DEVELOPMENT OF MODEL ARCHITECTURE TO MINIMIZE ERROR FOR SUPERVISED NILM OF RESIDENTIAL BUILDINGS

Course No.: EEE-400

Submitted By-

Md. Tawheedul Islam Bhuian

Student ID: 1606173

Md. Sadequl Islam

Student ID: 1606036

Submitted To-

Department of Electrical and Electronic Engineering

In partial fulfillment of the requirements for degree of Bachelor of Science in Electrical & Electronic Engineering

Supervised by-

Professor Dr. Md. Forkan Uddin

Department of Electrical and Electronic Engineering

Bangladesh University of Engineering & Technology

Dhaka, Bangladesh

May 2022

Certificate of Approval

We hereby would like to declare that this thesis titled “Development of model architecture to minimize error for supervised NILM of residential buildings.” has been composed and authored by Md. Tawheedul Islam Bhuian and Md. Sadequl Islam, under the supervision of Professor Dr. Md. Forkan Uddin, Department of Electrical and Electronic Engineering.

We would also like to declare the following

- This thesis has been written only to partially fulfill a requirement for the award of a graduation degree
- All contents of this document have been written by the authors, unless stated and referred.
- All external sources have been appropriately attributed

Authors

________________________ ____________________

Md. Tawheedul Islam Bhuian Md. Sadequl Islam

Signature of the Supervisor

Prof. Dr. Md. Forkan Uddin
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering & Technology

Acknowledgement

It fills us with profound joy to have completed our thesis work on the topic titled " Development of model architecture to minimize error for supervised NILM of residential buildings.". Firstly, we would like to thank Almighty Allah for giving us patience, integrity and most importantly, for keeping us in good health to have been able to put through effort in completing this manuscript

To accomplish this research, we had to take direct and indirect help from many people. Their invaluable contribution has guided us many times when we lost our path. Without their support, it would have been impossible to get the huge data and information we used to prepare this paper. We would like to convey our humble and heartfelt gratitude to our honorable thesis supervisor Dr. Md. Forkan Uddin, Professor, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology. His continuous support and guidance throughout this research work helped us from the beginning to the end of this research work. We are very grateful to him for being a resolute guide in our endeavor.

We are also thankful to our course teachers, many of our friends, and our beloved parents for their support and enthusiasm throughout the time.

Abstract

Non-intrusive load monitoring (NILM) is a technique that infers appliance-level energy consumption patterns and operation state changes based on feeder power signals. With the availability of fine-grained electric load profiles, there has been increasing interest in using this approach for demand side energy management in smart grids. NILM is a basically a regression problem where disaggregated power consumption across each appliance is predicted from the aggregated signal. Recently, deep learning-based techniques have shown to be a promising approach to solving this problem.

However, increasing appliances, multistate of appliances, and similar power consumption of appliances are three big issues in NILM. To address these problems, first of all we applied algorithm search across each appliance to find the best model that can accurately predict the disaggregated power consumption for different appliances.

This paper also introduces an innovative methodology to enhance prediction across each appliance. This approach is constructed from Long Short-Term Memory architecture with regression and classification sub-network. The proposed disaggregation approach outperformed the currently available state-of-the-art techniques taking mean square error (MSE) and mean absolute error (MAE) as key metrics.

We applied the proposed model on REDD dataset which is publicly available for research. Via our experimental results, we have confirmed that our model outperforms the reference model. Thus, we show that our combination between regression subnetwork and classification subnetwork can be a robust solution to overcome NILM's issues and improve the performance of predicting the disaggregated consumption across appliances.

Table of Contents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Energy is an essential component of all development programs. Without energy, modern life would cease to exist. However, energy resources all over the world are getting depleted. There are several energy-related problems that the world must solve. These energy problems can be grouped under the following three heads: environmental concerns, a large chunk of the population not having access to a modern form of energy, and potential for geopolitical conflict due to escalating competition for energy resources. Carbon dioxide levels, held responsible for climate change, are at their highest in 650,000 years

Scientists predict that if left unchecked, emissions of CO_2 and other greenhouse gases from human activities will raise global temperatures by 2.5◦F to 10◦F this century. The effects will be profound, and may include rising sea levels, more frequent floods and droughts, and increased spread of infectious diseases. Various initiatives have been taken for reducing carbon emissions, across different sectors. The buildings sector is particularly interesting from the viewpoint of reducing emissions. Across the world, buildings contribute significantly to the overall energy consumption In 2004, the total emissions from residential and commercial buildings were 39% of the total U.S. CO_2 emissions, more than the transportation or industrial sector. Furthermore, due to rapid urbanization, the contribution of buildings is only bound to increase.

Studies estimate the CO_2 emissions from buildings to grow faster than other sectors. Of this energy, residential buildings, or homes, can contribute up to 93% in some countries (like India). Thus, optimizing the energy usage of buildings can be an effective way to reduce carbon emissions. Among many ways to reduce energy consumption, energy disaggregation has greater advantages as it facilitates the overall demand response of the power system. Studies suggest that if people are provided feedback on their energy consumption, they can save up to 15% on their bills.

The breakdown of whole-house electricity consumption among the major end-uses is beneficial to increase the homeowners' awareness about the actual energy performance of houses. Traditional load monitoring techniques can be described as intrusive techniques due to the physical placement of sensors on individual appliances to gather end-use load data. Thus, this poses as a long-term intrusion onto the private life and property. More recently, researchers have developed non-intrusive techniques of load monitoring as an alternative to long-term intrusive metering. Non-intrusive techniques of load monitoring are based on the analysis of appliance energy signatures. An appliance signature gives information about the operating state of an individual appliance, using the monitored whole house electric demand. The main advantage of defining appliance signatures in terms of the whole-house load is that, afterwards, only a single monitoring point in the house is required to gather end-use load data. The appliance signature, like the building signature, is assumed to remain constant for the life of the appliance given that no modifications are made or malfunctions occur.

In this chapter, we provide a general overview of the smart grid, subsystems of the smart grid, applications of it in modern grid infrastructure, implementations, and research scopes around this topic. We also provide an overview of energy disaggregation, applications of it which align with the objectives of implementation of smart grid. Furthermore, we state our motivation and contribution to our work.

1.1 Smart Grid: Overview

A smart grid is a two-way communication network of electricity. Older electric grids had unidirectional electric flow from grid to consumer, but a smart grid integrated many intelligence systems into it and made it two ways communication from grid to consumer and consumer to grid [1]. It has advanced metering infrastructure (AMI) which measures all time electricity consumption rate, calculates demand of load in peak time, off-peak time. These help management systems and make the entire system more reliable and more flexible [1].

In this world of competition, electricity is the heart of the present era. Nothing can be possible without electricity. For this, we need more reliable, more available, and more efficient supplies of electricity [1]. Then a smart grid can be an option. The concept of a smart grid is complicated. A smart grid is an intelligent electricity network that deals with all actions of consumers connected to it and makes it more available, more informative, uses technologies in order to save energy, cost reduction, and increase transparency and reliability [2].

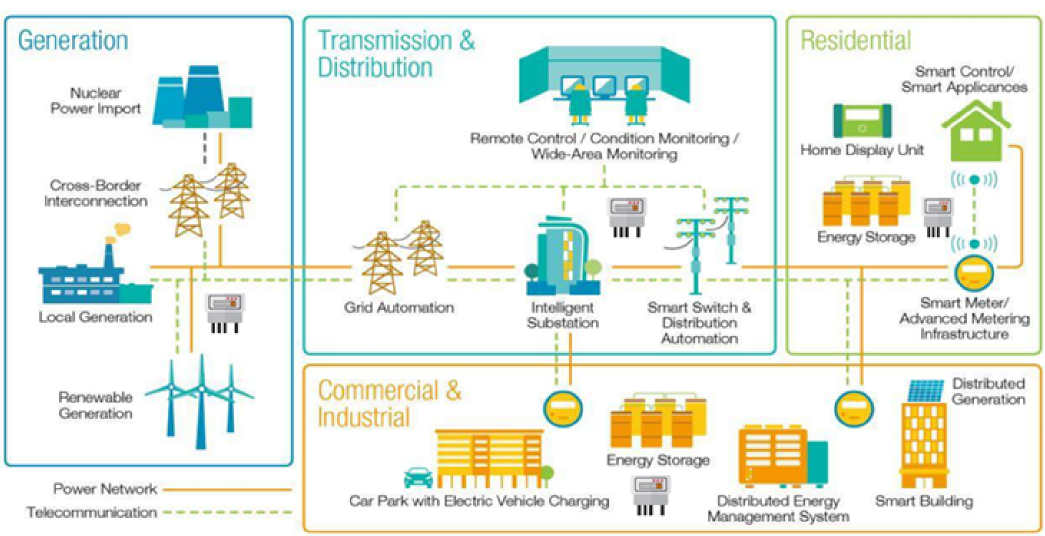


Figure 1: An Overview of Smart Grid

The easiest way to give an overview of a Smart Grid is by its characteristics. The Smart Grid is an upgrade to the current electrical power system, so it has all the functionality of our current power system plus several new functions such as

Self-healing: As the electricity network is getting complex day by day, it is more expensive and troublesome to correct the system faults and to get the system back to its normal operating state. This means that it can redirect and adjust the flow of electricity if an electrical transmission path is interrupted[3]. This is done by a smarter network that uses sensing, control, and communication technology to allow for real-time troubleshooting for unforeseen events. This will reduce the number of major blackouts and their severity will reduce the economic losses our society incurs during these blackouts.

Motivates Customers: Smart grid concept motivates the behavior of electricity usage of customers. The Smart Grid provides customers with more information and options about their electrical power consumption. Now the customer can make a better decision about their power usage, which not only saves them money but also promotes competition between power suppliers. This is done through two ways communication between consumers and suppliers.

Increases Power Quality: Smart grid integrates power electronics with electricity generation, transmission, and distribution systems. Various power electronics devices such as SVC, TSC, STATCOM, TSSC, TCVR, UPFC, IPFC, etc. have greater control over electricity transmission[4]. Some manufacturing processes are overly sensitive to voltage variations. Voltage fluctuations are caused millions of dollars per incident in the commercial field. By integrating power electronics, voltage, frequency, and power flow can be controlled resulting in an increase in the power quality of the electricity grid which is one of the key themes of a reliable grid[5].

Aggregated REs (Renewable Energy Sources): Renewable Energy sources are being added to the grid which ensures a large reserve of dispatchable capacity. Adding more Renewable Energy sources increases the power system reliability. REs is emerging with the demand side of the grid nowadays that's why more distributed energy sources (DERs) are being added to the smart grid technology. DERs include renewable technology such as solar photovoltaic systems, wind generations, and electric vehicles, but also encompass other resource capacities such as demand response (DR) programs, batteries, microgrids, and small generators[6].

Electricity Market: In the electricity market, the smart grid concept increases competition among suppliers. It has a concept called transactive energy. Transactive energy (TE) describes a system where anyone can trade electricity on the grid. Through this system, traditional consumers of electricity can generate electricity and sell their excess capacity back into the grid. As a result, consumers are encouraged to produce power at their premises[7].

Thus, the emerging concepts of smart grid can increase the overall efficiency and capacity over traditional grid.

1.2 Smart Grid Systems

A smart grid network infrastructure is under processing[8]. However, researchers give some infrastructure models. A smart grid has three major systems[1] followed by,

- Smart infrastructure system
- Smart management system
- Smart protection system

1.2.1 Smart infrastructure system

Smart Infrastructure System in a Smart Grid is basically the architecture of the system as a whole, to ensure the flow of bi-directional power and information between the two sides. Smart grid infrastructure includes automation technologies for smart power delivery; DGs, Storage, EVs (Electric vehicles); smart and faster sensing and measurement; integrated communications among all smart grid entities; automation technologies for smart power delivery; fault detection, energy pricing, and energy market; grid data management, data security or data protection; and energy policy under new smart grid environment.

These are the subsystem of a smart infrastructure system.

- The smart energy subsystem is responsible for advanced electricity generation, delivery, and consumption.
- The smart information subsystem is responsible for advanced information metering, monitoring, and management in the context of the SG.
- The smart communication subsystem is responsible for communication connectivity and information transmission among systems, devices, and applications in the context of the SG.

1.2.2 Smart management system

In this system, the management mechanism of power flow between the demand side and supply side is discussed. Demand Response (DR) and Demand Side Management (DSM) are two main management tools frequently called DSI. Demand-side management refers to the efficient use of local load and generation to support the operations, management, and power quality improvement. It can help defer investment in new infrastructure by reducing system peak demand. Demand-Side Management (DSM) are the utility activities that influence customers' usage of electricity[7]. DSM provides the planning, implementation, and monitoring of activities that are designed to motivate consumers to change their electricity usage patterns. Demand Responses (DR) are the mechanisms to manage the demand in response to supply conditions. Smart management systems provide some services such as peak clipping, load shifting valley filling etc.[9].

1.2.3 Smart protection System:

In this subsystem of smart grid systems reliability, security, failure protection, rapid fault analysis and privacy protections are discussed. Various protocols are used to ensure these functions. Smart sensors are embedded in this system which senses current, voltage, temperature, pressure, humidity, etc.[10]. The power system protection and control face both opportunities and problems as a result of the smart grid. The development of wide-area protection based on a wide-area measuring system is accelerated by rapid advancements in communication and measurement capabilities.

1.3 Smart Grid Components

Smart Grid components are a collection of intelligent appliances and heavy machinery that help with the generation, transmission, and distribution of electricity. These appliances are intelligent enough to comprehend how they work and how to use them.

1.3.1 Intelligent Appliances:

An appliance that includes embedded sensors and actuators, the intelligence, and communications to enable automatic or remote control and operation based on user preferences or external signals from a utility or third-party energy service provider. A smart device is generally connected to a smart meter and/or other devices via PLC (Power Line Carrier) or different wireless protocols such as Bluetooth, Zigbee, Wi-Fi, Li-Fi, 3G, etc., and can operate to some extent interactively and autonomously. [[11]]

1.3.2 Smart Energy Meters:

Smart meters establish two-way communication between utility server and consumers. Smart meters control appliances via wireless (HAN) or wired (PLC) communications channels. Unlike analog meters, smart meters have memory devices that record electricity consumption at each ten minutes or half hourly basis and send these data to the server via communication channels. It also detects the failure of systems and gives faster response to repair the faulty connections [[5]][12].

1.3.3 Smart Substations:

Substations are installed for monitoring and controlling critical and non-critical operational data such as power status, power factor performance, circuit breaker operation, security, transformer status, etc. It is an updated version of traditional substations with less instrumentation[13]. They can provide power for switchgear to change the configuration of the network, help isolate lines and clear faults before power can be restored safely. Smart substations contain transformers, capacitor banks, circuit breakers, network protected relays, current and voltage measuring devices, switches, and several others[14].

1.3.4 Integrated communication system:

An integrated communication system is a required object in smart grid technology. It must be able to transfer data as fast as possible. Depending on the requirement, various technologies are used in the smart grid. These are Supervisory Control and Data Acquisition System (SCADA), Energy Management System (EMS), Programmable Logic Controller (PLC), etc. Software and hardware are required for providing greater information to the consumers and enabling customers to participate in the energy market[15].

1.3.5 Phasor Measurement Units (PMU):

The phasor Measurement Unit (PMU) is referred to as Synchro Phasor. PMUs measure data in phasor forms. Phasor is defined by magnitude and phase angle. It collects data from various points of the smart grid and transmits them to the central data locations in a common time source for synchronizations. PMUs use GPS technology and are mainly used in a vast power system network to increase the visibility of data flowing[16].

1.3.6 Real-Time Monitoring and Sensing:

Real-time monitoring and sensing are an essential component of a smart grid. The power industry has only recently begun to move to real-time monitoring systems to provide up-to-date information using two-way communication. At the utility level advancements are occurring more rapidly on the transmission side as compared to the distribution side. The next-generation networked sensors have measuring and processing capabilities that help locate a faulted line and identify parts of the grid that might be susceptible to outages before they occur.

1.3.7 Power electronics and energy storage:

To increase the efficiency of power transfer and to improve the power quality, power electronics devices are needed components in smart grid technology. Various FACTS devices are used in the smart grid power system to greater control over power flows. HVDC, FACTS, and active filters, as well as integrated communication and control, ensure system flexibility, supply reliability, and power quality[4]. Energy storage is required for adding more Renewable Energy to the smart grid. It helps to extend the reliability and flexibility of the smart power system[17].

1.4 Applications of Smart Grid

A smart grid uses technology to improve the connectivity, communication, and automation among the components that are used in the power system. Previously grids had simple operating functions, all electricity was generated in a central unit and distributed over all users connected to it. The voltage transformers regulate the voltage level while delivering energy to the end-users. But in a smart grid, all activities are remotely monitored for better response. Various applications of the smart grid are discussed in the following sections.

1.4.1 AMI: Advanced Metering Infrastructure

In order to establish two-way communication between utility servers and consumers, an integrated system is required called advanced metering infrastructure. It integrates smart meters, communications networks, and data management systems that are required for two-way communication systems[18].

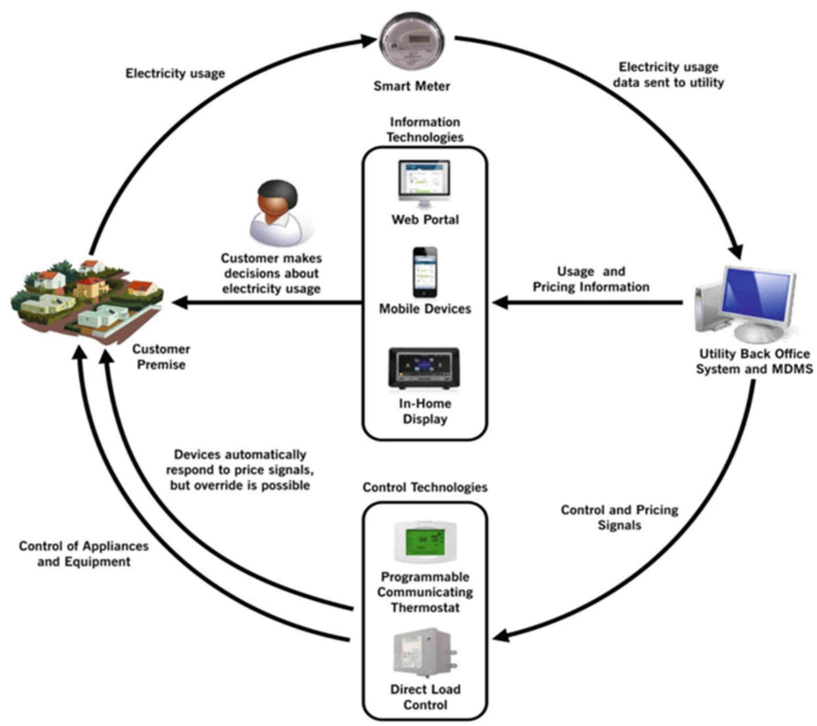


Figure 2: Consumers Technologies and AMI integrated system

AMI, when combined with customer technologies like in-home displays and programmable communicating thermostats, allows utilities to provide new time-based pricing programs and incentives to encourage customers to minimize peak demand and control energy use and expenses[6].

1.4.2 Electric Vehicles (EVs)

Vehicle-to-Grid (V2G) and Grid-to-Vehicle (G2V) communications are implied in the smart grid. Electric vehicles (EVs) are becoming more popular because of their benefits, which include increased efficiency, less environmental impact, and improved economic performance. Managing EV charging patterns is a critical step for EV uptake in global markets since it has a significant impact on the quality of electrical grid transmission and distribution. Furthermore, the integration of electric vehicles and electrical grids is critical, not only for charging management but also for allowing EVs to actively participate in grid support via Vehicle Grid Integration (VGI), which includes technologies such as vehicle-to-grid (V2G) and grid-to-vehicle (G2V)[1][19].

1.4.3 Distributed Energy Resources and Storage

DER stands for distributed energy resources, and they are a type of distributed generation. They refer to smaller energy sources or generation units found on the customer side of the electricity generation meter. Rather than relying on a centralized system, energy is created locally (mainly from renewable sources)[20]. Rooftop solar photovoltaic units and wind turbines are two examples. DER storage, on the other hand, refers to systems that store distributed energy for later consumption. DC-charged batteries and bi-directional inverters are used to accomplish this. It aids in the coordination of energy production, demand, and supply[3].

1.4.4 HAN: Home Area Network

HAN (Home Area Network) connects electricity consuming devices through a dedicated PLC or wireless LAN for remote monitoring and control by home's smart meter in response to signals from utility server. It enhances the capabilities of smart grid by using various network protocols. It also includes security features to safeguard customer information and the metering system. HAN enables centralized energy management and services as well as providing different facilities for the convenience and comfort of the household[21]. HAN objective is more oriented toward energy services i.e., shaping load curve, energy saving and helping grid.

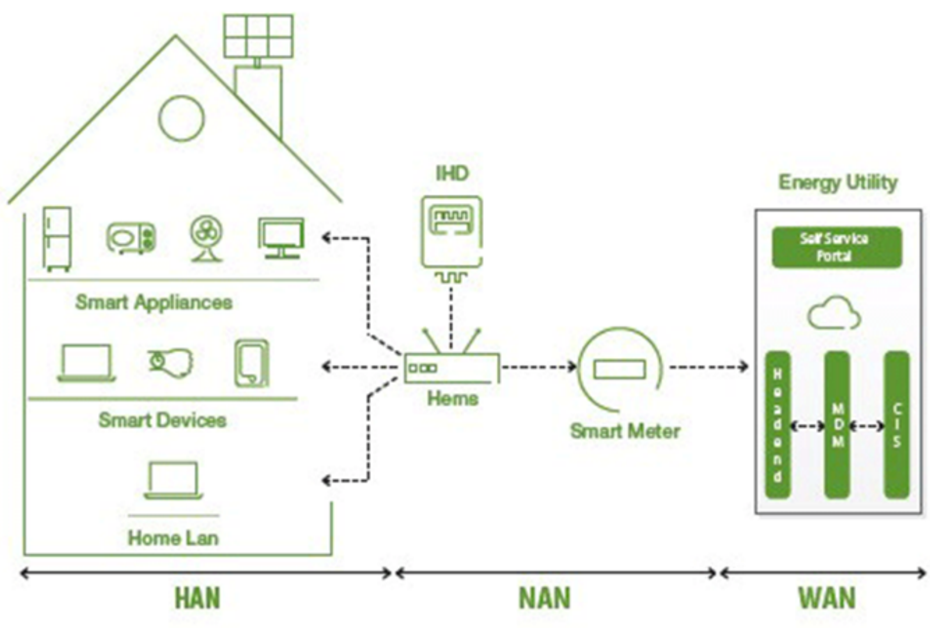


Figure 3: HAN Network

1.4.5 IoT (Internet of Things):

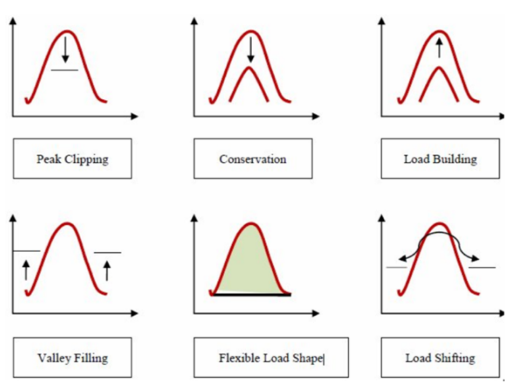
IoT means the Internet of Things. This refers to the ever-growing network of physical objects that feature an IP address for internet connectivity, and communication that occurs between these objects and other Internet-enabled devices and systems for the purpose of remote monitoring and operation by the owner's smartphone app. It is more oriented towards home automation[22].

- Smart grid IoT technology is widely employed in the supply chain to automate processes and boost efficiency[23].
- Adopt automated metering to monitor energy usage in real-time and dynamically respond to changing demand.
- Use environmental data and IoT technologies in renewable energy to optimize power production and maximize the use of green sources of energy.
- Monitor grid load and adopt data-driven strategy to minimize the risks of outages or overloads.

1.5 Demand Side Management (DSM)

Demand side Management is the utility activities that influence the consumer about their electricity usage. It allows customers to make their decisions regarding their electricity consumption and helps the energy provider to reduce the peak demand and reshape the demand curve. Sometimes instead of flattening the curve it is more desirable to follow the generation pattern. In each case, there is a need for control over customer energy use. DSM provides some mechanisms to balance between demand and generation[24].

The motivations for such services from demand side could be price based or incentive based. In the pre smart grid era such services, especially load shifting, valley filling and clipping, were exercised but technology was not modern as today. With the deployment of smart metering and the development of home area automaton technologies, domestic appliances can be controlled in a more intelligent way, therefore bringing more flexibility to the demand side.



DSM programs can be implemented in two ways. Price-based and incentive-based programs

1.5.1 Price-based DSM

Price-based DSI encourages customer load changes in response to changes in the electricity price [9]. The main Price-based programs are described below:

Time-Of-Use rates (TOUr): where a fixed pricing program is applied depending on the period of consumption.

Real-time pricing (RTP): the electricity price provided by RTP rates typically fluctuates hourly, reflecting changes in the wholesale electricity price. Customers are normally notified of RTP prices on a day-ahead or hour-ahead basis

Critical Peak Pricing (CPP): where utilities anticipate high wholesale prices or system emergency conditions for certain periods of time and predetermine electricity sales prices in order to address these situations.

1.5.2 Incentive Based DSM

Incentive-based DSI gives customers load modification incentives that are separate from, or in addition to, their retail electricity rates or in response to emergency condition in power system[25]. Some programs are given below:

Direct Load Control (DLC): where the operator or power distribution company can freely control, interrupt or postpone customer power consumption with a remote-control switch.

Interruptible/Curtailable rates (I/C): Interruptible/Curtailable rates (I/C): where customers, in exchange for lower prices, must reduce energy consumption in a brief period of time, which usually involves periods of high demand.

Demand Bidding Programs (DB): Based on wholesale electricity market prices, customers offer bids for curtailment.

Capacity Market Programs: where customers are guaranteed to contribute to meeting the needs of the grid when needed

Emergency Demand Response Programs (EDRP): where participating consumers respond voluntarily to emergency signals[26].

1.6 Active Research on Smart Grid and Challenges

Without advanced distributed network automation, the smart grid idea cannot be realized effectively. Many renewable energy generation units, such as solar, wind, and biomass, will be connected to distribution grids, as is to be expected[27]. A Smart Grid can potentially turn a point of electricity consumption into a point of generation. Different feeding technologies are employed for various tiny renewable energy units, which has an impact on the power quality fed to the Smart Grid. As a result, research into distributed network automation, demand response and power quality feature such as monitoring, control, and determining the source of disturbances may be the suitable areas of research in SG. This section discusses various solutions to the issues that smart grid deployment faces. Plug-in electric hybrid vehicles have been presented as a technique to store energy as energy storage devices. However, there have been issues expressed about their harmonic's generation to the power grid, as well as the period when they are plugged in or unplugged, creating voltage dips or swells and so compromising power quality. The hunt for a solution to this problem is ongoing.

The current electricity system is driven by demand; when demand is high, generation is high, and when demand is low, generation is low. Renewable energy units are all climate-driven, and their output is yet not matched to demand[28]. In order to accomplish peak shaving and minimize over-generation, incentives such as reduced energy tariffs during generally low demand hours must be used to match demand to power generation profiles.

1.7 Energy Disaggregation

A smart grid is a two ways communication network of electricity. Older electric grids have unidirectional electric flow from grid to consumer, but a smart grid integrated many intelligence systems into it and make it two ways communication from grid to consumer and consumer to grid. It has advanced metering infrastructure (AMI) which measures all time electricity consumption rate, calculates demand of load in peak time, off peak time[29]. These help the management system and make the entire system more reliable and more flexible. Under Smart metering & monitoring system in smart grid, there is a concept of energy disaggregation per home appliance. Energy disaggregation, also known as non-intrusive load monitoring (NILM), or non-intrusive appliance load monitoring (NIALM), can be broadly defined as a set of techniques used to obtain estimates of the electrical power consumption of individual appliances from measurements of voltage and/or current taken at a limited number of locations in the power distribution system of a building[16]. This is the process of decomposing a 'aggregated' energy signal into each individual device that contributes to the total. So, if you assess a house's power draw/energy consumption, the idea is to figure out how much of that should go to each appliance in the house (e.g how much is the refrigerator using, the TV, etc). Without requiring major infrastructure changes such as the addition of individual sensors on each device or power receptacle, disaggregation provides a feasible method for providing energy usage behavior data to the consumer, allowing them to identify behavioral trends or device malfunctions that lead to inefficiencies.

In intrusive monitoring system, consumed data of every appliance can be found by placing a meter to every end of all household appliances. But this meter buying cost will be extremely high and these meters also consume electricity for functioning itself. Therefore, many NILM algorithms came forward.

Disaggregation allows utility providers to strategically advertise items to consumers in addition to alerting consumers about ways to enhance energy efficiency. Companies are now routinely monitoring our internet activity and then presenting adverts that are tailored to our preferences. 'Personalized advertising' is the term for this. The disaggregation of energy data allows for a more consistent marketing of products to consumers. This raises the issue of user privacy as well as the issue of ownership when it comes to power consumption data. Furthermore, disaggregation provides opportunity for better control. Many systems in residential and commercial buildings, such as heating, ventilation, and air conditioning (HVAC) machines, use real-time measurements to implement control strategies. Disaggregation can give controllers information about system flaws such as device malfunction, which can lead to inefficient control. It can also offer data on energy consumption, which is useful for demand response systems.

NILM aims to help households understand how energy is used and, as a result, advise them how to manage energy effectively, allowing for energy efficiency, which is regarded one of the two pillars of sustainable energy policy (i.e., energy efficiency and renewable energy).

1.8 Motivation

The goal of this project is to overview, understand and interpret the recent advances in energy disaggregation, and to provide a new model architecture from which further research can be done. We will first survey the literature for datasets that are available, algorithms that can perform non-intrusive load monitoring, applications that can be deployed for practical usages of energy disaggregation. We will be implementing the algorithms and trying to verify the original results. We will be predicting appliance specific load signature pattern across each appliance based on their category. Developing such comprehensive methodology for detecting all appliances signature with minimal error is still a problem in this domain of research. Also, we learnt to extract hands on knowledge on requirements, scalability and accuracy of the methodology's others have presented before. Although disaggregation is a regression problem, this would be similar to high specificity and low sensitivity in an on/off appliance classification problem. Until now it hasn't been addressed as of yet. So, in the later part of our thesis, we implemented different experimentation pipeline for searching new model architecture that can optimize the current model in terms of minimizing errors for predicting the unseen power consumption data.

1.9 Contributions

The main contributions of the paper are:

- We have developed an end-to-end pipeline for energy disaggregation using public datasets available.
- Proposed methodologies to predict energy consumption across appliances inferring maximum consumption and adapting to the exact pattern with instantaneous changes of power.
- Based on our experimentation results, we proposed a model architecture with minimized error and accurate predictions for load signatures.

Chapter 2

Related Works

2.1 Background:

The idea of Non-Intrusive Load Monitoring was founded 30 years ago. MIT professor Hart proposed an algorithm of energy disaggregation for household appliances. This method is called Hart's method. Various algorithms have been proposed for solving NILM since the inception of the field [8]. These algorithms are classified into two categories. Supervised and Unsupervised[30]. Before 2010, no data sets were available for research purposes. In this time most algorithms were performed by Lab based experiment or via simulation[30]. These algorithms are unsupervised algorithms. Hart's method, Markov model, Hidden Markov model, Factorial Hidden Markov model etc. are the oldest unsupervised method of energy disaggregation. Most unsupervised methods model the on/off sequences of appliances using some variation of HMMs [31].

In the literature, the problem of non-intrusive load monitoring and the available hardware for non-intrusive load monitoring have been widely investigated. Non-intrusive load monitoring, according to the common agreement, is a means of providing consumers with information that makes them aware of their consumption and perhaps gives them with insight into how to improve their usage efficiency. Furthermore, non-intrusive load monitoring equipment is becoming more generally available. As a result, versatile and efficient disaggregation techniques are required.

In the literature, the problem of non-intrusive load monitoring and the available hardware for non-intrusive load monitoring have been widely investigated. Non-intrusive load monitoring, according to the common agreement, is a means of providing consumers with information that makes them aware of their consumption and perhaps gives them with insight into how to improve their usage efficiency. Furthermore, non-intrusive load monitoring equipment is becoming more generally available. As a result, versatile and efficient disaggregation techniques are required[30].

2.2 Disaggregation Algorithms

Energy disaggregation is the process where whole aggregated energy signals are separated into appliances specific data. To know the power-consuming characteristic profile of any appliance at any time each appliance-specific data is required. Energy disaggregation algorithms have generally been fairly basic, focusing primarily on detecting changes in a power signal from a restricted set of device states (e.g., off, on high, on low, sleep mode). The aim of energy disaggregation is to provide estimates of the actual power demand of each appliance at definite time from household aggregate power readings. Some different types of stochastic methods are used for this purpose[32]. They have separated them into two categories followed by:

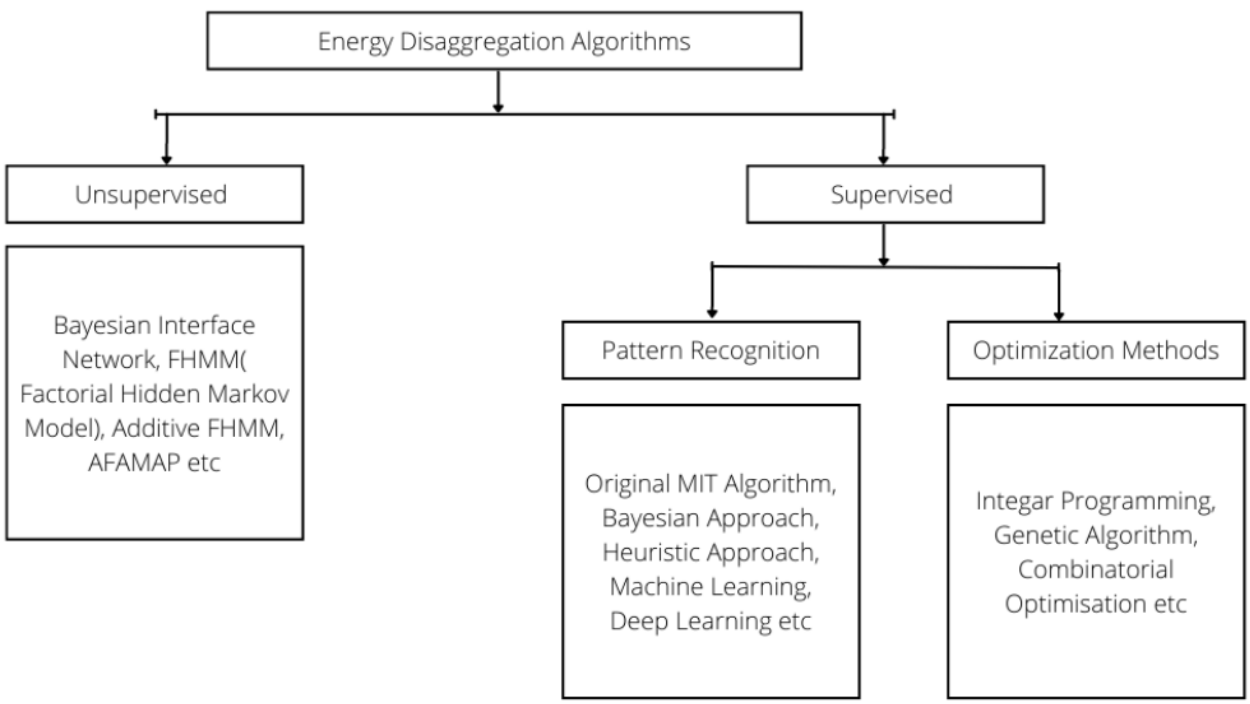


Figure 4: Energy Disaggregation Algorithms

2.2.1 Supervised Methods

In the supervised method, labeled datasets are required to train algorithms to get accurate results. This training dataset contains both correct and incorrect outputs, allowing the model to improve over time. The loss function is used to assess the algorithm's correctness, and it is adjusted until the error is suitably minimized. To train the classifiers, supervised approaches require a labelled dataset that includes the properties of several appliances. On-line and off-line techniques to system training are both viable options. Data is labeled based on real-time event detection and utilized to train the system concurrently in the on-line technique. Appliances are monitored in a specified environment and for a specific amount of time in the off-line training approach, and their associated signatures are labeled. It is costly and time-consuming to install a measurement equipment for each individual appliance in order to acquire the requisite tagged data. Hart et al. (1989) offered an alternate strategy in which appliances are turned on in order to detect them individually from the aggregated data. This approach was improved by using a smart phone to mark the operation of each gadget. Several datasets, including labelled data gleaned by examining the signatures of a variety of household equipment, are now publicly available. Researchers may use these free datasets to train their disaggregation algorithms without having to go through the time-consuming measurement approach stated above.

Pattern recognition and optimization approaches are the two primary groups of supervised disaggregation algorithms. The former depends on events to identify appliances, but the latter does not rely on events to detect them.

2.2.1.1 Pattern Recognition Methods

The original NILM algorithm proposed by Hart et al. (Hart et al., 1989) was a pattern recognition or event-based method. The pattern recognition methods commonly include the following three steps: event detection, feature extraction, and pattern matching. The event detection step in Hart's method was accomplished by using an edge detector to detect differences in steady state power levels. Additional criteria such as spiking, ramping, and small oscillating behavior, huge oscillation, and power fluctuations were proposed in the extended approaches proposed by other researchers in order to detect the events. Several other steady state and transient properties, which were described in the preceding section, have been used to detect events in addition to the aforementioned criteria. Following the detection, labeling, and time stamping of events, a set of signatures is extracted from the measured samples surrounding each event to characterize it[33].

2.2.1.1.1 Original MIT Method:

The original MIT algorithm is also known as Hart's method. A first energy disaggregation algorithm was proposed by Hart, where the aggregate power signal is decomposed to match the individual appliances' typical power demand curves (Commonly referred to as signatures).

In this method, all household devices are taken as constant power consumers. If a device is On, it will consume constant power and no power loss during off period. Total output power consumption at a time is taken as input in this method. Then it compares with all possible combinations the sum of output power of appliance. Then it will give the best possible combination of appliances which are actively used at this time. Following that, On/Off and FSM models are created based on the clusters of step changes. The Zero Loop Sum Constraint (ZLSC) and the Uniqueness Constraint (UC) are used to build the FSM model. As previously stated, the former states that the sum of state transition sequences in any loop is zero, whereas the latter states that there can only be one off state with a power level of zero in each cycle. The associated events are eliminated from the data as soon as the system learns one FSM, and the procedure is repeated by learning additional FSM from the remaining data. It's worth noting that some tolerance must be factored into the above techniques to account for probable errors caused by noise, load changes, or the operation of multiple small appliances at the same time. The Viterbi Algorithm (VA) is an optimum decoding technique that corrects errors when one symbol is distorted into another. Following the techniques outlined above, each appliance generates a distinct cluster in the P-Q plane. Household appliances could be identified in general by comparing each occurrence to a database of known appliance attributes. This data might be gathered through a training process or by analyzing historical data[34]. The limitation of Hart’s approach is that it cannot be usable for Finite State Machine (FSM). It is not able to decompose power signals made of simultaneous on/off events on multiple appliances.

2.2.1.1.2 Extension to MIT Method:

Other event base disaggregation algorithms were then developed to augment the original MIT technique, using additional features. Transient characteristics and a hybrid system that used both transient and steady-state signatures were used in these experiments. In addition, NordFord and Leeb created a method for detecting overlapping transients, which is a difficult problem in event-based algorithms[32].

2.2.1.1.3 Naive Bayes Approach

Naive Bayes is a classification approach that adopts the principle of class conditional independence from the Bayes Theorem. This means that the presence of one feature does not impact the presence of another in the probability of a given outcome, and each predictor has an equal effect on that result. For each appliance, The Bayes classifier is trained on power level during state change, and it can detect the specific state of an appliance. Particle filtering is one type of Bayesian filtering technique which uses successive Monte Carlo simulations.

2.2.1.1.4 Heuristic method

The histogram thinning strategy is used to cluster real and reactive power in this methodology. When comparing this method to the Bayesian approach, it is clear that Bayesian classifiers perform better in cases where appliances use stable electricity.

2.2.2 Supervised Machine Learning

Some probabilistic methodologies of machine leaning have been used. To handle the cases in which many electrical devices must be distinguished with high accuracy, supervised machine learning (ML) algorithms can be employed.

Supervised machine learning (ML) algorithms can be used to address scenarios when a large number of electrical devices must be recognized with high accuracy. These methods enable the training operation to be carried out utilizing a variety of features, such as state transitions and temporal data. Artificial Neural Networks (ANN), Hidden Markov Models (HMM), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two widely used machine learning algorithms in the literature. The number of signatures, kind, and a number of devices used to train ML classifiers all affect their performance. Machine learning algorithms are typically memory-heavy and computationally complex. Furthermore, as the number of input parameters grows, so does the amount of training and classification time required. The usefulness of HMM is limited due to complexity difficulties, and this complexity grows exponentially as the number of devices increases. Furthermore, replacing household appliances necessitates going through the learning process all over again (Suzuki et al., 2008). Both ANN and HMM require a considerable quantity of data to train and develop models for each individual appliance, a challenge that gets even more acute when dealing with a large number of devices. Input feedback, on the other hand, can help ANN become more adaptable and perform better.

2.2.2.1 Hidden Markov Model (HMM)

Hidden Markov model is a stochastic analysis. It follows Markov chain. A Markov chain is a mathematical system that changes from one state to another based on probabilistic criteria. In this chain, the output always depends on the present state no matter how the present state is attained. The future state is only the function of the present state. Suppose a Markov chain gives an output after N intermediate states of calculation. Here in Hidden Markov Model these intermediate models will be hidden. HMM only gives the final output states.

A first-order Markov chain is a sequence of conditionally dependent variables, where the variable at each time slice is dependent only on the variable immediately preceding it. An HMM is made up of a Markov chain of discrete variables, each of which is responsible for a corresponding observation. The HMM is a well-studied probabilistic model, and it is considered as a novel approach for appliance disaggregation. It has been successfully applied to the fields of speech recognition, natural modeling, and online handwriting recognition. According to the NIALM scenario, the simplest problem is one in which we wish to determine the state of a single multi-state appliance. A first-order Markov chain is a series of conditionally dependent variables in which each time slice's variable is only dependent on the one before it. A Markov chain of discrete variables makes up an HMM, each of which is responsible for a single observation. The HMM is a well-studied probabilistic model that is thought to be a unique approach to appliance disaggregation. Speech recognition, natural modeling, and online handwriting recognition have all been successfully implemented. The most basic challenge, according to the NIALM scenario, is determining the state of a single multi-state appliance.

2.2.2.2 Decision Tree

Decision Trees are a type of Supervised Machine Learning where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split. For decision tree regression, the decision or the outcome variable is Continuous. There are many algorithms out there which construct Decision Trees, but one of the best is called as ID3 Algorithm. ID3 Stands for Iterative Dichotomiser 3.

2.2.2.3 Artificial Neural Network (ANN)

As in recent years, energy consumption datasets are available, and various neural network models such as convolutional neural networks and recurrent neural networks have been investigated to solve the energy disaggregation problem. Neural network models can learn complex patterns from enormous amounts of data and have been shown to outperform the traditional machine learning methods such as variants of hidden Markov models.

2.2.2.4 Convolutional Neural Networks

Convolutional neural networks (CNNs) have achieved the state-of-the-art performance in many applications such as computer vision [19], speech and audio processing [24] and natural language processing [4]. With shared filters to capture local patterns of various signals, the number of parameters of a CNN is fewer than that of a fully connected neural network. Time domain CNNs have been applied to energy disaggregation, for example, in [32]. Similar to the two-dimensional CNN for computer vision [19], a time domain CNN consists of several convolutional layers, each of which contains several filters that are used to convolve with the output of the previous convolutional layers.

2.2.2.5 Recurrent Neural Networks

Recurrent neural networks (RNNs) have many successful applications in modeling temporal signals, e.g., audio and speech signal processing [8] and natural language processing [3]. Similar to the fully connected neural networks, each input sample is mapped to a hidden unit by a transformation matrix. In addition, there are connections between adjacent hidden units to carry on the information from previous samples. In a non-causal system, a RNN can be bidirectional so as to use information from both history and future. GRUs and LSTMs are different types of state-of-the-art RNN architectures used for time series forecasting, sequence prediction, machine translation etc. tasks.

2.2.3 Unsupervised Method

Unsupervised learning methods are basically dealt with unlabeled patterns in data sets containing data points that are neither classified nor labeled. Model itself finds hidden patterns and insights from the given data. Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have no corresponding output data for the model to predict to.

2.2.3.1 Factorial Hidden Markov Model (FHMM)

Hidden Markov models are a generalization of mixture models. The probability density over the observables defined by an HMM is a combination of the densities defined by each state in the underlying Markov model at every given time, step.

2.2.3.2 Additive Factorial Approximate Maximum A Posteriori (AFAMP)

AFAMAP works with two simultaneous models: the additive model and the difference model. The additive model represents each device's own contribution to the aggregate signal. On the other hand, the difference model corresponds to changes in the overall power consumption at each timestep under some assumptions, changes in the aggregated signal value can be explained by a single device turning on/off.

2.3 Datasets

REDD: First publicly available data set was the "Reference Energy Disaggregation Dataset (REDD)" in 2011 for the purpose of massive research. It took data from six households and has both aggregated and sub-metered power data. It has both high frequency and low frequency consumption data. This data set is mostly used in energy disaggregation as it was first dataset that was publicly available for NILM purpose. [7]

BLUED: The Building-Level fully labelled dataset for Electricity Disaggregation (BLUED) was introduced in 2012 (Anderson et al., 2012) [8]. The BLUED dataset comprises voltage and current measurements taken from a single household for a week at a sampling rate of 12 kHz.

SMART: The smart* project, released in 2012, consists of measurement data from three households located in Massachusetts (U.S). Although, the sub-metered data is just provided for one of the houses, the data set contains the data obtained from many additional sensors including aggregate electricity usage, with a sampling rate of 1 second, temperature and humidity data in indoor rooms and out-door weather data. [9]

GREEND: The GREEND dataset (Monacchi et al., 2014), introduced in 2014, includes appliance level power measurement data with the sampling rate of 1Hz from 9 households in Austrian region of Carinthia and the Italian region of Friuli-Venezia Giulia.

UK-DALE: UK Domestic Appliance-Level Electricity data set was released in 2014 which contains data from four households using both aggregate meters and individual appliance sub-meters. This data set includes the measured aggregate power data (with the sampling rate of 16 kHz) and appliance-level power consumption (with the sampling rate of 1/6 Hz) from five households for a duration of 655 days.

Table 1: Comparison on the mentioned datasets.

Dataset	REDD	BLUED	UK-DALE	WHITED	PLAID
Date	2011	2012	2017	2016	2014
Residential	Yes	Yes	Yes	Yes	Yes
Commercial	No	No	No	Yes	No
Industrial	No	No	No	Yes	No
High frequency sampling	15kHz	12kHz	16kHz	44.1kHz	30kHz
Low frequency sampling	0.5 Hz	N.A.	1 and 6Hz	N.A.	N.A.
Number of classes	N.A.	27	N.A.	47	11
Number of appliances	22	43	N.A.	109	235

2.4 Summary

Identification of Power consumption signature has been implemented in different schemes in literature. In general, NILM identification process requires implementing six stages of analysis and we can summarize the literature in following points:

1. Data Acquisition: It is required to gather information of the steady and transient states from the power waveforms of the device. A high frequency sampling stage captures information regarding transient events; meanwhile, a low frequency sampling stage gathers steady state information of the device.

2. Data processing: Data must be conditioned and processed in order to give meaningful information. This stage includes noise filtering, harmonic components separation, signal synchronicity, etc.

3. Event detection: Processing and storage of all the information is inefficient and impractical process, so it is important to detect the activation and deactivation of the device. It is necessary to establish a threshold crossing detection mechanism for the detection of transient.

4. Characteristic Extraction: Electric parameters, such as active power, reactive power, harmonic components and transient waveforms, can be extracted from the event detection and data processing stages. The identified characteristics are depending of the disaggregation method used for load identification.

5. Load classification or disaggregation: Using the characteristic information gathered from the processed data, along with a known pattern, the device disaggregation can be performed from the total energy consumption, that is, the device can be identified.

6. Energy calculation: By identifying an individual device, its operation pattern and energy consumption can be estimated.

Chapter 3

Proposed Methodology

3.1 Objectives

In this section, we outline the main experimentation procedure. We have designed these experiments with the following goals in mind:

- Find out if disaggregation is possible and can be done by producing useful results. We also want the experiments to give us insights into the requirements of energy disaggregation algorithms and their feasibility
- Interpret and understand the current NLM algorithms described in various papers. We have also implemented the current state-of-the-art energy disaggregation algorithm.
- We also developed a comprehensive methodology to predict energy consumption across appliances inferring maximum consumption and adapting to the exact pattern with instantaneous changes of power.

To do this we will train and test the algorithms with different power series from the REDD dataset.

3.2 Appliance Specific Algorithm Search

3.2.1 Dataset

As to the dataset used, we selected the real-world dataset "the Reference Energy Disaggregation Data Set (REDD)". It is a collection of power consumption data from 6 households. The REDD data set contains two main types of home electricity data: high-frequency current/voltage waveform data of the two power mains as well as the voltage signal for a single phase and lower-frequency power data including the mains and individual labeled circuits in the house. The dataset contains both house-level energy usage (aggregate readings) and appliance-level energy usage (appliance readings) of more than 22 different types of appliances. The data is logged at a frequency of about once a second for a main and once every three seconds for the circuits. In this thesis, we focus on the disaggregation of six types of appliances: Refrigerator, Microwave, Lighting, Kitchen Outlets, Dishwasher, and Washer Dryer for most values of total energy consumption.

3.2.2 Preprocessing

Data alignment: REDD dataset's low-frequency samples vary from 0.33Hz to 1Hz sample frequency. Also, the aggregate reading for house level usage was recorded at 1Hz where the appliance readings were taken once every three seconds for the circuit. Therefore, it was necessary for us to align multiple time series with different acquisition frequencies.

Data imputation: Due to hardware issues, there were smaller gaps present in the time series data. We needed to split the sequence so that the duration of missing values in the subsequence is less than 20 seconds. Then fill the missing values in each subsequence by a backward filling method.

Data filtering: In the filtering section of data preprocessing, we assumed zeros for the gaps that are caused by the appliance being switched off or malfunctioning.

Top-k appliances: It is often advantageous to model the top-k energy-consuming appliances instead of all appliances for the following three reasons. First, the disaggregation of such appliances provides the most value. Second, such appliances contribute the most salient features, and therefore the remaining appliances can be considered to contribute only noise. Third, each additional modeled appliance might contribute significantly to the complexity of the disaggregation task. So, we have selected the most energy-consuming appliances for our disaggregation task.

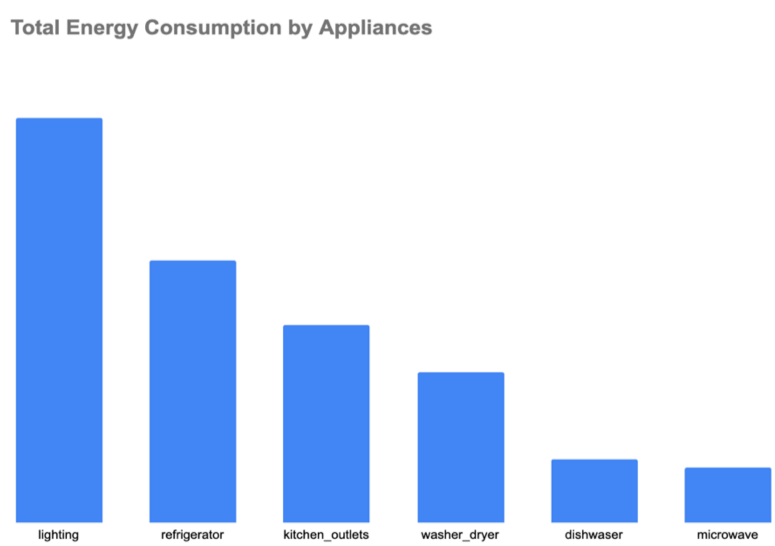
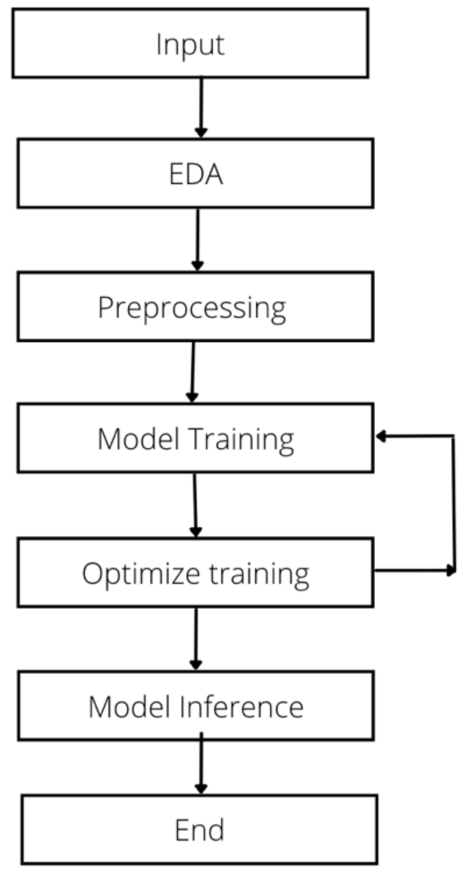


Figure 5: Top-6 appliances for most values of power consumption.

Oversampling: Oversampling is used to solve the problem of the overrepresentation of inactive windows and the irregularity of the active/inactive windows imbalance (described in the Dataset section). The process consists in replicating randomly picked active windows in each of the appliances to obtain a 50% - 50% class balance. The ratio between active/inactive windows is configurable in settings

3.2.3 Procedure

As we understand, energy disaggregation is a regression problem. That is to predict the disaggregated energy consumption data across each appliance with the aggregated power data given as inputs to the model. In the appliance category section, we have seen that various appliances have different load signatures (linear, finite state, multi-state, etc.) that can't be predicted by a single algorithm. Therefore, in this section, we first introduced an appliance-specific algorithm search on the following algorithms to predict the consumption data. The aim of implementing these algorithms is not to present the state-of-the-art disaggregation result but to understand and interpret the new algorithms on which the appliance-specific load signature pattern is predicted precisely.



We disaggregated the energy consumption across appliances with the following algorithms.

- Decision Tree Regression
- Random Forest Regression
- Linear Regression
- XG Boost Regression
- Artificial Neural Networks
- Long Short-Term Memory (Recurrent Neural Network)

We expected LSTM to perform better as it was mentioned in the reference paper. It is also obvious given that it is developed for time series forecasting problems. Next, we have trained the model with the preprocessed data after splitting into a train-evaluation-test set. We optimized the model by hyperparameter tuning and saved the model weights of the best performing model for inference. In inference we have predicted the energy consumption on the unseen test data and calculated associated error (MAE) for each algorithm and appliance.

3.2.4 Results

We have plotted the absolute error and squared error comparison for searching the best performing algorithm for a different algorithm.

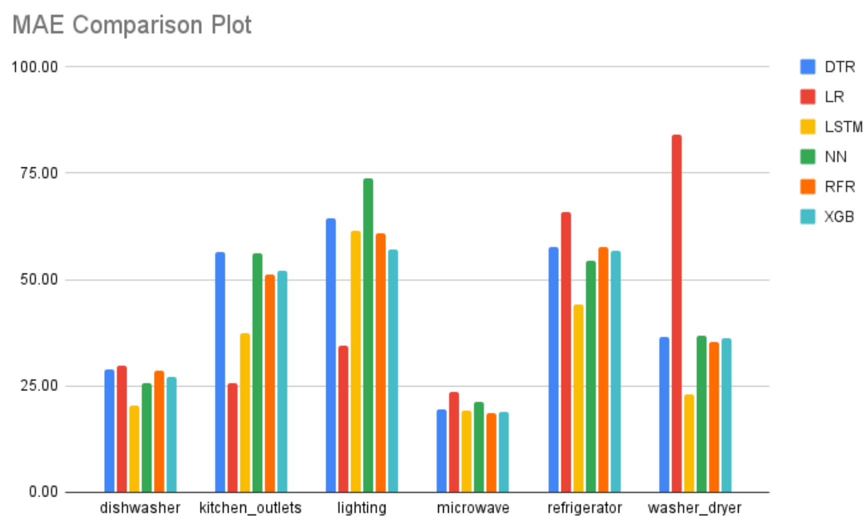


Figure 3.2

Figure 6: MAE Comparison Plot for specific algorithms across appliance

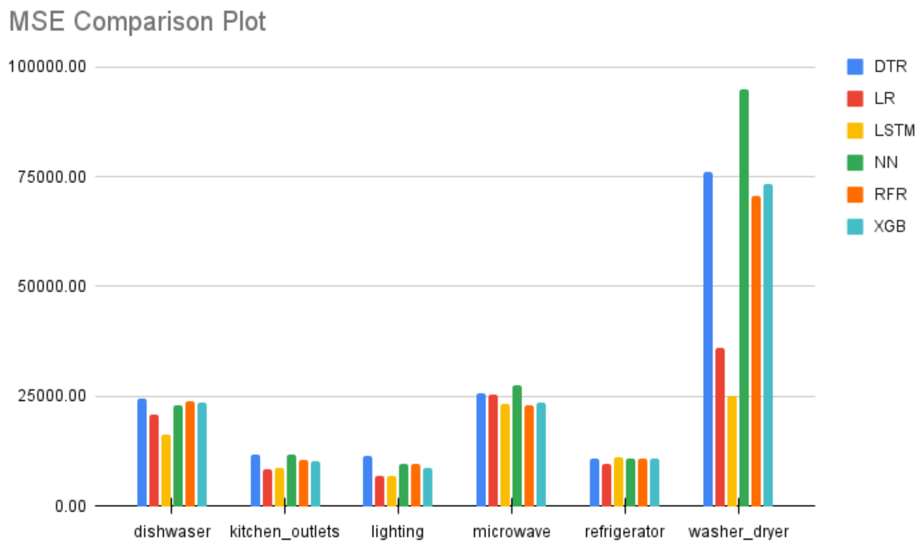


Figure 7: MSE Comparison plot for specific algorithms across appliance

We concluded our algorithm search for the best algorithm across each appliance in the following table:

Table 2: Best suggested algorithms against each appliance

Appliance Name	**Algorithm**
WASHER DRYER	LSTM (Long Short-Term Memory)
KITCHEN OUTLET	LR (Linear Regression)
LIGHTING	LR (Linear Regression)
MICROWAVE	DTR (Decision Tree Regression)
DISHWASHER	LSTM (Long Short-Term Memory)
REFRIGERATOR	LSTM (Long Short-Term Memory)

We visualized the actual vs predicted consumption pattern across test data to validate our algorithms and draw conclusions for further understanding of the energy disaggregation and improvements to focus on.

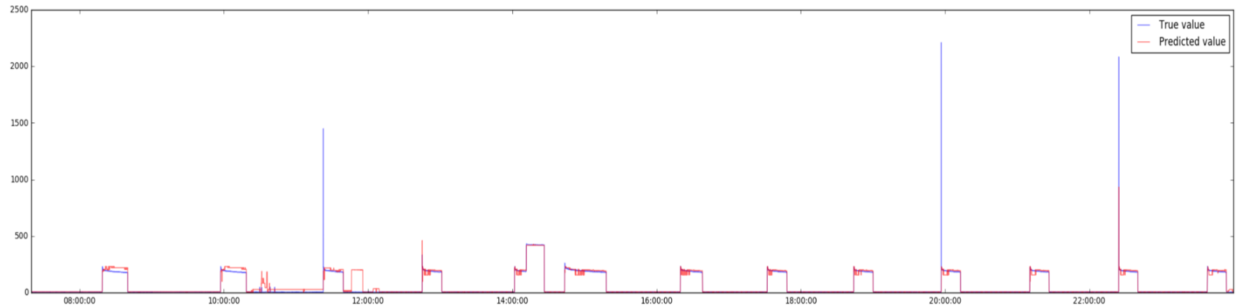


Figure 8: Actual vs Predicted Consumption pattern on the test data

3.2.5 Discussion:

During the experiments, the results were significantly better in train and evaluation than in test as we have seen for almost all algorithms. The explanation behind this outcome is that our model was trained with a dataset with a very low variation of patterns of appliances. This is because there were just 3 to 5 different types of the same appliance for training, to test the model for a totally different type of appliance. For example, the patterns of the fridge consumption in the training set were different from the pattern of the testing set, therefore, the model did not have a broad variety of load profile patterns to learn to infer from. In the following image, we can see the variation of one house from the training set and the house for the testing set for the fridge appliance. The variation is of the two houses is totally different, with this visualization our explanation is endorsed.

Figure 3.5

Figure 9: Box Diagram Plot for Fridge Consumption in House 1 & 2

Also at times, our initially proposed model didn't capture the adequate peaks of the consumption due to sudden state change (on/off). Also due to similar load profile of other appliances, our proposed models gave false state change which badly contributes in increasing our test error taking metrics into count. Both of these discussions are attested in the following-

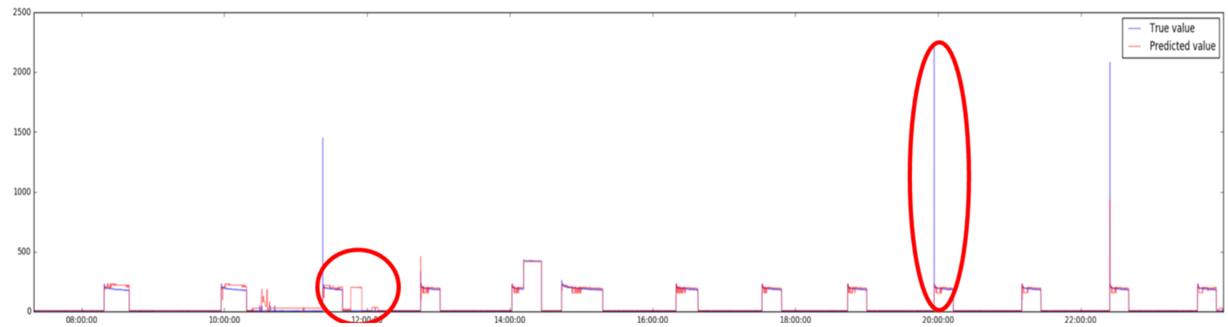


Figure 10: Actual Vs Predicted Consumption for Refrigerator

Also, for appliance with smaller state duration (e.g., Washer Dryer, Microwave), their consumption pattern largely depends on the few state changes in an entire day or even in months. Our searched models were incapable of predicting these load signatures and resulting error not acceptable for any implementations any further.

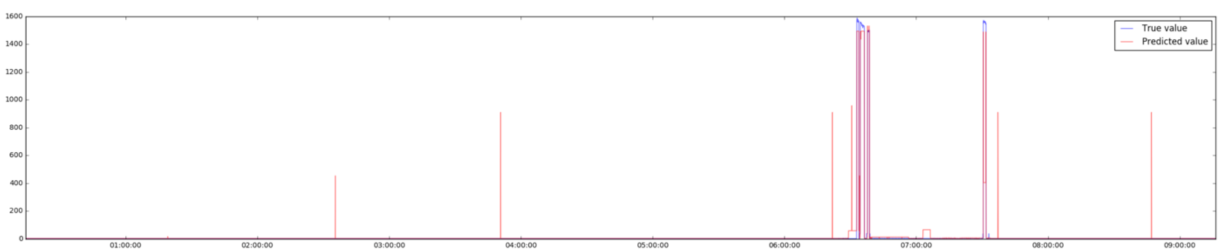


Figure 11: Actual Vs Predicted Consumption for Microwave

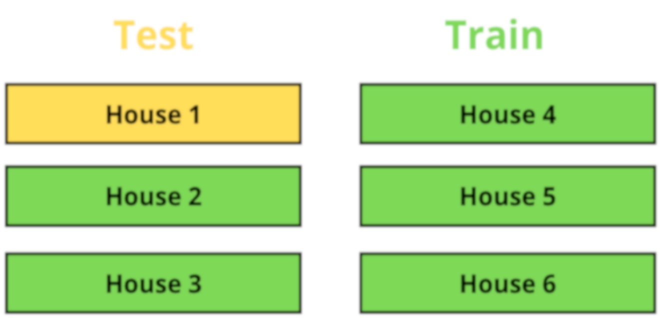
3.3 Proposed Optimization of LSTM Architecture

3.3.1 Dataset

The actual dataset of our model is a combination of two datasets. We found a deep learning research team from Seoul National University that had a pre-processed dataset that cleaned the data of the REDD dataset given in the Reference Paper. This dataset is used in their "Subtask Gated Networks for Non-Intrusive Load Monitoring paper".

On the other hand, in the REDD dataset, there is a high active/inactive windows imbalance. This irregularity is observed especially in the case of the dishwasher and the microwave. As is expected, due to the use of these appliances, much of the time a dishwasher and a microwave are not being used. Therefore, there is a high overrepresentation of inactive windows. We implemented an oversampling process described in the pre-processing section to solve the problem.

The dataset is split using house 2,3,4,5,6 to build the training set and house 1 as the test set.



3.3.2 Preprocessing

Initial project implementation was done using a raw REDD dataset and it was necessary to pre-process the data as described in "Subtask gated networks for non-intrusive load monitoring", See details in the preprocessing section above. We then generate sliding windows for the dataset across each appliance. Using a sliding window over the aggregated signal with a hop size equal to 1 sample.

Once authors from Seoul National University provided us with the same dataset as the Reference Paper, we disabled our data pre-processing. The main reason was to assure the same input data as the original paper to have the same, or similar, results.

Oversampling is used to solve the problem of the overrepresentation of inactive windows and the irregularity of the active/inactive windows imbalance (described in the Dataset section). The process consists in replicating randomly picked active windows in each of the appliances to obtain a 50% - 50% class balance. The ratio between active/inactive windows is configurable in settings.

After implementing oversampling, the number of windows used for train, evaluation and test is listed below

Table 3: number of windows used for train, evaluation and test

Appliance	Nº buildings train	Nº windows train	Nº windows eval	Nº buildings test
Dishwasher	5	289163	123927	1
Refrigerator	4	613167	262787	1
Microwave	3	82922	35538	1
Kitchen Outlet	5	329809	87101	1
Lighting	5	423876	134769	1
Washer Dryer	4	73987	23446	1

3.3.3 Model architectures

We've implemented three different model architectures:

- Regression and classification enabled
- Only regression is enabled.

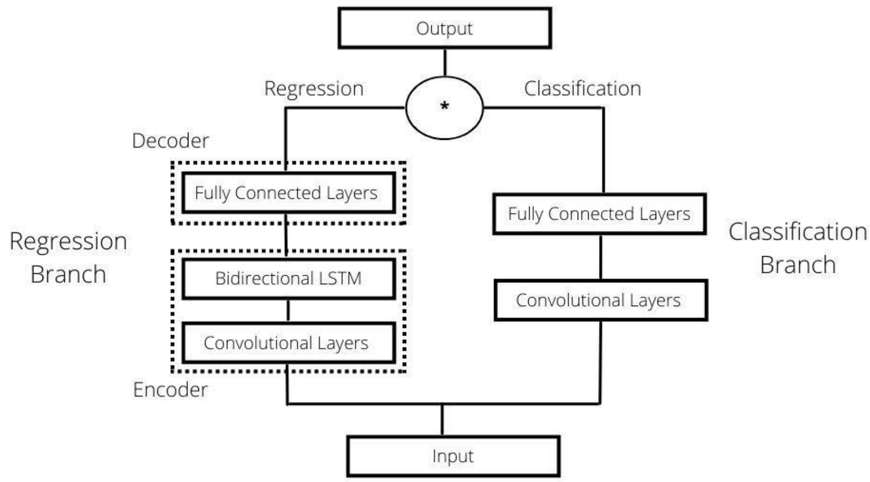
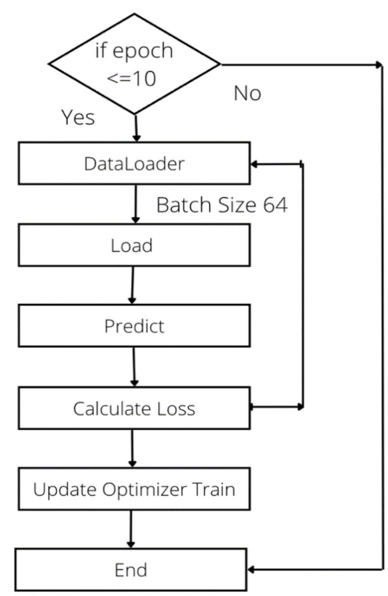


Figure 12: Block diagram of implemented model architecture

3.3.3.1 Train

We trained the model using the whole pre-processed train dataset and batches of size 64 via a data loader. At first, we set the epochs at 10 epochs, in most of the cases we found enough to do an initial analysis of model response and performance. Load -> predict -> calculate loss -> update optimizer train sequence is done per each of the train batches in each epoch. The common Load -> predict -> calculate loss validation sequence is done per each of the validation batches in each epoch. Loss function. An aggregated loss function is used for the joint optimization of both regression and classification network:



$$L=L_{out}+L_{clas}$$

Here,

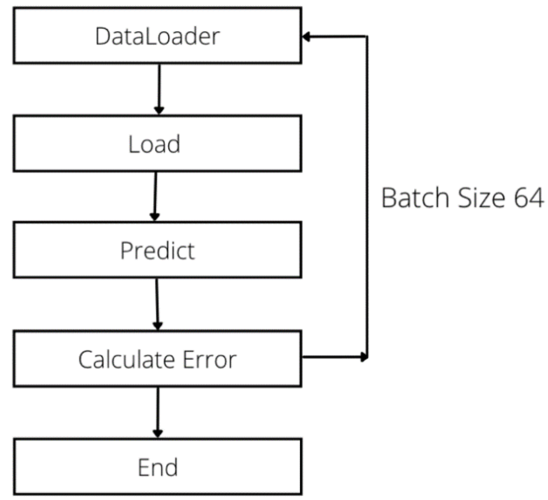
L_{out} is the Mean Squared Error (MSE) between the overall output of the network and the ground truth of a single appliance

L_{clas} is the Binary Cross-Entropy (BCE) that measures the classification error of the on/off state for the classification subnetwork.

3.3.3.2 Test

We run inference on our best model weights over the whole preprocessed test dataset using batches of size 64 via a data loader. Load -> predict -> calculate error test sequence is done per each of the test batches.

MAE (Mean Absolute Error) is used to evaluate the performance of the neural network. MAE is calculated after applying the prediction postprocessing described in the Postprocessing section. These are the metrics used in the Reference paper and are used as benchmarking criteria between the different experiments described below.



3.3.3.3 Experiments

The main goals of the experiments are to learn how to implement and optimize the deep learning-based energy disaggregation algorithm in the reference paper. Also, to understand and interpret the current NILM neural network described in the paper. Regarding experiments, we looked for understanding of these sub-networks (regression and classification) implemented and to conclude whether adding a classification sub-network improves the metrics for energy disaggregation. We proposed the three main architecture modifications evaluated in the experiments during the analysis of the reference paper. The experiments were not designed sequentially after processing the results of the previous experiment.

Main architecture modifications:

- Architecture 1 - Only regression enabled
- Modification 1 - Regression and classification enabled

We initially explored the data to have a first picture of the type and the amount of data available. We realized there was a high active/inactive windows imbalance in the case of dishwasher and microwave (as explained in the Dataset explanation). There would be enough total amount of windows to train the model, but not enough specific active windows to prevent a biased model. If no oversample was done the model would mainly predict null demand in inactive windows, which would be correct, but would fail to predict non-null demand inactive windows. Although disaggregation is a regression problem, this would be similar to high specificity and low sensitivity in an active/inactive appliance classification problem.

3.3.4 Architecture 1 - Only regression enabled

Hypothesis

In this architecture, we concatenate the output of the regression subnetwork with the output of the stack of convolutional layers, in the classification subnetwork. This concatenated vector is fed to the 2 fully connected layers on top of the classification branch. Both subnetwork outcomes are concatenated at the end to the outcome of the disaggregated consumption of the appliances.

Experiment setup

We have attached the experiment setup details in below chart. Each of the columns describe a specific option of the introduced network architectures:

Table 4: Experimental Setup Table(1)

Appliance	**Regression**	**Classification**
Refrigerator	TRUE	FALSE
Dishwasher	TRUE	FALSE
Microwave	TRUE	FALSE
Lighting	TRUE	FALSE
Kitchen Outlets	TRUE	FALSE
Washer Dryer	TRUE	FALSE

Results:

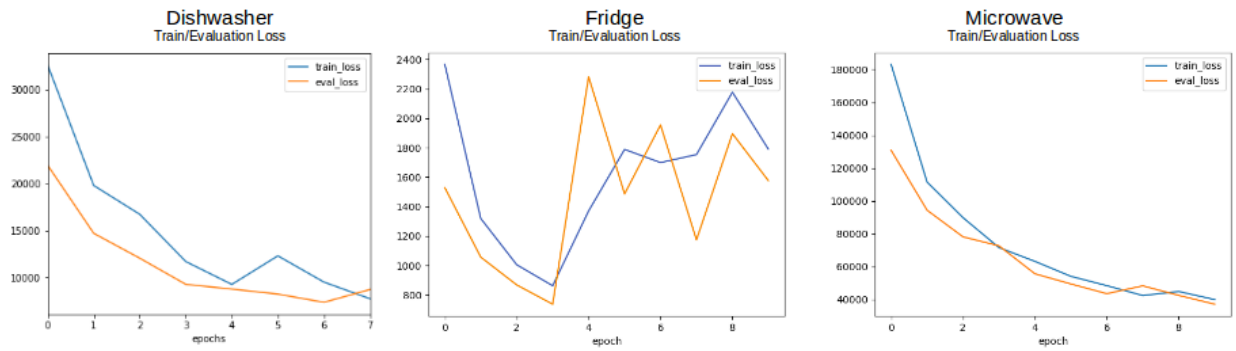


Figure 13: Train/Evaluation loss vs epoch curve

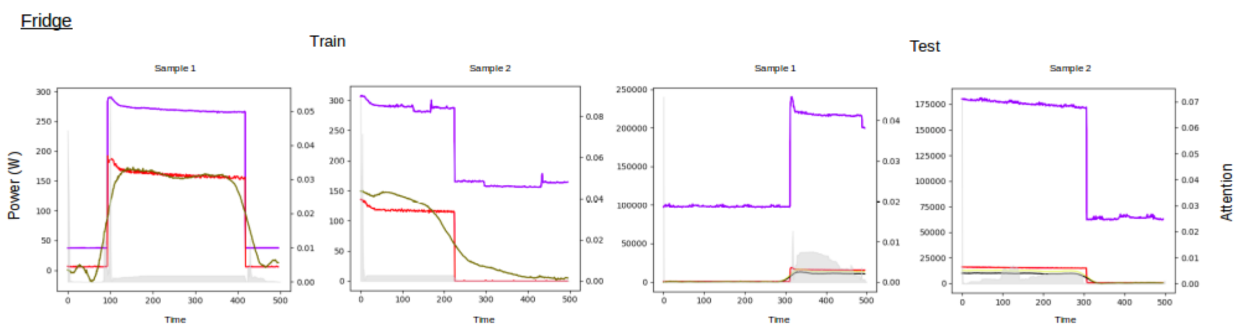


Figure 14: Train and test samples of fridge to evaluate disaggregation

For taking inference, we saved the model weights for the architecture across all appliances. We then predicted the energy consumption and calculated mean absolute error from the predicted sequence to compare the results:

Table 5: Error Comparison(Part 1)

Appliance	MAE	Nº Epoch	Training Time (Hrs.)
Refrigerator	44.14	10	32
Dishwasher	20.22	10	25
Microwave	19.15	10	18
Lighting	61.54	10	30
Kitchen outlets	37.36	10	19
Washer dryer	23.12	10	26

Discussion:

The results of this experiment are better than the simple LSTM architecture that was implemented in Part A. The main hypothesis was that the it would improve the performance as it was introduced earlier in a reference paper. Hence, the results were better than expected. It learns how to focus on peaks of consumption (much better than in Part A) and gives the model the ability to generalize better than what can be seen in Part A. Without the convolution branch, as it created a feature map for the inputs that was injected, model params increased and it trained the model without overfitting (as we can see that in fig 1) We conclude that after the peak the model expects a long-term change of consumption and in the microwave case it does not occur. That's the main difference between microwave and fridge. This hypothesis cannot be applied in the dishwasher, because of the peaks of other appliances during the time it is on (that produces noise). Lastly, the regression is more sensitive to changes and allows to catch the pattern of the input smoothly. But only having the regression model subtracts the model from the specification.

3.3.5 Modification 1 - Regression & Classification enabled

Hypothesis

The regression subnetwork infers power consumption, whereas the classification subnetwork concentrates on binary classification of appliance status (on/off). The network's representational power is improved by this classification branch, which allows it to detect points in the aggregated input sequence that contain important information for identifying appliance-specific patterns. The main hypothesis of this experiment is whether attention can detect the consumption pattern and replace what in previous experiments was the classification branch by modulating the output of the regression branch. By extracting the classifier branch, the model prediction is expected to detect the peaks but may predict values with the biggest difference to the input consumption than with the classification branch.

The designed architecture adopted to solve the NILM problem is based on a classical end-to-end regression network with its encoder-decoder components. Adding an attention mechanism in between the encoder and decoder. Apart from the main end-to-end regression network, an auxiliary end-to-end classification subnetwork is joined.

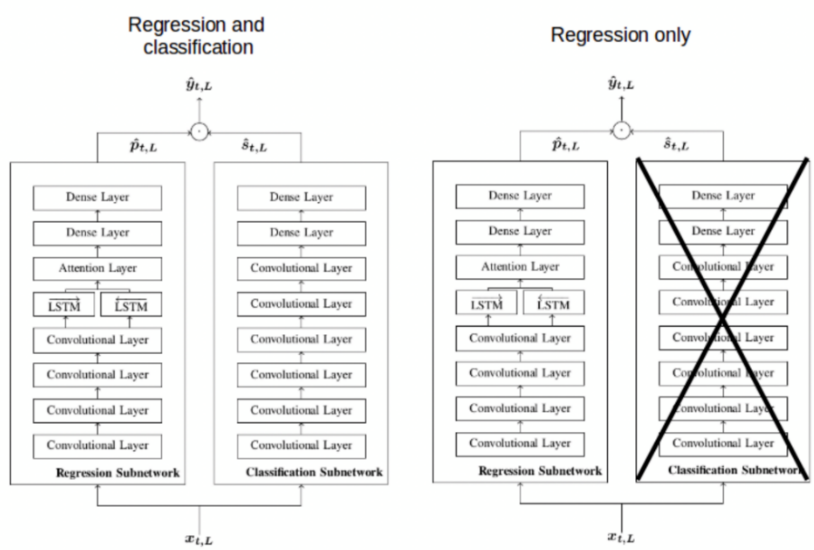


Figure 15: Paper Modification subnetworks

Both subnetworks have a different objective:

- Regression end-to-end network: allows the subnetwork to “implicitly detect and assign more importance to some events (e.g., turning on or off of the appliance) and to specific signal sections”.
- Classification end-to-end network: helps the disaggregation process by enforcing explicitly the on/off states of the appliances.

Experiment Setup

We have attached the experiment setup details in below chart. Each of the columns describe a specific option of the introduced network architectures:

Table 6: Experimental setup table (2)

Appliance	**Regression**	**Classification**
Refrigerator	TRUE	TRUE
Dishwasher	TRUE	TRUE
Microwave	TRUE	TRUE
Lighting	TRUE	TRUE
Kitchen Outlets	TRUE	TRUE
Washer Dryer	TRUE	TRUE

3.3.6 Results

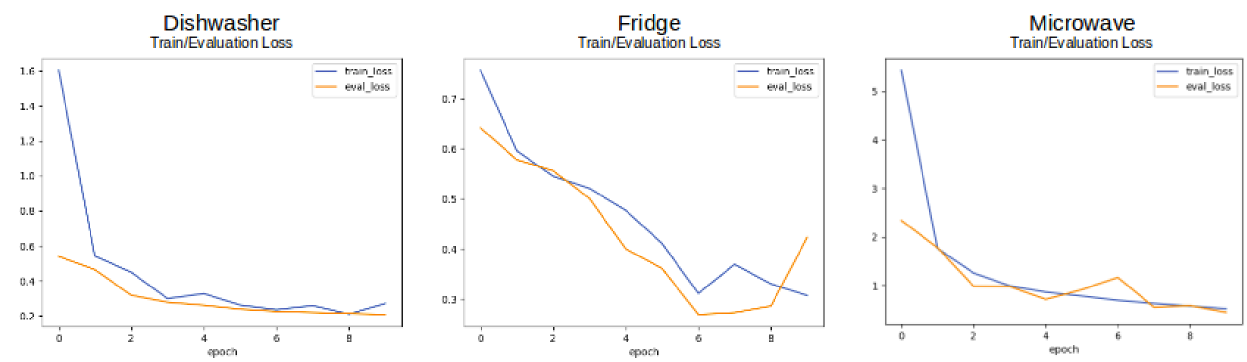


Figure 16: Train/Evaluation loss vs epoch curve

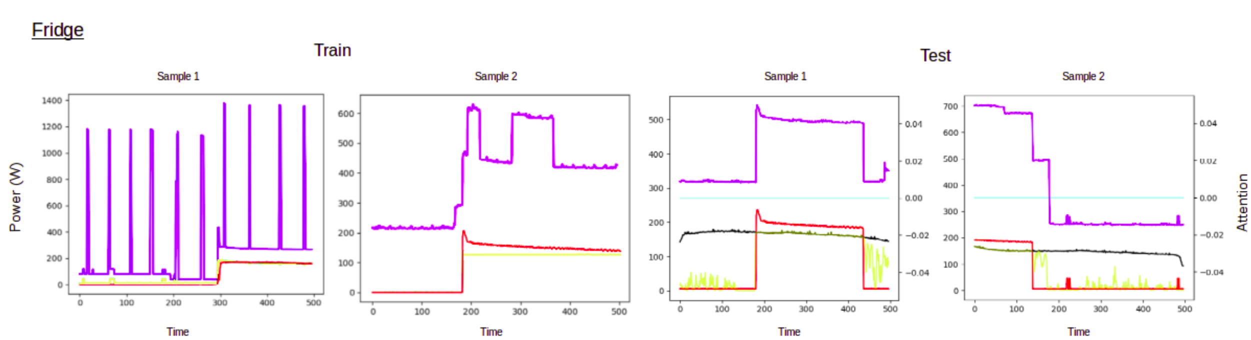


Figure 17: Train and test samples of fridge to evaluate disaggregation

Likewise for taking inference, we saved the model weights for the architecture across all appliances. We then predicted the energy consumption and calculated mean absolute error from the predicted sequence to compare the results

Table 7: Error Comparison (Part 2)

Appliance	MA E	Nº Epoch	Training Time (Hrs.)
Refrigerator	39.33	10	34
Dishwasher	18.22	10	25
Microwave	16.13	10	19
Lighting	52.69	10	32
Kitchen Outlets	32.44	10	21
Washer Dryer	22.15	10	27

3.3.7 Discussion:

As was described in the hypothesis the main goal of the regression branch is to predict the maximum expected demand of the appliance. As was also expected the classification branch is modulating the regression results to match the appliance load pattern. Classification has high specificity and low sensitivity.

In both cases, training and evaluation sets have good results but have less accurate results in test. Our hypothesis is that model does not generalize well due to the small number and variance of appliance patterns of the different train buildings.

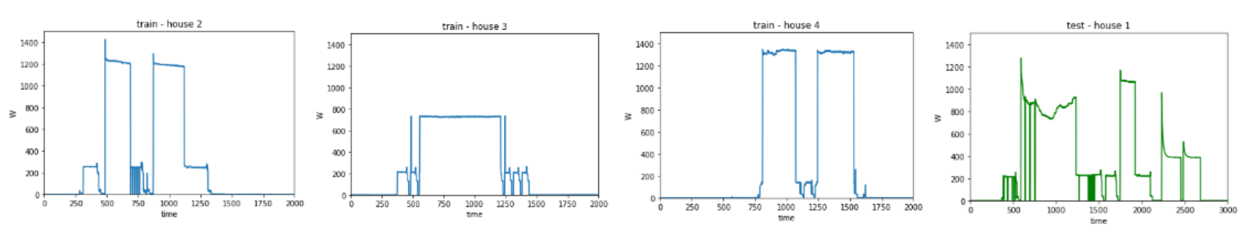


Figure 18: Samples of dishwasher consumption per building:

The classification network is in charge of modeling the patterns. As seen in the results, it is less accurate in the steady-state sections than expected. Hence, the instability, and in some cases, the highly sensitive response is also related to the overrepresentation issue.

In most cases, increasing the number of acquisition samples would not be a good solution to fix the instability issue as there would be more active windows but the same pattern. That's the case of appliances with components that do not depend on environmental factors (temperatures, etc.) like microwave or dishwasher. In the case of appliances with environmental factors, it would help to have also samples from different seasons. We implemented oversampling but it's similar to increasing the number of samples from the same appliance rather than new ones.

There's no more data available rather than the public dataset. As a solution, data augmentation cannot be easily implemented due to the lack of a database of appliance loads. In this case, it makes no sense to create synthetic aggregated scenarios mixing appliances from different buildings because they're already mixed in the training dataset and properly predicted in eval. In the classification branch, we hypothesize that in some cases adding noise would help to decrease highly sensitive responses.

Conclusion & Future Work

Classification branch is in charge of modeling the real consumption patterns of each window given. Regression branch is in charge of inferring the maximum consumption of the appliance in the input window. For Scenarios with a high simultaneity factor: Classification focus on State changes of appliances (switch on/off) and state duration. Therefore, adding classification sub-network improves the focus on the appliance pattern in case of the microwave and dishwasher. For scenarios with low simultaneity factor: classification focuses on the neighborhood, outside of the active section of the appliance to capture the peaks and valleys of load signature. Without a classification branch, the output is smoother and therefore it does not capture adequately the peaks of consumption. This is because the regression branch is not prepared to do both tasks of inferring the maximum consumption and adapting to the exact pattern with instantaneous changes of power. Our models took a big amount of time to be trained, the amount of data, the complexity of the forward and backward processes and the computational resources were the reason for that.

References

[1] R. Hassan and G. Radman, “Survey on smart grid,” in *Conference Proceedings - IEEE SOUTHEASTCON*, 2010, pp. 210–213. doi: 10.1109/SECON.2010.5453886.

[2] I. Vakulenko, L. Saher, O. Lyulyov, and T. Pimonenko, “A systematic literature review of smart grids,” in *E3S Web of Conferences*, Apr. 2021, vol. 250. doi: 10.1051/e3sconf/202125008006.

[3] M. ÇINAR and A. KAYGUSUZ, “Self-Healing In Smart Grid: A Review,” *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 7, no. 2, pp. 492–503, Dec. 2018, doi: 10.17798/bitlisfen.460164.

[4] X. Fang, S. Misra, G. Xue, and D. Yang, “Smart Grid-The New and Improved Power Grid: A Survey.”

[5] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” Feb. 2018, doi: 10.1109/TSG.2018.2818167.

[6] “Ekanayake, Janaka B., et al. Smart grid: technology and applications. John Wiley & Sons, 2012.”.

[7] “Nouredine Hadjsaïd and Jean-Claude Sabonnadière, Smart Grids, ISTE Ltd.,UK and John Wiley &Sons, Inc., USA, 2012.”.

[8] M. N. Meziane, A. Hacine-Gharbi, P. Ravier, G. Lamarque, J. C. le Bunetel, and Y. Raingeaud, “Electrical appliances identification and clustering using novel turn-on transient features,” in *ICPRAM 2017 - Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, 2017, vol. 2017-January, pp. 647–654. doi: 10.5220/0006245706470654.

[9] T. Logenthiran, D. Srinivasan, and T. Z. Shun, “Demand side management in smart grid using heuristic optimization,” *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1244–1252, 2012, doi: 10.1109/TSG.2012.2195686.

[10] D. R. Costianu, N. Arghira, I. Făgărăşan, and S. S. Iliescu, “A SURVEY ON POWER SYSTEM PROTECTION IN SMART GRIDS,” 2012.

[11] Z. Fan *et al.*, “Smart Grid Communications: Overview of Research Challenges, Solutions, and Standardization Activities,” Dec. 2011, doi: 10.1109/SURV.2011.122211.00021.

[12] C. S. Kudarihal, “Copy of 8. Electrical - IJEEER -SMART METERING AND - C S KUDARIHAL - OPaid _1_,” 2014. [Online]. Available: www.tjprc.org

[13] T. Baumeister, “Literature Review on Smart Grid Cyber Security,” 2010.

[14] A. Leonardi, K. Mathioudakis, A. Wiesmaier, and F. Zeiger, “Towards the Smart Grid: Substation Automation Architecture and Technologies,” *Advances in Electrical Engineering*, vol. 2014, pp. 1–13, Aug. 2014, doi: 10.1155/2014/896296.

[15] P. Y. Chen, S. M. Cheng, and K. C. Chen, “Smart attacks in smart grid communication networks,” *IEEE Communications Magazine*, vol. 50, no. 8, pp. 24–29, 2012, doi: 10.1109/MCOM.2012.6257523.

[16] W. Wang, Y. Xu, and M. Khanna, “A survey on the communication architectures in smart grid,” *Computer Networks*, vol. 55, no. 15. Elsevier B.V., pp. 3604–3629, Oct. 27, 2011. doi: 10.1016/j.comnet.2011.07.010.

[17] “Power Electronics Technology that Supports Smart Grid.”

[18] S. Aggarwal and N. Kumar, “Smart grid☆,” *Advances in Computers*, vol. 121, pp. 455–481, Jan. 2021, doi: 10.1016/BS.ADCOM.2020.08.023.

[19] G. Dileep, “A survey on smart grid technologies and applications,” *Renewable Energy*, vol. 146, pp. 2589–2625, Feb. 2020, doi: 10.1016/j.renene.2019.08.092.

[20] K. T. Ponds, A. Arefi, A. Sayigh, and G. Ledwich, “Aggregator of demand response for renewable integration and customer engagement: Strengths, weaknesses, opportunities, and threats,” *Energies (Basel)*, vol. 11, no. 9, Sep. 2018, doi: 10.3390/en11092391.

[21] B. Alohali, K. Kifayat, Q. Shi, and W. Hurst, “Group authentication scheme for neighbourhood area networks (NANs) in Smart Grids,” *Journal of Sensor and Actuator Networks*, vol. 5, no. 2, Jun. 2016, doi: 10.3390/jsan5020009.

[22] D. Baimel, S. Tapuchi, and N. Baimel, “Smart grid communication technologies-overview, research challenges and opportunities,” in *2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2016*, Jul. 2016, pp. 116–120. doi: 10.1109/SPEEDAM.2016.7526014.

[23] J. S. Zieliński, “Internet wszystkiego w sieci inteligentnej,” *Przeglad Elektrotechniczny*, vol. 91, no. 3, pp. 157–159, Jan. 2015, doi: 10.15199/48.2015.03.37.

[24] L. Gelazanskas and K. A. A. Gamage, “Demand side management in smart grid: A review and proposals for future direction,” *Sustainable Cities and Society*, vol. 11. pp. 22–30, Feb. 2014. doi: 10.1016/j.scs.2013.11.001.

[25] V. M. Laitsos, D. Bargiotas, A. Daskalopulu, A. I. Arvanitidis, and L. H. Tsoukalas, "An incentive-based implementation of demand side management in power systems," *Energies (Basel)*, vol. 14, no. 23, Dec. 2021, doi: 10.3390/en14237994.

[26] V. M. Laitsos, D. Bargiotas, A. Daskalopulu, A. I. Arvanitidis, and L. H. Tsoukalas, "An incentive-based implementation of demand side management in power systems," *Energies (Basel)*, vol. 14, no. 23, Dec. 2021, doi: 10.3390/en14237994.

[27] S. Tomkins, J. Pujara, and L. Getoor, "Disambiguating Energy Disaggregation: A Collective Probabilistic Approach," 2017.

[28] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," *Energy and Buildings*, vol. 30, no. 3, pp. 245–259, 1999, doi: 10.1016/S0378-7788(99)00007-9.

[29] I. Kaur, "Metering architecture of smart grid," *Design, Analysis, and Applications of Renewable Energy Systems*, pp. 687–704, 2021, doi: 10.1016/B978-0-12-824555-2.00030-7.

[30] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry, "A Dynamical Systems Approach to Energy Disaggregation," 2013.

[31] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-Intrusive Energy Disaggregation Using Non-Negative Matrix Factorization with Sum-to-k Constraint," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4430–4441, Nov. 2017, doi: 10.1109/TPWRS.2017.2660246.

[32] B. Najafi, S. Moaveninejad, and F. Rinaldi, "Data Analytics for Energy Disaggregation: Methods and Applications," in *Big Data Application in Power Systems*, Elsevier, 2018, pp. 377–408. doi: 10.1016/B978-0-12-811968-6.00017-6.

[33] J. P. Chavat, J. Graneri, and S. Nesmachnow, "Nonintrusive energy disaggregation by detecting similarities in consumption patterns," *Revista Facultad de Ingenieria*, no. 95, Apr. 2020, doi: 10.17533/UDEA.REDIN.20200370.

[34] M. Mejari, V. v Naik, D. Piga, and A. Bemporad, *Energy Disaggregation using Piecewise Affine Regression and Binary Quadratic Programming*.