a suitable design methodology but also provides an in-depth analysis of the rationale and significance of this research.

In the given context, evaluation is carried out by all the concerned academics and industry stakeholders including but not limited to Auckland University of Technology (AUT) supervisory team, PhD thesis proposal defense (AUT-PGR9<sup>15</sup>) examination team, and Genesis Energy Limited team. After a successful process of AUT-PGR9 and evaluation, a confirmation approval of this PhD research work has been granted. In this context, an official letter regarding confirmation of candidature can be found in the appendix A.1 of this thesis.

## **2.4 Proposed Methodology**

Based on the finding of the research gap analysis and the objectives of this study, this research work focusses on low sampling event-based non-invasive load disaggregation, i.e., NILM, system. Figure 7 presents a basic framework for an event-based NILM system employed in this study along with the techniques to be used at each stage, where in-depth discussion regarding the proposed or adopted algorithms and techniques will be presented in the later chapters of this thesis. As shown in Figure 7, an event-based NILM framework starts with aggregated load data acquisition (at a single metering point) followed by data pre-processing, event detection, and load feature extraction. Load identification based on the extracted load features is carried out at the classification stage. The proposed methodology is further employed to validate its effectiveness in real-world energy efficiency applications.

---

<sup>15</sup> PGR9 (Post Graduate Report 9), is a report/presentation that need to be successfully submitted/presented within the first year of PhD studies for confirmation of PhD candidature at AUT.

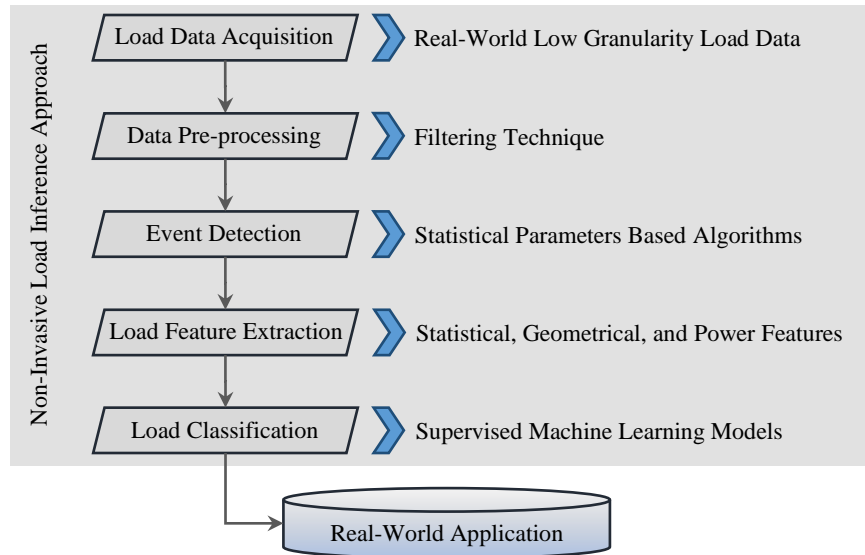


Figure 7 Proposed Event-Based Non-Invasive Load Disaggregation Approach

### 2.4.1 Load Databases

In terms of load data acquisition, the presented research work is entirely based on low data granularity, sampled at 1/60 Hz, i.e., 1-minute measurement interval. In the context of low sampling, it is worth noting that this research work is focussing on a data granularity that is 60 times lower than the data granularity of 1 Hz, which is generally used in the context of low sampling NILM system, also seen in Table 1. Further, to realize the real-world applications of non-intrusive load disaggregation, this research work is based on real-world load measurements that are acquired from real-world load databases from diverse geographic regions. Figure 8 presents the diverse geographic regions of electricity load databases employed in this research work.



Figure 8 Geographic Regions of Employed Load Databases

As shown in Figure 8, the load databases employed in this research work namely, Dataport<sup>16</sup> and NZ GREEN Grid<sup>17</sup> are from the United States of America (USA) and New Zealand (NZ), respectively. Both databases rely on low data granularity of 1/60 Hz, i.e., a 1-minute sampling rate, and are comprised of aggregated as well as segregated load data.

Dataport [20] is the world's largest<sup>18</sup> energy disaggregation research database [84], owned and operated by Pecan Street Inc.<sup>19</sup>, a research institution located in the USA. Dataport comprises electricity consumption profiles of nearly 1000 households in the USA. Each household contains both aggregated and individual appliance level power profile at a data granularity of 1/60 Hz. For commercial purposes, individuals and organizations can access Dataport contents by acquiring a license offered by Pecan Street. On the other side, for non-commercial educational purposes, Pecan Street offers free data access to current university faculty, staff, and students<sup>20</sup>.

The second database employed within the scope of this research work is a recently released<sup>21</sup> electricity dataset, namely NZ GREEN Grid Data [22]. This dataset comprises real-world load measurement of 45 NZ households, as part of the Renewable Energy and the Smart Grid (NZ GREEN Grid) project, a collaboration among the University of Canterbury and the University of Otago, New Zealand. For each household, the NZ GREEN Grid research database includes 1-minute mean electricity power, in watt, for aggregated (total incoming power) and individual power circuits, making it the first research database of its kind for New Zealand.

The proposed methodologies in this thesis are evaluated on both load databases. Moreover, for further robust assessment, it is ensured that load data are acquired from at least two different households of each database. The given criteria provide a robust assessment of the proposed methodologies in diverse and independent environments. This is also necessary for the classification stage to assess how well the employed classifiers generalize in an unseen and diverse environment. Therefore, within the scope of this thesis, data are acquired from two and five households of Pecan Street's Dataport and NZ GREEN Grid database, respectively. Further in terms of data acquisition duration, load data of up to 30 days are acquired and employed for digital simulation purposes. Details

---

<sup>16</sup> <https://www.pecanstreet.org/dataport/>

<sup>17</sup> <http://reshare.ukdataservice.ac.uk/853334/>

<sup>18</sup> <https://www.pecanstreet.org/work/energy/>

<sup>19</sup> <https://www.pecanstreet.org/>

<sup>20</sup> <https://www.pecanstreet.org/dataport/access/>

<sup>21</sup> NZ GREEN Grid project releases database for public use. Link: <https://www.otago.ac.nz/news/news/releases/otago695264.html>

of the employed households are given below, where further details in terms of data acquisition timeframe, number of days, and the number of data samples will be presented in the respective sections later in this thesis.

- Dataport, Pecan Street Inc.
  - Household ID 26
  - Household ID 3036
- NZ GREEN Grid Database
  - Household ID rf\_01
  - Household ID rf\_02
  - Household ID rf\_31
  - Household ID rf\_36
  - Household ID rf\_42

Moreover, the selection of the given households is based on the explicit availability/installation of the individual appliances/circuits that are under consideration for inference purposes within the scope of this research work. For example, for Dataport this research work targets the inference of electric vehicle charging and air conditioning unit, where for the NZ GREEN Grid database the inference of water heating is of primary focus. Further details regarding the selection of the aforesaid load elements can be found in the following chapters of this thesis.

#### **2.4.2 Performance Evaluation Metrics**

Performance metrics play a key role in the context of evaluating different methodologies. In the non-intrusive load disaggregation domain, there is no coherent way to evaluate the performance of the algorithm due to the diversity of performance metrics used by the research community to evaluate their methodologies, consequently, there is no standardized NILM performance metrics [34, 85]. Hence, in this research work, a set of diverse, well-known, and widely used performance metrics is employed for evaluation purposes. In the given context, the performance metrics namely, true positive (TP), false positive (FP), false negative (FN), precision, recall, f-score, and accuracy are adopted. All these performance metrics are defined and mathematically given as following [34, 55, 85, 86].

The terminologies of true positive, false positive, and false negative are explained in the form of a confusion matrix [55] and summarized in Table 4 [85].

Table 4 Confusion Matrix

		Ground-truth	
		Positive	Negative
Prediction	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Precision is a measure defined as the ratio of truly detected items and all detected items, in other words, how many selected items are relevant. It is given in (2).

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

The recall is a measure of item detection that occurred in reality or in other words how many relevant items are selected. It is mathematically given as in (3).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

F-score is defined as the harmonic mean of precision and recall and is mathematically given as in (4).

$$\text{F-Score} = \left( \frac{\text{Precision}^{-1} + \text{Recall}^{-1}}{2} \right)^{-1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

Accuracy is a measure defined as the fraction of predictions the algorithm identified correctly and is given as in (5).

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (5)$$

In addition to the aforesaid performance metrics, another performance metric employed for evaluation purposes is the Kappa index that is computed using both accuracy and expected accuracy, mathematically given as in (6) [87].

$$\text{Kappa Index} = \frac{\text{Accuracy} - \text{Expected Accuracy}}{1 - \text{Expected Accuracy}} \quad (6)$$

The expected accuracy is defined as the accuracy that any random classifier would be expected to achieve based on the confusion matrix, given in Table 4. Expected accuracy is mathematically defined as in (7) [87].

$$\text{Expected Accuracy} = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{TN} + \text{FN})(\text{TN} + \text{FP})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2} \quad (7)$$

As per definition, given in (6),  $\text{Kappa Index} < \text{Accuracy}$ , however, due to its inter-rater reliability; the degree of agreement among two or more raters, the Kappa index is a more robust measure to evaluate machine learning model performance. Moreover, the Kappa index of one machine learning model is directly comparable to the Kappa index of another machine learning model, being employed for a similar classification task. Landis and Koch [88] assigned labels in terms of agreement strength to different ranges of the Kappa values, presented in Table 5. It is evident from Table 5 that the higher the Kappa index value the better the agreement, where generally the Kappa index greater than 0.40 is desirable [87].

Table 5 Kappa Index Range and Corresponding Labels

<b>Kappa Index Range</b>	<b>Labels</b>
Less than 0	Poor
0 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.0	Almost Perfect

The details presented in Table 5 are also used as a performance benchmark for the classification stage within the scope of this research work.

## 2.5 Concluding Remarks

This Chapter presented a detailed research design of this PhD study starting with a literature review, followed by a comprehensive research gap analysis and research objectives. This Chapter also briefly discussed the proposed research methodology along with the details of the employed real-world load databases. Further different performance metrics to be employed for the evaluation purposes within this research work are also discussed in this Chapter. The details and analysis presented in this Chapter provide a groundwork for the modeling and simulations of the complete load disaggregation system towards successful non-invasive inference of distinct load elements.

Further, as evident from Figure 7, the proposed research is built-on different sub-blocks that include data pre-processing, event detection, feature engineering, load classification, and real-world applications. In this context, the following Chapters of this thesis will elaborate on all the details of the aforesaid sub-blocks of the proposed research work. As data acquisition is followed by data pre-processing and event detection, hence Chapter 3 presents the comprehensive details of employed techniques, proposed algorithms, digital simulations, and corresponding results and analysis in terms of data pre-processing and event detection.

## Chapter 3 Data Pre-processing and Event Detection

This chapter starts with the description of the data pre-processing technique, followed by the comprehensive details of the working principle, simulation studies, and corresponding results and analysis of the newly proposed event detection algorithms.

In this Chapter, the presented details in terms of working principles, simulation studies, and corresponding results and analysis regarding the proposed event detection algorithms are primarily based on [49, 89-91]. The said manuscripts have been published as follows:

1. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low Complexity Statistical Features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 751-759, 2020.
  - DOI: [10.1109/TIM.2019.2904351](https://doi.org/10.1109/TIM.2019.2904351)
2. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Intrusive Load Monitoring of Residential Water-Heating Circuit Using Ensemble Machine Learning Techniques," *Inventions*, vol. 5, no. 4, p. 57, 2020.
  - DOI: <https://doi.org/10.3390/inventions5040057>
3. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 2607-2612.
  - DOI: [10.1109/ISGT-Asia.2019.8881113](https://doi.org/10.1109/ISGT-Asia.2019.8881113)
4. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Event Detection Algorithm for Non- Intrusive Load Monitoring Systems," in *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 746-751.
  - DOI: [10.1109/ISGT-Asia.2018.8467919](https://doi.org/10.1109/ISGT-Asia.2018.8467919)

### 3.1 Data Pre-processing

As discussed in Chapter 2, this research work is based on the load measurements from diverse real-world load databases, consequently, measurement noise including impulses and ripples are inevitable. Therefore, pre-processing of the acquired load data is a pre-requisite for further simulations in terms of event detection. As an example, Figure 9 [22] graphically depicts a snapshot of the ground-truth power measurement profile of an appliance, i.e., fridge, from the NZ GREEN Grid database, where the phenomenon of impulses and ripples arising in real-world measurement can be seen. In this case, the



impulses and ripples are due to the fridge's motor starting and load characteristics/voltage fluctuation, respectively. For event detection, these measurement noises (impulses/ripples) in power signals can lead to false detection and need to be addressed using data processing algorithms before the event detection phase.

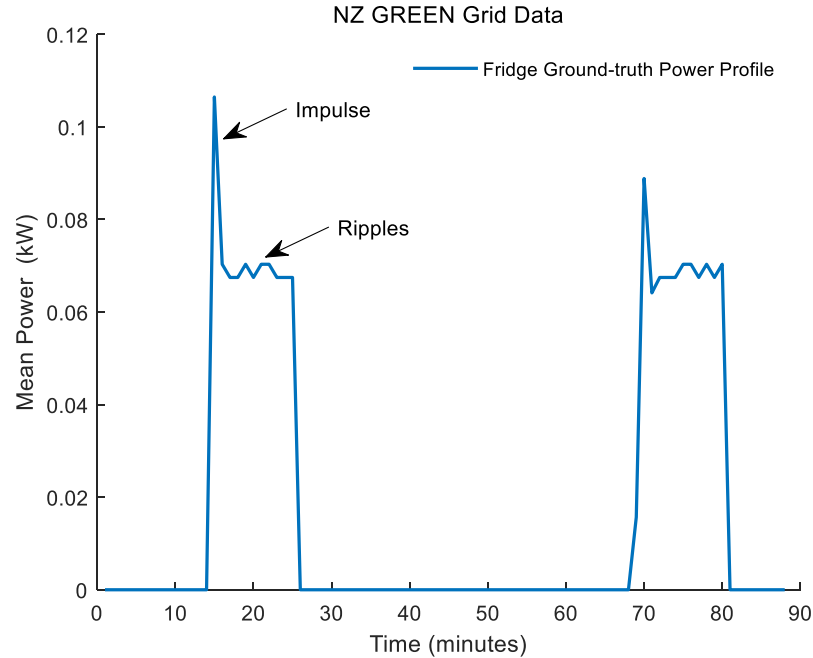


Figure 9 Impulse and Ripple Phenomenon

To address the aforesaid real-world measurement noises, this research work adopted a technique of median filtering. It is a nonlinear digital filtering technique used for signal noise removal. Median filtering has the capability to preserve the edges while eliminating the undesirable attributes of the signal. The said property of the median filtering technique makes it a more viable data pre-processing option for event detection in the context of non-intrusive load disaggregation, subsequently, it is extensively adopted by the research community in the NILM domain [67, 69, 71, 92]. A detailed working principle along with a graphical illustration of the median filtering phenomenon is presented in [92].

Figure 10 presents the simulation results of the employed median filtering technique in this research work for the data presented in Figure 9. It is evident from the presented results in Figure 10 that the median filtering technique successfully preserves the desired edges (for further processing in the context of event detection) while eliminating the undesirable attributes of the power signal.

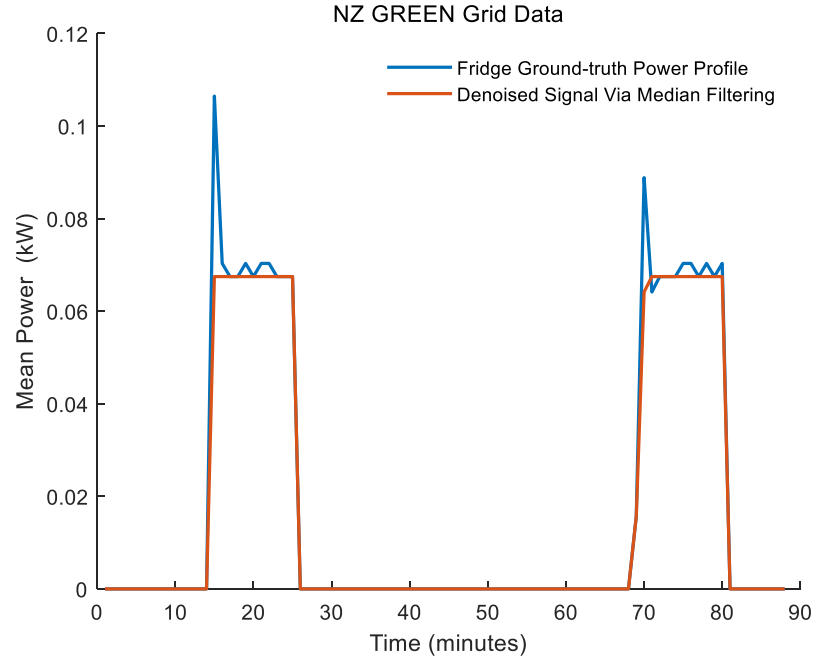


Figure 10 Outcome of Median Filtering Technique

### 3.2 Event Detection

In the context of event-based non-intrusive load disaggregation, this research work adopted the extended definition of an event by [41], i.e., an event is a portion of a signal envelop that deviates from the previous steady-state and lasts until the next steady-state has been reached [41]. A graphical depiction of events (turn-on and turn-off) with corresponding attributes is presented in Figure 11 for the same data as presented in Figures 9 and 10.

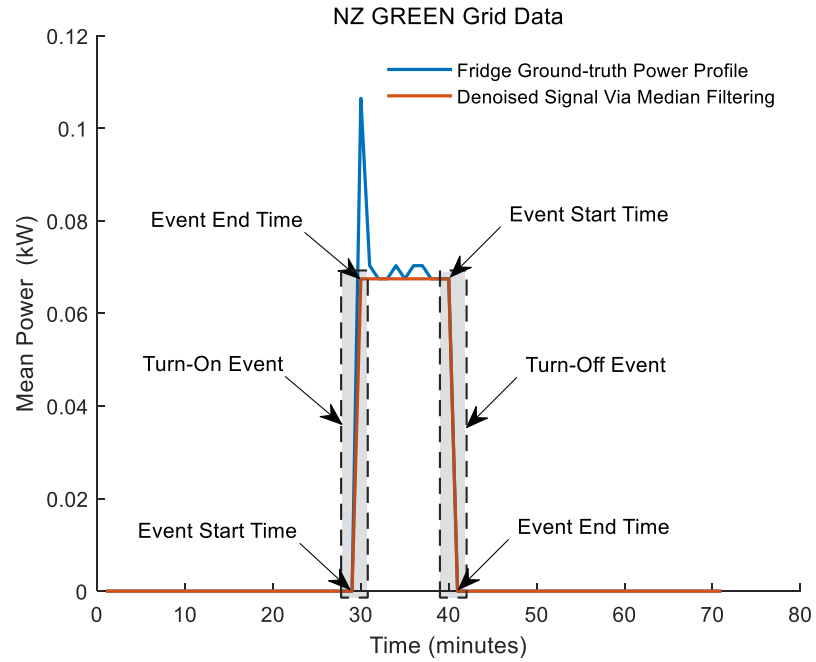


Figure 11 Graphical Depiction of Events

The purpose of the event detection algorithms is to accurately identify all the events (whether it relates to turning-on or turning-off of an appliance/circuit as shown in Figure 11) within the aggregate load profile. Numerous event detection algorithms are available in the existing NILM literature, discussed in Section 2.2.1, however, to contribute further to the existing state of the art on event detection algorithms for the NILM system, this research work proposes three new event detection algorithms.

### 3.2.1 Proposed Event Detection Algorithms

As discussed earlier: Section 2.2.1, numerous event detection algorithms are proposed by the research community using diverse input features, data granularity, and techniques. However, there are attributes, such as ease of implementation, computational efficiency, and robustness, that need to be further researched in the context of NILM event detection. Therefore, in this research work three different event detection algorithms, focusing on the mentioned attributes, are proposed. To encompass these attributes, the proposed algorithms are based on single input feature, iterative process, and straightforward statistical measures. These properties of the proposed event detection algorithms make them computationally efficient, simple to implement, and adaptable to diverse data.

Hence, for event detection within the aggregated pre-processed load profile, the proposed algorithms use a single input feature, i.e., real power, and iteratively track different statistical values, i.e., mean ( $\mu$ ), variance ( $\sigma^2$ ), and mean absolute deviation (MAD). Therefore, the algorithms are named as the mean sliding window (MSW) [89], variance sliding window (VSM) [90], and mean absolute deviation sliding window (MAD-SW) [90], respectively. It is worth noting that with the mentioned attributes, the proposed event detection algorithms not only advance the existing state of the art on NILM event detection but also significantly contribute to the low complexity and computationally efficient event-based NILM systems.

The basic working principle of all three proposed event detection algorithms are based on the same phenomenon, i.e., there must be a variation in the statistical values of the load data after each event as compared to the previous steady-state statistical values (can be visualized from Figure 11). To keep track of these statistical variations, a concept of sliding window, having width ' $\omega$ ', is employed that runs over the pre-processed aggregated power profile curve and track different statistical variations to detect different events within the aggregated load data. Figure 12 graphically depicts the working principle of the proposed event detection algorithms based on the sliding window concept for tracking different statistical parameters of aggregated load data.

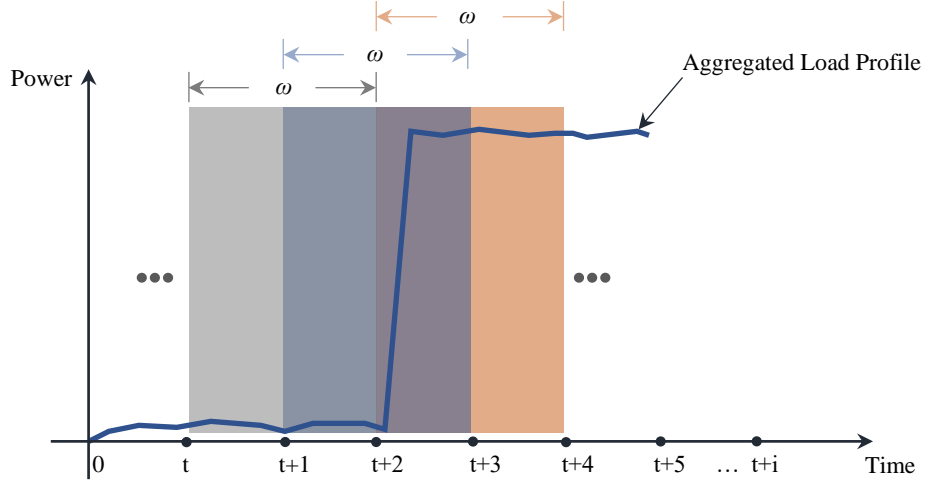


Figure 12 Algorithms' Working Principle Based on Sliding Window Concept  
Different color rectangles represent sliding windows that track the statistical features of the given power profile.

In terms of outcome, all three proposed event detection algorithms not only detect the events within the aggregated load profile but also identify the starting and ending time indices of each detected event.

### 3.2.1.1 MSW Algorithm

The MSW algorithm is based on the sliding window concept as discussed earlier and tracks the mean value of the aggregated load and takes the difference of the consecutive mean values to identify whether an event has occurred or not.

Suppose  $x$  represents the pre-processed aggregated load data, the mean value is mathematically given as in (8).

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

A detailed description of the MSW algorithm for the event detection within pre-processed aggregated load data based on the mean value, given in (8), is presented in Table 6 [49, 89].

Table 6 MSW Algorithm Description

<b>Input</b>	Pre-processed Aggregated Load Data
<b>Output</b>	Starting and Ending Time Indices of Detected Events
<ol style="list-style-type: none"> <li>1. Select sliding window width '<math>\omega</math>'</li> <li>2. Compute iteratively the mean values '<math>\mu_x</math>' of the input data using (8)</li> <li>3. Determine the difference among the consecutive means values</li> <li>4. Select a threshold value '<math>\delta</math>' for event detection</li> <li>5. Compare the results of step 3 with '<math>\delta</math>' to compute a thresholding signal indicating steady-state and transient portions</li> <li>6. Identify the edges using derivation function and extract the starting and ending time indices</li> <li>7. Post-processing for final approval of the starting and ending time indices and delay correction for ending time indices due to sliding window width.</li> </ol>	

### 3.2.1.2 VSW and MAD-SW Algorithms

The VSW and MAD-SW algorithms also rely on the sliding window concept but here the sliding window tracks the variance and mean absolute deviation, respectively, of the power signal to detect the events within the aggregate load profile. The corresponding two statistical features are given in (9) and (10), respectively.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n |x_i - \mu_x|^2 \quad (9)$$

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \mu_x| \quad (10)$$

Both algorithms take the pre-processed aggregated load data as an input and provide an output in the form of the starting and ending time indices of each detected event. The detailed working principle of the VSW and MAD-SW algorithms is presented in Figure 13 [90].

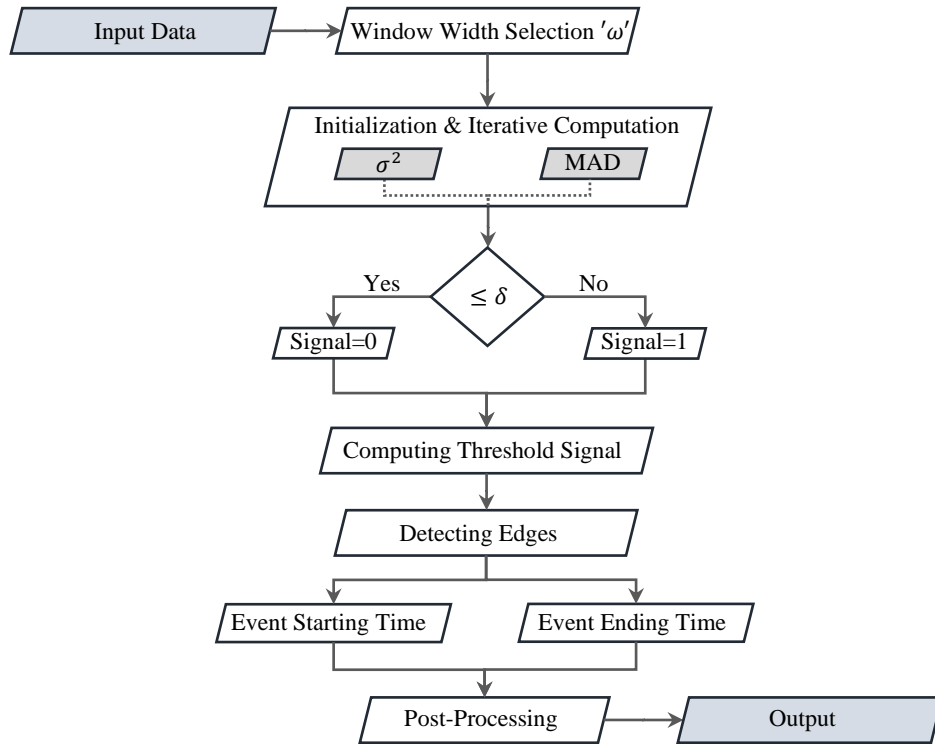


Figure 13 Working Principle of VSW and MAD-SW

The proposed event detection algorithms, within this research work, are not only simple to implement but also comprise low complexity and faster computation, as built on simple statistical features and iterative processes, respectively.

### 3.2.2 Event Detection Simulations and Evaluation

Comprehensive simulations are carried out in terms of testing, evaluations, and validation of the proposed event detection algorithms. For the said purposes, MATLAB is used as a simulation toolkit. All simulations are carried out on real-world load measurements and to further validate the robustness, the proposed algorithms are independently tested on two different electricity load databases having a diverse set of attributes, i.e., geographical location, availability of appliances/circuits, and installation configuration. Figure 14 presents the flow of simulation study carried out for event detection.

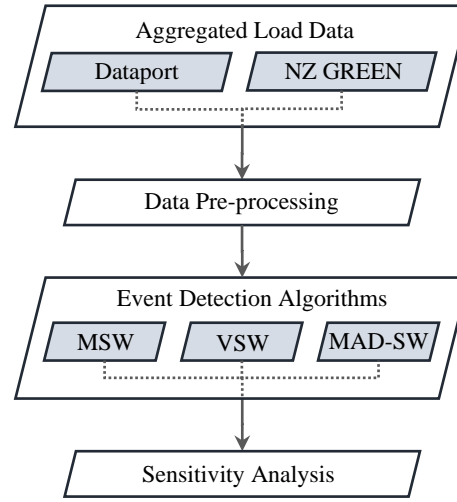


Figure 14 Event Detection Simulation Flow

It is worth noting that the starting time indices of the detected events are considered for evaluation purposes because it is the starting time of a given event which initiates that particular event, whether it is a turn-on or turn-off event, as shown in Figure 11. Further, for evaluation and validation purposes, events detected by the proposed algorithms will be considered as true events if and only if the starting time indices of the detected events exactly match with the starting time indices of the ground-truth events, available in both databases employed in this research work.

#### 3.2.2.1 Dataport

To evaluate and validate the performance of the proposed algorithms, real-world load measurements (aggregated and appliance ground-truth) for 15 days have been acquired from the Dataport database. Further, two different load elements, i.e., electric vehicle (EV) charging and air conditioning unit (AC), have been selected for non-intrusive load classification. Under the given conditions, EV and AC are selected due to their characteristics, i.e., they are not only high consumption load elements but also shiftable-interruptible [93] load elements. These attributes make EV and AC as high potential load

elements from the perspective of flexibility control and energy saving. Moreover, for the given low data granularity of the employed database, these high consumption load elements are more viable options for non-intrusive inference [4, 17].

Table 7 presents the details of different attributes of the Dataport database associated with digital simulations of the proposed event detection algorithms, i.e., MSW, VSW, and MAD-SW.

Table 7 Event Detection Simulation Attributes for Dataport

<b>Dataport, Pecan Street Inc.</b>	
Data Granularity	1/60 Hz, i.e., 1-minute sampling
Data (Household) ID	26
Data Acquisition Timeframe	June 18, 2014 – July 02, 2014
Total Number of Data Samples	21600
Ground-truth Events	334
Pre-processing Technique	Median Filtering
Window Width ' $\omega$ '	6 samples
Threshold Value	250 W

### 3.2.2.1.1 MSW Algorithm

As per the attributes presented in Table 7, comprehensive simulations for the MSW algorithm are carried out according to the strategy shown in Figure 14 and the working principle presented in Table 6. As per the simulation results, a total of 323 events are detected within the input load data. A portion (for visual clarity of event detection results) of the extracted results in terms of the starting and ending time indices of the detected events by the MSW algorithm is depicted in Figure 15.

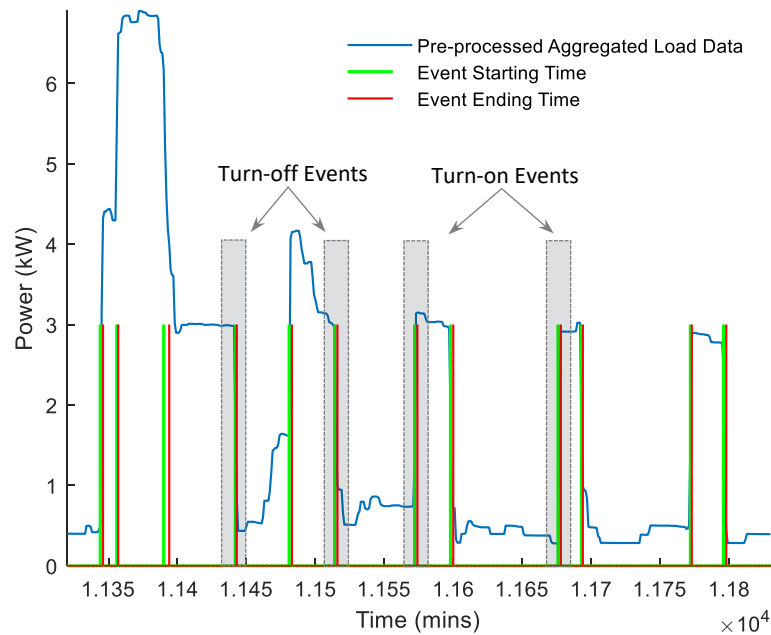


Figure 15 Event Detection Results by MSW Algorithm

Green and red vertical lines are used for the visual representation of event detection outcomes, i.e., starting and ending time indices, respectively.

As evident from Figure 15, the MSW algorithm not only successfully detected the events within the aggregated load data but also identified the starting (represented by green vertical lines in Figure 15) and ending (represented by red vertical lines in Figure 15) time indices of each detected event. To further validate the detected events by the MSW algorithm, it is benchmarked against the ground-truth of the load elements, i.e., EV and AC; the corresponding results are presented in Figure 16.

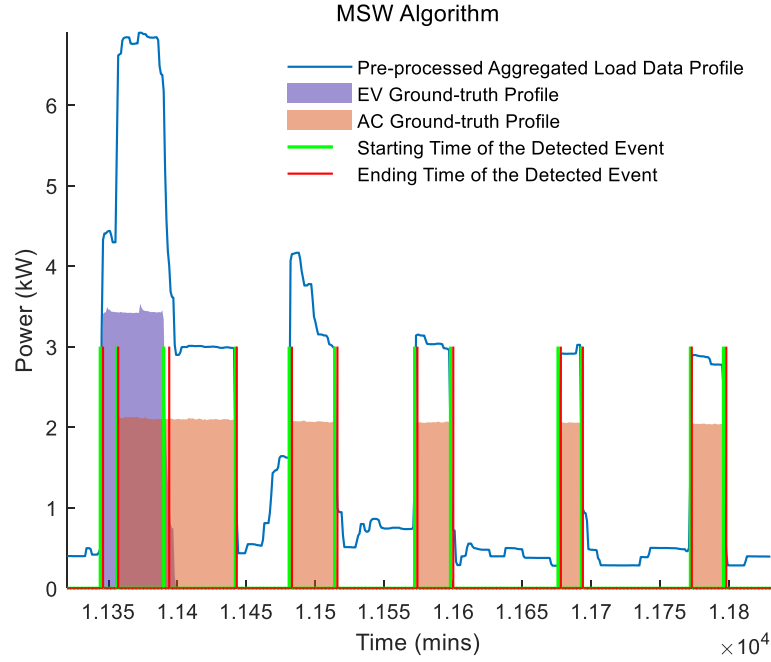


Figure 16 MSW Algorithm Results Validation

As seen in Figure 16, the MSW algorithm successfully detected the turning-on and turning-off events of the load elements under consideration. A complete overview of the MSW performance evaluation, for the simulation attributes presented in Table 7, is presented in Table 8 in terms of different performance metrics.

Table 8 MSW Algorithm Performance Results

<b>Total Detected Events</b>	<b>323</b>
True Positive	286
False Positive	37
False Negative	47
Precision	88.54 %
Recall	85.88 %

As seen in Table 8, for the given simulation attributes (Table 7) particularly  $\omega=6$ , MSW algorithm attained promising results, i.e.,  $> 85\%$ , in terms of both precision and recall performance metrics. It is worth noting that the proposed event detection algorithms are built on the concept of the sliding window, making the window width ' $\omega$ ' a key input parameter of the proposed algorithms that may substantially influence the performance of the proposed algorithms. Hence to investigate the influence of window width, a



comprehensive sensitivity analysis in terms of  $\omega$  has been carried out. The presented sensitivity analysis not only underlines the influence of  $\omega$  but also investigate the optimal value of  $\omega$ , under given conditions, for future simulation studies in terms of load feature extraction and load classification. Based on the performed sensitivity study, Table 9 presents the MSW algorithm performance results for different values of window width ' $\omega$ ', where the rest of the simulation attributes are kept constant as presented in Table 7.

Table 9 MSW Sensitivity Study Results

	Number of Samples	Window Width							
		2	3	4	5	6	7	8	9
MSW	Total Detected Events	423	384	370	341	323	312	303	272
	True Positive	162	194	221	268	286	276	238	162
	False Positive	261	190	149	73	37	36	65	110
	False Negative	171	139	112	65	47	57	95	171
	Precision (%)	38.29	50.52	59.73	78.59	88.54	88.46	78.54	59.55
	Recall (%)	48.64	58.25	66.36	80.48	85.88	82.88	71.47	48.64

As evident from the results presented in Table 9, by initially increasing the window width ' $\omega$ ', the numbers of false positive decreases, subsequently, true positive increases that lead to an increase in precision performance metric. Similarly, the numbers of false negative decreases initially leading to an increase in recall performance metric. But as seen from the presented results after a certain value of window width, i.e.,  $\omega=6$ , the numbers of false positive and false negative detections start ascending. Consequently, a descending trend is followed by the precision and recall performance metrics, respectively. The presented analysis and performance trend of the MSW algorithm in terms of precision and recall performance metrics against different window width can be seen in Figure 17.

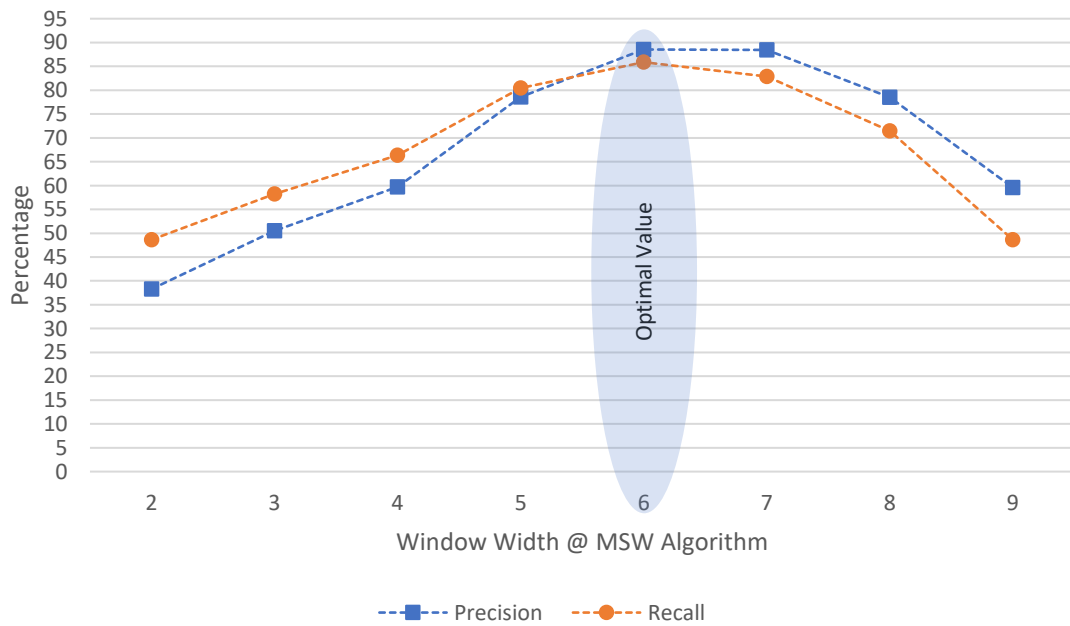


Figure 17 MSW Algorithm Performance Trend for Different Window Width ' $\omega$ '

Based on the extracted results presented in Table 9 and Figure 17, it is concluded that for the given conditions, the optimal window width for accurate event detection by MSW algorithm is 6 samples, i.e.,  $\omega=6$ , also highlighted in Figure 17.

### 3.2.2.1.2 VSW Algorithm

Comprehensive simulation studies have been carried out for the VSW algorithm according to the algorithm's working principle and event detection simulation strategy presented in Figure 13 and Figure 14, respectively. The simulation attributes are kept the same as presented in Table 7. For the given attributes, a total of 324 events were detected by the VSW algorithm. A portion of the extracted results in terms of detected events and corresponding starting and ending time indices by the VSW algorithm is presented in Figure 18. For validation purposes, Figure 18 also presents the comparison of the detected events with ground-truth events of the desired load elements, i.e., EV and AC.

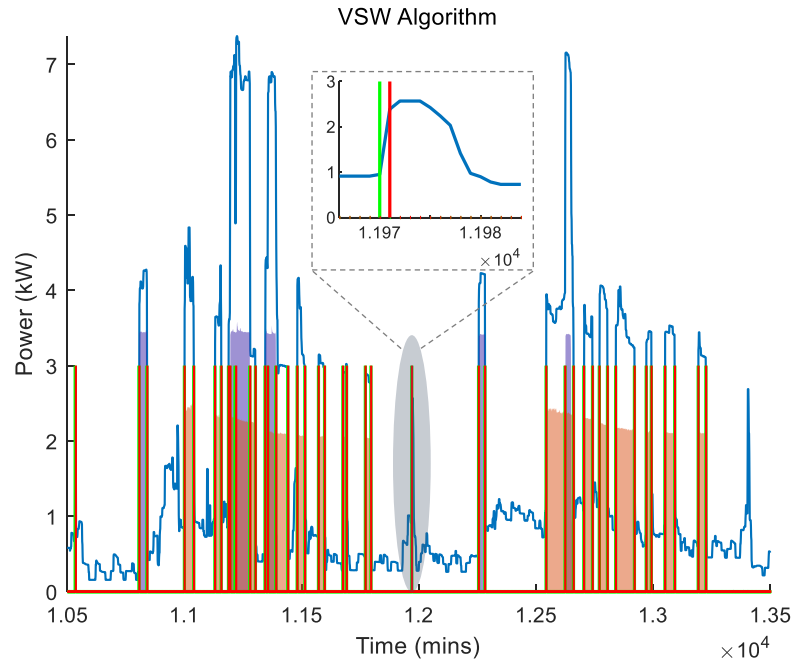


Figure 18 VSW Algorithm Event Detection Results

The blue line represents the pre-processed aggregated load data profile, where purple and orange color shaded areas represent the ground-truth profile of EV and AC, respectively. In terms of event detection, the green and red vertical lines represent the starting and ending time indices of the detected events by the VSW algorithm.

It is evident from the presented results in Figure 18 that most of the ground-truth events are detected precisely by VSW albeit with some false detections. Figure 18 also depicts one of the false detections in the form of a close-up visual, highlighted by the shaded area. For further detailed analysis, a portion of the extracted results is presented in Table 10 in the form of the starting time indices comparison between events detected by the VSM algorithm and ground-truth events.

Table 10 Comparison of Detected Events by VSW and Ground-truth Events

Sequence of Detected Events	Starting Time Indices of the Events		
	Detected by VSW Algorithm	Ground-truth	
		AC	EV
1	16	16	-
2	54	54	-
3	96	95	-
14	1101	1101	-
15	1135	1135	-
16	1185	-	1185
17	1238	1238	-
18	1272	1272	-
19	1281	-	1281
20	1298	-	1298
21	1318	1317	-
-	Not Detected	3715	-
-	Not Detected	4153	-
97	6210	6210	-
98	6234	-	6234
99	6247	-	6246
100	6502	-	6502
101	6615	-	6614
102	6643	6643	-
103	6686	6686	-
-	Not Detected	9780	-
147	9837	-	9836
148	9879	9878	-
149	9923	9922	-
150	9949	-	9949
-	Not Detected	-	10020
207	13968	13968	-
208	14030	14030	-
209	14083	14083	-
261	16945	Actual Event did not Occur	
262	16970	Actual Event did not Occur	
321	21058	-	21058
322	21083	-	21083
323	21103	21103	-
324	21160	-	21159

As evident from the presented results, most of the ground-truth events are precisely detected albeit with some false and misdetection, as shown in Table 10. For example, EV and AC trigger events (ground-truth) at time indices 10020 and 9780 respectively, but the VSW algorithm completely missed these events leading to false negatives, i.e., misdetections. Further, the VSW algorithm detects events at time indices 16945 and 16970, but no ground-truth event occurred at the said time indices thereby leading to false positive, i.e., false detection. Further as discussed earlier detected events will be considered as true (events) detection if and only if the starting time of the detected events exactly match with the starting time of the ground-truth events. Hence, the events detected at time indices 96, 1318, 6247, ..., are not considered as true events, i.e., true positive, due to the delay of a 1-time index. As seen in Table 10, this marginal delay variation may substantially decrease the number of true detections, consequently, degrades the event

detection performance. To address this issue, another parameter namely, delay tolerance, is introduced. However, this will be further discussed in section 3.2.2.2<sup>22</sup> and section 4.3.1<sup>23</sup> of this thesis along with a comprehensive sensitivity analysis to investigate the influence of delay tolerance factor on the performance of the proposed event detection algorithms.

Table 11 presents the overall performance results of the VSW algorithm, for the simulation attributes presented in Table 7, in terms of TP, FN, FP, precision, and recall.

Table 11 VSW Algorithm Performance Results

Total Detected Events	324
True Positive	261
False Positive	63
False Negative	70
Precision	80.55 %
Recall	78.85 %

As seen in Table 11, the VSW performance in terms of precision and recall is lagging compared to the MSW algorithm performance, as presented in Table 8. However, it is worth noting that both algorithms are different in terms of methodology where the presented simulation results are based on the same sliding window width, i.e.,  $\omega=6$ . And it is already established that  $\omega=6$  is an optimal window width for the MSW algorithm, as shown in Table 9, but for VSW it may vary. Therefore, similarly to the sensitivity analysis performed for the MSW algorithm, comprehensive simulations are carried out to identify the optimal  $\omega$  value that facilitates the VSW event detection performance. In this context, Table 12 presents the VSW algorithm performance results for different values of sliding window width where the other attributes of simulations are kept constant as presented in Table 7.

Table 12 VSW Sensitivity Study Results

		Window Width							
	Number of Samples	2	3	4	5	6	7	8	9
VSW	Total Detected Events	378	361	346	331	324	315	298	282
	True Positive	298	309	306	283	261	245	231	217
	False Positive	80	52	40	48	63	70	67	65
	False Negative	34	23	25	48	70	86	100	114
	Precision (%)	78.83	85.59	88.43	85.49	80.55	77.77	77.51	76.95
	Recall (%)	89.75	93.07	92.44	85.49	78.85	74.01	69.78	65.55

As evident from the results presented in Table 12, the VSW algorithm achieved the best performance at  $\omega=4$ . Further in terms of overall performance concerning  $\omega$ , the VSW algorithm follows a similar phenomenon as observed for the MSW algorithm. Initially

<sup>22</sup> For NZ GREEN Grid Database

<sup>23</sup> For Pecan Street's Dataport Database

the algorithm performance increase with an increase in window width due to decrease in false negative and false positive detections but after a certain value, here  $\omega=4$ , the performance starts degrading with an increase in window width due to the consistent rise in the numbers of FN and FP detection (as seen in Table 12). The following performance trend of the VSW algorithm for different window width ' $\omega$ ' can be seen in Figure 19.

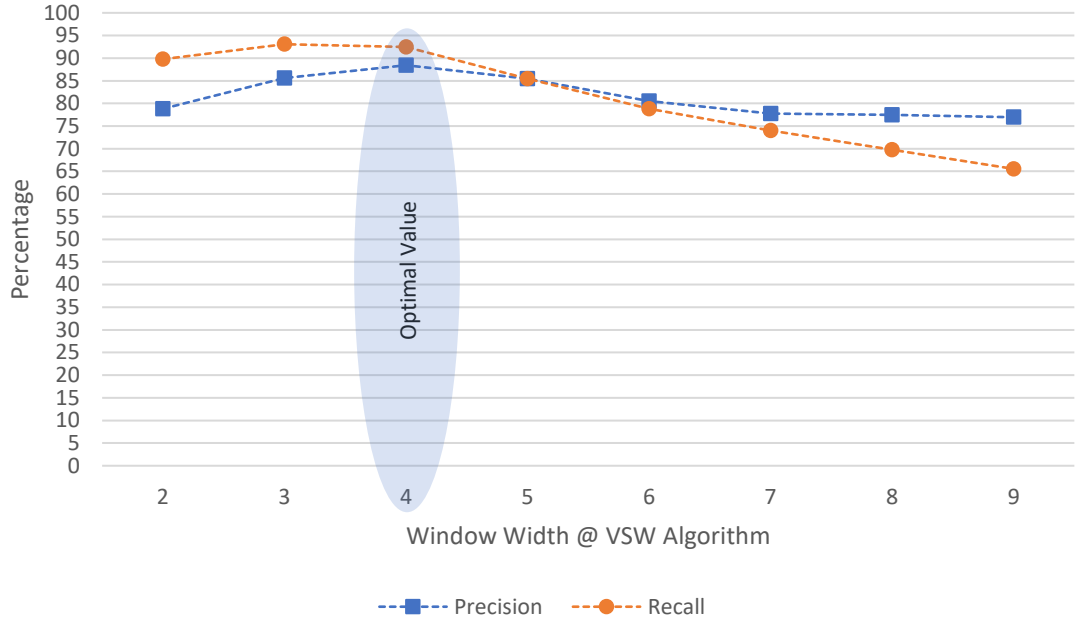


Figure 19 VSW Performance Trend for Different Window Width ' $\omega$ '

Based on the presented sensitivity analysis, it is concluded that the VSW algorithm performs best at a sliding window width of 4 samples (also highlighted in Figure 19) by achieving the performance results of 88.43% and 92.44% in terms of precision and recall performance metrics, respectively.

### 3.2.2.1.3 MAD-SW Algorithm

Similarly, to the MSW and VSW algorithms, comprehensive simulations are carried out for the MAD-SW algorithm as per the attributes presented in Table 7. For said simulation attributes, the MAD-SW algorithm detected a total of 343 events within the input aggregated load data. A portion of the said event detection results is shown in Figure 20, where the reference frame, i.e., time indices frame, is kept the same as presented for the VSW algorithm's results in Figure 18. Keeping the same reference frame for the graphical representation of the event detection result facilitates an in-depth comparison of different algorithms' performance. For example, likewise to the VSW algorithm most of the desired events are accurately detected by the MAD-SW algorithm, however, some variations in

terms of false detections are noted in the case of MAD-SW algorithm, highlighted by close-up in Figure 20, compared to Figure 18.

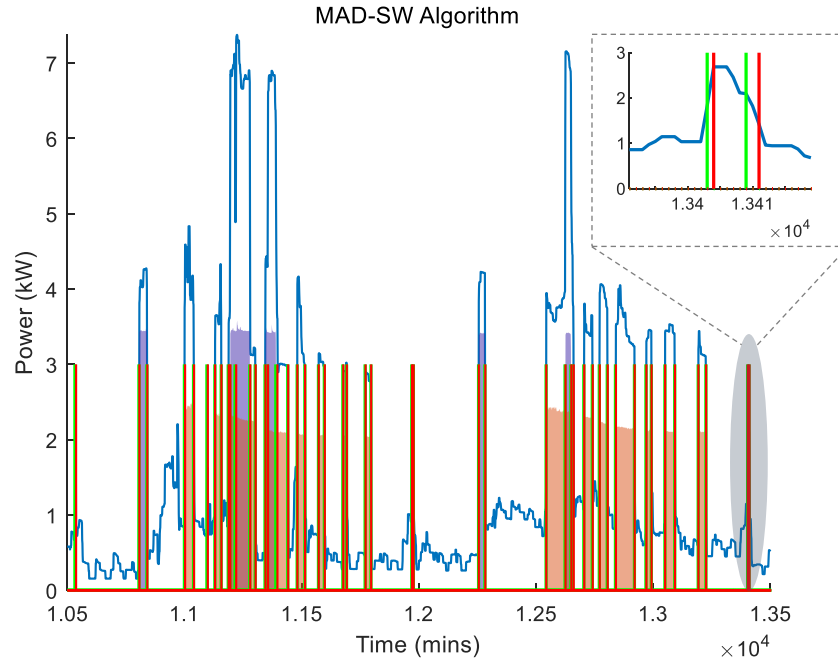


Figure 20 MAD-SW Algorithm Event Detection Results

The blue line represents the pre-processed aggregated load data profile, where purple and orange color shaded areas represent the ground-truth profile of EV and AC, respectively. In terms of event detection, the green and red vertical lines represent the starting and ending time indices of the detected events by the MAD-SW algorithm.

Overall, similarly to the MSW and VSW algorithm results presented in Figure 16 and Figure 18, respectively, it is also evident from the results presented in Figure 20 that the MAD-SW algorithm also detected the ground-truth events while successfully avoiding the detection of most of the low to mid (undesired) consumption peaks. This is anticipated and desired due to the pre-selected simulations' attributes and the granularity of the acquired load data [90, 94].

Similarly, to the VSW algorithm, Table 13 presents the validation of MAD-SW algorithm performance by highlighting the starting time indices of a portion of the detected events by the MAD-SW algorithm and its comparison with the ground-truth events of the load elements under consideration, i.e., EV and AC. From the presented results in Table 13, it is also evident that most of the individual appliances' ground-truth events within the aggregated load data are accurately detected albeit with some false and misdetection. Further, some events are not classified as true events just because of a marginal variation in the detected and ground-truth event starting time indices, e.g., the event detected by MAD-SW at time index 20170.

Table 13 Comparison of Detected Events by MAD-SW and Ground-truth Events

Sequence of Detected Events	Starting Time Indices of the Events		
	Detected by MAD-SW Algorithm	Ground-truth	
		AC	EV
1	16	16	-
2	54	54	-
3	95	95	-
20	1298	-	1298
21	1317	1317	-
22	1326	1326	-
23	1392	1392	-
24	1432	1432	-
76	4874	-	4874
77	4916	-	4916
78	5117	-	5117
79	5218	5218	-
-	Not Detected	5361	-
-	Not Detected	6800	-
149	9774	-	9774
150	9780	9780	-
151	9836	-	9836
152	9878	9878	-
216	13968	13968	-
217	14011	-	14011
218	14030	14030	-
219	14083	14083	-
220	14119	-	14119
317	19960	-	19960
318	19989	-	19989
319	20015	-	20014
320	20115	Actual Event did not Occur	
321	20170	20171	-
322	20251	20251	-
338	21058	-	21058
339	21083	-	21083
340	21103	21103	-
341	21125	Actual Event did not Occur	
342	21159	-	21159
343	21174	Actual Event did not Occur	

For the given simulation parameters, i.e., Table 7, the overall performance of the MAD-SW algorithm is presented in Table 14.

Table 14 MAD-SW Algorithm Performance Results

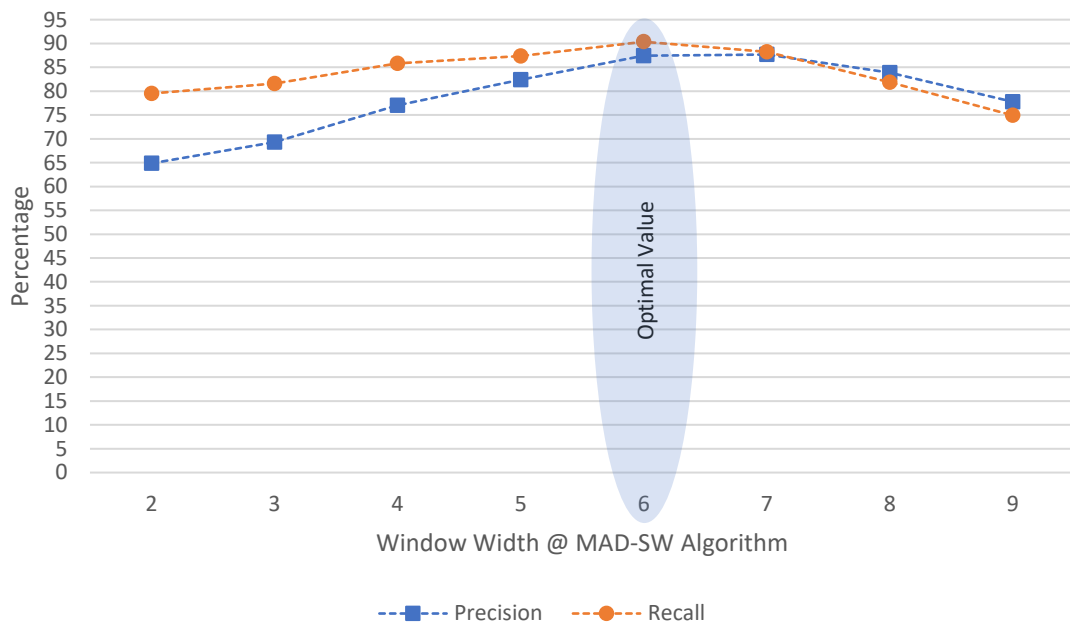
<b>Total Detected Events</b>	<b>343</b>
True Positive	300
False Positive	43
False Negative	32
Precision	87.46 %
Recall	90.36 %

Similarly, to the MSW and the VSW algorithms, a comprehensive sensitivity study for the MAD-SW algorithm in terms of window width is also carried out to investigate an optimal window width for the MAD-SW algorithm under given conditions. Table 15 presents the performance results of the MAD-SW algorithm at different window width.

Table 15 MAD-SW Sensitivity Study Results

		Window Width							
	Number of Samples	2	3	4	5	6	7	8	9
MAD-SW	Total Detected Events	407	391	370	352	343	333	323	319
	True Positive	264	271	285	290	300	292	271	248
	False Positive	143	120	85	62	43	41	52	71
	False Negative	68	61	47	42	32	39	60	83
	Precision (%)	64.86	69.30	77.02	82.38	87.46	87.68	83.90	77.74
	Recall (%)	79.51	81.62	85.84	87.34	90.36	88.21	81.87	74.92

From the presented result, a similar phenomenon and trend are noted for the MAD-SW algorithm as observed for the MSW and the VSW algorithms. This can be further validated from the graphical representation shown in Figure 21.

Figure 21 MAD-SW Performance Trend at Different Window Width ' $\omega$ '

From the presented sensitivity results, it is concluded that the MAD-SW algorithm performs best at  $\omega=6$  (highlighted in Figure 21) with precision and recall values of 87.46 and 90.36 percent, respectively.

#### 3.2.2.1.4 Inter-Algorithm Comparison

All three proposed event detection algorithms are individually evaluated on a real-world database: Pecan Street's Dataport. A sensitivity analysis in terms of window width ' $\omega$ ' has been carried out to investigate the optimal value for each algorithm that yields the best event detection performance. This section presents a comprehensive inter-algorithm comparative analysis to investigate which algorithm performs best under the given conditions. For the inter-algorithm comparative analysis, f-score, given in (4), is used because it takes both precision and recall performance metrics into account. Table 16



presents the proposed algorithms' performance results in terms of the f-score measure at different window width.

Table 16 Inter-Algorithm Comparison

Number of Samples	Window Width							
	2	3	4	5	6	7	8	9
MSW F-Score (%)	42.85	54.11	62.87	79.52	87.19	85.58	74.84	53.55
VSW F-Score (%)	83.94	89.17	90.39	85.49	79.69	75.85	73.44	70.79
MAD-SW F-Score (%)	71.44	74.96	81.19	84.79	88.88	87.95	82.87	76.30

As observed earlier, each algorithm performs best at a specific value of window width, which is also evident from the presented results in terms of the f-score measure. VSM algorithm attained the f-score of 90.39% at  $\omega=4$ , where MAD-SW and MSW algorithms achieved the best performance at  $\omega=6$ , i.e., f-score of 88.88 and 87.19 percent, respectively. It is evident that at respective optimal window width, each event detection algorithm performs well by attaining the minimum f-score performance of  $> 87\%$ . In terms of inter-algorithm performance at corresponding optimal window width, the MSW algorithm performance is lagging compared to the VSW and the MAD-SW algorithms, however, the performance variation is marginal.

To further analyze the event detection performance of the proposed algorithms for the complete set of window width values, i.e.,  $\omega=2, 3, 4, \dots, 8, 9$ , box plot representation is employed for evaluation purposes. Figure 22 presents the f-score performance results of all three proposed event detection algorithms, in the form of a box plot, for the entire set of window width values.

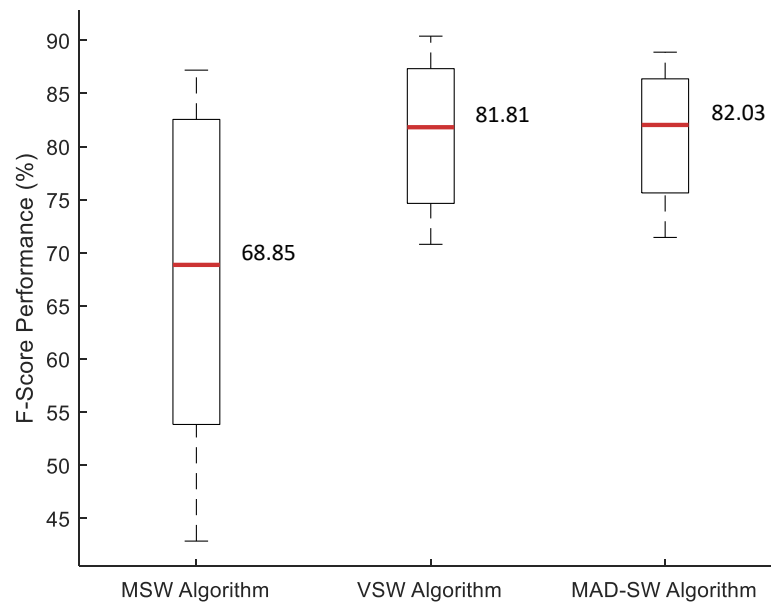


Figure 22 Inter-Algorithm Comparison in terms of F-Score  
The red horizontal lines and corresponding numeric values represent the median f-score performance.

Based on the presented results and corresponding analysis, it is concluded that the sliding window used to track different statistical features for identifying the events has a significant influence on the overall performance of the event detection. The window width is linked with the number of total events detected subsequently influencing the number of true positive, false positive, and false negative detections. Hence an optimal selection of window width ' $\omega$ ' is a key to the best performance of the proposed algorithms and it varies with the algorithm.

### 3.2.2.2 NZ GREEN Grid

To further validate the performance of the proposed event detection algorithms and evaluate its robustness, another real-world electricity load database: NZ GREEN Grid, having diverse individual circuits has been employed. In this context, the MAD-SW algorithm is selected to be further tested on NZ GREEN Grid, as the MAD-SW yields slightly more robust/stable performance compared to the VSW, as seen in Figure 22. Simulation studies are carried out as per strategy presented in Figure 14, however, to further validate the effectiveness of the proposed event detection algorithm, the input sample size has been doubled compared to the sample size of previous simulations performed for Dataport. Therefore, a real-world household with 30 days of load data has been selected for the simulation studies. The input data have been comprised of days from different months of the year to accommodate the diversity of the different seasons and corresponding consumption patterns. Table 17 [91] presents different attributes of the MAD-SW algorithm simulations for the NZ GREEN Grid database.

Table 17 NZ GREEN Grid Event Detection Simulation Attributes

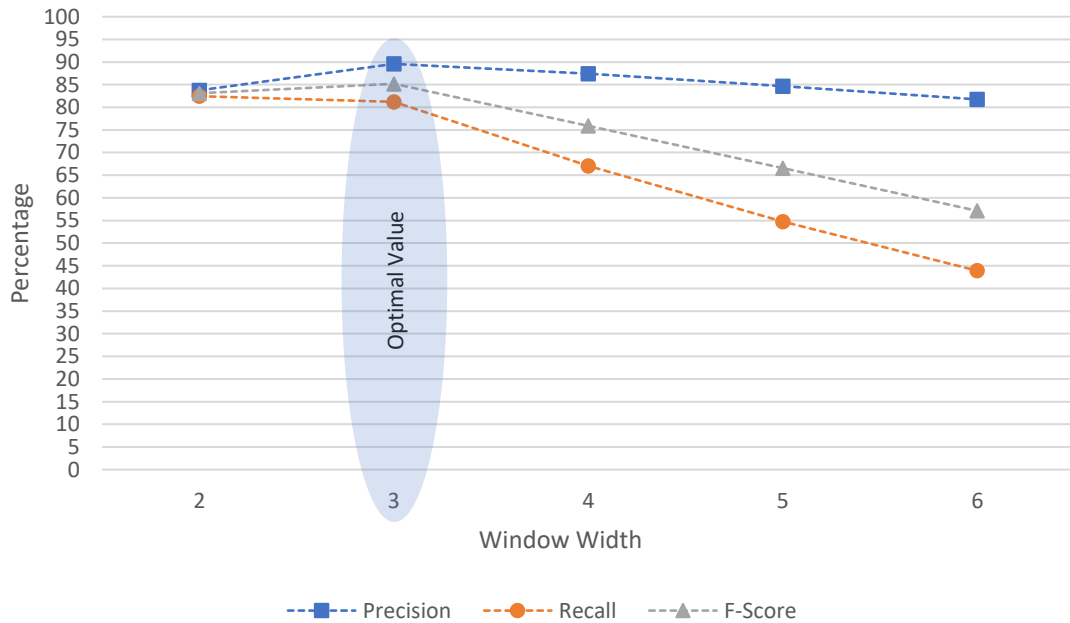
<b>NZ GREEN Grid Database</b>	
Data Granularity	1/60 Hz, i.e., 1-minute sampling
Data (Household) ID	rf_01
Data Acquisition Timeframe	2014: Mar. 11-15   Apr.11-13   May. 12-13 Jun. 12-15   Jul. 14-15   Aug. 11-15 Sep. 11-14   Oct. 11-15
Total Number of Days	30
Total Number of Data Samples	43200
Pre-processing Technique	Median Filtering
Threshold Value	150 W

As evident from the event detection simulations for the Dataport database, the proposed algorithms perform best at a certain optimal value of the sliding window width. Hence, to identify the optimal window width for the given condition, i.e., Table 17, a comprehensive sensitivity analysis is carried out. Table 18 [91] presents the results of the said sensitivity study in terms of different performance metrics, i.e., precision, recall, and f-score, given in (2)-(4), respectively.

Table 18 MAD-SW Sensitivity Analysis in terms of  $\omega$  for NZ GREEN Grid

MAD-SW Algorithm   NZ GREEN Grid Database					
Number of Samples	Window Width ' $\omega$ '				
	2	3	4	5	6
Total Detected Events	3651	3367	2853	2412	2005
True Positive	3058	3016	2495	2042	1639
False Positive	593	351	358	370	366
False Negative	651	698	1224	1684	2093
Precision (%)	83.76	89.58	87.45	84.66	81.75
Recall (%)	82.45	81.21	67.09	54.80	43.92
F-Score (%)	83.10	85.19	75.93	66.54	57.14

It is noted from the presented results that for the NZ GREEN Grid database the MAD-SW algorithm follows the same phenomenon and trend in terms of window width sensitivity as observed for Dataport database, discussed in section 3.2.2.1.1 and section 3.2.2.1.2. To visualize the corresponding performance trend, Figure 23 graphically depicts the corresponding results, particularly precision, recall, and f-score. It is evident from the results presented in Table 18 and Figure 23 that the MAD-SW algorithm performs the best at window width of 3, yielding the results of 89.58, 81.21, and 85.19 percent in terms of precision, recall, and f-score, respectively.

Figure 23 MAD-SW Sensitivity Analysis in terms of  $\omega$  for NZ GREEN Grid

It is worth noting that all the event detection simulation results presented so far are based on the definition that an event is considered to be true if and only if the starting time index of the detected and ground-truth event exactly matches (discussed earlier in section 3.2.2). But it is also noted in the simulation results, i.e., Table 10 and Table 13, and discussed in section 3.2.2.1.2, that some detected events by the proposed algorithms were not considered as true events due to a marginal variation in the starting time indices of

detected and ground-truth events. Figure 24 graphically depicts the said marginal variation phenomenon that arises in real-world load measurements for further understanding and visualization.

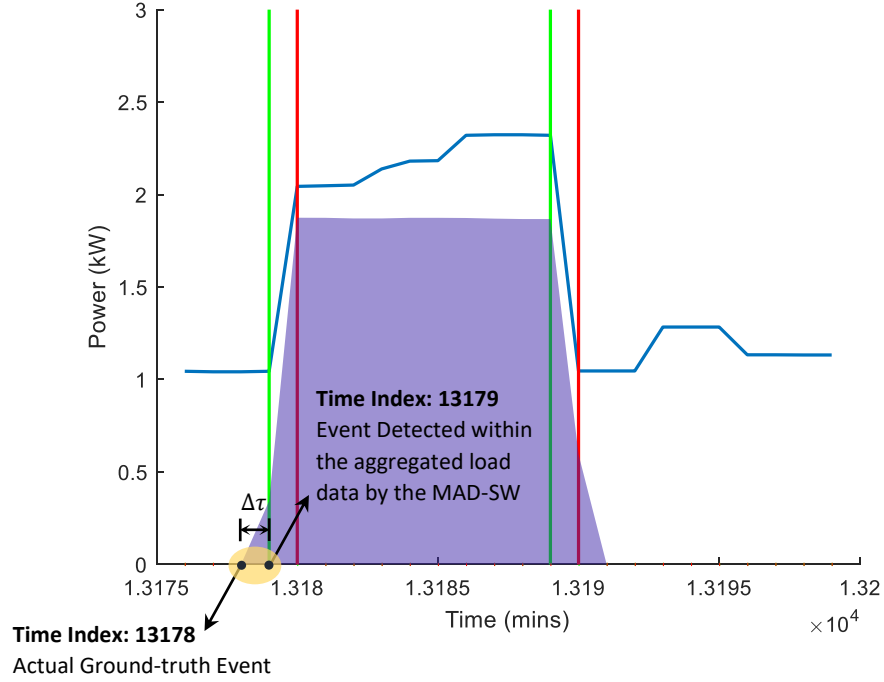


Figure 24 Delay Tolerance Phenomenon

NZ GREEN Grid Household ID. rf\_01: The blue continuous line presents the aggregated load profile, where the shaded area in purple color represents the ground-truth of the water heating circuit. The vertical lines in green and red represent the starting and ending time indices of the detected events by the MAD-SW algorithm.

To accommodate this time delay variation, as shown in Figure 24, a delay tolerance parameter ' $\Delta\tau$ ' is introduced as a post-processing step to facilitate the performance of the proposed algorithms.  $\Delta\tau$  is defined as a condition that must be fulfilled by each detected event to be considered as a true event, mathematically given as in (11).

$$|t_g - t_d| \leq \Delta\tau \quad (11)$$

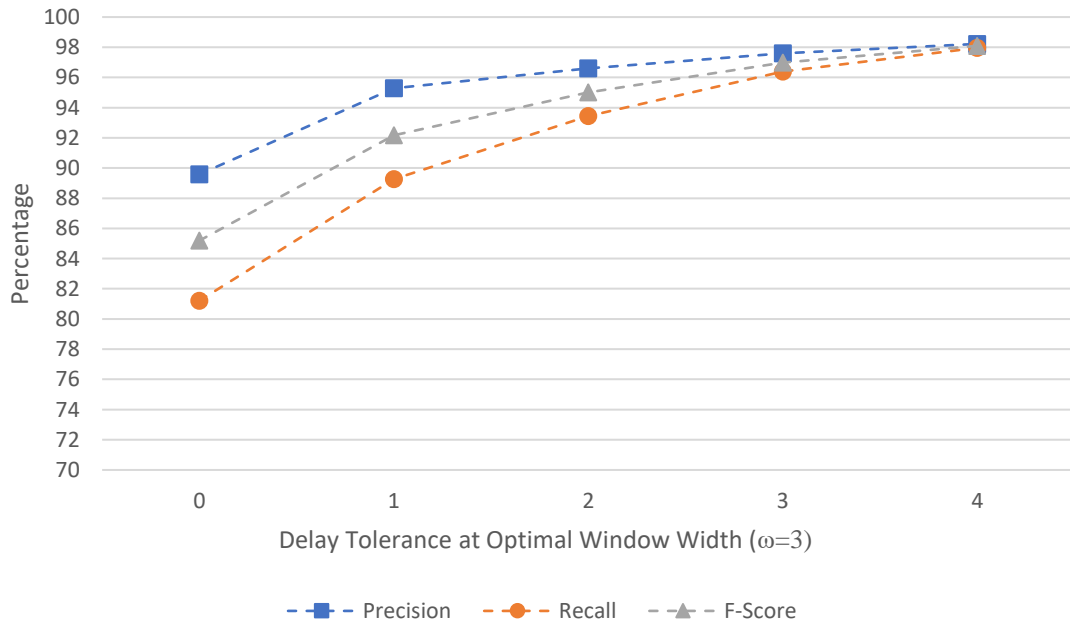
where  $t_g$  and  $t_d$  are the starting time indices of the ground-truth and detected event by the proposed algorithms, respectively. To investigate the influence of  $\Delta\tau$  on the performance of the proposed algorithms, a sensitivity analysis is further extended in terms of  $\Delta\tau$ . Based on the corresponding simulations, Table 19 [91] presents the performance results of the MAD-SW algorithm with the incorporation of delay tolerance factor. It is also worth noting that for the sensitivity simulations in terms of  $\Delta\tau$ , the window width is kept constant at  $\omega=3$  due to the best performance of the MAD-SW algorithm at the given value, as presented in Table 18 and Figure 23.

Table 19 MAD-SW Sensitivity Analysis in terms of  $\Delta\tau$  for NZ GREEN Grid

Window Width	3				
Total Detected Events	3367				
Delay Tolerance	0	1	2	3	4
True Positive	3016	3208	3253	3286	3307
False Positive	351	159	114	81	60
False Negative	698	386	228	123	69
Precision (%)	89.58	95.28	96.61	97.59	98.22
Recall (%)	81.21	89.26	93.45	96.39	97.96
F-Score (%)	85.19	92.17	95.01	96.99	98.09

It is evident from the presented results in Table 19 that the incorporation of delay tolerance significantly facilitates true positive detection. This is anticipated as per the definition of delay tolerance, given as in (11). It is further noted that with an increase in the variation margin, i.e., delay tolerance value, a consistent rise in true positive detection and decline in false positive and false negative detection have been recorded, that yields to a persistent increase in algorithms overall performance in terms of precision and recall, respectively.

The results presented in Table 19 are further graphically depicted in Figure 25 to visualize the performance trend of the MAD-SW algorithm with the incorporation of the delay tolerance ' $\Delta\tau$ ' factor.

Figure 25 MAD-SW Sensitivity Analysis in terms of  $\Delta\tau$  for NZ GREEN Grid

From the presented results, it is determined that  $\Delta\tau$  facilitates event detection performance and is directly proportional to the performance [67]. In terms of optimal  $\Delta\tau$  value selection, a trade-off exists between event detection performance and energy consumption estimation (as an application) at later stages. As high delay tolerance

increases event detection performance, but the higher delay tolerance will also impact the energy estimation, as higher the  $\Delta\tau$  value the higher is the error in the estimated and actual energy consumption. Therefore,  $\Delta\tau=2$  was chosen as the optimal value, as it is evident from the results presented in Table 19 that for  $\Delta\tau>2$ , the event detection performance improvement is marginal. Hence, for the selected optimal parameters, i.e.,  $\Delta\tau=2$  and  $\omega=3$ , the MAD-SW algorithm achieved the best performance result of 96.61, 93.45, 95.01 percent in terms of precision, recall, and f-score, respectively.

### 3.3 Concluding Remarks

This Chapter presented the detailed description of data pre-processing and event detection in terms of techniques, working principles of the proposed algorithms, simulations, and corresponding performance evaluations. In terms of event detection, the newly proposed event detection algorithms are evaluated and validated on distinct real-world electricity load databases having diverse load elements. From the presented results and corresponding analysis, it is concluded that the proposed event detection algorithms perform well and yield promising performance results.

Besides contributing to the existing state of the art on event detection, it is also anticipated that the proposed algorithms will further facilitate the research community to lead towards more robust and efficient event-based NILM systems particularly low data granularity based non-intrusive load disaggregation systems.

Event detection is a pre-requisite for feature extraction within the context of event-based non-invasive load inference approach, as presented in Figure 7. The extracted events need to be further processed to extract the load features, which acts as an input to load classification stage. Consequently, the following chapter, i.e., Chapter 4, is primarily based on the outcomes of this chapter, i.e., Chapter 3, towards further simulation in the context of feature engineering and load classification.

## Chapter 4 Feature Engineering and Load Classification

This chapter discusses the details of feature engineering and loads classification in the context of the non-intrusive load disaggregation. A detailed overview of feature extraction methodology and feature selection techniques adopted in this research work is presented. Further discussion on load classification in terms of supervised machine learning models and ensemble learning models are also part of this Chapter.

The details regarding methodologies, results, and corresponding analysis presented in this chapter are mainly based on [49, 58, 91]. The said research manuscripts have been published or submitted (explicitly mentioned where necessary) as follows:

1. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Intrusive Load Monitoring of Residential Water-Heating Circuit Using Ensemble Machine Learning Techniques," *Inventions*, vol. 5, no. 4, p. 57, 2020.
  - DOI: <https://doi.org/10.3390/inventions5040057>
2. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 2607-2612.
  - DOI: [10.1109/ISGT-Asia.2019.8881113](https://doi.org/10.1109/ISGT-Asia.2019.8881113)
3. A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Comparative Evaluation of Machine Learning Algorithms and Feature Space for Non-Invasive Load Disaggregation", *Journal of Modern Power Systems and Clean Energy* (Submitted: Under Revision)

### 4.1 Introduction

This section presents a brief introduction to feature engineering and load classification techniques that are employed for the proposed research methodologies, presented later in this Chapter.

#### 4.1.1 Feature Engineering

Due to exponential growth in electronic sensors deployment in real-world applications, the data dimensionality increases exponentially leading to the phenomenon of “curse of dimensionality” [95], coined by R. E. Bellman. The said phenomenon refers to the exponential data growth that ultimately leads to high data sparsity and excessive surges in the computational cost of a model. Feature engineering is a key area where different

techniques are employed to cope up with the curse of dimensionality. Feature engineering is a process to transform the raw data into more refined data, commonly referred to as features, that better represent the given problem to the models, subsequently improving models' performance. There are numerous ways to pursue the feature engineering process but based on the existing literature it can be broadly categorized into three, i.e., feature extraction, feature reduction, and feature selection.

Feature extraction is a technique where raw data are transformed into refined information. Feature selection is a process of selecting the most relevant subset of features within a given set of features. Feature reduction, on the other side, refers to the smart grouping of the refined data (features) to reduce the dimensionality of the feature space.

Both feature selection and feature reduction enable the feature dimensionality reduction but in different ways. Feature selection reduces feature space by eliminating irrelevant features from the set of extracted features, where feature reduction yields the same outcome but in terms of a combinatorial process where the actual features are modified. Feature selection has an advantage that the most significant features are selected without altering the actual information [96]. Therefore feature selection techniques are widely adopted by the research community to reduce feature space dimensionality, either to enhance classification models' accuracy or reduce computation time [97].

Based on the existing literature, feature selection techniques can be broadly categorized into three, i.e., filter method, wrapper method, and embedded method. The filter method, a learning model agnostic approach, is mainly based on numerous statistical assessments to identify the significance of the features. Contrary, the wrapper method is a model-dependent approach, based on a feedback method that employs a specific learning model and relies on its performance to evaluate the significance of the feature set [98]. The embedded method tends to optimize the performance of the learning model by extracting and utilizing the relative significance of each feature. A detailed discussion on the aforesaid feature selection techniques is presented in [96-99]. Figure 26 presents a semantic network of feature engineering and related well-known and widely used techniques.

#### **4.1.2 Load Classification**

The recent advancements in computational capabilities enable artificial intelligence to play a key role in many domains including non-invasive load disaggregation, i.e., NILM.



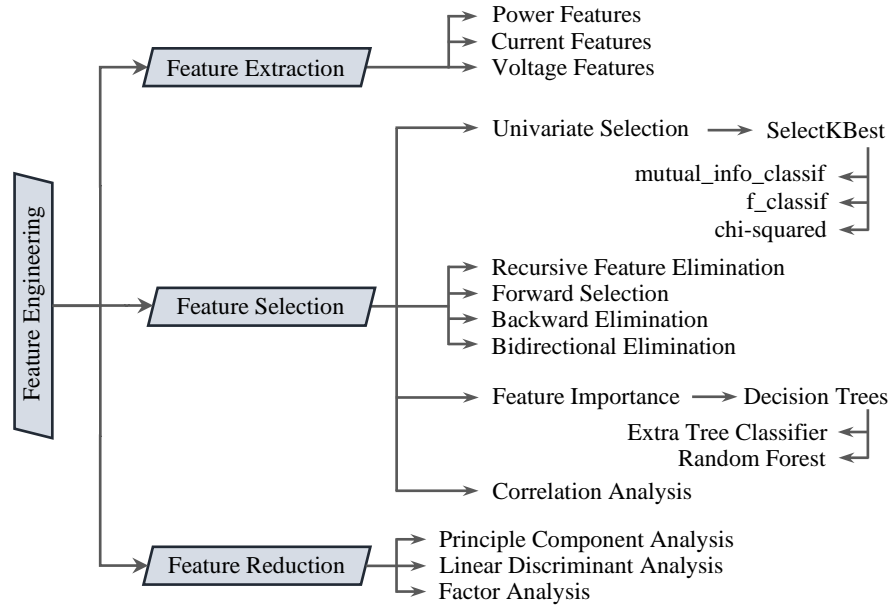


Figure 26 Feature Engineering Semantic Network

Machine learning is one of the artificial intelligence techniques that is widely used for load classification in the NILM domain. In the said context, eight well-known supervised machine learning models, namely support vector machine (SVM), logistic regression (LR), decision trees (DT), random forest (RF), k-nearest neighbors (k-NN), Naïve Bayes (NB), Gaussian process (GP), and multi-layer perceptron (MLP) artificial neural network, are employed for classification purposes in this research work. The aforesaid classifiers are selected due to their diverse attributes, i.e., different working principles, and varied strengths and weaknesses. Consequently, providing an opportunity to evaluate a diverse set of learning models and identify an optimal one for the given problem. A brief description of each employed learning model is presented below [58].

#### 4.1.2.1 Support Vector Machine

A support vector machine is a well-known supervised machine learning model, relying on a ‘margin’ concept, i.e., two data classes separated by a hyperplane [100]. Based on its accurate and robust technique, SVM is widely used by the research community and considered a must-try machine learning model [101]. SVM has not only good generalization capabilities in a wide range of applications but also has tolerance to irrelevant attributes and provides good performance for smaller training data sets [100, 101]. Most importantly, for non-invasive load disaggregation SVM is widely adopted and has established itself as a prominent learning model [3, 34, 37, 50, 102].

#### **4.1.2.2 Logistic Regression**

Logistic regression relies on the statistical models where a logistic curve is fitted to the dataset [48]. LR establishes a logit variable comprised of the natural log of the likelihoods of the class occurring or not. Later, the maximum likelihood estimation algorithm is used to estimate the probabilities [48]. It is also worth noting that despite its name, the LR model is used for classification purposes and has established itself on numerous real-world problems.

#### **4.1.2.3 Decision Trees**

Decision trees rely on a recursive hierarchical structure comprised of nodes and branches. Branches represent the decision rules, where internal and leaf nodes represent features (attributes) and outcomes, respectively. DT is a powerful classification model that is simple to understand and interpret.

#### **4.1.2.4 Random Forest**

Random forest is a combinatorial form of DT models and relies on the prediction of the said DT combinations. In this context, several DT models are trained, later each DT model votes for its preferred class, and the class with a greater number of votes is taken as a final prediction. RF models have been successfully deployed in a wide range of applications. RF models are faster to train and do not overfit irrespective of the number of trees being employed in combination [48].

#### **4.1.2.5 k-Nearest Neighbors**

k-nearest neighbors take the entire training data into account and classify new data according to the class of the majority of its k-number of nearest neighbors [48]. To obtain the nearest neighbors for each data point the algorithm measure the distance between the data points, generally Euclidean distance [48].

#### **4.1.2.6 Naïve Bayes**

Naïve Bayes is a probabilistic learning model established on Bayes theorem for conditional probabilities. It develops and optimizes a function, given that all the attributes in a database are independent. Generally, the maximum likelihood algorithm is used for the NB model's training [48]. NB demonstrates low train and prediction times and has been successfully deployed in complex practical applications [48].

#### **4.1.2.7 Gaussian Process**

The Gaussian process classifier is a supervised learning model designed to solve regression and classification problems. It relies on Laplace approximation and estimates

the conditional probabilities from the given sample. A detailed discussion on the GP classifier is presented in [103, 104]. GP classifier is successfully employed in a wide range of applications including but not limited to remote sensing image classification [103], electroencephalogram signals classification [105], and appearance-based gender classification [106].

#### 4.1.2.8 Multi-Layer Perceptron

A multi-layer perceptron is a supervised learning algorithm that uses a dataset for training to learn a function [107]. The multi-layer perceptron algorithm employs backpropagation technique for training purposes [107]. The MLP comprises at least three nodes, i.e., the input layer, hidden layer, and output layer. Any random classification problem can be learned even with one hidden layer, given that the hidden layer comprises enough units, further details can be found in [107, 108].

Further details of aforesaid supervised machine learning models are presented in [48, 100, 109, 110], where Table 20 [58] summarizes the strengths and weaknesses of the employed machine learning models.

Table 20 Classifiers Strengths and Weaknesses

	<b>Strength</b>	<b>Weakness</b>	<b>References</b>
SVM	Insensitive to data dimensionality, offers good generalization ability, versatile in terms of kernel selection	High algorithm complexity and memory requirements, rely on model parameters, poor interpretability	[100, 101, 107, 109]
LR	Parametric model, the capability to handle nonlinearity	Multicollinearity issues, required a large sample size	[109, 111]
DT	Good generalization capability, noise robustness, computationally faster, easy to interpret	Greedy construction process, overfitting issues, error propagation issue, prone to data dimensionality	[48, 100, 109, 111]
RF	Computationally fast, noise robustness, no parameter tuning, no over-fitting	The increasing number of trees slow down the model	[48, 109, 112]
k-NN	Suitable for multi-model classes, simplicity	Rely on k-value tuning, prone to noise and irrelevant features, data dimensionality issue, high memory requirement, poor interpretability	[48, 100, 109, 111]
NB	No parameters tuning, robust to missing values, computationally faster, requires low memory	Prone to data dimensionality	[48, 100, 109]
GP	Probabilistic approach, good performance in practice	High computational cost	[113, 114]
MLP	Non-parametric, robust to noise and irrelevant features	Require large training time, rely on input parameters, hard to interpret	[48, 109, 111, 115]

## 4.2 Research Methodologies

This section presents the details of the employed research methodologies in terms of feature engineering and load classification.

### 4.2.1 Feature Extraction and Reduction

As discussed in chapter 3, all proposed event detection algorithms provide output in the form of starting and ending time indices of the transient portion of the signal referred to as an event (as shown in Figure 11). The consecutive starting and ending time indices are linked together to extract the desired transient portions of the signal (events). These events are simply an indication of a variation within the aggregated load profile triggered by different individual appliances/circuits at different time indices within the aggregated load profile. These detected events do not enable the identification of any appliance explicit state, whether a specific appliance is turned-on or turned-off. In this context, different load features are extracted for each detected event that corresponds to the unique consumption pattern of the appliance explicit state. Later, these features are used as an input to the learning models to classify the explicit state of the specific appliance within the aggregated load profile.

The extracted load features within the scope of this research work are mainly based on statistical, geometrical, and power features. Initially, a feature set  $\mathcal{F}$  [49] comprising five distinct load features for each detected event is proposed, given in (12).

$$\mathcal{F} = \{\tau_{width}, P_{p2p}, \sigma, \sigma^2, \mu\} \quad (12)$$

where  $\tau_{width}$ ,  $P_{p2p}$ ,  $\sigma$ ,  $\sigma^2$ , and  $\mu$  represent width, peak to peak power magnitude, standard deviation, variance, and mean value of the detected events, and mathematically given as in (13)-(15), (9), and (8), respectively.

$$\tau_{width} = \tau_{Event\_End\_TimeIndex} - \tau_{Event\_Start\_TimeIndex} \quad (13)$$

$$P_{p2p} = P_{Event\_End\_Power} - P_{Event\_Start\_Power} \quad (14)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n |x_i - \mu_x|^2} \quad (15)$$

To facilitate the learning model in terms of classification performance and computational efficiency, feature space of  $\mathcal{F}$  is reduced using a combinatorial process and a new set of reduced number of features, having all the information of  $\mathcal{F}$ , is proposed, i.e., feature set  $\mathfrak{F}$  [58], and given in (16).

$$\mathfrak{F} = \{\mathcal{S}_E, C_{Disp.}, C_{var.}\} \quad (16)$$

where  $\mathcal{S}_E$ ,  $C_{Disp.}$ , and  $C_{var.}$  represent different geometrical and statistical properties of the

detected events, i.e., slope, coefficient of dispersion, and coefficient of variation, mathematically given in (17)-(19), respectively.

$$\mathcal{S}_E = \frac{P_{p2p}}{\tau_{width}} = \frac{P_{Event\_End\_Power} - P_{Event\_Start\_Power}}{\tau_{Event\_End\_TimeIndex} - \tau_{Event\_Start\_TimeIndex}} \quad (17)$$

$$C_{Disp.} = \frac{\sigma^2}{\mu_x} = \frac{\frac{1}{n} \sum_{i=1}^n |x_i - \mu_x|^2}{\frac{1}{n} \sum_{i=1}^n x_i} \quad (18)$$

$$C_{Var.} = \frac{\sigma}{\mu_x} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n |x_i - \mu_x|^2}}{\frac{1}{n} \sum_{i=1}^n x_i} \quad (19)$$

This research work also employed an extended feature set  $\mathcal{F}_{\text{exT.}}$  for different simulations and evaluation purposes, discussed in the later sections of this chapter. The extended feature set,  $\mathcal{F}_{\text{exT.}}$ , comprises nine distinct load features and is given as in (20).

$$\mathcal{F}_{\text{exT.}} = \{\tau_{width}, P_{p2p}, \sigma, \sigma^2, \mu, \mathcal{S}_E, \mathcal{M}, C_{Disp.}, C_{var.}\} \quad (20)$$

where  $\mathcal{M}$  represents the median absolute deviation value of the transient portion, i.e., event, and is given in (21).

$$\mathcal{M} = \text{median}[\text{abs.}(\mathbf{x} - \text{median}(\mathbf{x}))] \quad (21)$$

#### 4.2.2 Feature Selection

As discussed, feature selection techniques play a key role to facilitate the machine learning models in terms of computational efficiency and performance. However, opting for a specific technique among the filter, wrapper, and embedded method for feature selection is a crucial task. Within the scope of this research work, the feature selection is primarily intended to facilitate the learning models, hence in the given context, opting criteria are devised for overall feature selection (block) comprising of three key attributes: computational efficiency, robustness, and learning model performance.

In terms of computational efficiency, as discussed in Section 4.1.1, among all three feature selection methodologies, filter method is the only approach that is classifier agnostic and primarily relies on statistical assessments to identify features' intrinsic properties: significance. This property of filter method makes it more computationally efficient approach compared to classifier dependent approach (wrapper method) [99]. Moreover, filter method also established itself as a promising feature selection approach

for numerous datasets [97]. Hence, filter method is opted as one of the feature selection techniques for the feature selection (block), employed in this research work.

In terms of learning model performance, both wrapper and embedded methods have learning models' interaction. However, the wrapper method depends on a certain classifier's performance for feature significance, consequently, it is probable that the corresponding feature significance results may not be relevant if a different classifier is employed for classification [116]. Hence, within the scope of this research work, towards feature selection, wrapper method is not a viable option as eight diverse learning models/classifiers are employed for classification purposes. Moreover, the wrapper method is also criticized for its higher computational demands [97-99, 116]. Hence to accommodate classifier interaction within the intended feature selection (block), the embedded method is opted due to its capabilities to inherit the advantages of both filter method and wrapper method, i.e., computationally efficient and classifier interaction [99].

For the given problem, the selection of two diverse feature selection techniques: filter method (classifier-agnostic) and embedded method (classifier interaction), leads to a robust feature selection block that will not only strengthen the feature selection analysis but also significantly facilitate the learning model in terms of performance and computational requirements.

Further in the context of the filter and embedded methods, the techniques of univariate selection and feature importance have been employed, respectively. Univariate selection is performed using the *SelectKBest* module of Scikit-Learn<sup>24</sup> [107], where the feature importance technique is performed using the *ExtraTreesClassifier* of Scikit-Learn. It is also worth noting that within the *SelectKBest* module, two different variants of assessment, i.e., *f\_classif* and *mutual\_info\_classif*, are employed for evaluating the features' significance. The *f\_classif* computes the analysis of variance (ANOVA) f-value among the features and output vectors, where the *mutual\_info\_classif* estimates the mutual information (dependencies) among the features and the output vectors.

#### 4.2.3 Learning Models

In the context of NILM, most of the existing research employs either a single or two learning models for load classification purposes. In this context, some of the learning models are frequently explored with enhancements in terms of optimal tuning and

---

<sup>24</sup> Scikit-Learn, commonly referred to as sklearn, is a machine learning library for Python programming language. <https://scikit-learn.org/stable/>

classification performance. However, the selection of learning model is neither a case of “one size fits all” nor it is about the supremacy of a single model for a given problem. Subsequently, this research work employs eight diverse and independent machine learning models, discussed in Section 4.1.2. Employing diverse learning models not only facilitates the investigation of optimal/compatible learning model for the given problem but also smartly evades the supremacy of a single learning model over the others.

Further, the existing NILM research primarily relies on standalone/independent learning models for load classification and as mentioned that “one size fit all” is not a case, consequently, the performance of standalone learning models varies from case to case. To address these performance variations of standalone learning model, this research work explores an ensemble learning methodology (explicitly discussed in the respective section later in this thesis) to balance the performance of different standalone learning models.

The presented methodology of employing diverse standalone and ensemble machine learning models significantly contributes to the existing state of the art on low sampling NILM systems. As the presented comprehensive evaluation and corresponding comparative analysis of all the employed learning models are carried out on a data granularity of 1/60 Hz, which is 60 times lower sampling rate compared to the 1 Hz, which is mostly used in the context of low sampling NILM systems. Moreover, for simulation purposes, the details of the hyperparameters<sup>25</sup> of the employed learning models are presented in the Appendix A.2 of this thesis.

#### **4.2.4 Learning Models Evaluation**

Model evaluation is a key step in the machine learning domain and referred to as a process to estimate the generalization accuracy of the given model on unseen testing data. Therefore, before employing a learning model on a real-world diverse and independent set of training and testing load data, it is important to evaluate and validate the performance of the given learning model. In this context, cross-validation is one of the well-known techniques that evaluate the effectiveness of the given model, i.e., how well the model generalizes to an independent and unseen testing data. The cross-validation approach can be further broadly classified into holdout validation and k-fold cross-validation approaches.

The holdout validation method, also referred to as the train/test split method is an approach where the data is split into subsets of training and testing data, subsequently

---

<sup>25</sup> Hyperparameters are those parameters that are not learnt directly within the estimators.

utilized for training and testing of the learning model, respectively. The ratio of training and testing can vary but as per the existing literature, generally, it is kept at 70:30. On the other side, in the k-fold cross-validation approach, the data is split into k number of folds. The model is trained on k-1 folds and the remaining 1-fold is used for testing purposes. The process repeats itself until each k-fold serves as a testing set. The outcome is in the form of a mean value of evaluation metrics at each iteration. Figure 27 graphically depicts the basic working principle of the holdout and k-fold cross-validation approaches. In this research work, all the employed learning models for load classification are evaluated using a cross-validation approach.

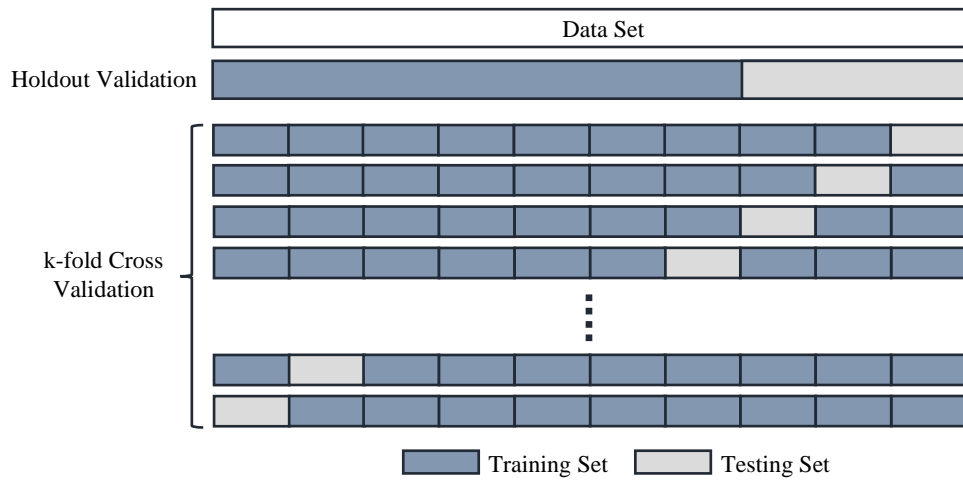


Figure 27 Working Principle of Cross-Validation Approaches  
For the given k-fold cross-validation depiction, 10 folds are considered, i.e.,  $k=10$

### 4.3 Simulations and Results

Based on the presented methodologies, Figure 28 presents a generalized simulation framework adopted in this research work for non-invasive load disaggregation: NILM.

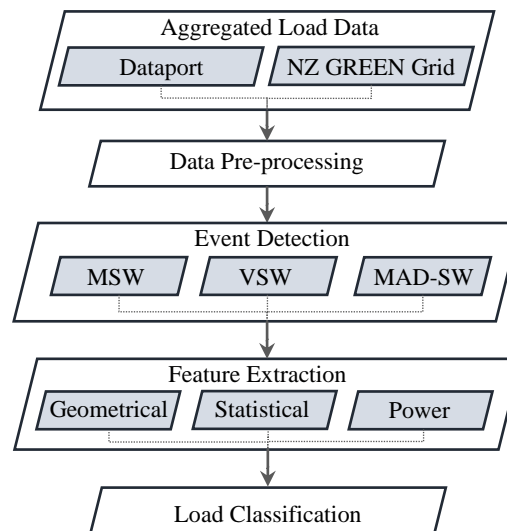


Figure 28 Generalized Employed Simulation Flow



It is worth noting that in this research work, data pre-processing, event detection, and feature extraction are carried out using MATLAB, as a simulation toolkit. Further, feature selection and load classification simulations are performed using Scikit-Learn [107]. The following sub-sections present the details of the simulation studies carried out based on each employed real-world load database within the scope of this research work.

#### 4.3.1 Dataport

According to the flow presented in Figure 28, comprehensive simulations are carried out on two real-world households' data acquired from Pecan Street's Dataport database. The presented simulations and corresponding results regarding feature extraction and load classification are based on events detected by the MSW algorithm with  $\omega=6$  and further incorporation of the delay tolerance factor. A detailed discussion regarding event detection simulations is already presented in section 3.2.2.1. Further, detailed sensitivity analysis results in terms of delay tolerance for Pecan Street's Dataport is presented in Appendix A.3 of this thesis. Table 21 presents the details of the parameters employed for event detection simulations in this section.

Table 21 Event Detection Simulation Parameters

<b>Dataport, Pecan Street Inc.</b>	
Data Granularity	1/60 Hz, i.e., 1-minute sampling
Event Detection Algorithm	Mean Sliding Window (MSW)
Pre-processing Technique	Median Filtering
Window Width ' $\omega$ '	6 samples
Threshold Value ' $\delta$ '	250 W
Delay Tolerance ' $\Delta\tau$ '	1-time index

Based on the presented parameters in Table 21, comprehensive simulations are carried out and the extracted event detection results for diverse input data in terms of different performance metrics are presented in Table 22. Details of different households and the corresponding load data acquisition are also presented in Table 22.

Table 22 Event Detection Results for Different Households of Dataport

	<b>Dataport, Pecan Street Inc. Household IDs</b>		
	Training Data	Testing Data	
	ID. 26	ID. 26	ID. 3036
Data Acquisition Timeframe	June 18 - July 02, 2014	August 01 - 04, 2014	June 18 - 21, 2014
Number of Days	15	4	4
Number of Samples	21600	5760	5760
Total Detected Events	323	99	231
True Positive	313	96	219
False Positive	10	3	12
False Negative	17	2	16
Precision (%)	96.90	96.97	94.80
Recall (%)	94.84	97.95	93.19
F-Score (%)	95.86	97.46	93.99

It is observed from the results presented in Table 22, particularly for Training Data ID. 26 (highlighted in brown color), that the overall performance of the event detection algorithm improves significantly as compared to the results presented in Table 8<sup>26</sup>. This performance improvement is facilitated by the incorporation of the delay tolerance factor that increases the numbers of true positive, from 286 to 313, subsequently, decreases the numbers of false positive and false negative detection, respectively. The same phenomenon of improvement in event detection performance, facilitated by a delay tolerance factor, is also observed for the NZ GREEN Grid data, as shown in Table 19.

Further, the data presented with a caption of training and testing data in Table 22, correspond to data used for the training and testing purposes of the employed classifier. It is seen in Table 22 that household ID. 26 is used for both training and testing purposes of the classifier, but it is worth noting that the testing data acquisition timeframe is different than the one used for training purposes, consequently, regardless of the same household, the testing data is completely unseen for the training phase of the employed classifier. Further, for testing purposes, another household, i.e., ID. 3036, is also employed to validate the robustness of the proposed approach.

For the given load data in Table 22, five distinct load features based on the feature set  $\mathcal{F}$ , given in (12), are extracted for each detected event. Table 23<sup>27</sup> [49] presents the extracted load features along with the corresponding time-indices of the detected events within the aggregated load profile of the Training Data ID. 26.

Table 23 Extracted Load Features ' $\mathcal{F}$ ' for Dataport ID 26

Detected Event Time-Indices	Extracted Load Features				
	$\tau_{width}$	$P_{peak\ to\ peak}$	$\sigma$	$\sigma^2$	$\mu$
16	2	1.836	1.055	1.114	3.051
55	1	-0.832	0.588	0.346	1.56
6780	3	3.37	1.583	2.507	5.802
6797	5	-4.892	1.950	3.803	4.162
21021	3	3.429	1.626	2.646	2.838
21045	2	3.039	1.531	2.346	6.093
21058	2	-2.852	1.543	2.381	4.998
21083	2	3.041	1.753	3.076	6.645
21103	2	-2.525	1.377	1.896	5.309
21160	3	-1.164	0.495	0.245	1.864

Later the extracted features are used as an input to the classification models. To validate the effectiveness of the proposed non-intrusive load inference approach on a real-world load dataset: Dataport, the k-nearest neighbor algorithm is employed as a learning model

<sup>26</sup> Simulation results presented in Table 8 and portion (highlighted in brown) of Table 22 correspond to the same load data.

<sup>27</sup> Table 23 present a portion of the extracted results in terms of feature extraction.

for load classification. The k-NN being a supervised learning model needs to be fed with the extracted features along with the corresponding labels<sup>28</sup> for training purposes, as shown in Figure 6.

For the given problem, the k-NN model is first evaluated using a cross-validation approach to validate its effectiveness for unseen testing data. For said purposes, holdout and k-fold cross-validation approaches are adopted, as discussed in section 4.2.4, and simulations are carried out on 15 days of load data, acquired from household ID. 26, given as training data in Table 22. Table 24 presents the corresponding evaluation results of the k-NN algorithm, where the nearest neighbor count value is taken as 5.

Table 24 k-NN Evaluation Based on Cross-Validation Approach

Attributes	Training Data ID. 26	
	Holdout Validation	k-Fold Cross-Validation
	Training Set = 70%   Testing Set = 30%	k = 10
Accuracy (%)	89.36	90.39

As seen in Table 24 the k-NN model attained promising generalization accuracy of  $> 89\%$  for the unseen testing data. To move forward, the given learning model is trained on the entire set of load data, i.e., Training Data ID. 26, and the trained model is further tested rigorously on a diverse and unseen set of real-world load data, i.e, Testing Data ID. 26 and 3036, given in Table 22. In this context, Table 25 presents the k-NN performance in terms of individual load element inference, i.e., EV and AC. It is worth noting that the results presented in Table 25 are based on the nearest neighbor count value of 5, i.e.,  $k=5$ .

Table 25 Appliance-Level Classification Results of Dataport

			Training Data ID. 26		
			Precision (%)	Recall (%)	F-Score (%)
Testing Data	ID. 26	AC Turn-off	98	98	98
		AC Turn-on	95	98	96
		EV Turn-on	80	67	73
		EV Turn-off	80	80	80
	ID. 3036	AC Turn-off	91	95	93
		AC Turn-on	97	90	93
		EV Turn-on	75	75	75
		EV Turn-off	50	75	60

As evident from the individual load classification results that k-NN generalizes well for the completely unseen testing data. But it is also observed that for EV inference, the results are not promising comparatively to AC inference, particularly for testing data ID. 3036. This is expected due to the diversity of the testing household and more importantly

<sup>28</sup> Labels are assigned to the features of each detected event. Labels information are extracted from the individual appliances' ground-truth profile that are available in the load databases being used in this research work.

the random selection of k-value, i.e.,  $k=5$ , for the k-NN algorithm. It is worth noting that for the k-NN algorithm the choice of k-value, i.e., the number of nearest neighbors, is of utmost significance in terms of algorithm's performance, as the optimal k-value reduces the noise effect on classification performance, discussed in Table 20. Therefore, to improve the k-NN based inference performance, parameter tuning in terms of k-value need to be carried out.

In the existing literature, the research community adopted different ways of selecting k-values, e.g., opting odd k-value to avoid confusion between two data classes, or opting  $k=\sqrt{n}$ , where  $\sqrt{\phantom{x}}$  and  $n$  refers to the square root function and the number of data samples, respectively. Despite the numerous proposed ways, the optimal k-value always varies and depends on the dataset under consideration. The best approach is to try different values and check which k-value provides the best performance for the given data. In this context, the elbow method<sup>29</sup> is one of the most widely used methodologies to identify the optimal k-value for the k-NN algorithm. Therefore, in this research work, simulations based on the elbow method are carried out to identify the optimal k-value for each testing data ID, under consideration. The elbow method simulation results in the form of the k-value vs. misclassification error rate for testing data IDs. 26 and 3036 are presented in Figures 29 and 30, respectively.

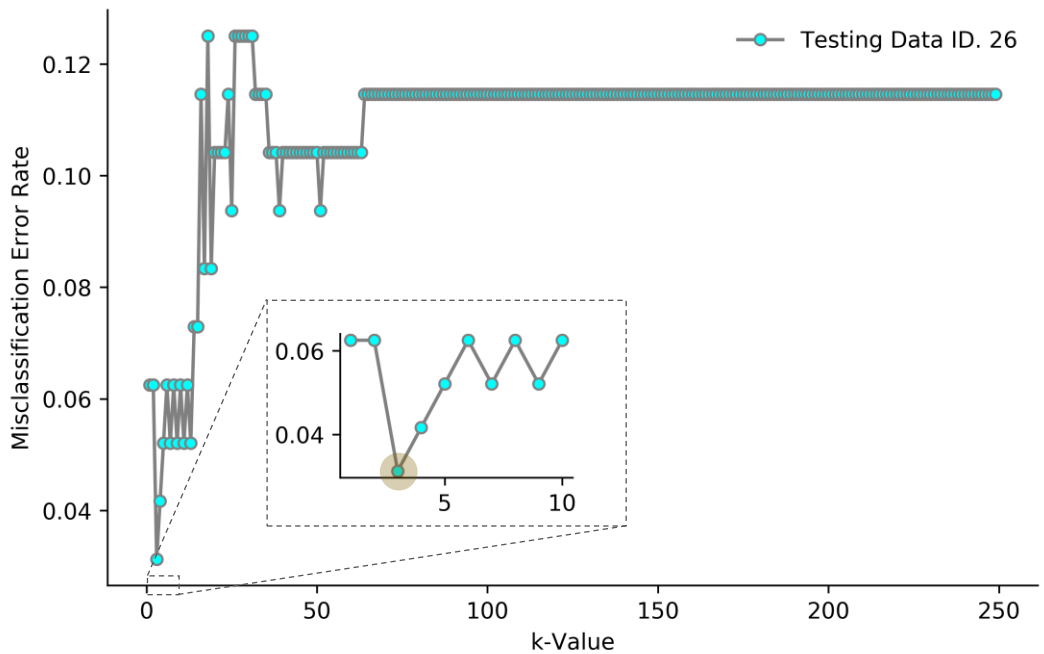


Figure 29 Optimal k-value Selection for Testing Data ID. 26

Close-up represents the optimal k-value (highlighted in light brown color) having a minimum misclassification error rate.

<sup>29</sup> Elbow method is a heuristic technique used to determine the optimal k-value by fitting the model with a range of k-values and plot the misclassification error rate as a function of the k-value.

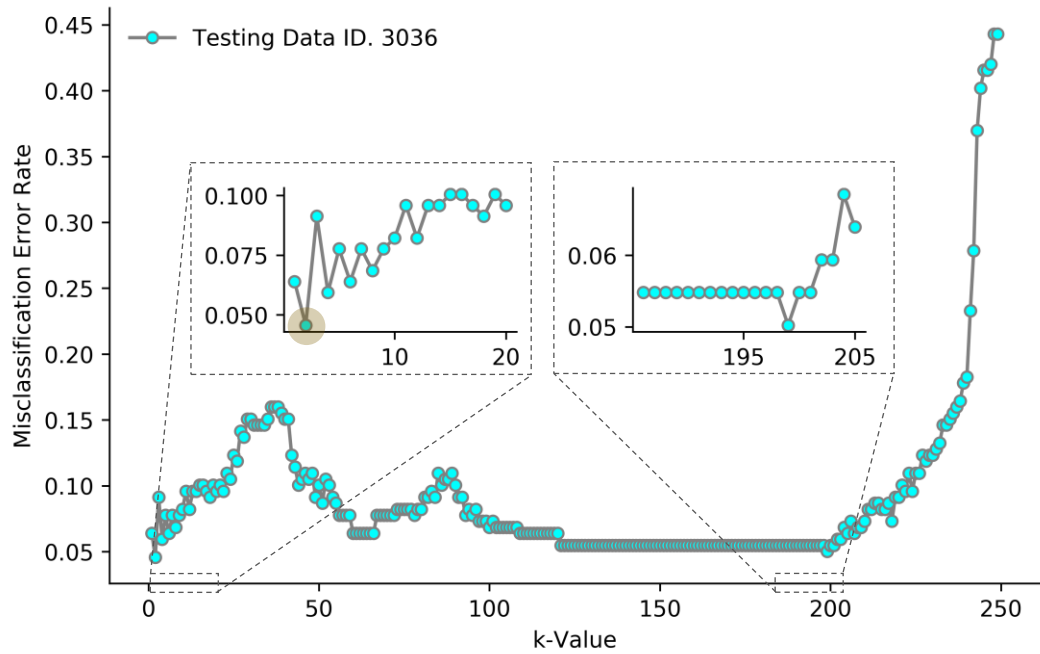


Figure 30 Optimal k-value Selection for Testing Data ID. 3036

Close-up represents different k-values having a minimum misclassification error rate, where the one highlighted in light brown color is the optimal k-value.

Based on the minimum misclassification error rate, the k-values of 3 and 2 are selected for the testing data IDs. 26 and 3036, respectively (also highlighted in the corresponding figures). To underline the influence of k-value, simulations for individual load elements inference are carried out again while incorporating the extracted optimal k-values. Based on the performed simulations, Table 26 presents the appliance level classification results of the k-NN classifier in combination with optimal k-values. To validate the performance improvement due to optimal k-values, Table 26 also represents the classification results presented in Table 25 for ease of comparison, where all the presented results are in percentages.

Table 26 Optimal k-value Based Inference Results Validation

			k-Nearest Neighbors Algorithm						
			Training Data ID. 26						
			k=5			Optimal k-value			Performance Improvement in F
			P	R	F	P	R	F	
Testing Data	ID. 26	AC Turn-off	98	98	98	100	98	99	1 %
		AC Turn-on	95	98	96	98	98	98	2 %
		EV Turn-on	80	67	73	83	83	83	10 %
		EV Turn-off	80	80	80	83	100	91	11 %
	ID. 3036	AC Turn-off	91	95	93	96	95	96	3 %
		AC Turn-on	97	90	93	96	97	97	4 %
		EV Turn-on	75	75	75	100	75	86	11 %
		EV Turn-off	50	75	60	60	75	67	7 %

P, R, and F represent Precision, Recall, and F-Score, respectively, and all the presented results are in percentage.

As evident from Table 26, with an optimal k-value, the k-NN model achieved a performance improvement of up to 11% (highlighted in brown) in terms of f-score<sup>30</sup> measure. It is established from the presented results that the k-value plays a key role in the k-NN algorithm and can significantly improve the performance, if an optimal/best k-value is selected.

For further in-depth analysis of the classification performance of the k-NN algorithm in terms of individual load elements' status prediction, i.e., turning-on and turning-off of EV charging and AC unit, Figure 31 depicts date-wise prediction results. It is evident from the presented results that during the four days, almost all the individual load elements operation states are precisely predicted by the k-NN algorithm except three instances that were misidentified by the k-NN classifier (highlighted in Figure 31). In terms of misclassification, it is seen in Figure 31 that AC turn-on and turn-off ground-truths are misidentified as EV turn-on and turn-off, respectively, further one ground-truth instance of EV turn-on is misclassified as AC turn-on by the k-NN algorithm. It is worth noting that the results presented in Figure 31 are based on testing household ID. 26 with an optimal k-value.

From the presented results and analysis, it is evident that the overall proposed non-intrusive load inference approach not only works well but also attained promising results in the context of low sampling real-world measurements. To further validate the effectiveness and robustness, the proposed non-intrusive load disaggregation approach is also tested on another real-world low sampling database: NZ GREEN Grid. Moreover, extended simulation studies are carried out in terms of different feature engineering techniques and machine learning algorithms. The extended simulations based on NZ GREEN Grid will further strengthen the presented analysis.

#### **4.3.2 NZ GREEN Grid**

Detailed discussion regarding event detection for the NZ GREEN Grid database has already been presented in chapter 3 of this thesis. This section mainly focusses on simulations studies related to feature engineering and load classification with further extended analysis in terms of feature space dimensionality and learning models' comparative evaluation. An ensemble learning approach is also presented in the NILM context.

---

<sup>30</sup> The f-score is selected for performance comparison because it considers both precision and recall performance metrics to calculate the score.

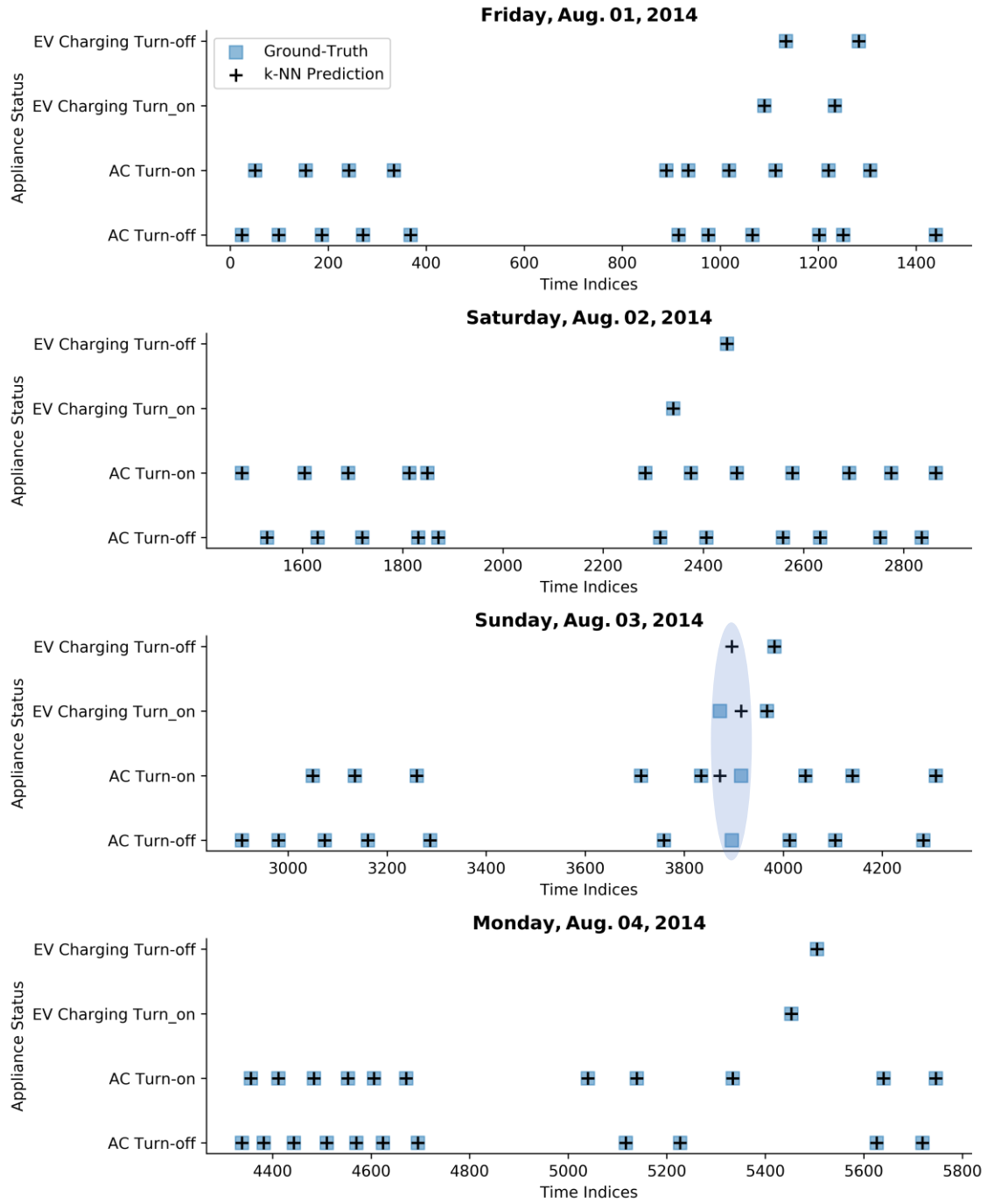


Figure 31 Comparison of Ground-truth and Predicted Load Element Status  
Date-wise inference results for Pecan Street-Dataport's Testing Data ID. 26 based on k-nearest neighbors algorithm with an optimal k-value, i.e.,  $k=3$ .

Similarly to Pecan Street's Dataport, the low data granularity (1/60 Hz) nature of NZ GREEN Grid also restricts the load inference (with reasonable accuracy) to high consumption load elements [81]. Therefore, within the NZ GREEN Grid database, the primary focus is the non-intrusive load inference of the water heating circuit. Besides data granularity<sup>31</sup>, the availability of a limited number of circuits in each household of the NZ

<sup>31</sup> Number of appliances precisely identified is directly proportional to the data granularity, i.e., high sampling rate data acquisition leads to a greater number of appliance/circuit disaggregation and vice-versa.

GREEN Grid also restricts the inference to a single load element, i.e., water heating circuit. However, the non-invasive inference of single load element, i.e., water heating circuit, is not only justified based on the aforesaid constraints, but it is also of utmost significance for the given database as well as in the larger perspective, as water heating is not only the significant load element of the given database but also, in general, it is one of the major load elements in the residential sector [117-119]. Furthermore, it is a flexible/interruptible load [120] having high potential towards many real-world energy efficiency applications, e.g., demand response [119, 121], power regulations [118], and peak shifting and frequency response [122]. Hence inference of only water heating circuits, being the focus of this research work, is not only more viable in the context of the given load database having low data granularity [4, 17] but would also accelerate the research and development of many real-world energy efficiency applications.

In the context of the NZ GREEN Grid database, five different real-world households are selected for simulation purposes, given that, all the selected households have dedicated water heating circuit installation, other circuits' configuration may vary<sup>32</sup>. For example, within the selected households for simulation purposes, household ID rf\_42 has a single circuit configured for laundry and freezer having a circuit label of "Laundry & Freezer\$4128". In contrast, household ID rf\_36 has two dedicated circuits for the said having the circuit labels of "Washing Machine\$4146" and "Kitchen Appliances\$4145". Similarly, household ID rf\_42 has a load circuit labelled as "Lighting (inc heat lamps)\$4129" where household ID rf\_36 has a load circuit labelled as "Lighting\$4149", which potentially implies that the latter has no heat lamps. Further details regarding different circuits and their installation configuration can be found in [123].

Table 27 presents the different attributes of the acquired load data, including but not limited to household IDs, data acquisition timeframe, and the number of acquired samples. Table 27 also presents the event detection results in terms of total number of detected events for each household under consideration. It is worth noting that the event detection results presented in Table 27 are based on the MAD-SW algorithm at optimal input parameters, i.e.,  $\Delta\tau=2$  and  $\omega=3$ , where the corresponding details of the event detection simulations and sensitivity analysis of the said input parameters have already been discussed and presented in chapter 3 of this thesis.

---

<sup>32</sup> The availability of individual circuits varies in different households, even in case of the same circuit, different households have diverse installation configuration.



Table 27 NZ GREEN Grid Household Attributes and Event Detection Results

	NZ GREEN Grid Database Household IDs.					
	rf_01	rf_02		rf_31	rf_36	rf_42
Data Acquisition	Sep. 01-07,	May. 11-30,	Jul. 01-10,	Sep. 01-07,	Jun. 21-27,	Jan. 07-13,
Timeframe	2014	2014	2014	2016	2017	2017
Number of Days	7	20	10	7	7	7
Number of Samples	10080	28800	14400	10080	10080	10080
Detected Events	808	1504	898	166	390	60

Data presented in brown color is used for the training purposes of the employed machine learning models where the rest of the presented load data is used for testing purposes.

#### 4.3.2.1 Classifiers and Feature Space Evaluation

According to the methodology presented in Figure 28, five distinct load features, i.e.,  $\mathcal{F}$ , are extracted for each detected event of different households, presented in Table 27. To further investigate the influence of feature space dimensionality on classification performance, a reduced feature set,  $\mathcal{F}$ , given in (16), having three distinct load features is also employed for the given load data, i.e., Table 27. Both load feature sets, i.e.,  $\mathcal{F}$  and  $\mathcal{F}$ , are independently used as an input for the training and testing purposes of the employed machine learning algorithms to predict the water heating circuit status. Figure 32 depicts the detailed flow of the simulations carried out for the said purposes.

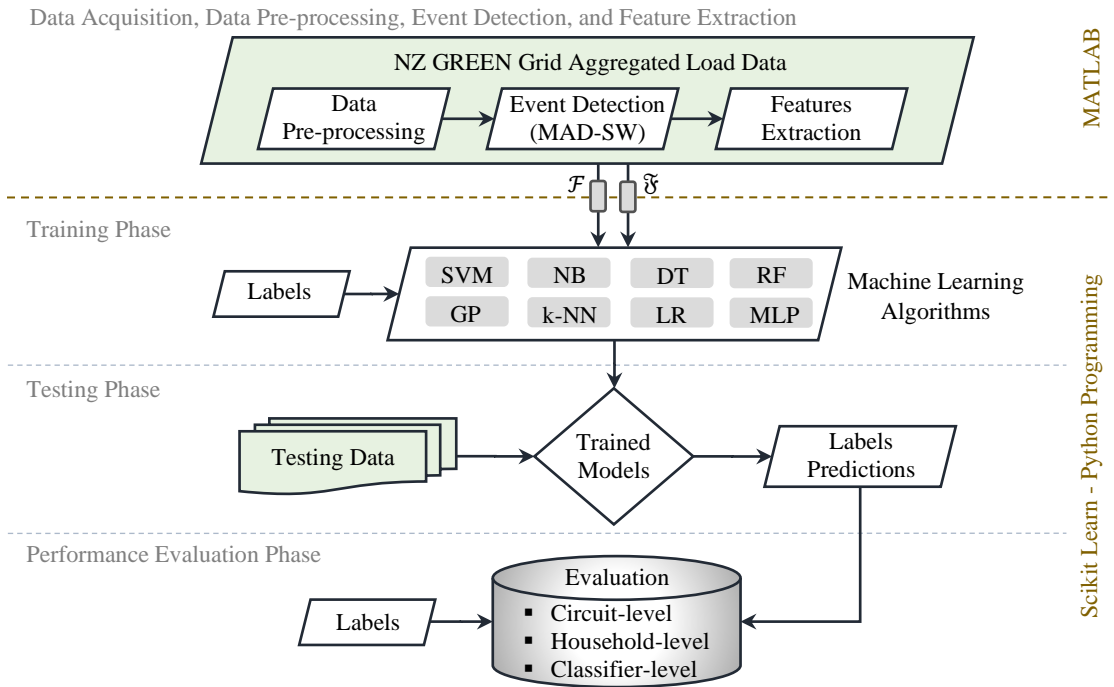


Figure 32 Non-Intrusive Load Disaggregation Simulation Flow

In terms of load classification, supervised machine learning algorithms discussed in section 4.1.2, namely, support vector machine, decision trees, random forest, logistic regression, k-nearest neighbors, Gaussian process, Naïve Bayes, and multi-layer perceptron, are independently tested and evaluated in combination with  $\mathcal{F}$  and  $\mathcal{F}$ , as an

input features, as shown in Figure 32. The employed classifiers are trained on 20 days of load data of household ID. rf\_02, having data acquisition timeframe of May. 11-30, 2014, as presented in Table 27. Later, the trained models are rigorously tested on the same household as well as on different households. It is worth noting that all the testing data are completely unseen for the trained models, even in the case of the same household as the data acquisition timeframe is different. Furthermore, in the context of NZ GREEN Grid data, the performance evaluation is carried out at three different levels, namely circuit-level, household-level, and classifier-level evaluation, as shown in Figure 32.

Based on the employed machine learning models, comprehensive simulations are carried out in combination with  $\mathcal{F}$ , as an input feature. Tables 28 [58] presents the circuit-level classification performance of the employed classifiers for all the testing households under consideration. As evident from the detailed performance results presented in Table 28 all the employed learning algorithms generalize well for the diverse and unseen real-world testing data. At the circuit-level evaluation, it is observed that household ID. rf\_02 achieved the best overall classification performance along with the most accurate water heating inference compared to other testing households, by all the employed classifiers. It is expected, even though the testing load data is completely unseen at the training phase, but the household is the same, i.e., ID. rf\_02. Consequently, similar attributes like occupancy, size, circuits' nature, installation configuration, and usage pattern make the prediction less complex. On the other side, due to variation in aforesaid household attributes the least overall and water heating circuit classification performance is recorded for household ID. rf\_36 and ID. rf\_01, respectively. Furthermore, for household ID. rf\_31, the water heating inference performance is recorded as 0% consistently by all the employed learning models, but it is worth noting that the corresponding water heating inference performance is due to the absence of the ground-truth activity of water heating circuit during the given data acquisition timeframe, i.e., Sep. 01-07, 2016, for the household ID. rf\_31, which is accurately predicted by all the employed classifiers. Further, it is also evident from the results in Table 28 that for the given conditions the MLP classifier based on neural networks concept outperforms other employed classifiers both at the circuit-level and overall inference, i.e., weighted average. The MLP classifier is followed by the SVM, LR, and GP classifiers with marginal variations in terms of performance. Further, under the given conditions, the decision trees algorithm lags in performance compared to other employed classifiers.

Table 28 Circuit-level Classifiers' Performance in Combination with  $\mathcal{F}$ 

		Feature Set, $\mathcal{F}$																							
		SVM			LR			DT			RF			k-NN			GP			MLP			NB		
Testing ID	Circuit Status	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
rf_01	WH Circuit Turn-Off	37	63	47	36	65	46	30	63	40	24	54	33	35	61	44	39	63	48	38	58	46	33	72	46
	WH Circuit Turn-On	24	76	37	32	78	46	24	65	35	28	64	38	29	78	42	32	75	44	32	58	42	32	78	46
	Misc. Circuit Turn-On	93	70	80	95	73	83	92	67	78	93	73	82	95	69	80	90	74	81	92	80	86	96	67	79
	Misc. Circuit Turn-Off	92	73	81	93	81	87	93	76	83	91	72	80	92	81	86	93	79	85	92	84	88	87	77	81
	Weighted Average	84	71	75	86	76	79	83	71	75	83	71	75	85	74	78	83	75	78	84	79	81	83	72	75
rf_02	WH Circuit Turn-Off	98	96	97	94	96	95	96	96	96	98	97	97	98	97	97	98	95	96	98	95	96	93	96	95
	WH Circuit Turn-On	94	96	95	93	96	95	92	92	92	92	96	94	92	96	94	94	97	96	94	95	94	94	96	95
	Misc. Circuit Turn-On	98	96	97	97	96	97	95	95	95	98	95	96	98	95	96	98	96	97	97	96	96	97	96	97
	Misc. Circuit Turn-Off	97	99	98	97	96	97	97	98	98	98	99	98	98	99	98	97	99	98	97	99	98	98	96	97
	Weighted Average	97	97	97	96	96	96	96	96	96	97	97	97	97	97	97	97	97	97	96	96	96	96	96	96
rf_31	WH Circuit Turn-Off	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	WH Circuit Turn-On	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Misc. Circuit Turn-On	100	80	89	100	80	89	100	82	90	100	80	89	100	81	89	100	80	89	100	81	90	100	80	89
	Misc. Circuit Turn-Off	100	69	82	100	71	83	100	57	73	100	66	79	100	64	78	100	67	80	100	74	85	100	67	80
	Weighted Average	100	76	86	100	77	87	100	73	84	100	75	85	100	75	85	100	75	86	100	79	88	100	75	86
rf_36	WH Circuit Turn-Off	78	67	72	84	80	82	71	74	73	75	79	77	78	64	73	83	73	78	84	78	81	83	82	83
	WH Circuit Turn-On	64	84	73	71	84	77	44	39	41	68	80	73	69	82	75	67	83	74	77	80	78	71	84	77
	Misc. Circuit Turn-On	76	56	65	70	65	72	44	49	47	75	61	67	77	62	69	76	58	66	79	76	77	80	62	70
	Misc. Circuit Turn-Off	69	76	72	79	84	81	70	67	69	76	71	73	70	79	74	74	83	78	78	84	81	77	82	79
	Weighted Average	72	71	70	79	78	78	57	57	57	73	73	73	74	73	73	75	74	74	79	79	79	78	77	77
rf_42	WH Circuit Turn-Off	83	100	91	83	100	91	50	100	67	50	100	67	50	100	67	83	100	91	83	100	91	62	100	77
	WH Circuit Turn-On	62	100	77	62	100	77	38	60	46	50	80	62	50	100	67	62	100	77	83	100	91	62	100	77
	Misc. Circuit Turn-On	92	88	90	100	88	94	91	80	85	95	84	89	100	80	89	100	88	94	100	96	98	100	84	91
	Misc. Circuit Turn-Off	100	88	94	100	96	98	100	80	89	100	80	89	100	80	89	100	96	98	100	96	98	96	88	92
	Weighted Average	92	90	90	95	93	94	87	80	82	90	83	85	92	83	85	95	93	94	97	97	97	92	88	89

P, R, F represent precision, recall, and f-score performance metrics, respectively, and all the presented results are in percentage.

WH: Water Heating | Misc.: Miscellaneous

To visually analyse the results presented in Table 28, Figures 33 and 34 present the least and best circuit-level inference performance attained by the employed classifiers for diverse testing households, i.e., rf\_01 and rf\_42, respectively. The presented graphical depictions are in the form of a normalized confusion matrix. The confusion matrix is used to evaluate the learning models' performance by visualizing the comparison of ground-truth and predicted labels. In Figures 33 and 34, the rows and columns of the confusion matrix represent the predicted labels by the employed learning models and ground-truth labels, respectively.

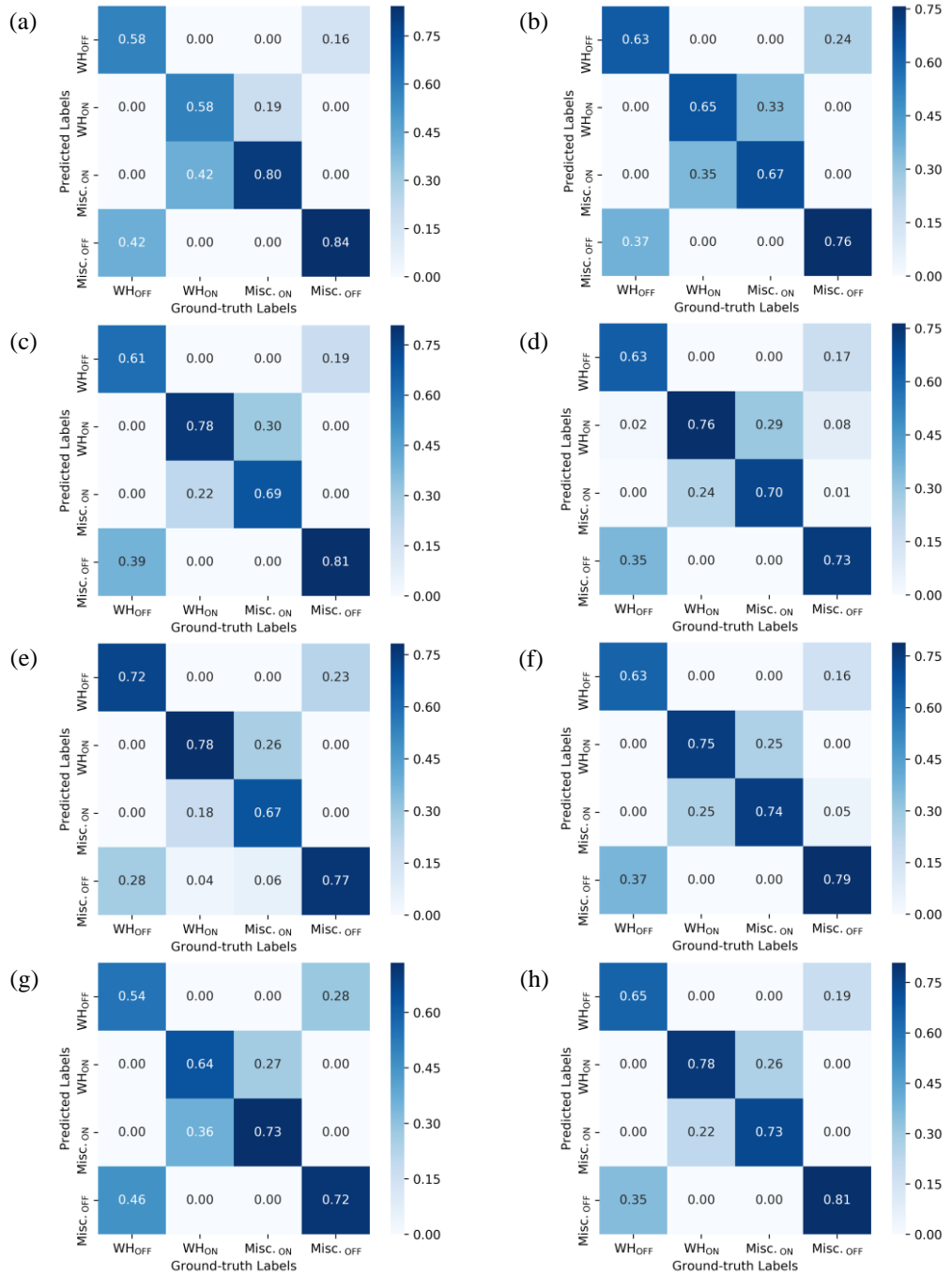


Figure 33 Circuit-Level Inference Results for Household ID. rf\_01  
(a) MLP, (b) DT, (c) k-NN, (d) SVM, (e) NB, (f) GP, (g) RF, and (h) LR

The symbols, i.e., WH<sub>OFF</sub>, WH<sub>ON</sub>, Misc.<sub>ON</sub>, and Misc.<sub>OFF</sub>, in Figures 33 and 34 represent the circuit operation status as, water heating circuit turn-off, water heating circuit turn-on, miscellaneous circuit turn-on, and miscellaneous circuit turn-off, respectively. The vertical color bars in Figures 33 and 34 represent the performance scale (normalized), where the darker color represents the best performance and vice versa.

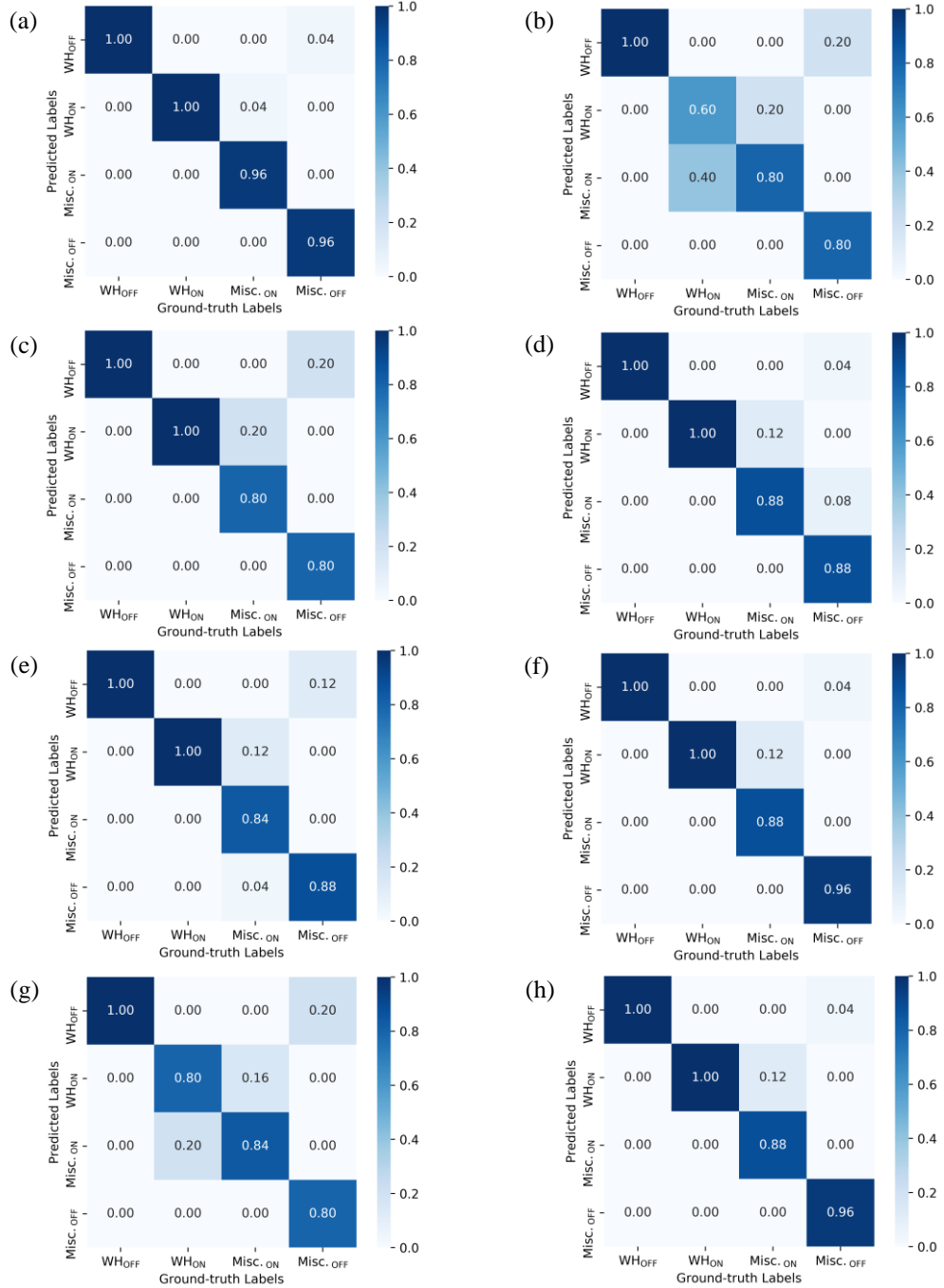


Figure 34 Circuit-Level Inference Results for Household ID. rf\_42  
(a) MLP, (b) DT, (c) k-NN, (d) SVM, (e) NB, (f) GP, (g) RF, and (h) LR

For household-level evaluation, the performance metrics of accuracy and Kappa index have been employed in this research work. In this context, Table 29 presents the performance results of all the employed learning models for each testing household within

the scope of this research work. It is worth noting that all the performance results presented in Table 29 are in percentage and based on input feature set,  $\mathcal{F}$ .

Table 29 Household-level Classifiers' Performance in Combination with  $\mathcal{F}$

	<b>SVM</b>	<b>LR</b>	<b>DT</b>	<b>RF</b>	<b>k-NN</b>	<b>GP</b>	<b>MLP</b>	<b>NB</b>	
rf_01	71.28	76.36	70.54	70.66	74.38	75.49	<b>78.96</b>	72.40	Accuracy
rf_02	<b>96.99</b>	95.87	95.54	96.77	96.65	<b>96.99</b>	96.32	96.10	
rf_31	75.90	76.50	73.49	74.69	74.69	75.30	<b>78.91</b>	75.30	
rf_36	70.76	78.20	57.43	72.82	72.82	74.10	<b>79.23</b>	77.43	
rf_42	90	93.33	80	83.33	83.33	93.33	<b>96.66</b>	88.33	
rf_01	58.42	65.01	57.31	57.13	62.35	63.37	<b>67.77</b>	59.65	Kappa Index
rf_02	<b>95.91</b>	94.40	93.94	95.61	95.46	<b>95.91</b>	95	94.71	
rf_31	58.36	59.25	54.43	56.59	56.46	57.47	<b>62.65</b>	57.47	
rf_36	61.03	70.93	43.21	63.72	63.77	65.48	<b>72.31</b>	69.90	
rf_42	84.87	89.91	70.73	75.60	76	89.91	<b>94.87</b>	82.64	

Bold numbers represent the best performance for the given household and all the presented results are in percentage.

It is also evident from the household-level performance results presented in Table 29 that the multi-layer perceptron classifier outperforms the other employed classifiers for most of the individual testing households, highlighted in bold in Table 29. Further, similarly to the precision, recall, and f-score performance metrics, presented in Table 28, the accuracy and Kappa index performance of all the employed classifiers, presented in Table 29, also varies with each testing household. It is anticipated due to the independent, diverse, and unseen set of household data employed for the classifiers' testing purposes. Further, as discussed earlier, the diverse circuit installation configuration in the employed households also contributes to the aforesaid variation in classification performance. All these factors yield a diverse set of consumption patterns associated with each testing household, subsequently not only making it harder to accurately predict but also lead to a variation in classifiers' prediction performance. However, it is worth noting that regardless of the aforesaid factors, all the employed classifiers attained the desirable Kappa index performance, i.e., >40% [87] for all the testing households under consideration.

In addition to circuit-level and household-level evaluation, a classifier-level evaluation based on the entire set of testing households under consideration is also carried out in terms of Kappa index and accuracy performance metrics. Figure 35 presents the corresponding results in the form of a boxplot representing the complete distribution of classifiers' performance in terms of maximum, minimum, median, and interquartile range values. It is seen in Figure 35 that, the MLP classifier has a clear edge over the other employed classifiers. On the downside, decision trees classifier shows lesser accuracy performance comparatively to the other employed classifiers under the given conditions.

However, it is established from Table 29 and evident from Figure 35b that in terms of the Kappa index performance measure, all the employed classifiers including the DT classifier attained the desired classification performance of  $>40\%$ . Furthermore, it is also worth noting that most of the attained performances by all the employed classifiers lie in the substantial region as seen in Figure 35b.

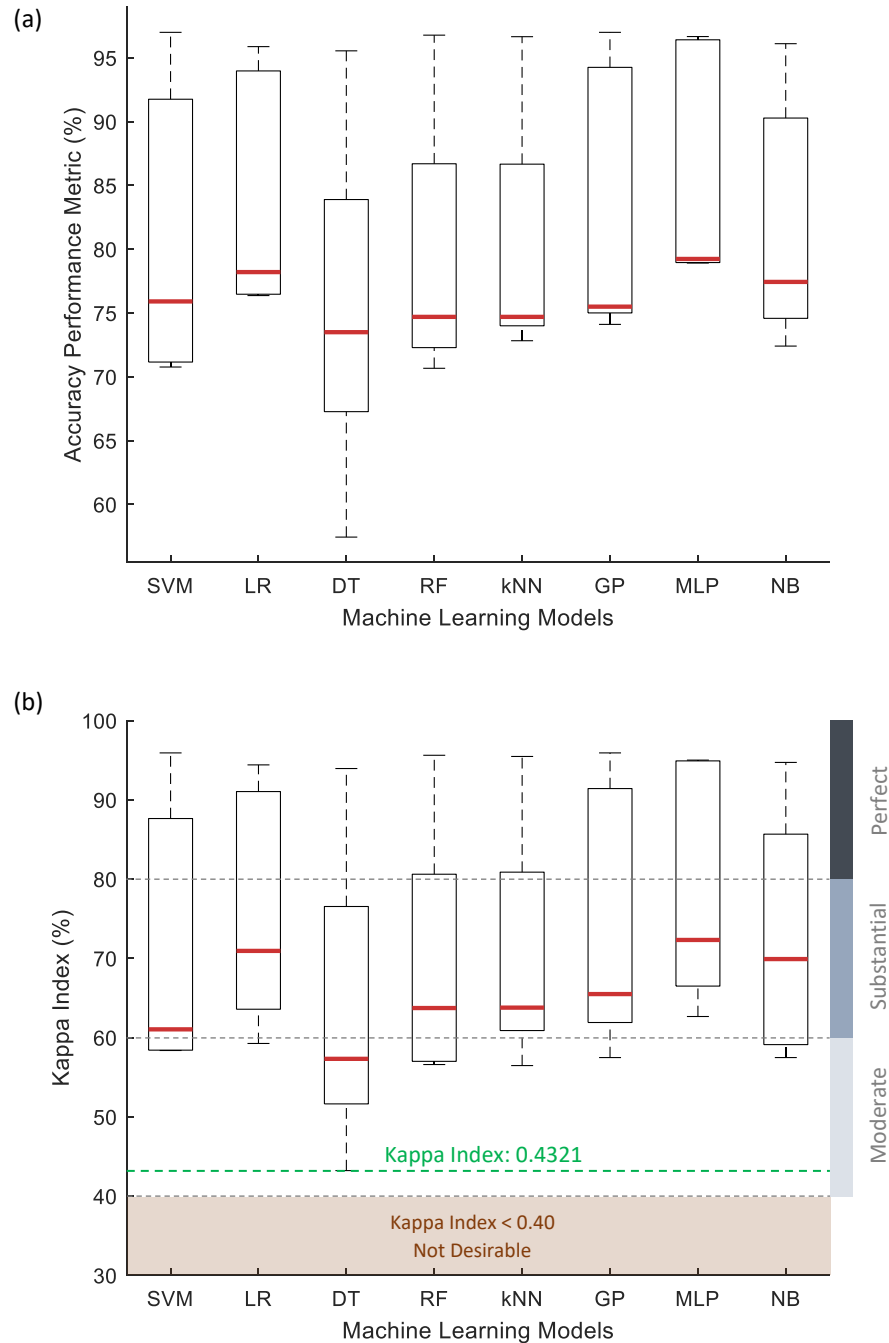


Figure 35 Overall Classifiers' Performance in Combination with  $\mathcal{F}$

(a) Accuracy (b) Kappa Index, all the presented results are based on the entire set of testing households. The red horizontal lines represent the median values where the horizontal dotted line in green represents the minimum Kappa index (normalized) attained by the employed classifiers, under the given conditions. The regions presented on the right side of Figure 35b are based on the description presented in Table 5.

#### 4.3.2.1.1 Reduced Feature Space Simulations

For the NZ GREEN Grid database, in addition to  $\mathcal{F}$  as input feature set, further simulations are also carried out to investigate the influence of features space dimensionality in the context of classification performance. For the said purpose, all the employed classifiers are further evaluated in combination with  $\mathcal{G}$ , given in (16). All the load data regarding training and testing of the classifiers and the simulation methodology are kept the same as presented in Table 27 and Figure 32, respectively. The corresponding circuit-level inference results are presented in Table 30, in terms of precision, recall, and f-score performance metrics.

From the results presented in Table 30, it is evident that irrespective of reducing feature space dimensionality, all the employed classifiers generalize well for the unseen diverse and independent testing data and achieved promising circuit-level classification performance. Even in some cases, a significant improvement has been recorded compared to the results presented in Table 28. For example, in the case of household ID 36, the DT model yields a 12% improvement in the weighted average circuit-level performance. This is expected as some of the employed learning models are susceptible to dimensionality issue, as discussed in Table 20.

Further analysis of the results presented in Table 30 shows that the overall performance trend in terms of the individual testing households and circuit-level performance variation is almost similar to the ones observed and discussed for classifiers' performance in combination with  $\mathcal{F}$ , presented in Table 28. For example, in terms of water heating circuit inference, similarly to the results presented in Table 28, the testing household ID. rf\_31 accommodate zero inference performance by all employed classifiers in combination with  $\mathcal{G}$ , however, the said zero inference result is accurate compared to the water heating ground-truth activity, as there were none for the given 7-days (testing) data acquisition. The corresponding rf\_31's water heating inference results, presented in Table 28 (for  $\mathcal{F}$ ) and Table 30 (for  $\mathcal{G}$ ), also validates the generalization capability of the employed learning models and corresponding approach. As the employed learning models precisely predicts no water heating circuit activity in the absence of ground-truth activity, even for a diverse and unseen testing household data.



Table 30 Circuit-level Classifiers' Performance in Combination with  $\mathfrak{F}$ 

		Feature Set, $\mathfrak{F}$																							
		SVM			LR			DT			RF			k-NN			GP			MLP			NB		
Testing ID	Circuit Status	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
rf_01	WH Circuit Turn-Off	41	35	38	44	35	39	25	54	34	32	54	40	34	40	37	40	33	37	40	32	35	30	32	31
	WH Circuit Turn-On	35	40	38	41	38	40	22	51	31	27	62	38	34	60	44	32	36	34	36	33	34	26	31	28
	Misc. Circuit Turn-On	90	88	89	90	91	91	90	72	80	92	74	82	93	82	87	90	88	89	89	90	90	89	86	87
	Misc. Circuit Turn-Off	89	92	91	90	93	91	90	73	81	92	81	86	90	87	88	89	92	91	89	92	91	89	88	88
	Weighted Average	83	83	83	83	84	84	81	70	74	83	75	78	83	80	81	82	82	82	82	83	83	80	79	80
rf_02	WH Circuit Turn-Off	99	90	94	99	88	93	96	95	95	98	95	96	99	95	97	95	86	90	94	85	89	88	85	87
	WH Circuit Turn-On	93	88	90	93	87	90	94	89	92	93	94	93	91	94	92	92	87	89	90	87	88	87	82	85
	Misc. Circuit Turn-On	93	96	94	92	96	94	94	96	95	96	96	96	96	95	95	92	95	94	92	94	93	90	93	91
	Misc. Circuit Turn-Off	94	100	97	93	99	96	97	97	97	97	99	98	97	99	98	92	97	95	91	96	94	91	93	92
	Weighted Average	95	94	94	94	94	94	95	95	95	96	96	96	96	96	96	93	93	93	92	92	92	89	89	89
rf_31	WH Circuit Turn-Off	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	WH Circuit Turn-On	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Misc. Circuit Turn-On	100	81	90	100	81	90	100	81	90	100	82	90	100	81	90	100	81	90	100	82	90	100	83	91
	Misc. Circuit Turn-Off	100	76	86	100	76	86	100	60	75	100	64	78	100	72	84	100	72	84	100	72	84	100	74	85
	Weighted Average	100	80	89	100	80	89	100	74	85	100	76	86	100	78	88	100	78	88	100	79	88	100	80	89
rf_36	WH Circuit Turn-Off	85	73	79	85	73	79	70	73	72	76	74	75	78	68	73	85	75	80	86	71	78	85	82	83
	WH Circuit Turn-On	78	76	77	80	73	76	68	71	69	68	80	73	74	80	77	79	76	77	80	72	76	81	80	80
	Misc. Circuit Turn-On	76	79	77	75	82	78	69	66	68	75	62	68	78	71	74	76	80	78	74	82	78	80	81	80
	Misc. Circuit Turn-Off	75	86	80	75	86	80	69	66	68	72	74	73	70	79	74	76	86	81	73	87	80	81	84	82
	Weighted Average	79	78	78	79	78	78	69	69	69	73	73	72	75	75	75	79	79	79	78	78	78	82	82	82
rf_42	WH Circuit Turn-Off	71	100	83	71	100	83	38	100	56	50	80	62	56	100	71	71	100	83	71	100	83	56	100	71
	WH Circuit Turn-On	83	100	91	83	100	91	56	100	71	50	80	62	57	80	67	83	100	91	83	100	91	71	100	83
	Misc. Circuit Turn-On	100	96	98	100	96	98	100	84	91	95	84	89	96	88	92	100	96	98	100	96	98	100	92	96
	Misc. Circuit Turn-Off	100	92	96	100	92	96	100	68	81	95	84	89	100	84	91	100	92	96	100	92	96	100	84	91
	Weighted Average	96	95	95	96	95	95	91	80	82	88	83	85	91	87	88	96	95	95	96	95	95	94	90	91

P, R, F represent precision, recall, and f-score performance metrics respectively and all the presented results are in percentage.

WH: Water Heating | Misc.: Miscellaneous

The employed classifiers in combination with  $\mathfrak{F}$  are also evaluated at household-level, and for the said purpose the performance metrics of accuracy and Kappa index have been employed. Corresponding performance results of all the employed classifiers in combination with  $\mathfrak{F}$  for each household, within the scope of this research work, are presented in Table 31.

Table 31 Household-level Classifiers' Performance in Combination with  $\mathfrak{F}$

	SVM	LR	DT	RF	k-NN	GP	MLP	NB	
rf_01	82.67	<b>84.28</b>	69.80	74.75	79.57	82.17	83.16	79.33	Accuracy
rf_02	94.43	93.76	95.10	<b>96.10</b>	95.87	92.65	91.64	89.30	
rf_31	79.51	79.51	74.09	75.90	78.31	78.31	78.91	<b>80.12</b>	
rf_36	78.20	78.20	69.23	72.56	74.61	78.97	77.69	<b>81.53</b>	
rf_42	<b>95</b>	<b>95</b>	80	83.33	86.66	<b>95</b>	<b>95</b>	90	
rf_01	72.00	<b>74.32</b>	55.63	62.23	68.22	71.16	72.48	66.94	Kappa Index
rf_02	92.40	91.48	93.32	<b>94.69</b>	94.39	89.96	88.60	85.42	
rf_31	63.58	63.58	55.44	57.96	61.73	61.73	62.53	<b>64.29</b>	
rf_36	70.96	70.96	58.94	63.40	66.17	71.98	70.28	<b>75.38</b>	
rf_42	<b>92.37</b>	<b>92.37</b>	71.65	75.20	80.16	<b>92.37</b>	<b>92.37</b>	85.12	
Bold numbers represent the best performance for the given household and all the presented results are in percentage.									

It is evident from the presented results in Table 31 that irrespective of the reduced feature set,  $\mathfrak{F}$ , as an input to the employed classifiers, not only all the employed classifiers yield promising performance but also accommodate improvement in most of the cases (comparing Table 29 and Table 31). Moreover, from the results presented in Table 31, it is also observed that no single learning model has a clear edge over the others for all the given testing households (best performance highlighted in bold). Except for the Naïve Bayes classifier that relatively outperforms the others for two testing households, i.e., IDs. rf\_31 and rf\_36. For testing households' IDs. 01 and 02, logistic regression and random forest attained the best performance, respectively. Furthermore, for testing household ID. 42, support vector machine, logistic regression, Gaussian process, and multi-layer perceptron classifiers attained the same classification performance (highlighted in bold in Table 31). Moreover, for  $\mathfrak{F}$  being an input feature set the DT performance improves, however, compared to the other employed classifiers it is still at the downside in terms of overall classification performance.

For  $\mathfrak{F}$  as an input feature set, a classifier-level evaluation based on the entire set of testing households is also carried out in terms of accuracy and Kappa index. Figure 36 presents the corresponding results in the form of the boxplot.

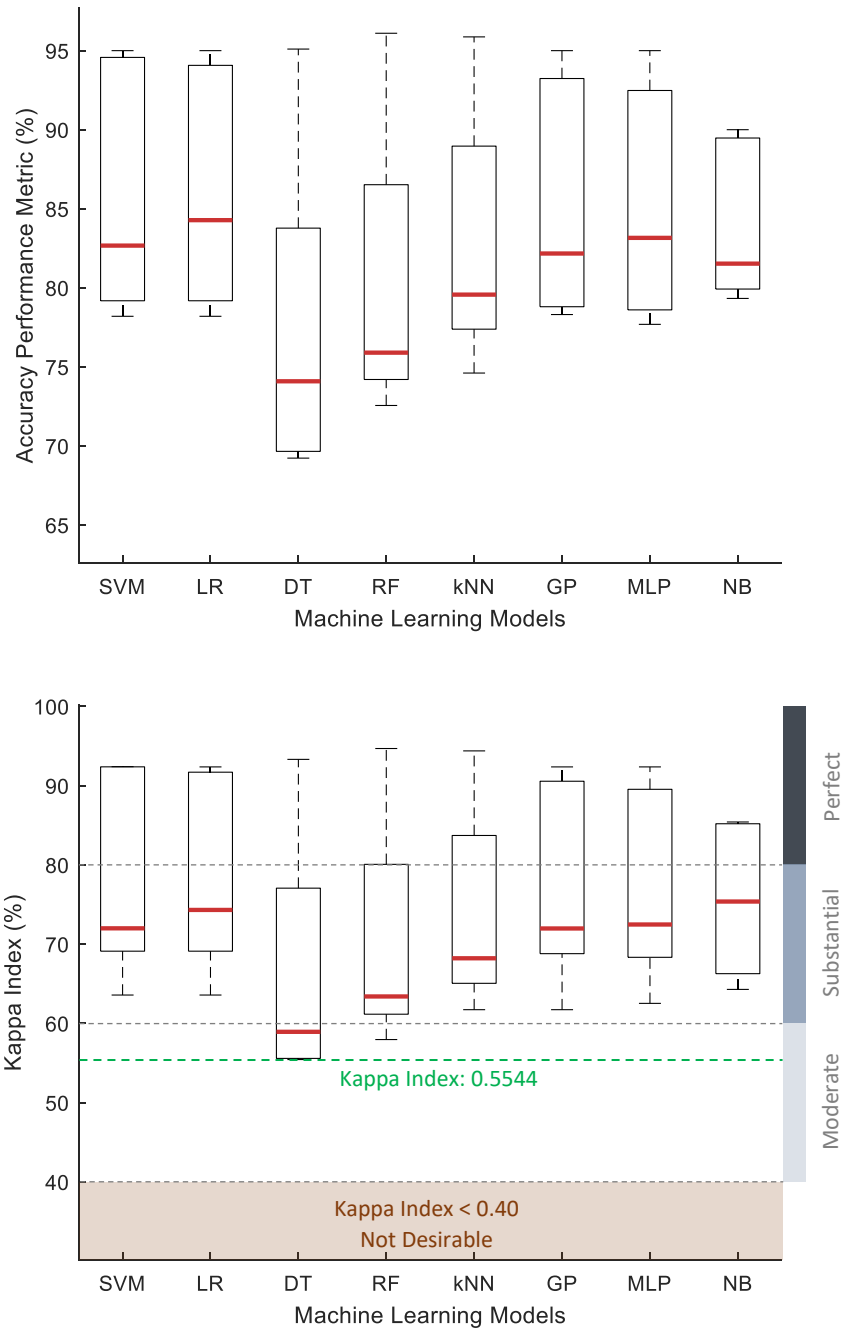


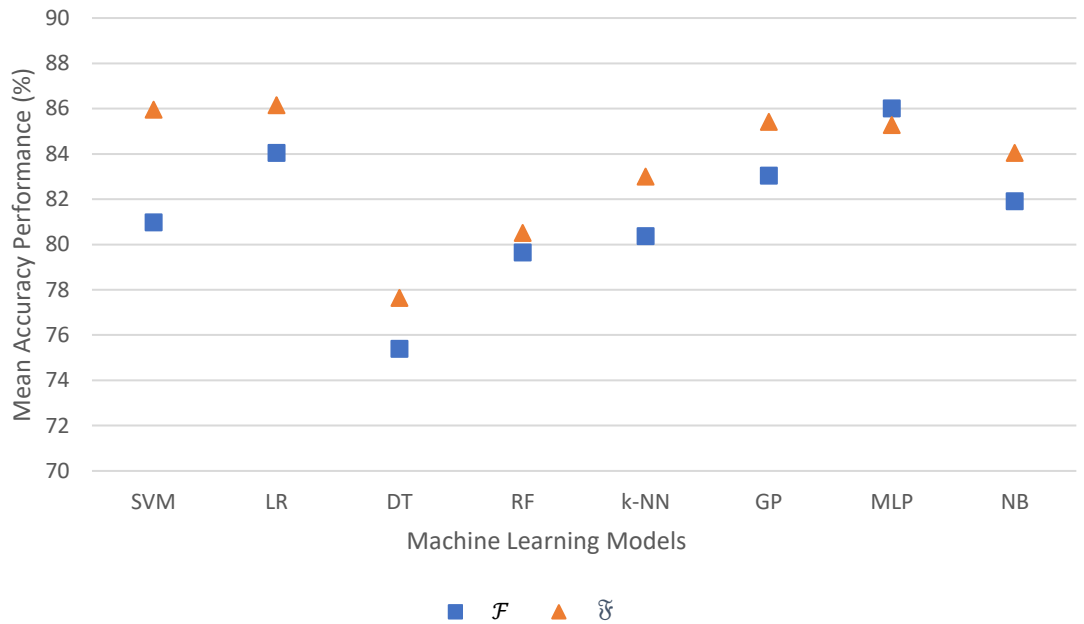
Figure 36 Overall Classifiers' Performance in Combination with  $\mathcal{F}$   
(a) Accuracy (b) Kappa Index, all the presented results are based on the entire set of testing households. The red horizontal lines represent the median values where the horizontal dotted line in green represents the minimum Kappa index (normalized) attained by the employed classifiers, under the given conditions. The regions presented on the right side of Figure 36b are based on the description presented in Table 5.

#### 4.3.2.1.2 Comparative Analysis

To underline the influence of feature space dimensionality, a comparative performance evaluation of the employed classifiers in combination with both feature sets, i.e.,  $\mathcal{F}$  and  $\mathcal{F}$ , is carried out. In this context, comparing the results of Figures 35b and 36b, it is noted that the least Kappa index performance recorded by any employed classifier (highlighted by a green dotted line) in combination with  $\mathcal{F}$  and  $\mathcal{F}$  is 43.21 and 55.44 percent,

respectively. This corresponds to around 12% of performance improvement in terms of the Kappa index, in case of reduced feature space.

To further underline the significance of feature space dimensionality based on the entire set of testing households, comparative analysis in terms of overall mean accuracy performance is carried out. Figure 37 presents the corresponding results in the form of a scatter plot. It shows that most of the employed learning models in combination with  $\mathfrak{F}$ , compared to  $\mathcal{F}$ , attained better classification accuracy. For example, in the case of the SVM, a performance improvement of up to 4.98% is achieved with the reduced feature set,  $\mathfrak{F}$ , compared to  $\mathcal{F}$ . It is also anticipated that feature space reduction may facilitate the learning models in terms of computational complexity.



**Figure 37 Feature Space Dimensionality vs. Classification Performance**  
The presented results are in the form of classifiers' mean accuracy performance based on the entire set of testing households, where  $\mathcal{F}$  and  $\mathfrak{F}$  represent the original and reduced feature sets, comprising of five and three distinct load features, respectively.

#### 4.3.2.2 Feature Selection Simulations

As observed in the previous section, reduced feature space dimensionality facilitates classification performance. In this context to further underline the significance of the load features in the context of classification performance, comprehensive simulations are carried out in terms of feature selection methodologies. The said will further strengthen the already presented analysis in terms of feature space dimensionality influence. The simulation flow adopted for the NZ GREEN Grid database, presented in Figure 32, is modified by incorporating the feature selection block, as discussed in 4.2.2, and is graphically depicted in Figure 38.

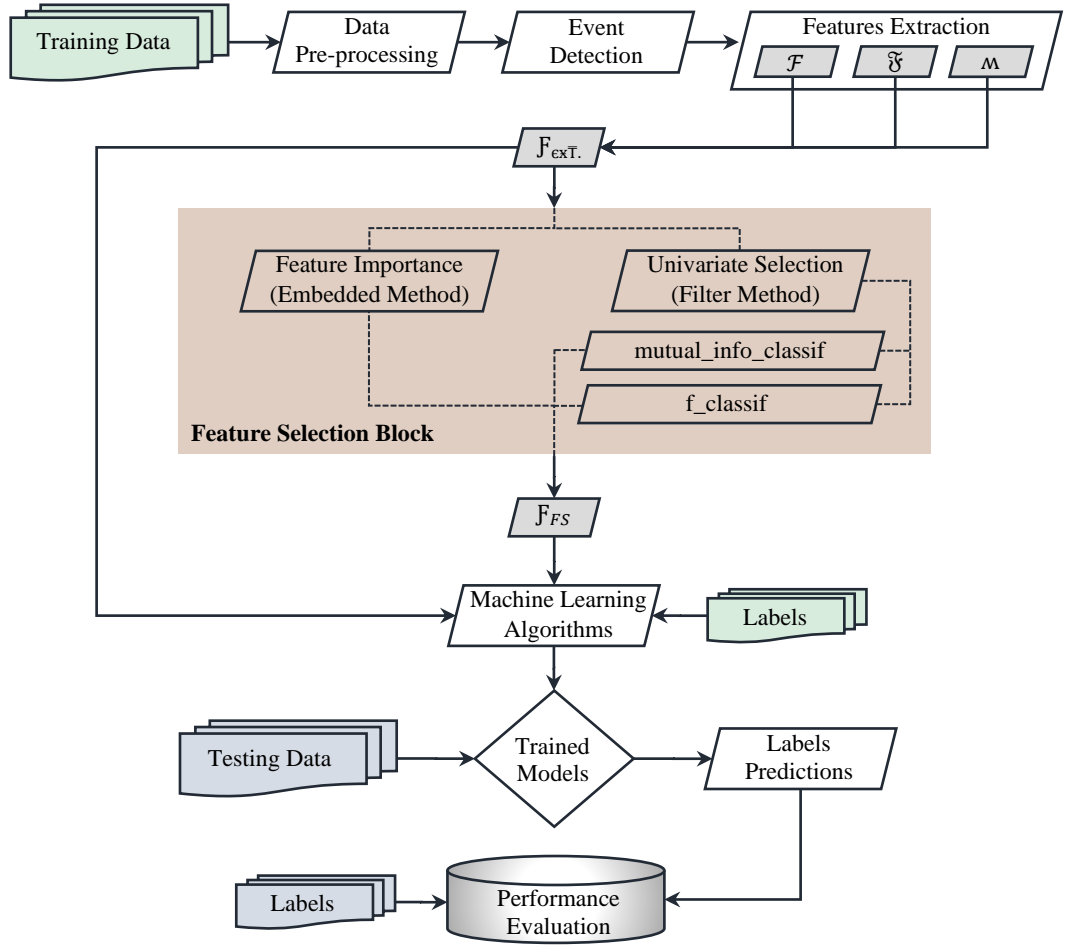


Figure 38 Feature Selection Based Non-Intrusive Load Inference Approach

As shown in Figure 38 and discussed in section 4.2.2, this research work adopted filter and embedded methods to investigate the most significant load features for the given problem. However, before plunging into feature selection simulations, it is worth noting that based on the results presented in section 4.3.2.1, particularly comparing Table 29 and Table 31, it is concluded that for the given problem the feature space dimensionality significantly facilitates the classification performance for testing households that are different, i.e., rf\_01, rf\_31, rf\_36, and rf\_42, than the one used for training purposes, i.e., rf\_02. As feature selection also aims to reduce the feature space dimensionality, therefore the simulations regarding feature selection will be mainly focussed on different testing households. The given selection of different testing households will further facilitate the analysis in terms of robustness due to the diversity factor of given testing households.

For feature selection simulations, an extended load feature set,  $F_{ext}$ , given as in (20), is extracted for all employed household data, presented in Table 27. A portion of the results in terms of extracted load features, based on  $F_{ext}$ , for a single household of the NZ GREEN Grid database is presented in Table 32.

Table 32 Extracted Load Features ' $F_{EXT}$ '

Load Features; NZ GREEN Grid Household ID rf_31								
$\tau_{width}$	$P_{p2p}$	$\sigma$	$\sigma^2$	$\mu$	$S_E$	$M$	$C_{Disp.}$	$C_{var.}$
1	0.3459	0.2446	0.0598	0.5015	0.3459	0.1729	0.1193	0.4877
1	0.3839	0.2715	0.0737	0.6626	0.383	0.1919	0.1112	0.4097
1	0.2803	0.1982	0.0393	0.4046	0.2803	0.1401	0.0971	0.4900
2	0.7394	0.3704	0.1372	0.7416	0.3697	0.3309	0.1850	0.4994
1	-0.349	0.2468	0.0609	0.8080	-0.349	0.1745	0.0753	0.3054
5	-0.272	0.6817	0.4647	1.6341	-0.054	0.6269	0.2843	0.4171
2	-1.714	0.8619	0.7429	2.0695	-0.857	0.6999	0.3589	0.4164
1	0.3661	0.2589	0.0670	1.4481	0.3661	0.1830	0.0462	0.1787
4	1.6134	0.6473	0.4191	2.4657	0.4033	0.5355	0.1699	0.2625
1	-1.333	0.9431	0.8895	2.5893	-1.333	0.6669	0.3435	0.3642
1	-0.361	0.2555	0.0653	1.3749	-0.361	0.1807	0.0475	0.1858
1	0.3901	0.2758	0.0760	1.3484	0.3901	0.1950	0.0564	0.2045

For each household under consideration, all nine distinct load features are evaluated using three different features selection techniques. The corresponding feature significance results are presented in the form of features significance score, defined as, the higher the score is the more significant is the feature and vice versa.

Tables 33 and 34 present the extracted results in the form of feature significance score based on the  $f\_classif$  and  $mutual\_info\_classif$  assessments, respectively. It is worth noting that both assessments lie in the domain of the filter methods.

Table 33 Features' Significance Score Based on  $f\_classif$  Assessment

Load Features	$f\_classif$ Assessment (Univariate Selection)			
	rf_01	rf_31	rf_36	rf_42
$\tau_{width}$	7.718722	2.661572	4.413996	8.987654
$P_{p2p}$	606.369197	222.908601	725.423027	158.846285
$\sigma$	24.099148	6.542112	93.511944	33.361798
$\sigma^2$	10.587415	3.499536	92.787630	52.573450
$\mu$	0.528617	11.276213	1.891491	0.064895
$S_E$	1399.447835	349.476223	755.473290	98.476328
$M$	22.084736	4.250438	45.677716	12.129802
$C_{Disp.}$	29.983604	3.237820	69.143266	61.318300
$C_{var.}$	7.151683	0.747307	46.206877	4.661424

Table 34 Features' Significance Score Based on  $mutual\_info\_classif$  Assessment

Load Features	$mutual\_info\_classif$ Assessment (Univariate Selection)			
	rf_01	rf_31	rf_36	rf_42
$\tau_{width}$	0.002936	0.078914	0.026197	0.167469
$P_{p2p}$	0.788226	0.650095	0.956332	1.130461
$\sigma$	0.067005	0.084969	0.215084	0.364605
$\sigma^2$	0.066411	0.080955	0.214801	0.357021
$\mu$	0.037242	0.043241	0.105265	0.089102
$S_E$	0.757102	0.650095	0.962622	0.935332
$M$	0.076684	0.051269	0.149710	0.067260
$C_{Disp.}$	0.043894	0.060418	0.217210	0.454361
$C_{var.}$	0.019504	0.067087	0.185740	0.171535

Besides the filter method, features evaluation is also carried out using the embedded method, i.e., features importance technique based on ExtraTreesClassifier. Table 35 presents the features significance score of each load feature for all the four diverse testing households based on feature importance assessment.

Table 35 Features' Significance Score Based on Feature Importance Assessment

Load Features	Feature Importance Assessment			
	rf_01	rf_31	rf_36	rf_42
$\tau_{width}$	0.016924	0.005606	0.027754	0.025618
$P_{p2p}$	0.235405	0.327899	0.260014	0.287648
$\sigma$	0.055907	0.020943	0.127056	0.114359
$\sigma^2$	0.054942	0.022241	0.095723	0.058623
$\mu$	0.045531	0.033672	0.060553	0.013738
$\mathcal{S}_E$	0.466498	0.507645	0.244609	0.380297
$\mathcal{M}$	0.048149	0.023126	0.064355	0.037655
$C_{Disp.}$	0.036820	0.035223	0.068357	0.064050
$C_{var.}$	0.039823	0.023645	0.051579	0.018014

From the results presented in Tables 33-35, it is evident that irrespective of the employed feature selection methodology, the load features based on geometrical and power features, i.e.,  $\mathcal{S}_E$  and  $P_{p2p}$ , emerge as the most significant load features for all the testing households, under the given conditions. The third most significant rank within  $\mathcal{F}_{EXT.}$  is equally shared by  $C_{Disp.}$  and  $\sigma$ , each having the frequency of 5 out of 12<sup>33</sup> instances. Further, statistical features like  $\mu$  and  $\mathcal{M}$  also emerged as the third most significant load feature by the filter method, however, each having the frequency of only 1 out of 12 instances. In this context, to better visualize the results presented in Tables 33-35, Table 36 highlights the features significance rank for all the testing households by different feature selection methodology. The columns of Table 36 present the corresponding load features' significance rank in the descending order.

Table 36 Load Features' Significance Ranking for all Testing Households

Filter Method								Embedded Method				Ranking
$f\_classif$				$mutual\_info\_classif$				feature importance				
rf_01	rf_31	rf_36	rf_42	rf_01	rf_31	rf_36	rf_42	rf_01	rf_31	rf_36	rf_42	
$\mathcal{S}_E$	$\mathcal{S}_E$	$\mathcal{S}_E$	$P_{p2p}$	$P_{p2p}$	$P_{p2p}$	$\mathcal{S}_E$	$P_{p2p}$	$\mathcal{S}_E$	$\mathcal{S}_E$	$P_{p2p}$	$\mathcal{S}_E$	1 <sup>st</sup>
$P_{p2p}$	$P_{p2p}$	$P_{p2p}$	$\mathcal{S}_E$	$\mathcal{S}_E$	$\mathcal{S}_E$	$P_{p2p}$	$\mathcal{S}_E$	$P_{p2p}$	$P_{p2p}$	$\mathcal{S}_E$	$P_{p2p}$	2 <sup>nd</sup>
$C_{Disp.}$	$\mu$	$\sigma$	$C_{Disp.}$	$\mathfrak{M}$	$\sigma$	$C_{Disp.}$	$C_{Disp.}$	$\sigma$	$C_{Disp.}$	$\sigma$	$\sigma$	3 <sup>rd</sup>
$\sigma$	$\sigma$	$\sigma^2$	$\sigma^2$	$\sigma$	$\sigma^2$	$\sigma$	$\sigma$	$\sigma^2$	$\mu$	$\sigma^2$	$C_{Disp.}$	4 <sup>th</sup>
$\mathfrak{M}$	$\mathfrak{M}$	$C_{Disp.}$	$\sigma$	$\sigma^2$	$\tau_{width}$	$\sigma^2$	$\sigma^2$	$\mathfrak{M}$	$C_{var.}$	$C_{Disp.}$	$\sigma^2$	5 <sup>th</sup>
$\sigma^2$	$\sigma^2$	$C_{var.}$	$\mathfrak{M}$	$C_{Disp.}$	$C_{var.}$	$C_{var.}$	$C_{var.}$	$\mu$	$\mathfrak{M}$	$\mathfrak{M}$	$\mathfrak{M}$	6 <sup>th</sup>
$\tau_{width}$	$C_{Disp.}$	$\mathfrak{M}$	$\tau_{width}$	$\mu$	$C_{Disp.}$	$\mathfrak{M}$	$\tau_{width}$	$C_{var.}$	$\sigma^2$	$\mu$	$\tau_{width}$	7 <sup>th</sup>
$C_{var.}$	$\tau_{width}$	$\tau_{width}$	$C_{var.}$	$C_{var.}$	$\mathfrak{M}$	$\mu$	$\mu$	$C_{Disp.}$	$\sigma$	$C_{var.}$	$C_{var.}$	8 <sup>th</sup>
$\mu$	$C_{var.}$	$\mu$	$\mu$	$\tau_{width}$	$\mu$	$\tau_{width}$	$\mathfrak{M}$	$\tau_{width}$	$\tau_{width}$	$\tau_{width}$	$\mu$	9 <sup>th</sup>

<sup>33</sup> Each feature has the evaluation frequency of 12 during the entire feature selection simulations, e.g., each feature corresponding to 4 households and evaluated by 3 assessment techniques (4×3=12).

Table 36 provides extra ease compared to Tables 33-35 in terms of identifying the significance rank of the given load features for each testing household. From the given results, it is seen that  $\mathcal{S}_E$  and  $P_{p2p}$  are unanimously ranked as the top two most significant loads features for all the households. On the other side, the load features, i.e., mean, and transient width, are frequently identified as the least significant load features under the given conditions. It is also observed that the extracted load features vary in significance rank among different testing households. It is anticipated due to the diverse set of testing households and associated diverse factors, i.e., independent circuits, installation configurations, and consumption patterns, etc. Even for the same testing household, the given load features are ranked differently by different feature selection assessments. This is also expected due to the different working principles of the employed feature selection methodologies, i.e., classifier-agnostic and classifier-interaction.

Based on the feature selection simulations and corresponding analysis, it is concluded that among all nine distinct load features for the given testing households, the top three load features are  $P_{p2p}$ ,  $\mathcal{S}_E$ , and  $C_{Disp}$ . or  $\sigma$ . Consequently, a new feature set  $\mathcal{F}_{FS}$ , given in (22), is formulated based on the analysis of the extracted feature selection results. It is worth noting that the newly formulated feature selection based feature set,  $\mathcal{F}_{FS}$ , comprises two distinct load feature based on the aforesaid top-ranked features while keeping the possible redundancy among the given load features at the minimum level.

$$\mathcal{F}_{FS} = \{\mathcal{S}_E, C_{Disp}\} \quad (22)$$

To further investigate the significance of the newly formulated feature set,  $\mathcal{F}_{FS}$ , the given results in (22) are validated in terms of classification performance. For the said purpose,  $\mathcal{F}_{FS}$  is independently employed as an input to the classifiers besides  $\mathcal{F}_{exT}$ . as per simulations methodology depicted in Figure 38. Table 37 presents the details of the household IDs linked with training and testing of the learning models. Table 37 also presents the details of input features to the classifiers.

Table 37 Feature Selection Simulation Data

		Training Data	Testing Data			
		rf_02	rf_01	rf_31	rf_36	rf_42
Data Acquisition		May. 11-30,	Sep. 01-07,	Sep. 01-07,	Jun. 21-27,	Jan. 07-13,
Timeframe		2014	2014	2016	2017	2017
Number of Days		20	7	7	7	7
Total Detected Events		1504	808	166	390	60
Classifier Input	$\mathcal{F}_{exT}$ .	1504×9	808×9	166×9	390×9	60×9
	$\mathcal{F}_{FS}$	1504×2	808×2	166×2	390×2	60×2



Based on the analysis presented in section 4.3.2.1, three learning models that attained the best and least classification performance compared to other employed models are employed to validate the results of feature selection methodologies. The employed learning models are multi-layer perceptron, Naïve Bayes, and decision trees. All three classifiers are independently evaluated in combination with  $F_{\text{exT}}$  and  $F_{FS}$ , as an input features, in terms of accuracy performance metric. Table 38 presents the corresponding simulation results in terms of classifiers' accuracy performance, in percentage. For further validation of the feature selection based results, Table 38 also presents an extended comparative performance analysis of the employed classifiers in combination with  $\mathcal{F}^{34}$  and  $\mathcal{F}^{35}$  along with  $F_{\text{exT}}$  and  $F_{FS}$ . It is also worth noting that while analyzing the classification performance results presented in Table 38, the feature space dimensionality perspective must be considered, i.e., feature space dimensionality is directly proportional to the model's complexity and computation demands. And under given conditions, the feature sets, i.e.,  $F_{FS}$ ,  $\mathcal{F}$ ,  $\mathcal{F}$ , and  $F_{\text{exT}}$  are comprised of two, three, five, and nine load features, respectively.

Table 38 Classifiers' Accuracy Performance Based on Feature Selection Methods

	Accuracy Performance Metric (%)											
	<b>Multi-Layer Perceptron</b>				<b>Decision Trees</b>				<b>Naïve Bayes</b>			
	$F_{\text{exT}}$	$F_{FS}$	$\mathcal{F}$	$\mathcal{F}$	$F_{\text{exT}}$	$F_{FS}$	$\mathcal{F}$	$\mathcal{F}$	$F_{\text{exT}}$	$F_{FS}$	$\mathcal{F}$	$\mathcal{F}$
rf_01	79.24	83.41	83.16	78.96	71.41	71.28	69.80	70.54	77.35	80.94	79.33	72.40
rf_31	77.10	78.91	78.91	78.91	75.90	71.68	74.09	73.49	76.50	78.91	80.12	75.30
rf_36	78.20	77.17	77.69	79.23	70.76	74.87	69.23	57.43	79.23	80.76	81.53	77.43
rf_42	95	95	95	96.66	78.33	80	80	80	86.66	91.66	90	88.33

In terms of  $F_{\text{exT}}$  and  $F_{FS}$  (highlighted in brown color), it is evident from the results presented in Table 38 that for most of the cases feature selection methodology facilitates the classifiers' performance. For example, for all testing households the NB classifier in combination with  $F_{FS}$  performs better than the NB in combination with  $F_{\text{exT}}$ . Likewise, in combination with  $F_{FS}$ , the classification performance of the MLP and DT classifiers also improves for different households. In a few cases, the MLP and DT classifiers perform better in combination with  $F_{\text{exT}}$ , however, it is also observed that the performance is either equivalent or having a marginal variation. But in the larger perspective of models' accuracy, complexity, and computational needs,  $F_{FS}$ , as an input feature has an edge over the  $F_{\text{exT}}$ . Further in the said larger perspective, it is also evident from the presented results in Table 38 that the employed classifiers with feature selection

<sup>34</sup> Based on the results presented in Table 29.

<sup>35</sup> Based on the results presented in Table 31.

perform better compared to the given classifiers in combination with  $\mathcal{F}$  and  $\mathcal{G}$ , which further underline the significance of the feature selection methodologies.

Based on the presented simulations and corresponding results, it is concluded that irrespective of having smaller numbers of load features,  $\mathbf{F}_{FS}$ , as an input feature to the classifiers, yields relatively better performance. This further establishes that classification performance is dependent on the relevance or significance of the input features rather than the number of input features.

#### **4.3.2.3 Ensemble Learning Approach**

As evident from Table 38 that each employed machine learning algorithm in combination with a given input feature performs differently for different testing data. For example, for testing data rf\_42 the MLP classifier outperforms the others, while for testing household rf\_36 the NB classifier has an edge over the other classifiers. It is anticipated due to the diverse nature of the employed classifiers, hence, to balance the performance of different classifiers for the given testing households, the concept of the ensemble learning is also investigated besides standalone learning models in the context of the proposed low-sampling event-based non-invasive load disaggregation approach. The ensemble learning refers to a combinatorial form of independent learning models, built to solve a given computational problem. The ensemble learning models are mostly used to enhance the classification performance and considered as a reliable approach in the said context [124]. However, it is also probable that the ensemble model does not lead to any improvement in terms of classification performance compared to the performance of the best learning model within the given ensemble model [125].

In the existing literature, numerous ensemble techniques are proposed, however, within the scope of this research work (based on diverse learning models having independent outcomes), a voting-based ensemble model with a majority voting concept is adopted. As voting based ensemble model with the majority voting concept is the optimal combination rule for the classifiers having independent outcomes [125]. In principle, a voting classifier merges diverse learning models and the final prediction is based on the voting system, i.e., hard vote or soft vote [107]. Hard voting and soft voting refer to the majority voting and average predicted probabilities, respectively. Figure 39 graphically illustrates a general framework of a voting classifier.

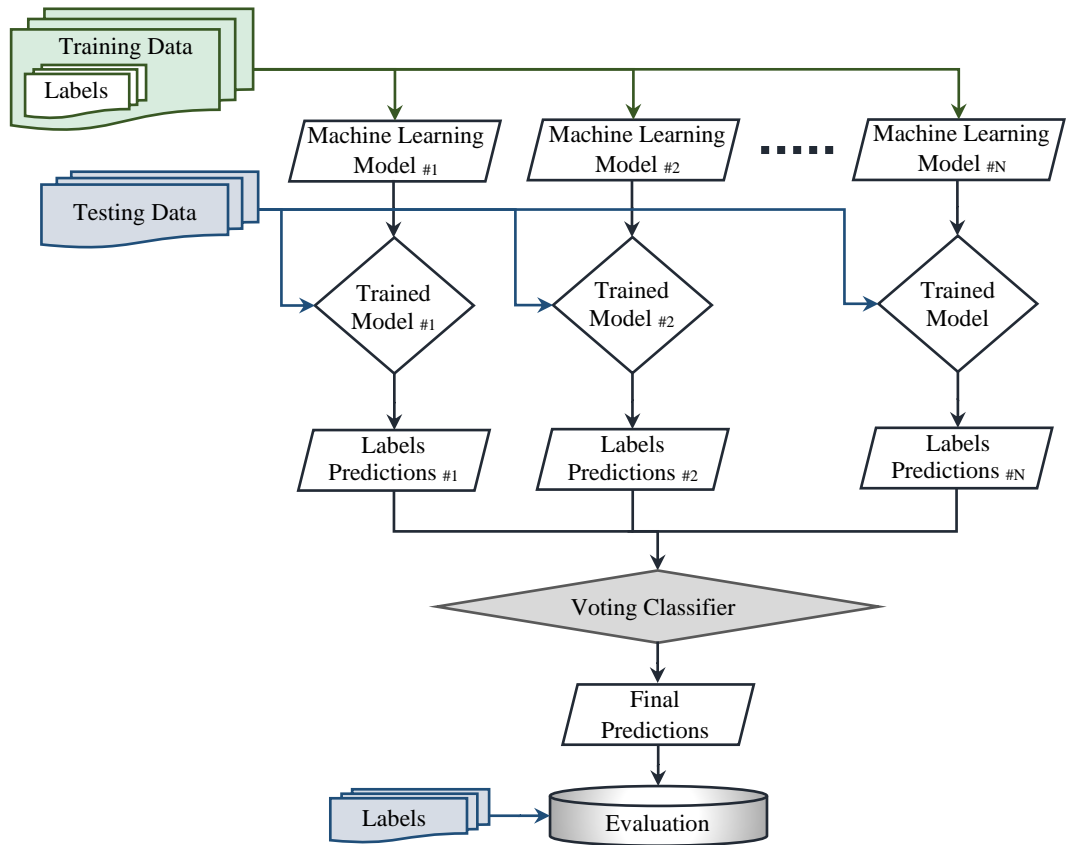


Figure 39 Voting Classifier Framework

In this research work, the voting-based ensemble learning model is built on a combination of three independent and diverse learning models namely, multi-layer perceptron, decision trees, and Naïve Bayes. It is worth noting that the members of the employed ensemble model are selected not only for their diversity but also for their corresponding classification performance, under the given conditions, as discussed thoroughly in section 4.3.2.1. It is expected that the said combination of standalone learning model, in the form of ensemble learning, will facilitate the classification performance by balancing the performance of the given machine learning models. For evaluation purposes, the proposed majority voting based ensemble model is rigorously tested in combination with all the extracted feature sets within the scope of this research work, i.e.,  $\mathcal{F}$ ,  $\mathcal{F}_T$ ,  $\mathcal{F}_{EXT}$ , and  $\mathcal{F}_{FS}$  given in (12), (16), (20), and (22) respectively. The said evaluation also provides an insight into the significance of the input features in the context of ensemble learning.

According to the methodology presented in Figure 39, comprehensive digital simulations are carried out on different testing data. The corresponding ensemble learning performance results are presented in Table 39 along with an extended comparison with the classification performance of respective members, i.e., the MLP, NB, and DT, on which the given ensemble learner is built. The extended comparative performance results are highlighted in brown color in Table 39, where all the results are in percentage.

Table 39 Ensemble Model Performance Comparison

	Accuracy Performance Metric (%)						
	Ensemble Model				MLP	DT	NB
	$F_{\text{exT.}}$	$F_{FS}$	$\mathcal{F}$	$\mathcal{F}$		$F_{FS}$	
rf_01	78.21	81.18	81.31	76.11	83.41	71.28	80.94
rf_31	77.71	79.51	78.31	75.90	78.91	71.68	78.91
rf_36	78.46	81.28	79.23	77.94	77.17	74.87	80.76
rf_42	93.33	93.33	93.33	93.33	95	80	91.66

For feature set,  $F_{FS}$ , as an input to the classifier, it is evident from the results presented in Table 39 (highlighted in brown color) that the employed ensemble learning model achieved better classification performance for two testing households, i.e., IDs. rf\_31 and rf\_36, compared to the performance of all its respective members. On the other side, for testing household IDs. rf\_01 and rf\_42, the ensemble model attained better performance compared to most of its respective members, i.e., DT and NB, but lags in performance compared to the best performing member, i.e., MLP. This is not unexpected, as discussed earlier that there is no guarantee that ensemble models always lead to an improvement in terms of classification performance compared to the performance of the best learning model within the given ensemble [125]. However, in terms of overall mean accuracy performance (extracted from the results highlighted in brown in Table 39) based on the entire set of testing households, the proposed ensemble learner leads with a result of 83.82% compared to 83.62, 74.45, and 83.06 percent for MLP, DT, and NB models, respectively.

Further in terms of intra-ensemble model comparison, i.e., in combination with different input features, it is evident from the results presented in Table 39 that irrespective of less number of input features, the ensemble model in combination with  $F_{FS}$ , facilitates the overall classification performance, except for testing household rf\_01 where the ensemble model in combination with  $F_{FS}$  lags in performance compared to the best performance but that is also marginal, i.e., 0.13% only.

The presented intra-ensemble comparative analysis can be further visualized in Figure 40 in the form of a scatter plot representing the accuracy performance of the employed ensemble learning model in combination with different input feature sets, i.e.,  $\mathcal{F}$ ,  $\mathcal{F}$ ,  $F_{\text{exT.}}$ , and  $F_{FS}$ .

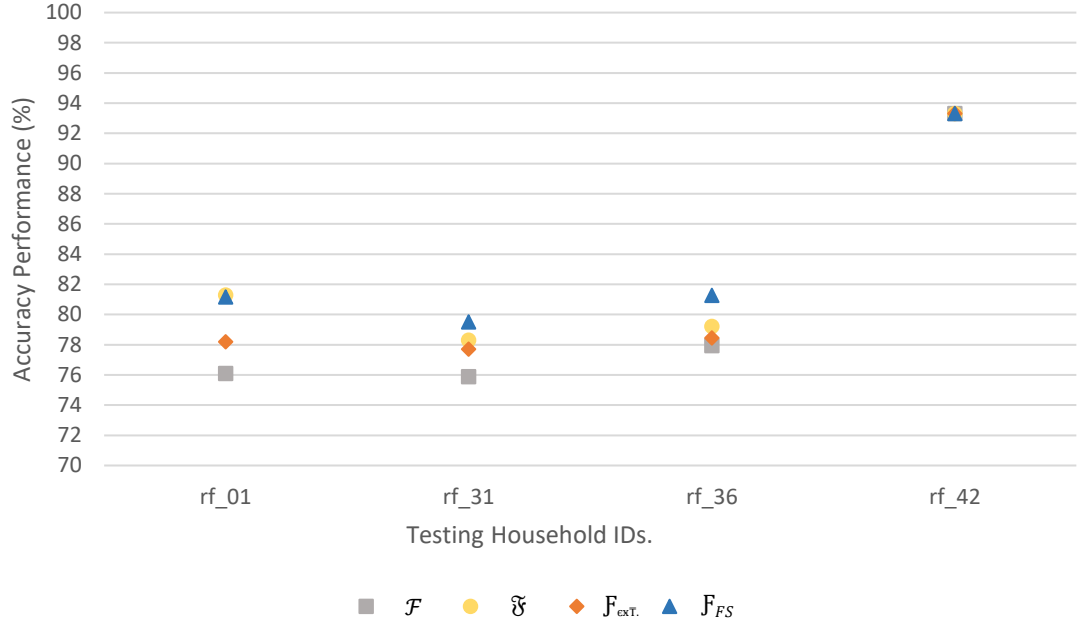


Figure 40 Intra-Ensemble Model Classification Performance

Based on the results presented in Figure 40 and the overall model's complexity and computational demand associated with the input feature space dimensionality, it would be fair to conclude that ensemble learner with feature selection has an edge over the given ensemble model without feature selection. Further, in terms of ensemble vs standalone configuration, it is evident from the analysis of the results that the ensemble model manages the classification performance improvement even in the presence of low performing member, i.e., the DT model, for the given problem. However, it is also worth noting that this improvement comes at a price of high complexity and computational demands due to the ensemble learning approach.

#### 4.4 Concluding Remarks

This chapter presented a complete flow of the proposed low sampling event-based non-intrusive load disaggregation approach, where comprehensive simulations are carried out on real-world load measurements acquired from the diverse load databases from different geographical regions. Based on the presented simulation studies and the corresponding analysis of the results, it is established that the proposed non-invasive approach works well and attained promising results in terms of load inference.

The research work presented in this chapter contributes to the existing state of the art on NILM in terms of feature space dimensionality and classification models, both in standalone and ensemble configuration. It is also expected that the presented analysis will significantly facilitate future research in the aforesaid domains, particularly in the domain of low sampling non-intrusive load inference approaches.

The proposed non-invasive load inference approach presented in this research work has a solid potential in real-life deployment towards energy-efficient systems. In this context, to validate the significance of the given research, a proof of concept in the form of real-world energy efficiency application is presented in the following chapter, i.e., Chapter 5.

## Chapter 5 Non-Intrusive Load Disaggregation Applications

This Chapter presents a brief outlook of the potential real-world applications assisted by non-invasive load disaggregation. Moreover, based on the presented research work a proof of concept is presented and validated in this Chapter.

The details presented in this chapter regarding the NILM assisted real-world application framework and corresponding case study are mainly based on [126]. The said research manuscript has been published as follows:

- A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Invasive Load-Shed Authentication Model for Demand Response Applications Assisted by Event-Based Non-Intrusive Load Monitoring", *Energy and AI*, p. 100055, 2021.
- DOI: <https://doi.org/10.1016/j.egyai.2021.100055>

### 5.1 State of the Art on Applications

Despite the availability of a range of research in the domain of non-intrusive load disaggregation and corresponding aspects, still, the existing literature is lagging in terms of its applications in the broader context of existing power grid systems [34]. This leads to the lack of actionable feedback regarding non-invasive load disaggregation applications in general and more specifically in a real-world scenario<sup>36</sup>. Recently, the research community came forward to address this shortcoming, e.g., Hernández et al. [60] briefly discuss the applications of non-intrusive load monitoring in the context of energy management and ambient assisted living. Zhuang et al. [62] also present an overview of the business applications of non-intrusive load monitoring.

Further, in the context of NILM actionable feedback, the authors of [8] recently presents a comprehensive review of the existing state of the art; exploring a wide-range of promising NILM applications including load scheduling, health monitoring, fault detection, demand response, and energy auditing. From the system perspective, the study [8] broadly classified the existing literature on NILM application into four categories namely, consumer-based applications, utility/policy-based applications, manufacturer-based applications, and miscellaneous applications. Figure 41 [8] graphically highlights numerous real-world applications of non-intrusive load disaggregation in the context of each aforesaid category.

---

<sup>36</sup> In the context of NILM, most of the existing research targets inference of a greater number of appliances, consequently utilizing high data granularity acquisition systems. However, in the real-world scenario, the existing metering infrastructure does not support high sampling data acquisition.

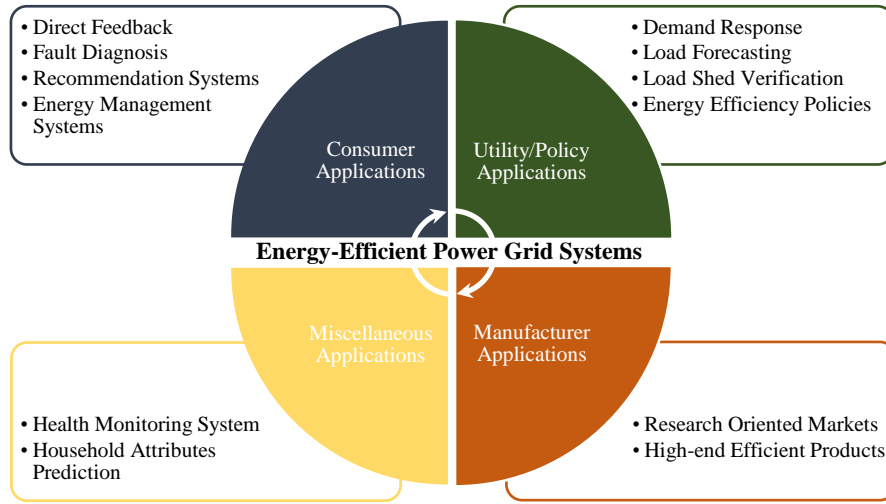


Figure 41 Categorization of NILM Applications

The potential real-world applications of the non-intrusive load disaggregation presented in Figure 41 are further elaborated in Table 40 [8], which presents a summary of the promising applications along with the corresponding aspects, i.e., respective study, application categories, and targeted appliances. It is evident from the existing literature that the incorporation of non-intrusive load disaggregation in real-world applications can play an active role in energy efficient systems [4]. However, it requires further research to be carried out to realize the real potential of non-intrusive load disaggregation particularly in the context of real-world deployment.

## 5.2 Proof of Concept

Rather than focusing on high data granularity towards inference of a greater number of appliances, this research work is primarily based on low data granularity, targeting the inference of high consumption, most importantly the flexible/interruptible appliances/circuits. This makes the presented research work more compatible and viable to be incorporated in real-world scenarios towards energy efficient systems. In this context, to further validate the effectiveness of the proposed non-intrusive load disaggregation approach in real-world deployment, a proof of concept in terms of non-invasive load-shed authentication framework for demand response (DR) program is presented. Towards effective DR programs, the proposed NILM assisted load-shed authentication model could facilitate both stakeholders, i.e., consumers and utility providers.



Table 40 State of the art on Non-Intrusive Load Disaggregation Applications

Study	Category	Targeted Applications	Targeted Appliances/Circuits
Lin and Tsai [11]	Consumer, Utility, Policy	Load scheduling, home energy management system (HEMS)	Major household appliances
Batra et al. [127]	Consumer	Malfunction or misconfiguration detection and provision of actionable feedback	Heating ventilation and air conditioning (HVAC), Refrigerators
Nation et al. [128]	Consumer	Fault detection	Water disposal system
Berges et al. [129]	Consumer, Utility, Policy	Energy audit	Household appliances
Liu et al. [130]	Consumer, Utility	Appliance scheduling	Air conditioner
Donnal et al. [131]	Consumer, Utility	Energy box platform; enables energy monitoring and control	Appliances like, air compressor, heating, shop equipment
Adabi et al. [132]	Consumer, Utility, Policy	Enabling energy management systems (EMS)	Vacuum cleaner, microwave, and leaf blower
Luo et al. [133]	Consumer	Personalized recommendation system (PRS)	Household appliances
He at al. [134]	Utility, Policy	Demand response (DR)	Circuit breaker level loads, e.g., HVAC, electric heat devices, EV charger & strip level loads, e.g., display devices (television, projector), resistive loads, etc.
Batra et al. [135]	Utility, Manufacturer, Miscellaneous	Predicting household properties	HVAC features are used as indicative of household properties
Alcala et al. [136]	Consumer, Miscellaneous	Health monitoring	Kettle (as an indicator of household occupants' routine)
Alcala et al. [137]	Consumer, Miscellaneous	Homecare monitoring system	BLUED [28] and PLAID [26] datasets are used to evaluate the performance of event detector and classifiers respectively.

### 5.2.1 Demand Response Program

Demand response is a key element of modern days power grid systems, particularly in the presence of increasing numbers of distributed renewable energy sources, where it plays a vital role in terms of improving grid reliability [138]. Demand response refers to approaches that alter the consumers' energy consumption either by reducing, shedding, or shifting loads to maintain an equilibrium between power supply and demand. Han and Piette [139] present a comprehensive overview of demand response programs and categories it into two main types: incentive-based DR and time-based rates DR. The former refers to the programs where the consumers get preferential tariff rates for a non-DR time due to altering their energy consumption patterns in times of system contingencies. The latter is built on rising price signals and the corresponding reduction of consumers' energy consumption [139]. The two types of DR can be further split into numerous categories, summarized in Figure 42 with further details in [139].

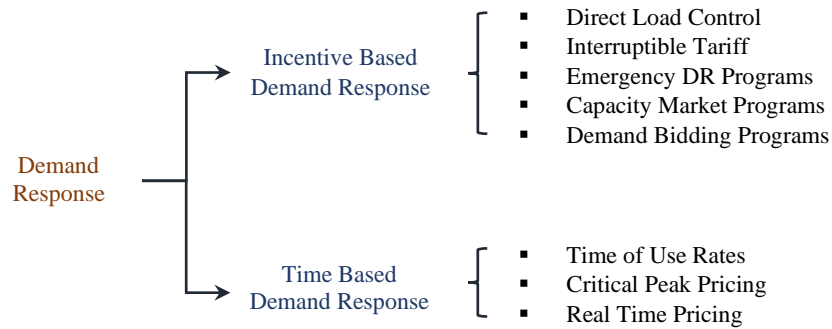


Figure 42 Demand Response Categories

The controlling strategies are the key factors towards the effective deployment of demand response programs. In this context, Piette et al. [138] classified the DR control into three categories: manual, semi-automated, and automated DR programs. Manual DR, here referred to as user-assisted demand response (UDR), is a technique where individual loads are controlled physically by the consumer, based on the received DR instructions. Semi-automated DR refers to an approach where a preprogrammed DR strategy is initiated by a person using a centralized control system [138]. Automated demand response (ADR) is a DR strategy that is completely non-invasive in terms of human involvement.

Demand response programs are widely adopted in different geographical regions. Samad et al. [140] present different case studies that are based on demand response programs for smart buildings and microgrids in different geographical regions. In New Zealand, a ripple control system for water heating circuits was introduced in the 1950s [141]. In the

recent past, Transpower<sup>37</sup> New Zealand, which owns and operates the national grid, started its early demand response trials in 2007. Later in 2011, Transpower acquired a DR management system to further scale up its DR programs [142].

### **5.2.2 Non-Intrusive Load Disaggregation Assisted Demand Response**

This section presents the details of the proposed DR framework assisted by the proposed non-intrusive load disaggregation approach, as a potential application of the presented research work in this thesis.

#### **5.2.2.1 Rationale and Significance**

Demand response has an effective and proven role in the sustainability and reliability of the power grid systems. One of the widely used demand response approaches is the load curtailment, commonly referred to as load shedding approach. In this approach, pre-selected consumers' loads are curtailed in time of need, i.e., during the peak demand or system contingencies. But since long this approach lacks a trust-worthy business model towards effective deployment due to a tradeoff that exists among consumers and utility providers. For example, if the consumers' loads are controlled directly (remotely) by the utility, there is a possibility that the consumers may face unpleasant situations due to curtailment of their loads by the utility at the time of their requirement, e.g., water heating circuit. Likewise, if the consumer's loads are controlled by consumers themselves based on demand response instructions received from the utility provider, there is a possibility that the consumers do not follow the DR instruction while getting incentives from the utility providers based on their mutual agreement as per established DR program. To address this tradeoff, it is necessary to work out a DR program that facilitates both parties, i.e., empowers the consumers in terms of controlling their load and enable the utility providers to effectively monitor the consumers' DR compliance, subsequently, eliminate the freeloaders<sup>38</sup>. In this context, this research work presents a DR business model assisted by non-intrusive load disaggregation, inspired by the early work presented in [143], that can facilitate all the concerned stakeholders.

Based on the aforesaid rationale, this research work presents a demand response program incorporated with a non-invasive load-shed authentication framework, that is assisted by low-sampling event-based non-intrusive load disaggregation approach proposed in this thesis. For evaluation and validation purposes, real-world load data from NZ GREEN

---

<sup>37</sup> <https://www.transpower.co.nz/>

<sup>38</sup> Freeloaders refer to those consumers who acknowledge but do not execute the received demand response instructions from the utility provider, while at the same time getting benefits from the incentives offered by the utilities based on the terms of their already existing demand response agreements.

Grid is used, where a water heating circuit is utilized as a potential DR load, as it is not only the most significant load element of a typical household but also has a high potential towards demand response [121]. Further, the water heating circuit has a proven history in DR programs, e.g., in New Zealand, water heating based ripple control dates back to the 1950s<sup>39</sup>, for balancing the demand and supply by controlling the former entity.

The presented NILM based DR application in this thesis, based on low-sampling data granularity, will not only significantly contribute to the existing literature in terms of broader applications of NILM but also provide a reliable non-invasive trust model for the concerned stakeholders towards effective deployment of demand response programs.

### 5.2.2.2 Problem Formulation and Methodology

As discussed earlier the time-series load consumption can be considered as an algebraic sum of  $i$  numbers of loads' consumption at a single metering point, given as in (23).

$$Load_{Metering\ Point} = \sum_{x=1}^i Load_x(t) \quad (23)$$

With the only information of the aggregated load:  $Load_{Metering\ Point}$ , the task of the NILM is to identify the operation state of the individual loads:  $Load_x(t)$ , where  $x=1, 2, 3, \dots, i$ . In the given context, assume, a household having  $A$  appliances and a single day has  $T$  timeslots, where  $A=\{1, 2, 3, \dots, a, \dots, A\}$  and  $T=\{1, 2, 3, \dots, t, \dots, T\}$ . Under the given conditions,  $T$  equal to 1440, as the given database (NZ GREEN Grid) has the data granularity of 1/60 Hz, i.e., 1-minute sampling rate. Further, the given household has a DR enrolled appliance  $a$ , given that  $a \in A$ . For non-invasive load-shed authentication purposes in DR program, the DR enrolled appliance  $a$ , water heating circuit under given conditions, is monitored and inferred, using the proposed NILM approach, for  $D$  days, where  $D=\{1, 2, 3, \dots, d, \dots, D\}$ . In this context, (23) can be represented as in (24).

$$Load_{Metering\ Point} = Load_a(t) + n(t) \quad (24)$$

where  $n(t)$  is a measurement noise comprised of acquisition noise along with other household appliances' consumptions not under consideration. The proposed non-intrusive load disaggregation approach will monitor and estimate the operation status  $s$  of the DR enrolled appliance  $a$  for  $D$  days, mathematically defined as in (25).

$$s_a^{t,d} = \begin{cases} 0, & \text{Appliance Turn-Off} \\ 1, & \text{Appliance Turn-On} \end{cases} \quad (25)$$

<sup>39</sup> <https://www.transpower.co.nz/keeping-you-connected/demand-response/demand-response-journey-so-far>

where  $s_a^{t,d}$  denotes the operation status, i.e., turn-on and turn-off, of the given DR enrolled household appliance  $a$ , i.e., water heating circuit, at timeslot  $t$  in day  $d$ . The information provided, i.e.,  $s_a^{t,d}$ , by the non-intrusive load disaggregation approach plays a vital role in terms of non-invasive load-shed authentication, as it can be used to authenticate whether an appliance is turned-off or turned-on in reality according to the DR instruction provided by the utility provider. To accomplish this, a complete framework is proposed for the DR program with non-invasive load-shed authentication capabilities assisted by the non-intrusive load disaggregation approach presented in this research work. The proposed framework is depicted in Figure 43 and is comprised of different blocks/components namely, non-intrusive load disaggregation, smart demand response hub, in-house display, controllable relays, metering device, and communication gateway.

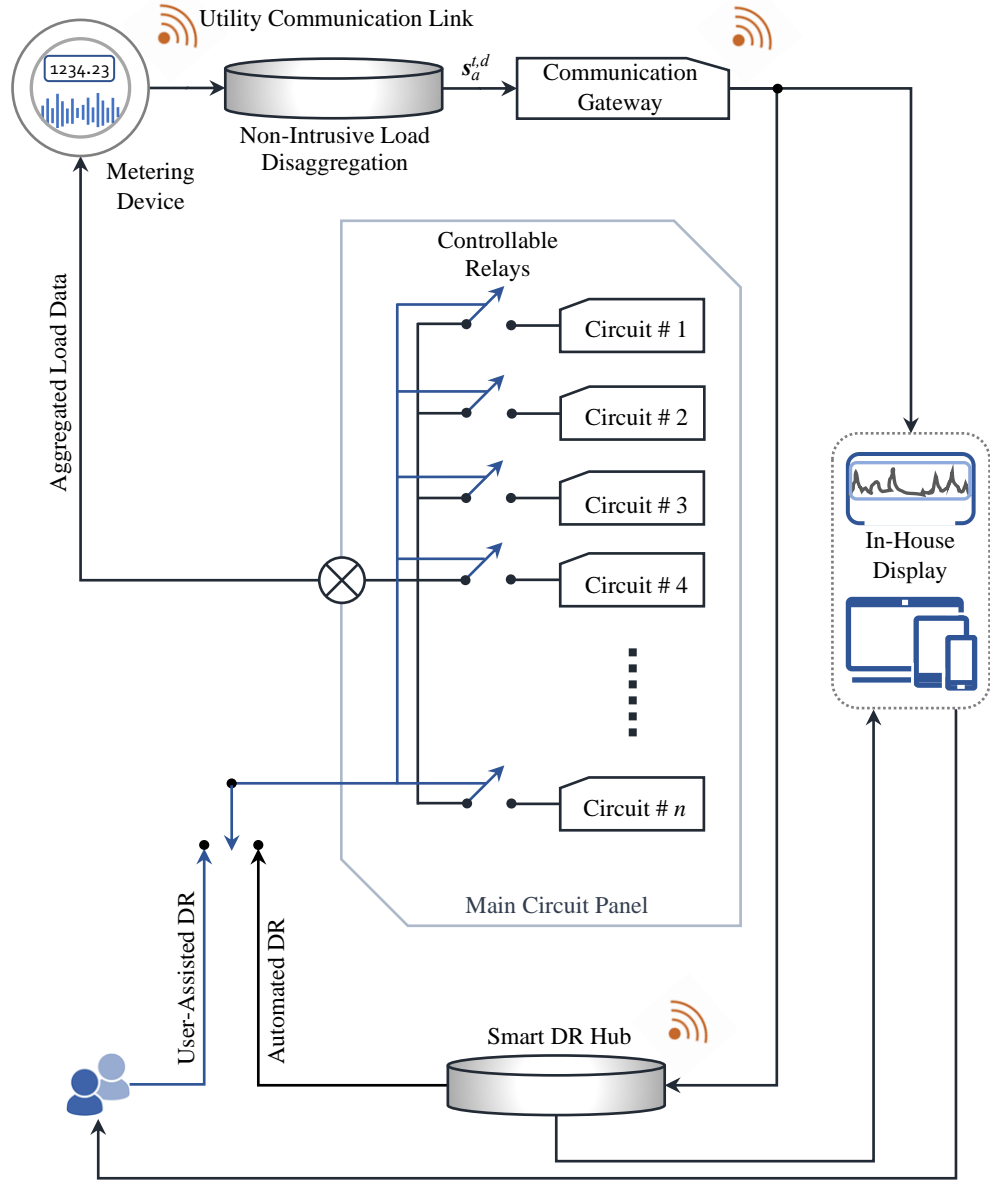


Figure 43 Proposed Application Framework  
Demand response with non-invasive load-shed authentication assisted by non-intrusive load disaggregation.

Each component has dedicated functionality, e.g., non-intrusive load disaggregation block takes aggregated load data from a single metering device and provides output in the form of individual load circuits operation status, i.e.,  $s_a^{t,d}$ . The smart demand response hub manages all the information concerning DR strategies implementation among the concerned stakeholders. Dedicated controllable relays are employed to control, i.e., connect or disconnect, different loads according to received DR instructions. Communication gateway is used to communicate the outcome of non-intrusive load disaggregation, i.e., individual load operation status,  $s_a^{t,d}$ , among different components of the presented framework as well as concerned stakeholders. The in-house display comprises a centralized display unit within the consumers' premises and other smart portable devices having the capability to enable the consumers to interact with information related to the DR. Moreover, the sequential flow of information among the concerned stakeholder, i.e. utility provider and consumer, towards establishing DR program and further execution of DR instruction and non-invasive load-shed authentication, in case of UDR, is graphically presented in Figure 44.

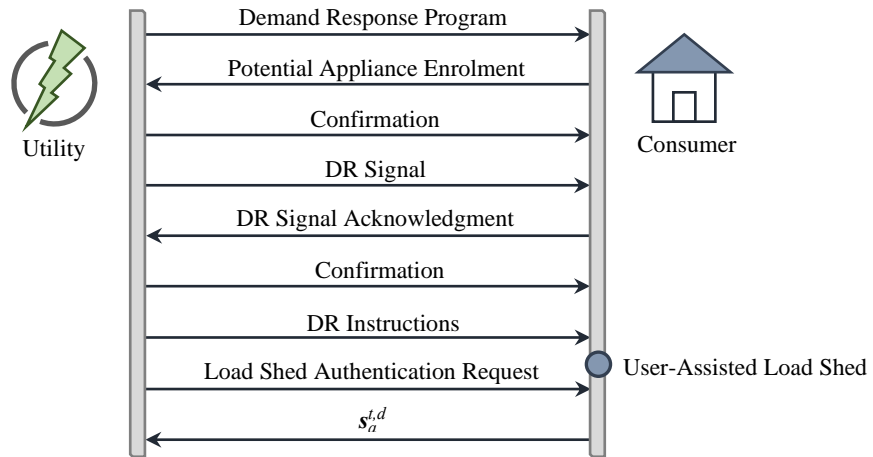


Figure 44 Information Flow Between Consumer and Utility

### 5.2.2.3 Case Study and Validation

To validate the proposed framework, a case study is carried out based on a real-world scenario. A single household of NZ GREEN Grid database, having a dedicated water heating circuit installed in its premises is considered. Assume the utility provider approaches the household to take part in the demand response program with certain incentives. The household accepts the offer and enrolls its most significant load element, i.e., water heating, in the DR program. As per the DR agreement, the utility provider agrees that the given household will control the enrolled load element manually as per received DR instructions, given that the household already acknowledged the DR signal

and the utility provider sent a confirmation for that, as shown in Figure 44. In return, the household agrees that the utility provider can access the non-invasive load-shed authentication information assisted by the non-intrusive load disaggregation system installed at the premises of the given household.

Consider, on the first day of a month the utility provider has a scheduled maintenance, and to maintain the reliability of services a DR signal is sent to the participating households. The given household (under consideration) acknowledges the DR signal and in return, the utility provider sends a confirmation. Later, the DR instruction is sent to the given household, i.e., to curtail the enrolled load (water heating) from 06:30 to 07:30 in the morning of the given day. Now according to the DR agreement, it is the responsibility of the household to curtail its load at the instructed time and day, as per DR instructions. On the other side, to validate whether the household executed the DR instruction or not, the utility provider can access the non-intrusive load disaggregation outcome of the given household for non-invasive load-shed authentication purposes.

Comprehensive simulations are carried out based on the proposed low-sampling event-based non-intrusive load disaggregation approach to retrieve the individual load inference information,  $s_a^{t,d}$ , of the given household. The said information can be further used by the utility provider to authenticate whether the DR instructions are executed or not. For the given household, Figure 45 presents the extracted results in the form of  $s_a^{t,d}$ , related to the day of scheduled maintenance and DR implementation.

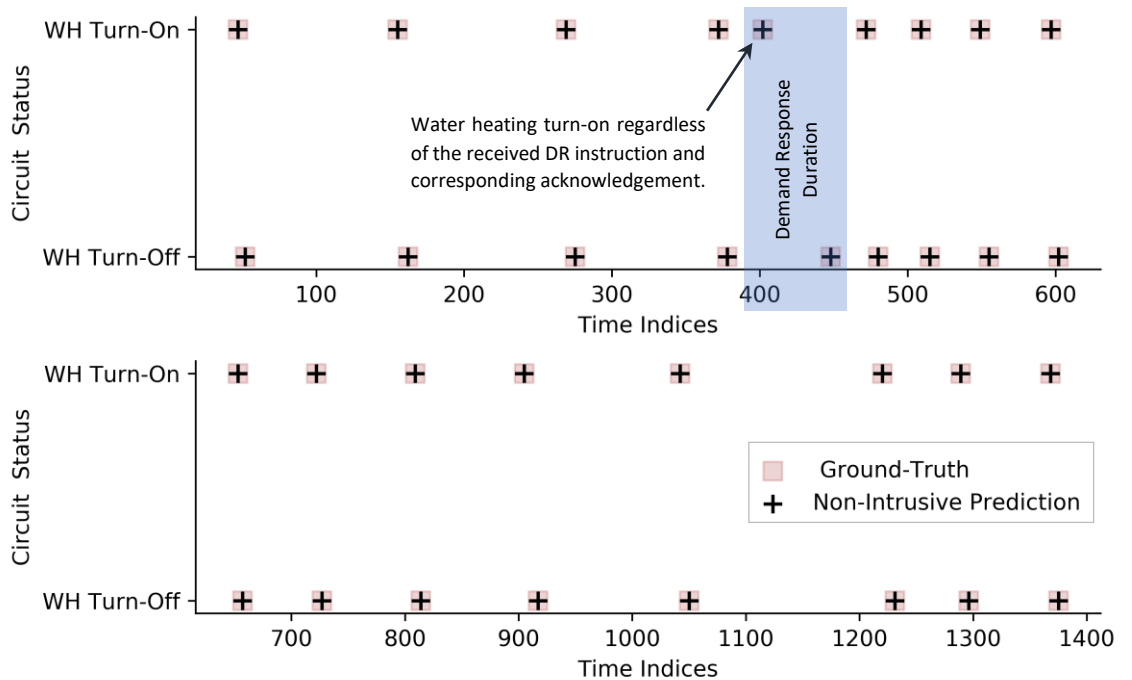


Figure 45 Water Heating Circuit Operation Authentication

Figure 45 also validates the non-invasive load inference results by benchmarking it against the actual ground-truth water heating activity. The non-intrusive water heating inference results (Figure 45) are further elaborated in Table 41, by presenting the extracted results in 24-hours time format along with the operation duration of the water heating circuit of the given household.

Table 41 Water Heating Operation Inference Results

Water Heating Turn-On		Water Heating Turn-Off		Duration (minutes)
Time Indices	24-hours Format	Time Indices	24-hours Format	
47	00:47	52	00:52	5
155	02:35	162	02:42	7
269	04:29	275	04:35	6
372	06:12	378	06:18	6
<b>402</b>	<b>06:42</b>	<b>448</b>	<b>07:28</b>	<b>46</b>
472	07:52	480	08:00	8
509	08:29	515	08:35	6
549	09:09	555	09:15	6
597	09:57	602	10:02	5
653	10:53	657	10:57	4
722	12:02	727	12:07	5
809	13:29	814	13:34	5
905	15:05	917	15:17	12
1042	17:22	1050	17:30	8
1220	20:20	1231	20:31	11
1289	21:29	1296	21:36	7
1368	22:48	1375	22:55	7

In terms of load-shed authentication, it is evident from the non-intrusive load disaggregation based inference results that irrespective of the received and acknowledged DR signal, the given household operated the water heating during the DR timeframe, as highlighted in Figure 45 (shaded region) and Table 41 (brown color). Hence, it is established that the non-intrusive load disaggregation information, i.e.,  $s_a^{t,d}$ , can be successfully employed to effectively monitor the DR compliance in a non-invasive way, subsequently enable the utility providers to effectively eliminate the freeloaders. Further, the given information i.e.,  $s_a^{t,d}$ , is not only beneficial for the utility provider in terms of load-shed authentication but also facilitates the households in terms of keeping track of their individual circuits operation status, using the in-house displays or smart portable devices as shown in Figure 43. Providing such information to the households will further facilitate effective load monitoring and management.

It is worth noting that the presented case study and the corresponding analysis of the results in this section are explicitly intended for the proof of concept and validating the proposed NILM application. Therefore, it does not imply that the given household is a freeloader.



### 5.3 Miscellaneous Applications

Besides non-invasive load-shed authentication in the context of UDR, the non-intrusive load disaggregation outcome, i.e.,  $s_a^{t,d}$ , can also facilitate the automated DR programs. As the extracted inference results in the form of consumption patterns, i.e., time of use, and operation duration (presented in Table 41) can significantly facilitate the utility companies to identify the potential DR loads and formulate more efficient and effective ADR strategies. Furthermore, the NILM assisted inference can also contribute to the load forecasting systems, consequently, facilitates the utility companies and policymakers to formulate more efficient and effective strategies/policies. For example, it is observed from the results presented in Figure 45 and Table 41 that the water heating circuit of the given household is mostly operated (not only in terms of frequency but also in terms of duration) in the morning and evening time-slots. From a larger perspective, such NILM assisted feedback regarding the consumption patterns will be highly valuable towards energy efficiency and conservation programs.

Further, in terms of identifying the DR potential load elements and corresponding patterns, the non-intrusive load disaggregation outcome, i.e.,  $s_a^{t,d}$ , can also facilitate the load shifting techniques in the context of demand side management (DSM). As load shifting of high consumption load elements, like water heating circuit, will not only significantly flatten the peak load demand but also facilitates the consumers in terms of savings [144].

### 5.4 Concluding Remarks

This chapter presented a brief overview of non-invasive load disaggregation applications along with a proof of concept in terms of real-world energy efficiency application, i.e., non-invasive load-shed authentication based demand response framework. The presented application is assisted by the proposed non-invasive load disaggregation approach, presented in this thesis. Based on the presented overview, proposed demand response framework, and corresponding analysis of the results, it is established that the non-intrusive load disaggregation has a solid potential toward real-world energy efficiency applications.

Furthermore, it is also established that irrespective of less number of load elements to be disaggregated with low data granularity, the proposed low-sampling event-based non-intrusive load disaggregation approach has a high potential for real-world deployment towards energy efficient systems.

Despite promising potential and leap forward of non-invasive load disaggregation domain, there are constraints and open challenges that limit the advancement of the aforesaid domain. Therefore, in addition to concluding remarks regarding each aspect of this thesis, the next chapter, i.e., Chapter 6, also highlights different open challenges in the NILM domain and provides an outline of the future research directions accordingly.

## **Chapter 6 Conclusion and Future Work**

This Chapter presents a brief synopsis of the research work carried out during the three years of this PhD study. This Chapter also presents comprehensive conclusive remarks regarding different aspects of the presented research work. Furthermore, in the context of non-intrusive load disaggregation, a brief discussion on the challenges limiting the development of the said field and an outline of the future research scope is also the key focus of this Chapter.

### **6.1 Synopsis and Conclusion**

This research work is primarily focussed on the implementation of a non-intrusive load disaggregation approach that can further contribute towards energy efficient systems. The proposed approach is built on the event-based working principle and all the results are validated on low sampling data granularity, subsequently making it more viable option for the existing metering infrastructure in terms of load measurements and real-world energy efficiency applications.

#### **6.1.1 Event Detection**

In the context of event-based non-intrusive load disaggregation, three new low-complexity and computationally fast event detection algorithms, namely mean sliding window (MSW), variance sliding window (VSW), and mean absolute deviation sliding window (MAD-SW), are proposed in this research work. The proposed algorithms track statistical features using the iterative process to detect the events within the aggregated load measurements. The proposed algorithms not only detect the events but also return the time occurrence of the corresponding detected events. Comprehensive simulations including sensitivity analysis in terms of different input parameters are carried out, where the extracted event detection results are validated using the available ground-truth load measurements. Based on the results and corresponding analysis, it is established that all three proposed event detection algorithms attained very promising results. It is also worth noting that the robustness of the proposed event detection algorithms is further evaluated using two diverse real-world load databases hailing from New Zealand and the United State of America.

#### **6.1.2 Feature Engineering**

This research work explored the given load data in the context of feature extraction, feature reduction, and feature selection. A total of nine distinct load features based on

statistical, geometrical, and power features are extracted and thoroughly analyzed in the context of feature reduction and feature selection. To investigate the significance of feature space reduction and feature selection, comprehensive digital simulations on real-world load data are carried out in this research work. The extracted results are further validated using classification performance. In the context of feature space reduction, it is concluded that the combinatorial process of distinct load features, i.e., intelligently grouping distinct load features, not only reduces the feature space dimensionality but also facilitates learning models in terms of classification performance. The same phenomenon is noted for feature selection with a further observation that, it is the relevance of the individual load feature, not the number of load features that facilitate the classification performance. Overall, under the given conditions, it is concluded that the feature reduction and feature selection methodologies significantly facilitate the employed classifiers in terms of classification performance. Further, it may also facilitate the classification models in terms of complexity and computational demands due to reduced feature set (dimensionality) as an input to the classifiers.

### **6.1.3 Classification**

For load classification, this research work adopted the machine learning domain, more specifically the supervised approach, where eight diverse supervised machine learning models are employed. The employed machine learning models are first evaluated using a holdout/train-test and 10-fold cross-validation approaches and later trained on a single real-world household and tested on a diverse and independent set of real-world households, including the same household as used for training purposes and completely different households. It is worth noting that all the testing data (including the same household as of training) are completely unseen in the training phase. Comprehensive simulation studies are carried out to extract circuit-level, household-level, and classifier-level performance of the employed learning models. A comparative analysis of the employed learning models is also part of this research work. For the given conditions, it is concluded that the multi-layer perceptron, support vector machine, logistic regression, and Gaussian process classifiers outperform other employed models. On the other side, the decision trees classifier lags in performance compared to the other employed classifiers, under the given conditions. It is worth noting that the comparative evaluation of different machine learning models presented in this research work does not mean to undermine the one or another learning model but to investigate the optimal learning model under the given conditions. As in the machine learning domain, ‘one size fits all’ is not

the case, hence the quest is to investigate an optimal classifier for a certain application under a given condition, rather than exploring the supremacy of one learning model over the others.

To further address the shortcoming of the combinatorial learning model techniques in the non-intrusive load disaggregation domain, this research work also explored the ensemble learning approach, in addition to employing standalone learning models, for load classification. It is observed that the given ensemble model built on three diverse learning models, i.e., multi-layer perceptron, decision trees, and Naïve Bayes, balanced the performance of all its members and manages some reasonable classification performance improvement compared to most of its respective individual members. But it is also worth noting that the performance improvement enabled by the presented ensemble model came at a price of higher model complexity and computational demands, which leads to a trade-off among overall performance and model complexity and computational demands. Hence it is fair to conclude that the selection of ensemble learning over the standalone learning model exclusively depends on the sensitivity of the targeted application or the choice of the end-user; to prefer performance over complexity or vice versa.

#### **6.1.4 Applications**

Despite leap forward in the field of non-intrusive load disaggregation, the research in terms of its actionable feedback, i.e., applications, is lagging, consequently, its real-world deployment is limited. To address this shortcoming, a real-world application framework based on non-intrusive load disaggregation is presented as a proof of concept. The presented framework targets the demand response program incorporated with non-invasive load-shed authentication that is solely assisted by the proposed research work, i.e., low-sampling event-based non-intrusive load disaggregation approach. In this context, based on the presented case study for a real-world household, it is concluded that the research work presented in this thesis has a high potential for real-world deployment toward energy efficient systems.

Finally, the proposed low-sampling event-based non-intrusive load disaggregation approach and corresponding application framework presented in this research work established that the non-intrusive load disaggregation performed at low sampling data granularity is not only viable but can also be effectively incorporated in real-world applications. Consequently, it is concluded that the non-intrusive load disaggregation approach has a solid potential towards energy efficient systems and further research and development in the said domain will unlock the actual potential of the smart grid concept.

## 6.2 Potential Challenges and Future Research Scope

Despite the recent development in the non-intrusive load disaggregation domain, still, it faces many challenges that limit its real potential. In this context, some of the key factors are generalized load features, availability of standardized load disaggregation databases, and performance metrics for evaluation.

The existing load space is too diverse and complex, making it hard to define a universal load feature space of interest. In the existing literature, there are load features which can be employed with confidence towards load inference, however, formulating a generalized load feature space that could be employed for load inference across the board is still an open research problem. As the electrical characteristics of the existing loads are quite diverse and complex, even in case of similar appliances the characteristics may vary due to their diverse manufacturing technologies. This diversity and variation in load space make generalized load features formulation harder, consequently, making accurate load classification more difficult. In the given context, the load feature analysis presented in this research can be extended to further formulate load features that are applicable to effectively model more load elements, subsequently enabling further robust and precise non-intrusive load disaggregation.

Another factor that limits the advancement of the non-intrusive load disaggregation is the open availability of adequate load disaggregation databases. In terms of measurement attributes, the available load disaggregation databases are quite diverse and in some cases inadequate, consequently limiting the development of non-intrusive load disaggregation algorithms [145]. Hence, building a load disaggregation database with further detailed and refined measurement information for a wide range of load elements would be valuable for future algorithms development. Further, in the given context, future research needs to be more focus on low data granularity, as load databases with low data granularity measurements will not only provide ease in terms of computational needs: acquisition, storage, and processing, but also enable the research community to realize the real-world potential of NILM applications by developing and evaluating low sampling NILM algorithms.

Further in the context of low data granularity, load classification techniques need to be further researched and to this end, the recent development of computational capabilities and advanced learning technologies can play a significant role. In the given context, this research work provides comprehensive evaluation of diverse learning models not only in a standalone/independent configuration but also in terms of ensemble learning. However,

this research can be further extended and more advanced learning models in both configurations: standalone and ensemble, can be explored toward accurate non-invasive load inference.

Moreover, as discussed in Chapter 5 of this thesis, non-intrusive load disaggregation has solid potential towards efficient and sustainable energy systems. Hence, to further unlock the real-world potential of the non-intrusive load disaggregation, future research must be focussed on the real-world actionable feedback: applications, of non-intrusive load disaggregation towards, rather than exclusively focusing on existing algorithms' performance enhancements and inference of more individual appliances/circuits.

## References

- [1] R. Hassan and G. Radman, "Survey on smart grid," in *IEEE SoutheastCon 2010 (SoutheastCon), Proceedings of the*, 2010, pp. 210-213: IEEE.
- [2] V. Amenta and G. M. Tina, "Load Demand Disaggregation based on Simple Load Signature and User's Feedback," (in English), *Sustainability in Energy and Buildings: Proceedings of the 7th International Conference Seb-15*, vol. 83, pp. 380-388, 2015.
- [3] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey," *Sensors (Basel)*, vol. 12, no. 12, pp. 16838-66, Dec 6 2012.
- [4] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, no. Supplement C, pp. 213-234, 2013/01/01/ 2013.
- [5] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473-484, 2014.
- [6] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation for low-sampling rate data," in *2nd International Non-intrusive Appliance Load Monitoring Workshop, Austin, TX*, 2014.
- [7] Y. F. Wong, Y. A. Şekercioğlu, T. Drummond, and V. S. Wong, "Recent approaches to non-intrusive load monitoring techniques in residential settings," in *Computational Intelligence Applications In Smart Grid (CIASG), 2013 IEEE Symposium on*, 2013, pp. 73-79: IEEE.
- [8] A. U. Rehman *et al.*, "Applications of Non-Intrusive Load Monitoring Towards Smart and Sustainable Power Grids: A System Perspective," in *2019 29th Australasian Universities Power Engineering Conference (AUPEC)*, 2019, pp. 1-6.
- [9] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," (in English), *Ieee Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1445-1454, Jul 2008.



- [10] D. Egarter and W. Elmenreich, "Load disaggregation with metaheuristic optimization," *Energieinformatik, Karlsruhe, Germany*, 2015.
- [11] Y. H. Lin and M. S. Tsai, "An Advanced Home Energy Management System Facilitated by Nonintrusive Load Monitoring With Automated Multiobjective Power Scheduling," (in English), *Ieee Transactions on Smart Grid*, vol. 6, no. 4, pp. 1839-1851, Jul 2015.
- [12] H. Altrabalsi, V. Stankovic, J. Liao, and L. Stankovic, "Low-complexity energy disaggregation using appliance load modelling," (in English), *Aims Energy*, vol. 4, no. 1, pp. 1-21, 2016.
- [13] N. F. Esa, M. P. Abdullah, and M. Y. Hassan, "A review disaggregation method in Non-intrusive Appliance Load Monitoring," (in English), *Renewable & Sustainable Energy Reviews*, vol. 66, pp. 163-173, Dec 2016.
- [14] G. W. Hart, *Nonintrusive Appliance Load Data Acquisition Method: Progress Report*. MIT Energy Laboratory, 1984.
- [15] G. W. Hart, "Nonintrusive Appliance Load Monitoring," (in English), *Proceedings of the Ieee*, vol. 80, no. 12, pp. 1870-1891, Dec 1992.
- [16] E. J. Aladesanmi and K. A. Folly, "Overview of non-intrusive load monitoring and identification techniques," (in English), *Ifac Papersonline*, vol. 48, no. 30, pp. 415-420, 2015.
- [17] K. Basu, V. Debusschere, S. Bacha, A. Hably, D. Van Delft, and G. J. Dirven, "A generic data driven approach for low sampling load disaggregation," (in English), *Sustainable Energy Grids & Networks*, vol. 9, pp. 118-127, Mar 2017.
- [18] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA, 2011, vol. 25, pp. 59-62.
- [19] S. Makonin, B. Ellert, I. V. Bajic, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Sci Data*, vol. 3, p. 160037, Jun 7 2016.
- [20] "Pecan Street Inc. Dataport 2020" [Online]. Available: <https://www.pecanstreet.org/dataport/>
- [21] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in

*Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, 2014, pp. 80-89: ACM.

- [22] B. Anderson *et al.* New Zealand GREEN Grid Household Electricity Demand Study 2014-2018 [Online]. Available: <http://reshare.ukdataservice.ac.uk/853334/>
- [23] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci Data*, vol. 2, p. 150007, 2015.
- [24] A. S. Uttama Nambi, A. Reyes Lua, and V. R. Prasad, "Loced: Location-aware energy disaggregation framework," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 45-54.
- [25] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, "GREEND: an energy consumption dataset of households in Italy and Austria," in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, 2014, pp. 511-516: IEEE.
- [26] J. Gao, S. Giri, E. C. Kara, and M. Bergés, "PLAID: a public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract," in *proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, 2014, pp. 198-199: ACM.
- [27] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study," *Sci Data*, vol. 4, p. 160122, Jan 5 2017.
- [28] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Berges, "BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research," in *Proceedings of the 2nd KDD workshop on data mining applications in sustainability (SustKDD)*, 2012, pp. 1-5.
- [29] M. Ribeiro, L. Pereira, F. Quintal, and N. Nunes, "SustDataED: A public dataset for electric energy disaggregation research," in *ICT for Sustainability 2016*, 2016: Atlantis Press.
- [30] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, "It's Different: Insights into home energy consumption in India," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, 2013, pp. 1-8.

- [31] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, "A comparison of non-intrusive load monitoring methods for commercial and residential buildings," *arXiv preprint arXiv:1408.6595*, 2014.
- [32] Smart-Grid Smart-City Customer Trial Data [Online]. Available: <https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details>
- [33] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart\*: An open data set and tools for enabling research in sustainable homes," *SustKDD*, August, vol. 111, no. 112, p. 108, 2012.
- [34] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem," *arXiv preprint arXiv:1703.00785*, 2017.
- [35] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, "Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview," in *Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on*, 2015, pp. 1175-1180: IEEE.
- [36] J. Y. Yu, Y. C. Gao, Y. X. Wu, D. Jiao, C. Su, and X. Wu, "Non-Intrusive Load Disaggregation by Linear Classifier Group Considering Multi-Feature Integration," (in English), *Applied Sciences-Basel*, vol. 9, no. 17, p. 3558, Sep 1 2019.
- [37] S. M. Tabatabaei, S. Dick, and W. S. Xu, "Toward Non-Intrusive Load Monitoring via Multi-Label Classification," (in English), *Ieee Transactions on Smart Grid*, vol. 8, no. 1, pp. 26-40, Jan 2017.
- [38] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—Part I: Basic concept, structure, and methodology," *IEEE transactions on power Delivery*, vol. 25, no. 2, pp. 551-560, 2010.
- [39] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of VI trajectory based load signatures for non-intrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 870-878, 2014.
- [40] A. L. Wang, B. X. Chen, C. G. Wang, and D. Hua, "Non-intrusive load monitoring algorithm based on features of V–I trajectory," *Electric Power Systems Research*, vol. 157, pp. 134-144, 2018.

- [41] B. Wild, K. S. Barsim, and B. Yang, "A new unsupervised event detector for non-intrusive load monitoring," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 73-77: IEEE.
- [42] M. Azaza and F. Wallin, "Evaluation of classification methodologies and Features selection from smart meter data," (in English), *Proceedings of the 9th International Conference on Applied Energy*, vol. 142, pp. 2250-2256, 2017/12/01/ 2017.
- [43] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in NILM," (in English), *Energy and Buildings*, vol. 151, pp. 98-106, Sep 15 2017.
- [44] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE transactions on Consumer Electronics*, vol. 57, no. 1, 2011.
- [45] C. Klemenjak and P. Goldsborough, "Non-Intrusive Load Monitoring: A Review and Outlook," *arXiv preprint arXiv:1610.01191*, 2016.
- [46] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-60, Jul 17 2015.
- [47] M. Perez-Ortiz, S. Jimenez-Fernandez, P. A. Gutierrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A Review of Classification Problems and Algorithms in Renewable Energy Applications," (in English), *Energies*, vol. 9, no. 8, p. 607, Aug 2016.
- [48] A. C. Lorena *et al.*, "Comparing machine learning classifiers in potential distribution modelling," (in English), *Expert Systems with Applications*, vol. 38, no. 5, pp. 5268-5275, May 2011.
- [49] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 2607-2612.
- [50] S. Su *et al.*, "Non-intrusive load monitoring of air conditioning using low-resolution smart meter data," in *Power System Technology (POWERCON), 2016 IEEE International Conference on*, 2016, pp. 1-5: IEEE.

- [51] X. Wu, Y. C. Gao, and D. Jiao, "Multi-Label Classification Based on Random Forest Algorithm for Non-Intrusive Load Monitoring System," (in English), *Processes*, vol. 7, no. 6, p. 337, Jun 2019.
- [52] K. H. He, D. Jakovetic, B. C. Zhao, V. Stankovic, L. Stankovic, and S. Cheng, "A Generic Optimisation-Based Approach for Improving Non-Intrusive Load Monitoring," (in English), *Ieee Transactions on Smart Grid*, vol. 10, no. 6, pp. 6472-6480, Nov 2019.
- [53] F. M. Wittmann, J. C. Lopez, and M. J. Rider, "Nonintrusive Load Monitoring Algorithm Using Mixed-Integer Linear Programming," (in English), *Ieee Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 180-187, May 2018.
- [54] D. Egarter and W. Elmenreich, "Load disaggregation with metaheuristic optimization," in *Energieinformatik*, 2015, pp. 1-12.
- [55] R. Machlev, J. Belikov, Y. Beck, and Y. Levron, "MO-NILM: A multi-objective evolutionary algorithm for NILM classification," (in English), *Energy and Buildings*, vol. 199, pp. 134-144, Sep 15 2019.
- [56] W. Kong, Z. Y. Dong, B. Wang, J. Zhao, and J. Huang, "A practical solution for non-intrusive type II load monitoring based on deep learning and post-processing," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 148-160, 2019.
- [57] J. Cho, Z. Hu, and M. Sartipi, "Non-Intrusive A/C Load Disaggregation Using Deep Learning," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, 2018, pp. 1-5: IEEE.
- [58] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Comparative Evaluation of Machine Learning Algorithms and Feature Space Reduction for Noninvasive Energy Disaggregation," *Journal of Modern Power Systems and Clean Energy*, 2020 (Submitted; under revision).
- [59] S. S. Hosseini, K. Agbossou, S. Kelouwani, and A. Cardenas, "Non-intrusive load monitoring through home energy management systems: A comprehensive review," (in English), *Renewable & Sustainable Energy Reviews*, vol. 79, pp. 1266-1274, Nov 2017.
- [60] A. Hernandez, A. Ruano, J. Urena, M. G. Ruano, and J. J. Garcia, "Applications of NILM Techniques to Energy Management and Assisted Living," (in English), *Ifac Papersonline*, vol. 52, no. 11, pp. 164-171, 2019.

- [61] A. Ruano, A. Hernandez, J. Urena, M. Ruano, and J. Garcia, "NILM Techniques for Intelligent Home Energy Management and Ambient Assisted Living: A Review," (in English), *Energies*, vol. 12, no. 11, p. 2203, Jun 1 2019.
- [62] M. Zhuang, M. Shahidehpour, and Z. Li, "An Overview of Non-Intrusive Load Monitoring: Approaches, Business Applications, and Challenges," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 4291-4299: IEEE.
- [63] "Low Cost NIALMS Technology," Electric Power Research Institute, Technical Report TR-198918-V1, Sept. 1997, vol. 1.
- [64] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," (in English), *Energy and Buildings*, vol. 30, no. 3, pp. 245-259, Aug 1999.
- [65] M. L. Marceau and R. Zmeureanu, "Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings," (in English), *Energy Conversion and Management*, vol. 41, no. 13, pp. 1389-1403, Sep 2000.
- [66] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. Moura, "Event detection for non intrusive load monitoring," in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*, 2012, pp. 3312-3317: IEEE.
- [67] M. N. Meziane, P. Ravier, G. Lamarque, J.-C. Le Bunetel, and Y. Raingeaud, "High accuracy event detection for Non-Intrusive Load Monitoring," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, 2017, pp. 2452-2456: IEEE.
- [68] K. Basu, A. Hably, V. Debusschere, S. Bacha, G. J. Driven, and A. Ovalle, "A comparative study of low sampling non intrusive load dis-aggregation," in *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, 2016, pp. 5137-5142: IEEE.
- [69] A. A. Girmay and C. Camarda, "Simple event detection and disaggregation approach for residential energy estimation," in *Workshop on Non-Intrusive Load Monitoring (NILM), 2016 Proceedings of the 3rd International*, 2016.

- [70] L. Kong, D. Yang, and Y. Luo, "Non-Intrusive Load Monitoring and Identification Based on Maximum Likelihood Method," in *Energy Internet (ICEI), IEEE International Conference on*, 2017, pp. 268-272: IEEE.
- [71] D. Egarter and W. Elmenreich, "Autonomous load disaggregation approach based on active power measurements," in *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, 2015, pp. 293-298: IEEE.
- [72] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 4340-4343: IEEE.
- [73] H. A. Azzini, R. Torquato, and L. C. da Silva, "Event detection methods for nonintrusive load monitoring," in *PES General Meeting/ Conference & Exposition, 2014 IEEE*, 2014, pp. 1-5: IEEE.
- [74] C. Laughman *et al.*, "Power signature analysis," *IEEE power and energy magazine*, vol. 99, no. 2, pp. 56-63, 2003.
- [75] L. De Baets, J. Ruysinck, D. Deschrijver, and T. Dhaene, "Event detection in NILM using cepstrum smoothing," in *3rd International Workshop on Non-Intrusive Load Monitoring*, 2016, pp. 1-4.
- [76] T. Zia, D. Bruckner, and A. Zaidi, "A hidden markov model based procedure for identifying household electric loads," in *IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society*, 2011, pp. 3218-3223: IEEE.
- [77] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial Intelligence and Statistics*, 2012, pp. 1472-1482.
- [78] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, 2011, pp. 747-758: SIAM.
- [79] H. Altrabalsi, J. Liao, L. Stankovic, and V. Stankovic, "A low-complexity energy disaggregation method: Performance and robustness," in *Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on*, 2014, pp. 1-8: IEEE.

- [80] J. Kim, T. T. Le, and H. Kim, "Nonintrusive Load Monitoring Based on Advanced Deep Learning and Novel Signature," *Comput Intell Neurosci*, vol. 2017, p. 4216281, 2017.
- [81] K. Basu, V. Debusschere, S. Bacha, U. Maulik, and S. Bondyopadhyay, "Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach," (in English), *Ieee Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 262-270, Feb 2015.
- [82] S. Gupta, M. S. Reynolds, and S. N. Patel, "ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 139-148: ACM.
- [83] K. Basu, V. Debusschere, A. Douzal-Chouakria, and S. Bacha, "Time series distance-based methods for non-intrusive load monitoring in residential buildings," (in English), *Energy and Buildings*, vol. 96, pp. 109-117, Jun 1 2015.
- [84] O. Parson *et al.*, "Dataport and NILMTK: A building data set designed for non-intrusive load monitoring," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 210-214: IEEE.
- [85] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," (in English), *Energy and Buildings*, vol. 116, pp. 96-103, Mar 15 2016.
- [86] A. Godil, R. Bostelman, W. Shackleford, T. Hong, and M. Shneier, "Performance metrics for evaluating object and human detection and tracking systems," *No. NIST Interagency/Internal Report (NISTIR)-7972*, 2014.
- [87] Y. Sakiyama *et al.*, "Predicting human liver microsomal stability with machine learning techniques," *J Mol Graph Model*, vol. 26, no. 6, pp. 907-15, Feb 2008.
- [88] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159-74, Mar 1977.
- [89] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Event Detection Algorithm for Non- Intrusive Load Monitoring Systems," in *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 746-751.
- [90] A. U. Rehman, T. T. Lie, B. Valles, and S. R. Tito, "Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low



- Complexity Statistical Features," (in English), *Ieee Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 751-759, Mar 2020.
- [91] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Intrusive Load Monitoring of Residential Water-Heating Circuit Using Ensemble Machine Learning Techniques," *Inventions*, vol. 5, no. 4, p. 57, 2020.
  - [92] M. Liu, J. Yong, X. Wang, and J. Lu, "A new event detection technique for residential load monitoring," in *Harmonics and Quality of Power (ICHQP), 2018 18th International Conference on*, 2018, pp. 1-6: IEEE.
  - [93] M. Vasirani and S. Ossowski, "A collaborative model for participatory load management in the smart grid," in *Workshop on AI Problems and Approaches for Intelligent Environments*, 2012, p. 21.
  - [94] G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation of domestic smart meter readings using dynamic time warping," in *Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on*, 2014, pp. 36-39: IEEE.
  - [95] R. E. Bellman, *Adaptive control processes: a guided tour*. Princeton university press, 2015.
  - [96] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*, 2014, pp. 372-378: IEEE.
  - [97] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," (in English), *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16-28, Jan 2014.
  - [98] A. Janecek, W. Gansterer, M. Demel, and G. Ecker, "On the relationship between feature selection and classification accuracy," in *New challenges for feature selection in data mining and knowledge discovery*, 2008, pp. 90-105.
  - [99] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data classification: Algorithms and applications*, p. 37, 2014.
  - [100] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," (in English), *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, pp. 3-24, 2007.
  - [101] X. D. Wu *et al.*, "Top 10 algorithms in data mining," (in English), *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, Jan 2008.

- [102] M. Figueiredo, A. de Almeida, and B. Ribeiro, "Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems," (in English), *Neurocomputing*, vol. 96, pp. 66-73, Nov 1 2012.
- [103] Y. Bazi and F. Melgani, "Gaussian Process Approach to Remote Sensing Image Classification," (in English), *Ieee Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 186-197, Jan 2010.
- [104] X. Sun, X. Wang, Y. Liu, and J. Wu, "Non-intrusive sensing based multi-model collaborative load identification in cyber-physical energy systems," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2014, pp. 1-6.
- [105] B. Wang, F. Wan, P. U. Mak, P. I. Mak, and M. I. Vai, "EEG signals classification for brain computer interfaces based on Gaussian process classifier," in *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*, 2009, pp. 1-5.
- [106] H. C. Kim, D. Kim, Z. Ghahramani, and S. Y. Bang, "Appearance-based gender classification with Gaussian processes," (in English), *Pattern Recognition Letters*, vol. 27, no. 6, pp. 618-626, Apr 15 2006.
- [107] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," (in English), *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825-2830, Oct 2011.
- [108] I. D. Longstaff and J. F. Cross, "A pattern recognition approach to understanding the multi-layer perception," *Pattern Recognition Letters*, vol. 5, no. 5, pp. 315-319, 1987.
- [109] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 1310-1315: IEEE.
- [110] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," (in English), *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159-190, Nov 2006.
- [111] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *J Biomed Inform*, vol. 35, no. 5-6, pp. 352-9, Oct-Dec 2002.

- [112] W. Yan, "Application of Random Forest to Aircraft Engine Fault Diagnosis," in *The Proceedings of the Multiconference on "Computational Engineering in Systems Applications"*, 2006, vol. 1, pp. 468-475.
- [113] P. Morales-Alvarez, A. Pérez-Suay, R. Molina, and G. Camps-Valls, "Remote sensing image classification with large-scale Gaussian processes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1103-1114, 2017.
- [114] T. N. A. Nguyen, A. Bouzerdoun, and S. L. Phung, "A Scalable Hierarchical Gaussian Process Classifier," (in English), *Ieee Transactions on Signal Processing*, vol. 67, no. 11, pp. 3042-3057, Jun 1 2019.
- [115] A. Subasi and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," *Comput Methods Programs Biomed*, vol. 78, no. 2, pp. 87-99, May 2005.
- [116] C. Lazar *et al.*, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 9, no. 4, pp. 1106-19, Jul-Aug 2012.
- [117] "Electricity in New Zealand," Electricity Authority, New Zealand 2018, Available: <https://www.ea.govt.nz/about-us/media-and-publications/electricity-new-zealand/>.
- [118] Y. Yang, M. Zengqiang, X. Zheng, and D. Chang, "Accommodation of curtailed wind power by electric water heaters based on a new hybrid prediction approach," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 3, pp. 525-537, 2019.
- [119] M. Wu, Y. Q. Bao, J. L. Zhang, and T. Z. Ji, "Multi-objective optimization for electric water heater using mixed integer linear programming," (in English), *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 1256-1266, Sep 2019.
- [120] Z. M. Haider, K. K. Mehmood, M. K. Rafique, S. U. Khan, S. J. Lee, and C. H. Kim, "Water-filling algorithm based approach for management of responsive residential loads," (in English), *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 1, pp. 118-131, Jan 2018.

- [121] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 742-750, 2013.
- [122] T. Clarke, T. Slay, C. Eustis, and R. B. Bass, "Aggregation of Residential Water Heaters for Peak Shifting and Frequency Response Services," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 22-30, 2019.
- [123] B. Anderson *et al.*, "NZ GREEN Grid Household Electricity Demand Study: 1 minute electricity power (version 1.0)," Centre for Sustainability, University of Otago, Dunedin2018, Available: <http://www.otago.ac.nz/centre-sustainability/>.
- [124] F. Leon, S.-A. Floria, and C. Bădică, "Evaluating the effect of voting methods on ensemble-based classification," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2017, pp. 1-6: IEEE.
- [125] R. Polikar, "Ensemble learning," in *Ensemble machine learning*: Springer, 2012, pp. 1-34.
- [126] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Non-Invasive Load-Shed Authentication Model for Demand Response Applications Assisted by Event-Based Non-Intrusive Load Monitoring," *Energy and AI*, p. 100055, 2021/02/11/ 2021.
- [127] N. Batra, A. Singh, and K. Whitehouse, "If you measure it, can you improve it? exploring the value of energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 191-200: ACM.
- [128] J. C. Nation *et al.*, "Nonintrusive monitoring for shipboard fault detection," in *2017 IEEE Sensors Applications Symposium (SAS)*, 2017, pp. 1-5: IEEE.
- [129] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring," (in English), *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844-+, Oct 2010.
- [130] T. Liu, Y. Liu, Y. Che, S. Chen, Z. Xu, and Y. Duan, "SHE: smart home energy management system for appliance identification and personalized scheduling," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 247-250: ACM.

- [131] J. S. Donnal, J. Paris, and S. B. Leeb, "Energy Applications for an Energy Box," (in English), *Ieee Internet of Things Journal*, vol. 3, no. 5, pp. 787-795, Oct 2016.
- [132] A. Adabi, P. Manovi, and P. Mantey, "Cost-effective instrumentation via NILM to support a residential energy management system," in *Consumer Electronics (ICCE), 2016 IEEE International Conference on*, 2016, pp. 107-110: IEEE.
- [133] F. J. Luo, G. Ranzi, W. C. Kong, Z. Y. Dong, S. Wang, and J. H. Zhao, "Non-intrusive energy saving appliance recommender system for smart grid residential users," (in English), *Iet Generation Transmission & Distribution*, vol. 11, no. 7, pp. 1786-1793, May 11 2017.
- [134] D. He, W. Lin, N. Liu, R. G. Harley, and T. G. Habetler, "Incorporating non-intrusive load monitoring into building level demand response," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1870-1877, 2013.
- [135] N. Batra, R. Baijal, A. Singh, and K. Whitehouse, "How good is good enough? re-evaluating the bar for energy disaggregation," *arXiv preprint arXiv:1510.08713*, 2015.
- [136] J. Alcalá, O. Parson, and A. Rogers, "Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 225-234: ACM.
- [137] J. M. Alcala, J. Urena, A. Hernandez, and D. Gualda, "Sustainable Homecare Monitoring System by Sensing Electricity Data," (in English), *Ieee Sensors Journal*, vol. 17, no. 23, pp. 7741-7749, Dec 1 2017.
- [138] M. A. Piette, G. Ghatikar, S. Kiliccote, D. Watson, E. Koch, and D. Hennage, "Design and Operation of an Open, Interoperable Automated Demand Response Infrastructure for Commercial Buildings," (in English), *Journal of Computing and Information Science in Engineering*, vol. 9, no. 2, p. 021004, Jun 2009.
- [139] J. Han and M. A. Piette, "Solutions for summer electric power shortages: Demand response and its applications in air conditioning and refrigerating systems," *Refrigeration, Air Conditioning, and Electric Power Machinery*, vol. 29, no. 1, pp. 1-4, 2008.


- [140] T. Samad, E. Koch, and P. Stluka, "Automated Demand Response for Smart Buildings and Microgrids: The State of the Practice and Research Challenges," (in English), *Proceedings of the Ieee*, vol. 104, no. 4, pp. 726-744, Apr 2016.
- [141] S. Heinen and P. Richards, "Towards customer-centric energy utilities-A granular data-driven bottom-up approach to understanding energy customer trends," *The Electricity Journal*, vol. 33, no. 9, p. 106836, 2020.
- [142] Transpower. (2019). *Demand Response*. Available: <https://www.transpower.co.nz/keeping-you-connected/demand-response>
- [143] D. C. Bergman, D. Jin, J. P. Juen, N. Tanaka, C. A. Gunter, and A. Wright, "Nonintrusive load-shed verification," *IEEE Pervasive Computing*, no. 1, pp. 49-57, 2010.
- [144] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand Side Management in Smart Grid Using Heuristic Optimization," (in English), *Ieee Transactions on Smart Grid*, vol. 3, no. 3, pp. 1244-1252, Sep 2012.
- [145] C. Shin, S. Rho, H. Lee, and W. Rhee, "Data Requirements for Applying Machine Learning to Energy Disaggregation," (in English), *Energies*, vol. 12, no. 9, p. 1696, May 1 2019.

# Appendix

## A.1 Confirmation of Candidature

**GRADUATE RESEARCH SCHOOL**

Private Bag 92006 Auckland 1142  
+64-9-921-9907  
grs@aut.ac.nz



Ref: 16935632  
21 November 2018

Attique Ur Rehman  
2A/16 Burton Street  
Grafton  
Auckland  
NEW ZEALAND

Dear Attique Ur,

Re: Confirmation of Candidature

Congratulations!

I am pleased to inform you your PGRS was approved by the Faculty of Design & Creative Technologies and was noted at the University Postgraduate Board at their meeting held on 20 Nov 2018. You have now completed all conditions placed on your provisional admission to your programme, and the Board will now confirm your candidature in the Doctor of Philosophy.

The completion of this milestone marks a significant point in the career of every doctoral student, and represents the successful passage from provisional to confirmed candidature. It demonstrates your maturity as a doctoral researcher capable of contributing an original contribution to your field of enquiry. It also demonstrates that your project is of a suitable scope and standard for your degree, and that you have the capacity, resources, and potential to complete your research at the required level.

Data Collection

If your research does not require ethics approval, you may now begin data collection. If your research does require ethics approval, you may begin data collection once you have ethics approval.

Expected Submission Date

Currently, we are expecting that you will submit for examination on 20 Jul 2021. We are currently updating our records and if you feel this is incorrect, please contact your faculty contact person listed below.

Business Cards

As a recognition of this milestone, the University would like to provide you with your own AUT business cards for you to use when attending conferences and networking with other researchers. We have attached the 'AUT Business Card Order Form' for you to complete and return by email only (no hard copies required) to Scott Pilkington (pgrresearch@aut.ac.nz).

The University will cover this initial printing expense, however, reprints will be at the candidate's expense. Please contact the Graduate Research School when a reprint is required.

[www.aut.ac.nz/study-at-aut/postgraduate-study/graduate-research-school](http://www.aut.ac.nz/study-at-aut/postgraduate-study/graduate-research-school)

**Faculty Contacts**

Your primary supervisor is Tek Tjing Lie  
Your secondary supervisor is Brice Valles  
The Associate Dean (Postgraduate) is Rosser Johnson, ext 7818  
Your faculty doctoral contact person is Angela Anderson, ext. 6761, doffice@aut.ac.nz

**Graduate Research School Contact**


Your enrolment contact at the Graduate Research School is Jessica Yamamoto, ext. 8220, [jessica.yamamoto@aut.ac.nz](mailto:jessica.yamamoto@aut.ac.nz)

**Congratulations Again**

On behalf of all staff involved in the programme we would like to acknowledge the challenge of undertaking research at this level as well as the commitment and application which are required to pass this significant milestone in your research career.

If you have any questions, please feel free to contact me.

Yours sincerely



Martin Wilson  
Manager, Graduate Research School  
[martin.wilson@aut.ac.nz](mailto:martin.wilson@aut.ac.nz)  
+64-9-921-9999 ext 8812

cc: Tek Tjing Lie, Brice Valles, Angela Anderson DI, Jessica Yamamoto

## A.2 Learning Models Parameters

For simulation purposes, the details of the employed learning models' parameters are presented in Table 42 [58]. Further details of the presented parameters can be found in Scikit-Learn [107].

Table 42 Employed Learning Models' Parameters

Learning Models	Parameter Details
Support Vector Machine	C=1.0, kernel='rbf',
Multi-Layer Perceptron	activation='relu', hidden_layer_sizes=(100,), solver='sgd'
Decision Tree	min_samples_leaf=1, min_samples_split=2, splitter='best', criterion='gini'
Random Forest	criterion='gini', min_samples_leaf=1, min_samples_split=2, n_estimators=10
Naïve Bayes	var_smoothing=1e-09
Gaussian Process	max_iter_predict=100, multi_class='one_vs_rest'
Logistic Regression	C=1.0, max_iter=100, penalty='l2'
k-Nearest Neighbors	p=2, n_neighbors=5, weights='uniform'
Voting Ensemble	voting='hard'

### A.3 Sensitivity Analysis for Dataport

Based on simulation parameters presented in Table 21, comprehensive sensitivity analysis is carried out for Dataport's load data in terms of delay tolerance factor. The corresponding results along with the details of the employed household are presented in Table 43.

Table 43 MSW Sensitivity Analysis in terms of  $\Delta\tau$  for Dataport

<b>Dataport, Pecan Street</b>					
Household ID 26					
Data Acquisition Timeframe	June 18 - July 02, 2014				
Number of Days	15				
Number of Samples	21600				
Total Detected Events	323				
<b>Delay Tolerance</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
True Positive	286	313	314	315	315
False Positive	37	10	9	8	8
False Negative	47	17	16	14	14
Recall (%)	85.88	94.84	95.15	95.74	95.74
Precision (%)	88.54	96.90	97.21	97.52	97.52
F-Score (%)	87.19	95.86	96.17	96.62	96.62

It is evident from the results presented in Table 43 that with the increase in delay tolerance value, the true positive detection also increases, subsequently, a consistent improvement in event detection performance has been recorded. Hence, similarly to the sensitivity results presented in Table 19 and corresponding analysis for the NZ GREEN Grid database, it is also evident from the results presented in Table 43 that the incorporation of delay tolerance significantly facilitates the event detection performance for Pecan Street's Dataport.