UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in Engineering and Applied Sciences

XXXV Cycle

SSD: ING-INF/01 Electronics

DOCTORAL THESIS

Sensor-Based Systems, Data Acquisition, Analysis, and Prediction Algorithms for the Energy Data of Smart Homes

Supervisor:
Prof. Valerio Re

Co-Supervisor:
Sabrina Romano

Author:
Asad Hussain
Student ID: 1067642

ACADEMIC YEAR 2021 / 2022

Contents

List of Figures

List of Tables

Abstract

The residential sector contributes around 30-40% of global energy use. Managing and understanding home electricity consumption behavior is a vital and open challenge for many researchers. Many initiatives and research have been undertaken to reduce energy use and electric load on individual households. Among the different energy conservation measures, changing energy usage behavior has received significant attention since it significantly impacts daily energy use without building alterations. The Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA) is also working to achieve similar goals mainly focusing on energy efficiency, climate and environment improvement, and renewable energy sources.

The work presented in this thesis is part of a joint research initiative between the University of Bergamo and ENEA. There were two primary research objectives. The first goal was data aggregation from sensors and integration with DHOMUS (an ENEA storage platform). This objective has been achieved by developing a low-power modular platform and an IoT application for environmental parameter acquisition using the sensing device and integrating it with the DHOMUS platform. The second goal was the energy consumption disaggregation of the acquired data in a way that it could be used to facilitate the consumers to understand the energy consumption behavior. It is achieved by developing an energy consumption disaggregation method, in addition, we also devised appliance-level statistical analysis based on feature extractions. Furthermore, we developed hourly energy consumption forecasting based on machine learning algorithms.

To evaluate the effectiveness of the developed approaches, data acquired from DHOMUS with two different time-span datasets have been used. The results indicate that the developed solution proved effective in enhancing the energy knowledge of users in an efficient manner which will eventually contribute toward energy consumption savings.

Chapter 1

Introduction

1.1 Background

Sustainable energy systems must ensure sustainable development by providing affordable and uninterruptible energy to consumers. Hence, knowledge and understanding of energy consumption in the residential sector are indispensable for energy preservation and energy efficiency, which can only be possible with the help of consumer participation. Major economic expansions, industrialization, and population growth have resulted in escalated energy consumption and environmental deterioration, posing a threat to long-term development. Total domestic energy consumption is estimated to account for 30-40% of worldwide electricity generation [1], which is expected to rise as more appliances and electronic devices are utilized. Moreover, buildings contribute to one-third of worldwide final CO2 emissions; therefore, lowering consumption and emissions in this sector is necessary. Thus, energy efficiency has become one of the most critical challenges. According to statistics, immersing the customer in energy conservation and providing awareness of the peculiarities of energy use is beneficial and makes a considerable difference in energy waste reduction [2].

Consumers should be given individual load-level consumption information rather than aggregated consumption information. By raising customer knowledge of personal appliance use, revealing individual appliance-level energy information can reduce energy consumption [3]. A home energy management system (HEMS) provides a practical solution by enabling long-term, innovative energy-saving applications. HEMS can give consumers visual feedback through energy use statistics, utility-initiated automation and control, load forecasting, and an optimum load scheduling horizon. With smart meters and sensors that can measure the real-time power consumption of each appliance, monitoring power consumption and supporting HEMS applications have become easier with the help of signal processing methods [4]. The Internet of Things (IoT) adds a new degree of machine-to-machine communication to the interactions between humans and applications. The advancement of new communication technologies and intelligent devices that function by parsing signals has increased the ability to manage different equipment. Smart Home (SH) devices can store data, respond to users' prompts, provide feedback to users, and issue alarms. The synergy of these two fields, termed SH-IoT, promises to create an energy-optimized environment by connecting devices and giving flexibility in home administration and monitoring [5].

The Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA) is working on different research goals, primarily in energy efficiency, climate and the environment, and renewable energy sources. In terms of research on energy efficiency, ENEA developed a smart home network project. Further, a smart home platform called DHOMUS, an acronym for Data HOMes and Users, is a cloud part of the network where data from smart homes is collected in a database. Moreover, different energy consumption services needed to be integrated into the DHOMUS platform.

The work in this thesis is part of a project that the University of Bergamo and ENEA are working on together. There were two main research targets. The first goal was data aggregation from sensors and then integrating that with DHOMUS for data visualization. The second was to build energy consumption services based on user feedback and collected data so that it could be used in optimizing the customers' energy management.

1.2 Research Problem

Energy consumption analytics-based energy conservation has been the most intriguing research area for the researcher for the last two decades. Much existing literature provides better analytic using various methods, and more need to be explored since the consumer product market is evolving very fast. Also, machine learning techniques advancements have revolutionized how data is analyzed to meet a specific requirement.

Data acquisition and energy consumption-related services are two significant parts of this research. A low-power modular platform is required to be developed to effectively collect ambient factors for the first part of environmental condition monitoring. Further, integrating this hardware platform with the DHOMUS for data storage and visualization is required.

For the second target, energy consumption-based services must be designed for residential customers to make them aware of their" energy data" and understand how much energy they use and for what purposes. It also aims to help them marginalize electricity consumption and costs, thus reducing the environmental impact and making the residential customer an active subject contributing to the stability of the national electricity grid.

Because new sensors are coming out on the market and new ways to analyze data are being made, the consumer scene is dynamic and changing quickly. Therefore, this DHOMUS platform must be constantly updated to meet these new challenges. New data acquisition procedures and practices must be devised with better data analytics techniques, such as overall home energy consumption disaggregation, appliance-level energy use pattern analytics, and energy demand forecasting.

1.3 Research Objective

This research comprises four key objectives:

i. To develop a low power modular platform for wearable and IoT applications.

ii. To devise an approach that can disaggregate whole house energy consumption into specific energy sectors for feedback to smart homes users in DHOMUS platform.

iii. To develop a method to extract energy used patterns at the appliance level in order to facilitate the consumers by providing feedback in the DHOMUS platform.

iv. A comparative analysis of predictive models to identify a suitable model for the smart home platform.

1.4 Philosophy of the Work

To save energy, it is essential to measure electricity consumption and provide an accurate and more detailed insight into the behavior of consumers and service providers. Therefore, it is necessary to develop models by direct involvement of the consumers by providing them feedback at the appliances level rather than only overall energy consumption.

1.5 Contribution of Research

The main contributions of our research work are summarised as follows:

i. A detailed and comprehensive review of smart homes is presented to identify the different innovative home services and technologies.

ii. A low power modular platform for wearable and IoT applications is developed. This contribution is included in chapter 4 of this dissertation.

iii. Data-driven disaggregation method for electricity-based energy consumption for smart homes platform is proposed as the second contribution. This contribution is included in chapter 5 of this dissertation.

iv. This contribution is the energy consumption patterns extraction technique for household appliances for the DHOMUS platform. Chapter 6 of this dissertation is based on this contribution.

v. To achieve the last objective, we performed a comparative analysis of machine learning and deep learning methods in order to identify a suitable forecasting model for home energy consumption forecasting.

1.6 Research Scope

This thesis mainly focuses on providing energy services to smart homes by developing approaches based on data acquired from sensors and questionnaires.

This dissertation covers a comprehensive review of smart homes and different services. Therefore, this thesis could be an initial point for new researchers to gain knowledge about smart homes. In addition, our work regarding methods development for energy consumption disaggregation, appliance level data analysis, machine learning, and deep learning-based energy consumption forecasting for real data is viable for future real world projects and studies.

1.7 Thesis Organization

The dissertation is organized into the following chapters:

Chapter 1 briefly introduces home energy consumption, research objectives, list of contributions, and research scope.

Chapter 2 discusses the Smart home's characteristics, services, technologies, and communication protocols.

Chapter 3 briefly introduces concepts of energy consumption disaggregation, appliances operating state, energy consumption patterns,fundamental components of time series, and forecasting methods. In the end a number of studies related to our work and the existing state-of-the-art limitations are discussed.

Chapter 4 presents the development of a low power modular platform for wearable and IoT applications for smart homes.

Chapter 5 discusses the data-driven disaggregation method for electricity-based energy consumption for smart homes.

An energy consumption patterns detecting technique for household appliances for smart homes platform is presented in chapter 6.

Chapter 7 provides a a comparative analysis of machine learning and deep learning methods in order to identify a suitable forecasting model for home energy consumption forecasting.

Chapter 8 concludes the thesis by reviewing the outcomes and discussing the scope of future study in this research area.

Chapter 2

Comprehensive Literature Review on Smart Home Services and Technologies

This chapter presents a literature review of smart home services and technologies. Some definitions, historical background knowledge, and conceptual explanation are initially given. Afterward, different smartness levels of smart homes, smart home appliances, sensors, and some primary services are discussed; at the end of the chapter, information about wireless communication protocols adopted by smart home platforms is discussed.

2.1 What is a Smart Home?

The smart home is an accommodation hosting many electrical and electronic devices that assist occupants in better living. Smart home appliances help users control and monitor all aspects with a single click. The following section will define smart home technologies with their historical, conceptual, characteristic, and hierarchical perspectives.

2.1.1 Smart Home Definition

There is no proper definition in the literature regarding smart homes researchers and technologists have their definitions and explanation of smart homes [6, 8, 9, 10, 11, 12, 13, 14, 15, 7], so using this freedom, we have compiled our definition to the best of our knowledge.

Definition: A basic home is called a smart home when we add sensors, actuators, wireless communication technologies, and software systems to existing infrastructure. As a result, it provides occupants with comfort and control with the following:

a. Monitoring of house environmental factors (Temperature, Humidity, Pressure, Air Quality, bio particle, Oxygen level)

b. Monitory and environmental benefits by energy consumption monitoring, visualization, forecasting, and efficiency using automatic control along with damage control by early anomaly detection.

c. Basic health care services.

2.1.2 Smart Home Technologies Historical Background

Smart home inventions have a more extended history than we may recognize or comprehend. The thought of accommodation that could be more comfortable and lavish living can be followed back to the 1890s and mid-1900s when wealthy individuals started utilizing the electrical power that made their homes automated for better and adequate living [15]. Figure 2.1 shows that Thomas Edison invented automated colored lighting for homes in early 1910; afterward, that was used to promote public advertising for New York Edison. Additionally, the Rural Electrification Administration in the United States effectively announced various advanced electric products during the 1930s for the electrification of rural farms. In 1956 a drive, “Live Better Electrically, “was set up by General Electric and Westinghouse with the idea that the homes would be awarded a gold medallion if the overall appliances could convert to electric power [16].

In 1965, Gordon Moore presented a law for integrated circuits, also known as Moore’s law. Which stated that the density of integrated circuits on a single chip would double yearly for the next decade; in the same study, Moore anticipated using IC in personal portable communication gadgets, vehicles, and computers. In

Figure 2.1: Smart homes evolution diagram from archive , Source:[16]

[17], the acknowledgment of Moore's law, alongside the arrival of the internet, has remarkably transformed how individuals used computers and telephones earlier. With the miniaturization of the electronics and IT revolution, the mechanical age's transformation into the information age has increased the use of PCs and telephones presented in the Majesco Disruption Model shown in Figure 2.2.

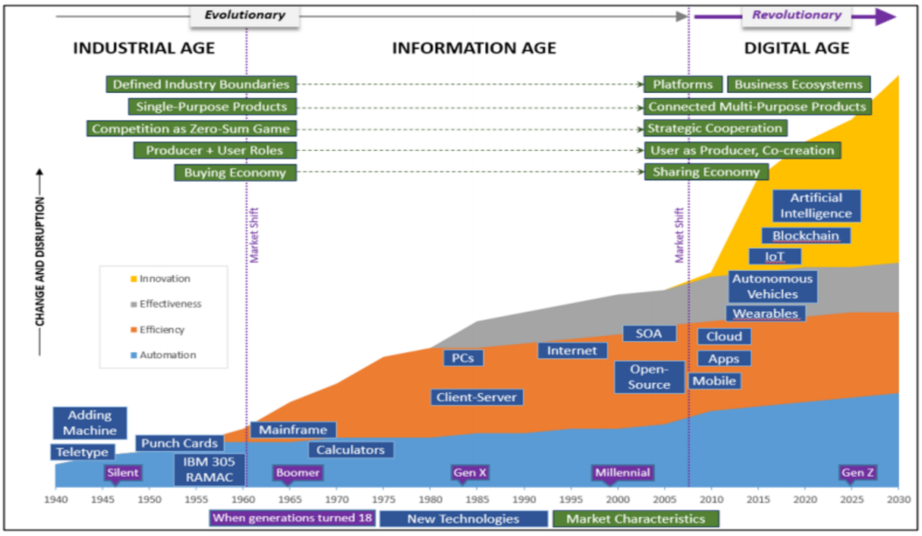


Figure 2.2: Majesco Disruption Model, Source:[17]

In the past seventy years, computers have evolved from research information processing machines to message processors to cutting-edge devices with limitless potential, from work to entertainment, visual computerization to gaming. The same was true with telephones. With the development of the cell phone, phones have evolved from simple devices for making calls and sending instant messages to small-scale PCs with great processing power that can govern every aspect of an individual's life. The continual change in size and performance of these devices has resulted in a high turnover of devices, with customers replacing their PCs and cell phones every couple of years on average. Since the last decade of the 20th century, smart houses have reemerged as fundamentals for automating and enhancing the enjoyment of homes by making them more energy-efficient and emitting less carbon dioxide.

2.1.3 Conceptual Explanation of Smart Home

Smart home technologies are highly in demand because they significantly impact society's evolution. Primarily fundamental questions arise: what is a smart home? And what is the purpose of a smart home? It can be answered from various angles by keeping the operational view in the discussion; Smart home technologies can assist the occupants by sensing information from the environment and acting on them accordingly, enhancing satisfaction and quality of life [16]. Subsequently, in light of the subservient view, a smart home is an optimally-managed building energy system allowing information and price-responsive adjustments to behavior and will give a monetary benefit. Lastly, the Socio-technical idea is that the technological revolution toward modernization swamps domestic life.

2.1.4 Smart Home Smartness Levels

The level of smartness is relative among various smart home technology systems, and we can segregate them into different levels for better understanding, as shown in Figure 2.3.
Level 0: Classical home or dumb home can be a house with no technologies installed in it, and this type of home can be found in some remote areas far from the cities.
Level 1: The house has an isolated level of smartness that can be graded as a level one smart home. Since it contains a few essential technology-based devices, such as radio, television or solar system, etc., users need to handle these technologies manually and are not interconnected.
Level 2: Home begins to use technologies grouped and coordinated in a way to invoke better services for smart the home that is heat (can be a smart meter along with heat pump and thermometer for display purpose) or entertainment (maybe a TV combined with a web switch, sound system, PC, and cell phone all integrated).
Level 3: The level three home obtained slightly more significant automation, with

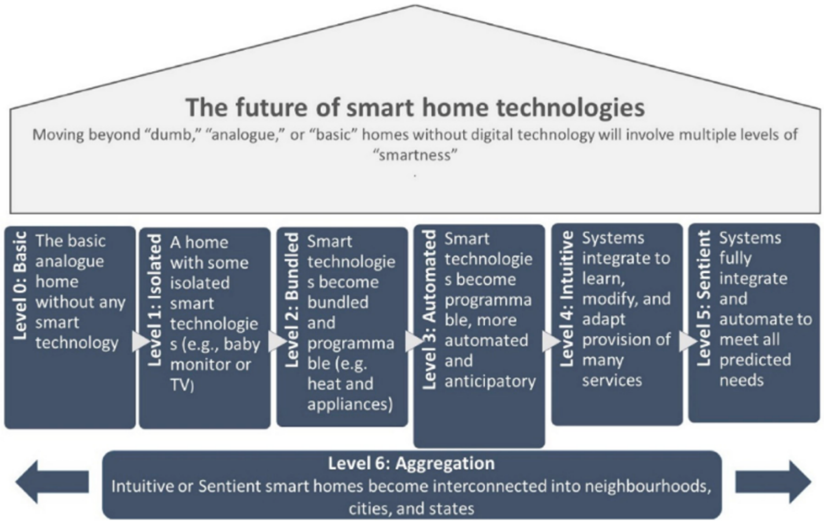


Figure 2.3: Levels of smartness with smart home technologies, Source:[16]

frameworks starting to interconnect and even predict specific necessities, for example, turning on lights and other household gadgets a couple of seconds advance an inhabitant gets back. A level three house can likewise be modified to meet particular inclinations over numerous devices, incorporating distinctive temperatures in various rooms.

Level 4: A level 4 smart home system starts to learn for itself and adjust its arrangement of services to set, for example, turning the lights on if a tempest is coming or turning them off when the sun comes out. At this level, sensors and screens can empower innovation to know the state of the home, and feedback can encourage some adapting. Hence, it becomes increasingly independent and can adjust to what it thinks you need.

Level 5: A level 5 home turns out to be practically aware and can consequently meet and even foresee all house needs. After the evolution at this most excellent level, smart homes can autonomously provide additional seating, portability, cultivation, and lighting). Homes at this level would most likely start conversing with occupants. This kind of home can be called an artificially intelligent home or a fully smart home.

Level 6: Level, six of the smart homes, is the aggregation process of 5th level homes to make society and cities bright where a smart home is connected by different complex networks, which could be the future of smart homes.

2.2 Smart Home Services

A smart home system consists of different services which may not be unique. The famous services grouped are assistive and management services, as shown in Figure 2.4.

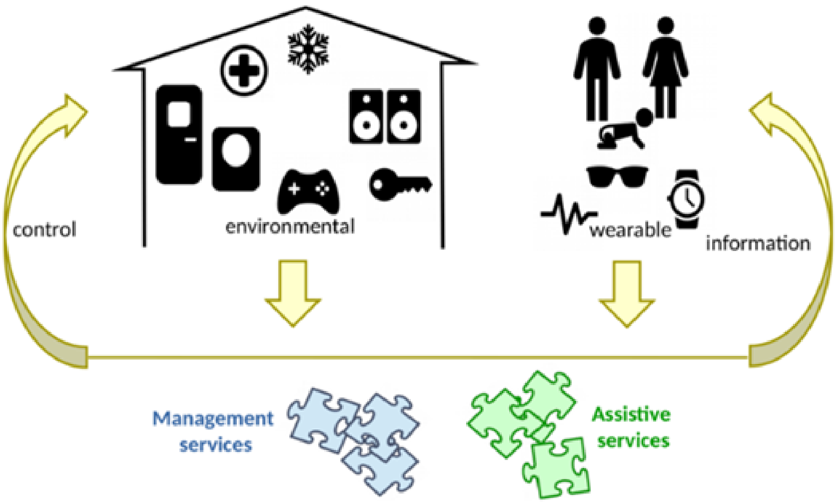


Figure 2.4: Components of smart Home Services, Source:[18]

2.2.1 Smart Home Assistive Services

This service is fundamental for any smart home, which is added to comfort the occupants' needs, mainly the routine actions and activities performed in the house [19]. For example, suppose an occupant does not like the noise of the washing machine while listening to music or watching Tv. In that case, the smart home assistive service can quickly help the person by simply turning off the washing machine and adjusting the timing accordingly. Assistive service may likewise be customized to some special needs of the user as they desire. Furthermore, assistive services can be categorized into two classes: ambient service and health service.

2.2.1.1 Smart Home Indoor Environmental Services

Better indoor environmental conditions are essential because it is directly related to the health of the occupants of the home or building. The key features of the indoor environment are temperature, humidity, and a smart indoor environment service that can efficiently control air quality [20, 21]. For example, if the temperature rise in a room exceeds a specific limit, a fan or Ac can be automatically turned on. However, if the air quality is not good, a window can be opened, or an air purifier can be used in this case. Furthermore, natural light can also be a factor in the indoor environment, and window blinds can also be controlled under this service.

2.2.1.2 Smart Home Health Services

Smart home health service (SHHS) is the most important service among smart home services. This becomes more critical when the inhabitants of the home are aging as they need more care and to do such a large number of skilled personnel are required. World Health Organization is working to improve the quality of life of aging people [22]. Besides this, with the evolution in health technology, smart home health services are also progressing to cater the health issues. In [23], it is described that the health of an individual is directly related to their habits and behaviors during the day, and the use of health care services can reasonably manage it. However, this service can help disabled occupants sit at home for a long time independently, providing peace of mind for other family members. Telemedicine is one of the essential features of SHHS, which remotely provides medical services to patients at their doorstep [24]. Moreover, the patient can have video consultations with doctors for medical advice, remote diagnostics and some minor treatments, and medical image sharing. The prime goal of SHHS is that an occupant can measure temperature, pulse, and blood pressure remotely and send it to the specialist and get advice without any delay.

2.2.2 Smart Home Management Services

Smart Home Management Services (SHMS) handle explicit functionalities of the home. Safety and security are the main features of SHMS, which is responsible for events like intrusion or gas leakage detection, and fire alarms [25]. Beside safety and security, one most important SHMS is energy consumption efficiency [26]. Energy consumption efficiency in SHMS can be achieved by monitoring and controlling windows and solar panels' orientation according to the sun. Furthermore, automatic control of home appliances according to peak hours can also reduce energy consumption.

2.2.2.1 Smart Home Safety and Security Services

Smart Home Safety and Security Services (SHSSS) are essential for any smart home. Therefore, the few SHSSS scenarios are if the occupant leaves home, the doors, windows, and shutters can be closed, and all water taps are closed correctly. Furthermore, if something is not addressed properly, this service can alert the user and notify a possible safety warning [27]. SHSSS is more helpful for elderly occupants who feel safer and secure using these services. In addition, remote access control is a crucial service feature for smart home residents. Moreover, a smoke detector is almost used in every SHSSS near the kitchen because this is the most venerable home area and can generate alter for the user [28].

2.2.2.2 Smart Home Energy Services

Smart Home Energy Services (SHES) have importance beyond a smart home and its occupant but are equally beneficial for an ordinary user. Primarily energy consumption and its monitoring is every one demand, and most smart homes are equipped with energy meters for this purpose. Energy is a significant resource that needs to be managed and tracked well for optimal use; as far as a smart home is concerned, the indoor environment is key to the consumption of energy [29]. Furthermore, SHES is also helpful in changing the energy consumption behavior of the residents for energy saving [30].

2.3 Smart Home Technologies

Smart Home Technologies covers the vast knowledge of modern ear technologies, which is the backbone of smart home systems. All smart home services are implemented with the collaboration of modern technologies, which are electrical appliances and electronic devices, sensors, actuators, communication technologies, and software applications, as explained in subsections.

2.3.1 Smart Home Devices and Appliances

Over the years, smart home devices and appliances are famous among users, giving them almost every service discussed earlier. Home appliances are televisions, washing machines, and vacuum cleaners that are controllable with specific smartphone applications. Whereas IoT devices are those devices that connect these home appliances to the internet to make them smart, such devices are LED bulbs and power sockets [27]. Table 2.1 illustrates some most common devices along with their functionalities. Smart thermostats need to control the temperature of the HVAC system with additional capabilities of adaptation with varying parameters. Smart lights comprise simple lighting with embedded modules to manage their functionalities to user needs. Smart plugs lie between power plugs and appliances that consume energy for monitoring energy consumption. Furthermore, smart hubs combine several smart devices within the smart home system.

Table 2.1: Smart Devices

Smart Devices	Functionality	Service
Thermostats	Temperature control	Energy, Environmental
Smart Lights	Adaptive lighting	Energy, Environmental
Plugs	Energy consumption record	Energy
Hubs	Connection and integration to Smart Home System	Energy, Environmental
Smart Water Heaters	Transformation of old heaters with advanced functionalities.	Energy, Environmental
Dishwasher, Washing Machine, Refrigerator, Cooker	Daily routine work	Assistive living and comfort
TV,radio,Home,cinema	Leisure	Comfort
Wearable devices	Remote therapy	Health

2.3.2 Sensors for Smart Home

The indoor environment of a home or building is important for the resident's health and comfort, as well as monitoring of electrical energy consumption for monitory and controlling of environmental damaging factors. That can only be possible if we have correct information about all contributing parameters. This is only possible when we have good sources or devices for collecting data. So, many sensors available for smart homes will give the occupants great comfort and satisfaction, making their lives healthy and productive. Furthermore, for more detail, we can divide sensors into many categories such as occupancy sensors, indoor ambient parameters sensors, smart switches, smart plugs, energy meters, and behavioral sensors. A few well-known sensors are given in the next subsections.

2.3.2.1 Image based Senors

Image-based sensors are a combination of three cameras that is luminance camera, a visible light camera, and an infrared camera. This sensor works with electromagnetic radiation, and the captured information is stored in the matrix [31].

2.3.2.2 Motion Sensors

Motion sensors are of different types. The most common indoor occupancy presence types are passive infrared (PIR), photo sensors, Ultrasonic sensors, and microwave dopplers [32].

2.3.2.3 Threshold and mechanical sensors

These sensors give us the number count of people within the building, office, or house when somebody interacts with a door or window [32, 33]. A list of mechanical and threshold sensors is available. The well-known sensors are infrared beams (IR), piezoelectric mats, door badges, and reed contacts. Reed contacts sensor is mainly used to detect the opening and shutting of windows and doors; it is an inefficient energy sensor.

2.3.2.4 Temperature and humidity sensor

Temperature and humidity sensors are environmental sensors used to collect the indoor climate information of a home or a building. These sensors have a wide variety in the market with different specifications and costs. However, the precision of the read values is essential, and it may affect the cost of the sensor with precision l +/- 0.1 K counted as good [34] while purchasing.

2.3.2.5 Photometric sensor

A photometric sensor is also known as a light sensor used to fix the light intensity based on sunlight inside the smart home [35]. However, the sensor's working is subject to luminance level and placement place. This sensor controls our lighting system to save much energy.

2.3.2.6 CO_2 sensor

The CO_2 sensor is a dual-function sensor that can be used as the number count of people inside the home by correlating the CO_2 emission with the number of people present[36]. The second use of the sensor can be as an environmental sensor to find the quality of the home environment.

2.3.2.7 Particulate matter (PM) sensor

A particulate matter sensor is an environmental sensor that senses particles in less than 2µ of indoor air. In [37], the authors presented that feeling low concentration pollutants is not an easy task for the sensor, so few PM sensors use piezoelectrical crystals to measure dust particles.

2.4 IoT Communication Protocols for Smart Home

This section presents the main wireless communication protocols for smart home environments. The Internet of Things (IoT) brings enhanced efficiency and convenience to practically every aspect of the urban world. Communication protocols for IoT-enabled smart homes are essential for maximizing the use of these most recent advancements.

2.4.1 IoT Physical Layer Network Protocols

The network protocols of IoT have two main categories: data communication from devices to the gateway and the other is from the gateway to the cloud, or called an application-level protocol. However, few protocols lie in both categories. The explanation of a few of the protocols is described in the proceeding subsections.

2.4.1.1 Bluetooth and BLE

Bluetooth is a 2.4GHz wireless network for personal communication. Network providers choose the 2.4GHz network for delivering private networks since it is less expensive and has a more excellent range than other networks. Bluetooth low energy (BLE) is a new and improved form of Bluetooth that links IoT devices. For communication, BLE uses less power than regular Bluetooth. Electronic devices that may operate as a data transfer hub between IoT devices and the cloud are often employed with BLE-enabled devices. As a result, BLE is an excellent choice for IoT wearables [38]. BLE is frequently used in fitness and health trackers and smart home products such as door locks. Smartphones can quickly receive data from BLE-enabled IoT wearables. BLE can be used with beacon technology to provide customer assistance, such as in-store navigation in the retail setting. Beacons are miniature transmitters that use Bluetooth Low Energy (BLE) to send signals to neighboring IoT devices. Beacons can make location-based searching and navigation considerably more accessible and more precise by emitting signals to adjacent IoT devices.

2.4.1.2 Low Power Wide Area Networks

LPWANs (Low Power Wide Area Networks) are new protocols developed for IoT applications, as shown in Figure 2.5. Still, they can also be used by other devices to communicate over long distances. Cellular networks can provide a wide-area communication network, but cellular networks have a high communication cost due to their high-power consumption [39]. LPWANs provide communications over a broad area using small, low-cost batteries that can endure prolonged, making them a cost-effective alternative to cellular networks.

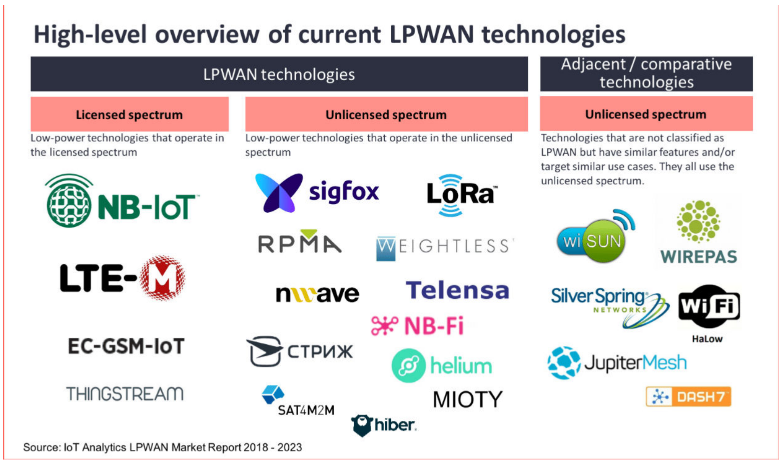


Figure 2.5: LPWANs Technologies

Different varieties of licensed (NB-IoT, LTE-M) and unlicensed (MIOTY, LoRa) LPWANs exist, each with its features. While power consumption is a significant issue for licensed LPWANs, unlicensed LPWANs encounter issues such as quality of service (QoS) and scalability.

LPWANs, in general, may connect practically any sort of sensor and enable data sharing between them and the cloud. IoT sensors can benefit from LPWANs for a variety of applications. Sensors, for example, can allow for remote monitoring of anything. LPWANs, on the other hand, can only transport tiny blocks of data over the network at a time and cannot send massive amounts of data.

2.4.1.3 Mesh protocols

A mesh is a densely connected network of devices made up of devices arranged in a mesh topology. Mesh topology is a networking infrastructure that allows all connected devices to work together to move and share data.

One of the most widely utilized mesh protocols for IoT applications is ZigBee. It's a low-power, short-range protocol often used to connect many Internet of Things devices. Compared to LPWANs, ZigBee allows to LPWANs, ZigBee allows for substantial data transfers in a single instance but at a lower power efficiency due to the mesh topology [40].

Z-Wave is an interoperable, wireless, RF-based communications technology developed for control, monitoring, and status reading applications in residential and light business contexts. It is a low-Power RF communications technique that provides complete mesh networks without requiring a coordinator node.

2.4.1.4 WiFi/WiFi HaLOW

Due to its pervasiveness in both industrial and residential settings, everyone would know what WiFi is. The majority of IoT devices, however, do not use WiFi. WiFi, except a few applications such as digital signage and security cameras, is not a viable alternative for IoT connectivity [41]. The WiFi network's application in IoT devices is limited owing to its short-range, high-power consumption, and lack of scalability. WiFi HaLow, a lesser-known WiFi variant, is now available for IoT devices. WiFi HaLow extends the range of WiFi and reduces battery consumption. Industries, on the other hand, have been less supportive of WiFi HaLow due to the network's lack of security.

2.4.1.5 RFID

RFID (radio-frequency identification) is a technology that uses radio waves to send data packets across a network in a small region. In IoT devices, embedding an

RFID chip is simple. RFID readers can then read the tags and provide information about the products [42]. Inventory management is one of the most prevalent RFID applications. Businesses can keep track of the number of products in stock by adding RFID tags to all products and connecting them to IoT devices. As a result, RFID can aid in better stock planning, resulting in improved supply chain management. Smart home IoT devices can also benefit from RFID tags. For example, a smart washing machine that reads RFID tags could be controlled.

2.4.2 IOT Application Layer Network Protocols

2.4.2.1 Constrained Application Protocol (CoAP)

While the existing Internet infrastructure is free and open to every IoT device, it is typically too heavy and power-hungry for most IoT applications. Constrained Application Protocol (CoAP), developed by the IETF Constrained RESTful Contexts working group and released in 2013, was created to convert the HTTP model for use in restricted device and network environments.

CoAP is a secure communication protocol for HTTP-based IoT systems that uses the User Datagram Protocol (UDP) to provide secure communication between endpoints. UDP's broadcasting and multitasking capabilities allow it to send data to several hosts while maintaining communication speed and minimal bandwidth utilization. It is an excellent fit for wireless networks used in resource-constrained M2M scenarios. The RESTful design provides a request/response interaction model between application endpoints. Furthermore, CoAP uses the fundamental HTTP get, post, put, and delete methods, eliminating uncertainty during client interaction [43].

Quality of Service is a feature of CoAP that allows you to manage the messages you send and identify them as "confirmable" or "non confirmable," indicating whether the recipient should return an "ack" or not. CoAP also enables content negotiation and a resource discovery technique, both noteworthy features. CoAP uses Datagram Transport Layer Security (DTLS) for secure message exchange in the transport layer and sending IoT data. CoAP meets all the requirements for a very light protocol to suit the needs of battery-operated or low-energy devices. Overall, when it comes to existing web service-based IoT systems, CoAP is an excellent match.

2.4.2.2 Message Queuing Telemetry Transport (MQTT)

Message Queuing Telemetry Transport is a lightweight publication/subscription type (pub/sub) messaging protocol that is the most extensively accepted standard in the Industrial Internet of Things to date. MQTT's architecture is simple and lightweight, making it ideal for battery-powered devices with low power consumption. It is based on the TCP/IP protocol. It has been built specifically for unreliable communication networks to address the problem of the increasing number of small, low-power items that have appeared in the network in recent years [44].

The subscriber, publisher, and broker model is used in MQTT. The publisher's role in the model is to collect data and transmit it to subscribers via the mediation layer, which is the broker. On the other hand, the broker's responsibility is to assure security by double-checking publisher and subscriber authorization.

2.4.2.3 Extensible Messaging and Presence Protocol (XMPP)

This communication IoT protocol for message-oriented middleware was developed in 1999 by the open-source Jabber community and was originally intended for real-time communications. It is based on the XML language. It allows two or more network clients to exchange structured yet extendable data in real-time [45].

XMPP has been widely used as a communications technology since its debut. It has since been utilized in the Internet of Things, thanks to creating a lightweight XMPP specification called XMPP-IoT. It has an open community supported standard, and its addressing and scalability capabilities make it ideal for consumer-oriented IoT installations.

2.4.2.4 Data-Distribution Service (DDS)

The publish-subscribe paradigm guided the development of the DDS protocol. The DDS protocol for real-time M2M communication, developed by the Object Management Group (OMG), allows for scalable, reliable, high-performance, and interoperable data sharing between connected devices regardless of hardware or software platform. To deliver high-quality QoS and assure interoperability, DDS provides broker less architecture and multitasking [46].

The Data-Centric Publish-Subscribe layer (DCPS) and the optional Data-Local Reconstruction layer makes up the DDS protocol's architecture (DLRL). While the DCPS layer distributes data to subscribers in a resource-aware, scalable, and efficient manner, the DLRL provides an interface for DCPS functionality, transmitting data between IoT- connected items.

DDS is used in some Industrial Internet of Things installations, such as air traffic control, smart grid management, autonomous vehicles, transportation systems, robotics, power generation, and healthcare services, despite not being a conventional IoT solution. DDS can manage data interchange between lightweight devices and general interconnection of large, high-performance sensor networks high-performance sensor networks in general. It's also capable of sending and receiving data from the cloud.

2.4.2.5 Advanced Message Queuing Protocol (AMQP)

AMQP is an open standard publish/subscribe type protocol that started in the financial services sector in 2003. While it has gained some traction in information and communication technology, its application in the IoT market is still somewhat limited. Message orientation, queuing, routing (including point-to-point and publish-and-subscribe), dependability, and security are all described in the AMQP specification [47]. The most significant advantage of AMQP is its robust communication model. Although functional, AMQP can ensure complete transactions, which isn't always what IoT applications want. AMQP is not suitable for sensor devices with limited memory, power, or network bandwidth due to its heaviness; however, for specific IoT use cases, it may be the only protocol viable for end- to-end applications, such as heavy industrial machinery or SCADA systems, where the devices and networks are typically much more capable.

2.5 Discussion of the Review

The primary objective of this comprehensive literature review was to investigate existing smart home domain studies to identify smart home services and technologies used for the development of different smart home platforms. The literature review finding shows that smart home services are mainly grouped into two classes, the first one is smart home assistive services and the second one is smart home management services. The increasing demand for energy worldwide requires more efficient and scalable energy-related services, and smart home platforms are needed to meet the user requirement. Furthermore, this comprehensive literature review provides detailed information regarding technologies (sensors, appliances, wireless communications protocols) used in the literature to choose appropriate technologies for smart home platforms.

2.6 Summary

This chapter has provided a detailed and comprehensive review of smart homes to identify the different innovative home services and technologies available in the literature from both technical and theoretical points of view. The effort is made to summarize the literature on all aspects of a smart home. The evolution of the simple home to the smart home has been discussed by keeping the historical background in view. Afterward, how different homes vary from others based on smartness level; six smartness levels are defined with the help of state-of-the-art studies.

Different smart home services such as assistive service have been discussed following the smartness levels. Assistive service is composed of an indoor environment and health services. Moreover, smart home management services are summarized as a collection of safety security and, most importantly, Home energy consumption management services.

Smart home service can only be acquired by using technologies available for smart homes and buildings, and these technologies have rapid growth, and it's hard to summarize them at this stage. However, few smart home devices and appliances have been covered, along with fundamental sensors used in smart home technologies. Lastly, an effort has been made to summarize IoT communication protocols used in the domain of home area networks.

Chapter 3

Background Concepts and Related Work

This chapter discusses the fundamental concepts of energy consumption disaggregation methods, appliance use patterns, time series data, and prediction methods with different horizons. We start by discussing appliance load monitoring, followed by appliance operating state information and usage patterns. After that, we present the time series' fundamental components and the basic concept of forecasting , forecasting approaches, and forecasting horizons. Finally, we discuss the number of studies related to our work and the existing state-of-the-art limitations.

3.1 What is Energy disaggregation?

Energy disaggregation is a process of dividing the overall energy acquired from the energy meter or aggregated energy into individual devices (Tv, washing machine, dishwasher, etc.) or sectors (lighting, kitchen, cooling, heating, etc) that contribute to the total energy consumed.

In literature, energy disaggregation is also known as load disaggregation, which aims to calculate individual devices' power demand from the dwelling's overall electricity consumption [48].

3.2 Appliance Load Monitoring

Appliance load monitoring assesses individual appliances' electrical energy consumption and operating status by analyzing the aggregated load recorded by the building's main power meter. These can provide information to both the user and the utility, such as the kind of load, power consumption details, and the operational conditions of the equipment [49]. Moreover, different appliance load monitoring techniques are available in the literature with their various advantages two most prominent techniques will be discussed in this section.

3.2.1 Intrusive Load Monitoring (ILM)

The intrusive load monitoring system is a standard metering method that measures an appliance's energy use by mounting sensor to each device in the household. So, it implies entering the house, which is why the system is called intrusive [49, 50].

The intrusiveness level of ILM can be associated with the number of appliances connected to a single set of the sensor. In literature ILM is divided into three subsets; the first is a sub-metering system where one meter is assigned to a zone of appliances. The second uses smart plugs for each device and connects to one meter for monitoring, and the last uses smart appliances, each connected to a dedicated meter [50].

3.2.2 Non-Intrusive Load Monitoring (NILM)

Non-intrusive load monitoring does not require any intrusion into the house for power consumption monitoring of different appliances. NILM is an efficient method of estimating energy usage and the status of individual appliances based on examining the aggregate load recorded by the building's main power meter. NILM evaluates variations in voltage and current entering a building to determine which appliances are in operation and their energy usage [51] [52][53].

The NILM procedure is divided into data acquisition, feature extraction, and appliance classification [54, 55]. The data acquisition measures aggregated load usage (voltage current, etc.) sampled at a reasonable rate to detect the appliance pattern. Feature extraction is capturing or computing the appliance's signature from obtained data, which may include data processing and optimization [56]. The signatures gathered from the feature extraction module are used to perform a load classification algorithm.

3.3 Appliances Operating States and Use Patterns

Appliance operating state information and usage patterns can assist consumers in understanding how they utilize their equipment. Additionally, appliance usage patterns might be used to detect abnormal appliance usage or to develop an intelligent appliance control system [57].In the literature, different appliance states and patterns have been used. Some of them are discussed in this section.

3.3.1 ON-OFF State

On-off states correspond to a home appliance's two primary states: switched on and operational or turned off and not functional. This notion is fundamental to home appliances, from basic equipment like lights and fans to more complicated appliances like refrigerators and washing machines. Typically, the on-off state is controlled by a switch or button on the instrument itself or by a remote control or smart device linked to the machine [58].

3.3.2 Standby State

Standby mode is when an appliance is turned on but not actively in use. It is also known as "idle" or "sleep" mode. Even while in standby mode, the device consumes some energy to keep its components working, even if it is not doing any relevant functions. According to the United States Department of Energy, standby power use can account for up to 10% of a household's total energy consumption [59]. Most household appliances are embedded with standby mode; televisions, computers, printers, and gaming consoles are the most common. Few appliances, such as refrigerators and air conditioners, also have a standby mode that allows them to consume less energy when not actively cooling [60].

3.3.3 Energy Consumption Patterns

Appliance energy consumption patterns indicate how much energy household appliances utilize over a specific period. These energy consumption patterns might vary based on various factors, including the kind of appliance, its energy efficiency rating, the size of the appliance, and how frequently it is used [61]. Recognition of appliance energy usage patterns is critical for various reasons. Initially, it assists homeowners in determining which appliances spend the most energy and, as a result, where energy savings may be realized. Moreover, it can assist families in identifying devices that should be replaced with more energy-efficient variants. Finally, analyzing energy use patterns can assist people in lowering their energy expenditures, and total carbon emission [62].

3.3.3.1 Peak hours energy consumption Patterns

Peak hours energy consumption refers to the hours of the day when the power demand is high. These peak hours are usually in the early morning and late evening, when people get up and go home from work [63]. Knowing peak-hour energy consumption trends is critical for utility companies to correctly manage supply and demand and supply and demand, as well as for legislators to establish energy-efficient laws.

3.3.3.2 Long and Short energy consumption cycles

Washing machines and dishwashers are among the most energy-intensive household equipment, and their energy consumption varies based on the cycle used. The length of the washing cycle is one element that influences energy use. Longer cycles require more energy, and the customer must know about this information. Most washing machines provide a variety of cycle settings, such as standard, heavy-duty, and rapid wash cycles. Each cycle is tailored to a particular load and soil level, and its duration can range from 15 minutes to many hours. Longer cycles often demand more water, energy, and hotter water temperatures [64].

3.4 What is a Time Series?

A time series, according to mathematics, is a collection of data points that have been indexed (or listed or graphed) over time. These data points are often made up of successive measurements taken from the same source over time and are used to track change over time [65, 66]. Since time is an element of all that can be observed, time series data may be found just about anywhere. As more and more of the physical world is instrumented, sensors and other devices generate a never-ending stream of time series data. Weather data, annual rainfall measurements, electricity-based energy consumption, and heart rate monitoring are a few examples of time series data.

3.4.1 Components of a Time Series

Typically, time series comprises three components: trend, seasonality, and irregular components, often known as residuals [66].

3.4.1.1 Trend

A trend is a typical movement of a time series during the observation period, excluding seasonal and irregular fluctuations. This component is also termed long-term variation in some literature[66]. Although there are numerous trends in time series, linear, exponential, and parabolic trends are the most common.

3.4.1.2 Seasonality

This component highlights alterations that occur at regular intervals and may provide meaningful information when periods demonstrate similar patterns. It combines the effects with the time, amplitude, and direction in a relatively reliable manner [66].

3.4.1.3 Irregular components

Residual values remain after calculating and removing the trend and cyclic variations. Sometimes, these levels can be high enough to obscure the trend and seasonality. In this instance, these residuals are referred to as outliers, and robust statistics are typically employed to account for them. These variations may have numerous causes, rendering prediction nearly tricky. However, if this origin can be recognized or modeled, it might be viewed as a precursor of change in trends[66].

3.5 What is forecasting?

Forecasting refers to acquiring the knowledge of the future, immediate future, and long-term future as much as we can. We are most interested in forecasting because we want to quantify the uncertainty in our regular operations ahead of us. Additionally, the conceptual model is used in forecasting to anticipate or compute future occurrences or developments. This model might be informal, based on personal experiences and social, cultural, and economic aspects, or formal and complicated, incorporating numerical measurements, equations, simulations, and forecasts. Forecasting is critical in many industries, from finance and economics to weather and climate research. It entails studying existing data to assess the likelihood of future events [67].

3.6 Forecasting Horizons

The energy consumption forecasting horizon describes the period for which energy consumption is predicted. The forecasting horizon can differ depending on the energy consumption forecast's specific application and purpose. In literature energy consumption forecasting horizons are generally categorize in three horizons [68, 69].

- Short-Term Forecasting Horizon: Short-term energy consumption forecasting generally refers to forecasting for the next few hours, days, or weeks.

- Medium-Term Forecasting Horizon: Medium-term energy consumption forecasting mainly refers to forecasts for a few months or years into the future.

- Long-Term Forecasting Horizon: Long-term energy consumption forecasting typically refers to predictions for several years to several decades into the future.

The forecasting horizon can vary in length based on the unique demands and goals of the energy user or producer, as well as the availability and dependability of the data utilized for the prediction. It should be noted that estimating energy consumption is a complicated and difficult operation, and the accuracy of the prediction can be affected by a range of factors.

3.7 Forecasting Methods

There are two fundamental approaches used to make predictions about future events: qualitative and quantitative methods.

3.7.1 Qualitative method

The qualitative forecasting approach, also known as the judgemental method, includes producing forecasts based on subjective assessments given by experts or forecasters. Because these assessments are frequently dependent on personal knowledge, intuition, and experience, qualitative forecasting outcomes might be subjective and misleading. As a result, this technique is non-mathematical and does not rely on mathematical models. Despite its limits, qualitative forecasting can be useful in instances when data is few or unavailable, or where elements such as human behavior or political events are difficult to define [70].

3.7.2 Quantitative method

Quantitative forecasting is a systematic procedure that employs mathematical approaches to assure objectivity and consistency in its outcomes. This method eliminates the need for personal opinions and intuition to create predictions; instead, it uses massive amounts of data and numerical information [71]. Quantitative forecasting approaches include causal and time series methods which are discussed in this section.

3.7.2.1 Causal method

Causal forecasting obtains a forecast of the quantity of interest by relating it directly to one or more other quantities that drive the quantity of interest.This forecasting method is used when the cause and effects are known for the prediction [72].

3.7.2.2 Time-Series method

Time-series analysis is a statistical approach for analyzing and extracting useful information from time-series data. Time-series data is a set of observations gathered at regular time intervals, such as hourly, daily, weekly, monthly, or yearly. There are many different methods that can be used for time-series analysis, such as: descriptive statistics, trend analysis, seasonal analysis, moving averages and neural networks [73].

3.8 Related Work

The scientific community has been actively participating in the development of intelligent sensor-based systems/home energy management systems to improve the data acquisition and energy efficiency of smart buildings. Here we have listed the most relevant works that propose hardware platforms and methods in the specific context of smart homes. The state-of-the-art covers primary research areas: (i) wearable environmental monitoring systems, (ii) energy consumption disaggregation for home energy management systems, (iii) appliance energy use patterns methods, and (iv) forecasting models for buildings energy consumption.

Literature related to objective one: Here we presented some notable studies related to our first research objective and identified the research gap, The study [74] describes a wearable environmental monitoring system (WEMS) design aimed at healthcare applications. The system consists of many hardware components encapsulated by a wearing bracelet. The authors used the Arduino Pro Mini microcontroller platform, which employs the ATMega328P MCU. TMP102 from Texas Instruments is used as the system's temperature sensor, while DHT11 from Adafruit Industries is utilized as the humidity sensor. A gateway that works as an aggregator is responsible for receiving the sensor system's sensed data and transmitting control data to the system. The SlapOS platform is the cloud platform for storing the collected data for subsequent processing and visualization.

A WEMS is implemented in [75] to develop features for recognizing user activities and predicting location based on acquired environmental data. The MCU used is Atmel's ATmega1284, which serves as the logical core of the system. The temperature and humidity sensor HTU21D product of TE connectivity is used for ambient parameters. AMS's TSL2560 light-to-digital converter measures ambient visible light, while ST Microelectronics's LIS3DH accelerometer quantifies acceleration variation and orientation in the wearable device. Microchip's RN-XV Wi-Fi unit is incorporated as the communication module, and a 3.7 V Lithium-Thionyl Chloride battery is utilized to power the system. Wi-Fi transmits the acquired data to a local server and the internet cloud for further processing.

The design and development of wearable systems for active monitoring of multiple environmental and physiological data in asthma patients are presented in [76]. The system includes a bracelet, a chest patch, and a portable spirometer. Many environmental factors, including O3, VOCs, ambient temperature, relative humidity, and mobility, were monitored with the wristband. The ADXL362 accelerometer from Analog Devices Inc. measures motion. For data acquisition and transmission, Texas Instruments' CC2541 system-on-chip (SoC) solution is utilized. Lithium-ion polymer (Li-Po) batteries are used to power the devices.

In [77], the authors design a wearable sensor system for monitoring air quality called EnviroSensor to analyze personal exposure to air pollution. This research describes the development of two variants of the proposed sensor system. The initial version has a DHT22 sensor for measuring air temperature and relative humidity, an MQ-131 sensor for measuring ozone concentration, an MQ-7 sensor for determining carbon monoxide concentration, and a Shinyei particle sensor unit estimating particulate matter. An Arduino Uno microcontroller controls all components, and a 13750 Logger Shield GPS module calculates location. The power source for the device is a 9V battery. A micro SD card temporarily saves the captured data, which is then transferred to the computer's storage for further processing. In the later version, MQ-7 is removed while MQ-131 and Shinyei sensors are replaced with SPEC Ozone sensor and GP2Y101 sharp dust sensor because previous modules consume more power.

The MuSe platform [78] is a compact 25 mm x 25 mm form factor PCB that integrates sensing, processing, storage, and communication units. It leverages the latest technologies in sensing and processing to provide advanced capabilities while maintaining low power consumption. The platform includes a 32-bit microcontroller, a 3D geomagnetic System-in-Package (SiP), and a high-capacity 128 Mbit serial NOR flash memory for hours of continuous data logging without the need for an active radio connection. The processing core of the MuSe platform is the STM32F411 high-performance Cortex-M4 microcontroller by STMicroelectronics, which features a current consumption per MHz as low as 100 uA/MHz and a scalable floating-point unit (FPU) core capable of operating at a frequency up to 100 MHz. The MuSe platform is equipped with the LSM9DSI geomagnetic module, a System-in-Package (SiP) manufactured by STMicroelectronics, which combines a ±8 g 3D accelerometer, a ±2000 dps gyroscope, and a ±16 Gauss 3D magnetometer in a compact 3.5 mm x 3 mm package. In addition, the MuSe platform embeds two more high-performance modules, a ±4000 dps 3D gyroscope (ITG3701 by InvenSense) and a ±400 g 3D accelerometer (H3LIS331DL by STMicroelectronics) for detecting high angular rates and accelerations.

For the tasks for which these research-graded platforms were built, they were among the best options and proven to perform the role well. However, the selected devices have several limitations that may prevent their use in other environments. The WEMS [74, 75, 76] lacks onboard data storage modules, whereas the storage capacity of [78] only allows short acquisition sessions. As far as wireless communication modules are concerned, WEMS [74, 77] do not have any onboard modules, and [77] lacks low power consumption modules.

Chapter 4 describes the development of a new environmental and inertial platform called Winter, aimed at compensating for these limitations and potentially replacing the platforms discussed.

Literature related to objective two: Here we presented some notable studies related to our second research objective and identified the research gap, Energy monitoring can be disaggregated (non-Intrusive load monitoring -NILM or Intrusive load monitoring -ILM), depending on the monitoring location and application context [79]. In smart homes, sensors connect via wired or wireless ways, and multiple protocols are designed to allow communication. With the Internet of Things (IoT), this real-time data may be transferred to remote controllers for additional analysis, training, and prediction applications, providing a cost-effective energy management system option providing a cost-effective preference for energy management systems [79].

A novel ILM approach for load monitoring with the objective of developing an IoT-based appliance classification is presented in [80]. The developed IoT architecture consists of appliance, perception, network, middleware, and application layers. The appliance recognition module's primary duty is to label sensor data and enable the implementation of various home applications. With the disaggregation of appliance consumption, several smart grid applications, such as demand response and load planning, could be implemented.

According to [81], the significant energy savings can be achieved by monitoring energy consumption at the device level. Hence, the analysis of energy at the device level is performed by energy disaggregation, which is the extraction of energy consumption at the appliance level from one or many smart meter measurements. NILM [51] refers to the challenge of measuring just the aggregated usage using a single sensor (smart meter) per residence or building.

Researchers have demonstrated that combining temporal information and actual power levels facilitates load disaggregation [82, 83]. Hence, Artificial Neural Networks (ANN) [84] and Hidden Markov Models (HMM) [85] perform well in the task of load disaggregation due to their capacity to incorporate temporal and appliance state change information into their learning. Nevertheless, the complexity of HMM models increases exponentially with the number of target appliances, limiting the utility of this learning strategy. In addition, each time a new appliance class is introduced, the entire model must be retrained.

In existing state of the art following research challenges are observed.

- To develop reliable algorithms with great generalizing ability.
- To build hybrid NILM models incorporating user feedback and techniques supporting continuous learning.
- Applicability in real world.

In addition, The NILM method requires sensors that acquire high-frequency data, but this type of sensor has two disadvantages: they are much more expensive than those used in the experimentation under consideration, and they generate a large amount of data to manage. The experimentation carried out, on the other hand, took as a starting point the use of low-cost sensors to guarantee the replicability and real diffusion of the proposed model.

In the light of above-mentioned challenges, we proposed a generic and reliable algorithm for energy consumption disaggregation. In the development of the proposed algorithm, we incorporated user feedback in order to develop a robust energy consumption disaggregation system.

Literature related to objective three: Here, we presented several significant works related to our third research objective and highlighted limitations in the literature. Many approaches for analyzing electricity consumption data has been described in the literature [86] so that end-user applications can be designed on top of these techniques. For time series data, these approaches are primarily concerned with event detection [87] and classification [88]. Generally, the energy consumption pattern is the activation (ON) and deactivation (OFF) of appliances during power distribution over time.

Smart meter data analysis includes analyzing smart meter data to identify appliance usage patterns. Smart meters capture energy consumption data with a high temporal resolution (every 15 minutes, for example), allowing them to detect the operating patterns of individual appliances. Smart meter data analysis can provide accurate information on appliance energy consumption, but it needs advanced analytics and may not be as precise as other approaches [86].

A study [89] investigated the elements that influence the energy consumption of individual household appliances. In the first stage, context-rich data from six households across the United States was collected. In the second stage, a rule mining algorithm was developed to detect significant connections between energy usage and four essential parameters: an hour of the day, the day of the week, the use of other home appliances, and user-provided annotations of activities such as working or cooking. The analysis verifies the hypothesis that most devices exhibit a consistent daily or weekly usage pattern, which is false for all devices.

There are a few clustering algorithms for recognizing the operating patterns of appliances in nonintrusive load monitoring (NILM) systems that have been developed. Hart, who pioneered the concept of NILM and examined the algorithm and characteristics of nonintrusive household appliance load monitoring, proposed one of the earliest approaches. This approach decomposes the aggregated load data of household appliances [51]. Since then, numerous research attempts have been made to enhance NILM's pattern recognition. K-means, K-nearest neighbor, enhanced ISODATA, and artificial neural networks are some prominent methods [90, 91, 3, 92]. These techniques have provided valuable insights for determining the operation modes of electrical appliances. Unfortunately, when there are several home appliances, these methods may not produce good recognition results.

Smart metering, on the other hand, has been promising in several applications, such as energy modeling [93, 94, 95, 96, 97, 98], behavior characterization [99, 100, 101], grid infrastructure technical evaluations [102, 103, 104, 105, 106], and end-user engagement [107, 108, 109].

From the literature, it has been observed that most of work reported is related to overall load-level consumption information, cost-aware appliance scheduling, load balancing, and a scheduling algorithm based upon energy meter data. To the best of our knowledge, we have not found simple models aimed at extracting patterns such as the number and duration of operations, cycle disaggregation for appliances that have cyclic operation (e.g., washing machine, dishwasher), and energy consumption throughout various time periods, which can be easily integrated into IoT platforms dedicated to residential users.

Literature related to objective four: we presented some significant works related to our last research objective to find the best models for energy consumption forecasting for the DHOMUS platform. Rodrigues et al. [110], used ANN to predict daily and hourly short-term load and household electricity consumption. They consider apartment location, occupants' numbers, electric appliance consumption, and hourly meter system to train the model. In this work, a feed-forward ANN and the Levenberg-Marquardt algorithm produced better forecasting results.

In [111], authors demonstrated that the forecasting granularity and one-day-ahead load forecasting accuracy for residential customers can be affected by the calendar variables and the scaling of the training set. Statistical analysis has demonstrated that the regression trees approach significantly surpasses ANN, and SVR techniques despite the similarity of average RMSE for all techniques.

A study [112] perform a comparative analysis of six data mining techniques, including SVM and RF, for estimating Energy Use Intensity (EUI) for commercial office buildings in the US. They also estimated the plug loads and lighting loads of HVAC based on the 2012 CBECS microdata. According to experimental results, SVM and RF provided better predictive accuracy based on a large number of outliers in the CBECS dataset.

Authors in [113] performed a comparative analysis to analyze the performance of statistical and ML models in predicting energy consumption in non-residential smart buildings by utilizing the electrical energy consumption data collected from thirteen smart buildings located on a university campus in Spain. The authors demonstrated that highly accurate prediction accuracy can be reached in favor of strategies based on ML approaches and the historical window's optimal size optimization.

Wang et al. [114] proposed a RF-based prediction model for predicting hourly building energy. To evaluate the performance of the proposed model, the hourly electricity consumption of two educational buildings in North Central Florida was predicted. The training of the RF was performed by considering different input variables with the intention to search the feature space that has a critical impact on the prediction model's performance. According to experimental results, RF showed better performance in comparison with RT, and SVR models.

Based on the literature findings, ML-based forecasting models have been studied, combined with the promising results achieved in the energy consumption forecasting domain. Furthermore, extensive experiments have been performed to compare ML forecasting models to find a suitable model to deploy in IoT Platform.

3.9 Summary

In this chapter, we have briefly introduced concepts of energy consumption disaggregation methods, also known as appliance load monitoring (ILM, NILM). After this, we discussed the appliances operating states (on-off, standby), followed by energy consumption patterns of household appliances, such as peak hours energy and long and short energy consumption cycles. We then discussed the fundament components of time series (trend, seasonality, irregular components). Then basic forecasting methods (qualitative, quantitative) have been discussed. We then discussed works related to this thesis. The analysis of the literature highlights the limitations of existing studies.

Chapter 4

Winter: A Novel Low Power Modular Platform for Wearable and IoT Applications

This chapter introduces a multifunction, ultra-low power device to avoid commercial IoT systems' typical limitations. The first function could be a stand-alone system with embedded sensing, data processing, storage, and communication capabilities. The second is a motherboard for micro expansion boards that can be attached to its top to expand the basic functionality. Lastly, the chapter presents the device's performance and IoT use cases.

4.1 Background

Intelligent and wearable devices are becoming the driving force of the Internet-of-Things (IoT) era due to the integration of wireless technologies, Micro-Electro-Mechanical Systems (MEMS), and the Internet. The capacity to monitor different parameters using multiple sensing units enables such systems to employ for various purposes, ranging from recording human movements for fitness and rehabilitation purposes to monitoring environmental indicators for assessing the quality of life [115]. In this scenario, the rapid growth of machine learning and big data applications has increased the quantity and type of data required. It is especially true for medical research projects, where selecting the best instruments has become a critical factor

[116]. The most important characteristics of a device should be the availability of raw data that could be handled by in-house algorithms and the capacity to collect a variety of data types over lengthy periods.

In this work, a novel wearable platform has been developed in response to the constraints of the wearable devices used in earlier research. The platform has two main features, first is an efficient monitoring system ideal for edge computing and real-time data analysis. The second is a long-term data logger, offering up to several days of continuous data log owing to low power consumption and large memory.

4.2 System Architecture

Winter, an abbreviation for Wearable Inertial TrackER, is the product of a design intended to give a system-on-board with sensing and processing capabilities, low power consumption, (relatively) large storage capacity through an on-board SD connection, and wireless transmission capabilities supplied by a BLE module, all in a 32 x 20 mm^2 form size (see Figure 4.1). Figure 4.2 provides a visual representation of the system's general block architecture as well as the communication interfaces that were implemented [117].

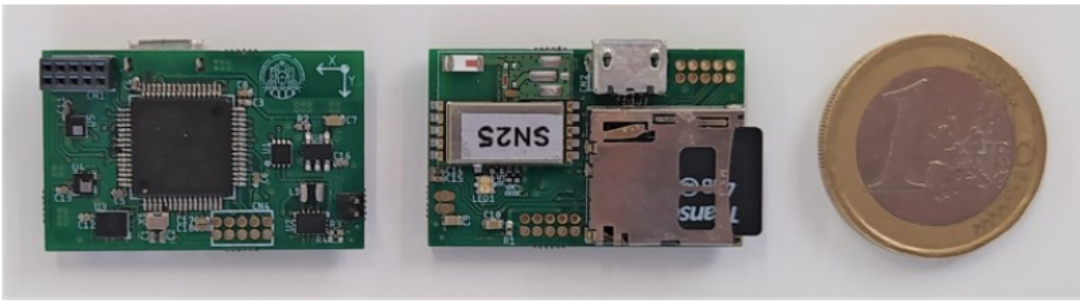


Figure 4.1: The Winter device

4.2.1 Processing

STMicroelectronics is responsible for producing the ultra-low power STM32L475RG microcontroller unit (MCU), which serves as the central processing unit (CPU) of Winter. It incorporates high-speed storage (1 MB of Flash memory and 128 kB of SRAM), a low-power RTC (Real-Time Clock), and a comprehensive set of upgraded I/Os and peripherals. It is based on the ARM Cortex-M4 32-bit architecture. In addition to this, the ABS06-1-T SMD crystal manufactured by Abracon was connected to the microcontroller. This allowed for the generation of a clock signal at 32.768 kHz with a frequency tolerance of 10 ppm (parts per million), which allowed for the creation of a timebase that was both more accurate and precise.

4.2.2 Sensing

The platform comprises two sensing modules, one module is used for environmental parameters monitoring, and the other is for inertial parameters measurement. The first one is the LSM6DSL, a system-in-package provided by STMicroelectronics featuring a 3D digital accelerometer and a 3D digital gyroscope with full-scale ranges of ±2/±4/±8/±16 g and ±125/±245/±500/±1000/±2000 dps, respectively. The module works at 0.65 mA in high-performance mode and enables always-on low-power features [3]. The configurable event-detection interrupts allow efficient and reliable contextual awareness, implementing hardware recognition of free-fall events, 6D orientation, click and double-click sensing, activity or inactivity, and wakeup events. In particular, the click detection will be used to monitor external input and thus generate a wakeup interrupt to the microcontroller.

The second sensing module combines the HTS221 and LPS22HH (STMicroelectronics), where HTS221 is an ultra-compact sensor for relative humidity (rH) and temperature. It includes a sensing element and a mixed signal ASIC to provide the measurement information through digital serial interfaces [4]. The HTS221 has a relative humidity range of 0–100 % (accuracy: ±3.5 % in the 20–80 % rH interval) and a temperature range from -40 to 120 °C (accuracy: ±0.5 °C in the 15–40 °C interval). When the module operates at 1 Hz ODR, the power consumption is 2 µA. The LPS22HH is a piezoresistive absolute pressure sensor that works as a digital barometer. However, its best working range is between -40 °C to +85 °C; among other key features, the current consumption is 4 µA, and the pressure accuracy is 0.5 hPa.[https://www.st.com/en/mems-and-sensors/lps22hh.html].

4.2.3 Storage

The MicroSD card connector on the Winter platform makes it feasible to save raw data and processed findings on an external memory card. The data transfer was performed at a rate of up to 48 MHz utilizing the MCU's 4-line SDMMC communication interface (8-bit mode). The saved data could be accessed by removing the

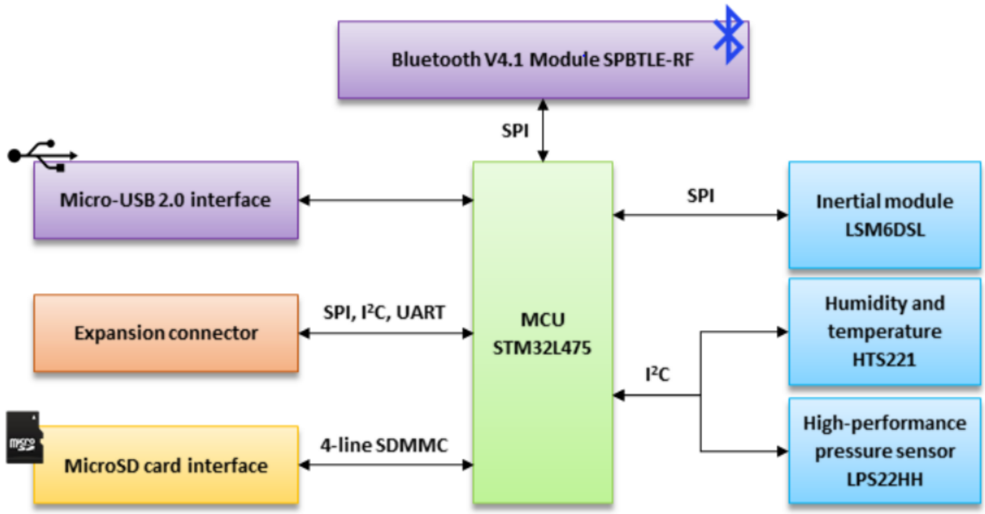


Figure 4.2: Block diagram of the Winter platform

SD card or USB.

4.2.4 Connectivity

Through its communication with external devices, Winter can (1) transmit the outcomes of its internal data analysis algorithms and (2) be remotely configured to operate in any of its available operating modes. Two unique methodologies were employed to make this type of communication feasible. The first solution is the SPBTLE-RF from STMicroelectronics, a Bluetooth Low Energy master/slave network processor module compatible with Bluetooth version v4.1. Alternatively, Winter features a high-speed Micro-USB 2.0 port that permits data transfer when utilized with an external device.

4.2.5 Power Management

The device is powered by a 3.7 V, 210 mAh lithium polymer battery, scaled down to the 3 V operational voltage employing a high-efficiency step-down converter (TPS62740 manufactured by Texas Instruments). Microchip's MCP73831 charging controller supervises the process of recharging the battery through its micro-USB port. The MAX17048X+ was incorporated and connected via I^2C interface to the MCU to monitor the battery charging level.

4.2.6 I/O

On the underside of the Winter platform, two 10-pin connections are placed. The first is a basic debug hookup that should be disabled before release. The second allows the Winter platform to expand its sensing and communication capabilities by offering an expansion board connection. In reality, the system's power supply and all of the most popular interfaces (SPI, I2C, UART) may be accessed via this connection by any extension board put into the top of Winter. In such a scenario, although the Winter platform would function as a motherboard, giving the necessary power and providing primary storage and communication, an extension board might be built with only the essential components, lowering its size.

4.3 Firmware and Performance

4.3.1 Finite-State Machine

Figure 4.3 illustrates the finite-state machine (FSM) implemented on the MCU of the Winter, while Table 4.1 specifies the state information.

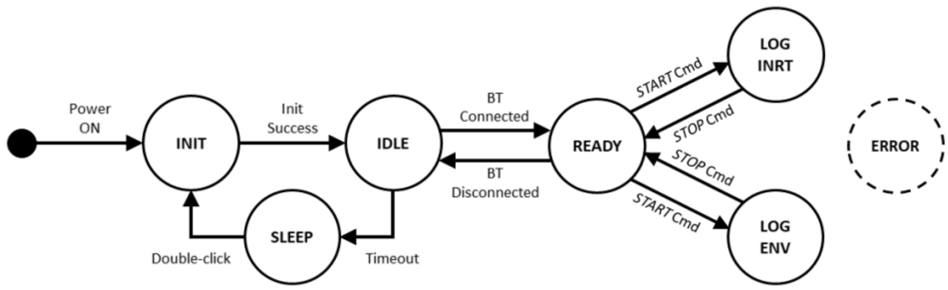


Figure 4.3: Finite-State Machine of the Winter platform

After the startup sequence has been completed successfully (INIT state), the platform enters the IDLE state, defined by the inertial sensor being disabled, the environmental sensing components being powered on, and the Bluetooth (BT) module being powered on prepared to receive a connection. If a link is not formed within a specific time frame (1 minute), the system enters the SLEEP state: the Bluetooth module is deactivated, and the microcontroller is put to Stop 2 mode. A double-click event recognized by the inertial module initiates the INIT state of the system.

An active Bluetooth connection switches the FSM from the IDLE state to the READY state, where the platform is prepared to communicate with the master de-

Table 4.1: List of Winter's FSM states (Inrt: inertial; Axl: accelerometer; Env: environmental)

Name	MCU Mode	Inertial Sensor	Env Sensor	Data Storage	Bluetooth Status
INIT	Run	-	-	-	-
IDLE	LP Run	OFF	ON	Inactive	Advertising
SLEEP	Stop 2	Axl ON	OFF	OFF	OFF
READY	LP Run	OFF	ON	Inactive	Connected
LOG_INRT	LP Run/Run	ON	ON	Active	Connected
LOG_ENV	Run/Stop 2	OFF	ON	Active	Connected

vice. Winter may transition between two distinct log states using the BT command set. The first one is referred to as LOG INRT, and it involves the logging of inertial data from the selected sensors (only accelerometer, only gyroscope, or both) on the SD card at a specified frequency; the MCU alternates between the Low-Power Run mode (collection phase) and the Run mode (SD writing/analysis phase) to reduce the average power consumption. The second is known as LOG ENV: when the inertial sensors are switched off, the MCU is kept in Stop 2 mode and is awakened (Run mode) every minute for the period necessary to collect environmental parameters and store them on the SD card. Both states can function with an active BT connection or the BT module operating in Advertising mode.

There are two methods to get the data collected by Winter. The first needs the user to remove the SD card: all the logged data (inertial or environmental, depending on the mode previously employed) may be transferred to any device with an SD support (e.g., a computer or a smartphone), and the SD card can then be erased and reused for future collection sessions. Second, Data are transmitted to the outside world via the BLE characteristics system. This only applies to environmental data conducted in all states when such sensors are operational. The system does not currently support real-time data streaming, although it might be incorporated in future firmware updates.

During each FSM's distinct stages, the operation schedule is based on the MCU management's internal interrupt handling system. For each task, the timebase represented by the system clock is used to raise internal interrupts after a particular period has passed; next, the relevant call-back function is activated based on the interrupt priority (i.e., the task's priority); and lastly, the code related to interrupt handling is run. Specifically, the primary functions performed by the system are as follows (the order defines the priority from highest to lowest): (1) reading the data collected by the onboard sensors from their respective registers; (2) logging the data onto the SD card (if present); and (3) handling the Bluetooth connection by interpreting incoming commands and updating the user-readable characteristics.

4.3.2 Power Consumption

A preliminary evaluation of the total power usage was conducted using the provided testing software. Table 4.2 provides the results of the measurements. The 210 mAh battery permits up to 3 days of continuous inertial data logging (with both accelerometer and gyroscope ODRs set to 416 Hz) or up to 2 months of environmental monitoring (with one-shot measurements per minute).

Table 4.2: Average power consumption of the states (power supply: 3.7 V, 210 mAh battery)

State	Current Drawn	Power Consumption	Autonomy
IDLE	0.966 mA	3.574 mW	9 days
SLEEP	0.045 mA	0.167 mW	6 months
READY	1.046 mA	3.870 mW	8 days
LOG_INRT	2.881 mA	10.660 mW	3 days
LOG_ENV	0.171 mA	0.633 mW	$\tilde{2}$ months

4.3.3 Data Comparison

Accelerations collected by Winter were compared to those collected by MuSe, a research-grade inertial platform used by authors in other works [78, 118], both in static and dynamic conditions shown in Figure 4.4. Acceleration signals collected by the two platforms are almost identical on each axis, thus confirming the possibility to use Winter in place of the MuSe one when the collection of inertial data is needed. Some negligible differences can still be observed in static conditions: this can be ascribed to the different 0 g-offset values of the devices' inertial modules (±90 mg for MuSe's LSM9DS1, ±40 mg for Winter's LSM6DSL). However, such differences will be reduced once the self-zeroing procedure is implemented.

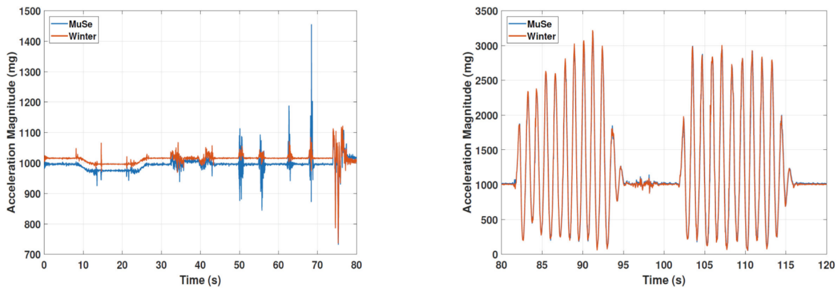


Figure 4.4: Comparison of acceleration magnitude collected by MuSe and Winter devices in quasi-static (left) and dynamic (right) conditions

4.4 IoT Application: Indoor Environmental Monitoring

An IoT-centric application was developed in order to test the system and give a use case for the Winter platform. Specifically, the described device was utilized to continually gather data regarding the environmental parameters of a room, including temperature, humidity, and air pressure. The system is depicted in Figure 4.5 and consists of the following components: (1) a Winter device used to collect environmental data; (2) a Raspberry Pi 3 connected to the Internet via an Ethernet cable and functioning as a gateway; and (3) an open-source IoT application called ThingSpeak that provides both a cloud server and a customizable graphical front-end. This free and ready-to-use cloud-based IoT analytics platform solution allows storage, visualizing, and analyzing live data streams. ThingSpeak delivers real-time visualizations of data sent to ThingSpeak by the devices. ThingSpeak is an Internet of Things open data platform. Using an MQTT, the device can communicate with ThingSpeak; for data privacy, a private mode is available on the platform.

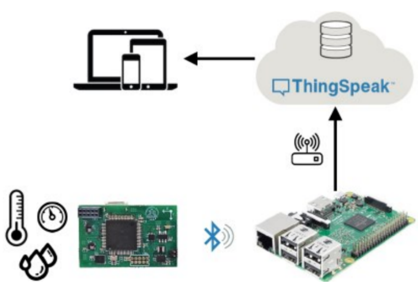


Figure 4.5: The environmental monitoring system based on Winter device

The availability of a BLE module on the Raspberry platform made it possible to develop an ad hoc Python script for a periodic readout of environmental data; Winter made them accessible as features of the standard Environmental Sensing Service (ESS) of the BLE protocol.

The data were then transmitted to the ThingSpeak cloud to aggregate and display real-time values. First, a dedicated channel was built on the web platform; next, the given API keys and function calls were included in the Python script to enable the batch job to transmit data packets through the HTTP protocol: ThingSpeak's temperature, pressure, humidity, and battery level chart for numerous monitoring sessions is depicted in Figure 4.6.

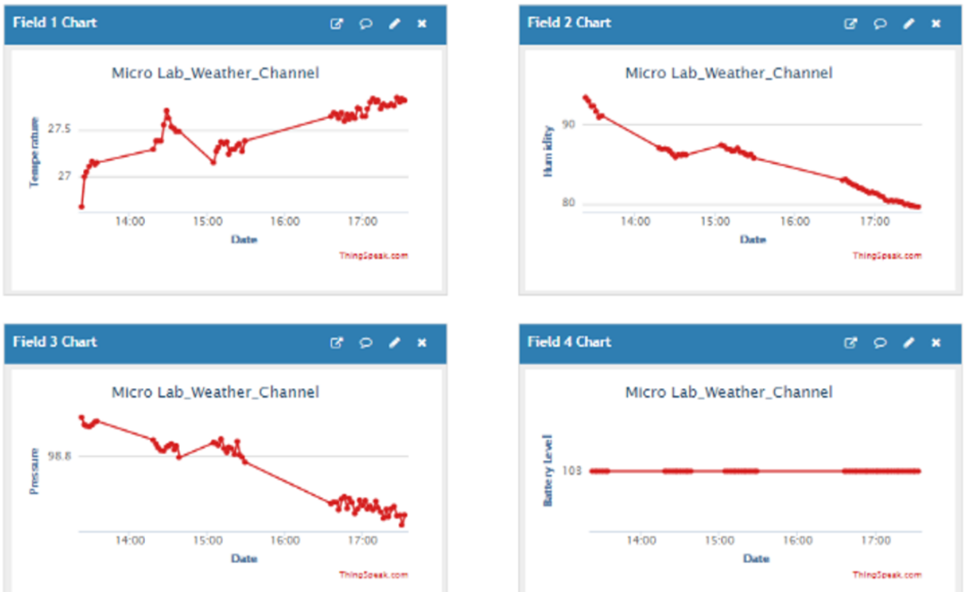


Figure 4.6: Environmental parameters for monitoring session

4.5 Data Acquisition at Microelectronics Lab (Unibg) transferred to DHOMUS platform using Rest API for Visualization

The previous section used a Winter device for data acquisition and visualization, MQTT-based cloud Thingspeak. However, we replaced the MQTT with REST API to communicate with DHOMUS. In the second use case, it was a business requirement of our industrial partner. To integrate the Winter environmental parameter with the DHOMS platform to visualize and analyze live data streams. To see how the device works and how people use it for environmental, assistive, and energy-related services on the smart home platform. Moreover, data from the sensing device to the gateway is transferred using a low-energy Bluetooth protocol. Our Use case data are temperature, humidity, air pressure, and battery of the Winter device. Furthermore, Raspberry Pi is our gateway that collects the Winter device's data and is transmitted to the cloud every two minutes using REST API.

4.5.1 Data Transmission Architecture

Using a Raspberry Pi as a gateway and a REST API-based DHOMUS cloud platform to live to stream the measured parameters of the modules is a fundamental goal of the Winter product. Figure 4.5 illustrates the entire data transfer and display process among various communication protocols, but the DHOMUS platform replaces Thinkspeaks in this setting.

The initial step is to send environmental data from Winter to the Raspberry Pi, which is made possible by the onboard presence of BLE, as well as the use of Linux shell scripting and the Python programming language as a software tool. To discover and connect a peripheral using the BLE protocol, the Python module "bluepy.btle" was utilized. The next step was gaining access to services that included the device profile's generic attributes.

The next step is to acquire the values read by the sensor, which may be done by invoking the device's environmental sensing module's characteristic handles. Furthermore, these values were obtained as bytes and required to be converted to readable form, which was accomplished using Python packages (Struct, binascii) and the function "int. from bytes(Value, byte order)."

4.5.2 Data Visualization

DHOMUS platform cloud service makes it possible to aggregate, visualize, and analyze data streams in the cloud. Table 4.3 shows the information mapping for the object available on the cloud. Figure 4.7 presents the live data visualization every minute and a graph of Temperature and Humidity during the day.

Table 4.3: Information mapping objects information

Sr No.	Data	apioId	objectId
1	Temperature	unibg	11
2	Humidity	unibg	11
3	Pressure	unibg	11
4	Battery	unibg	11

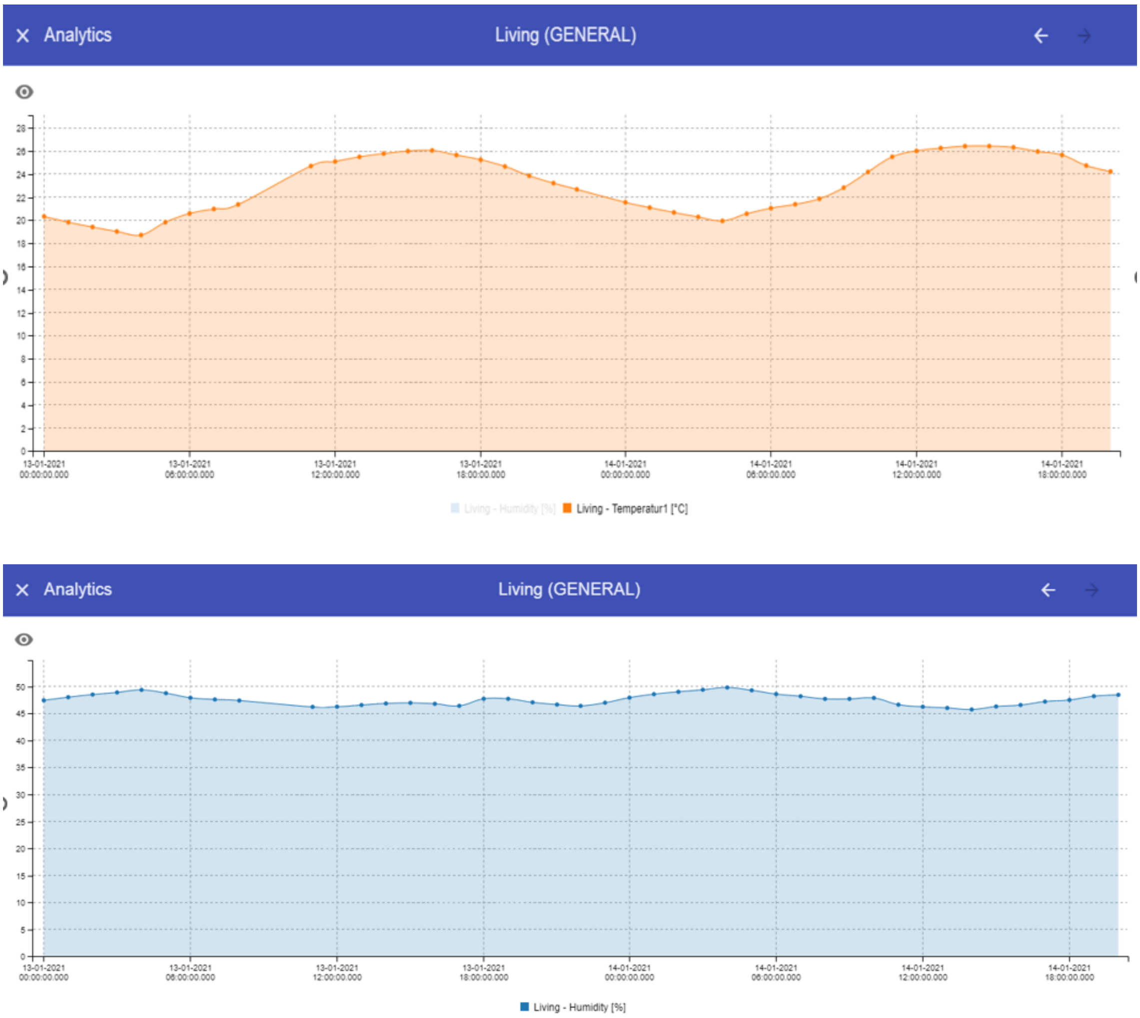


Figure 4.7: Temperature and Humidity Visualization

4.6 Summary

In this chapter, to fulfill our first objective, We developed a research-graded device with the most critical feature, which is the availability of raw data that is not present in commercially available devices. Moreover, the device is equipped with environmental sensing and inertial models. The system architecture of the device is available in section 4.2, with an explanation of all its parts. Subsequently, the firmware part and power management blocks are discussed.

Secondly, An IoT application is developed as a use case to test the device for environmental parameter acquisition and store them on the cloud using two different platforms. Section 4.4 provided indoor environmental monitoring use case using the ThinksSpeak cloud. Section 4.5 provides detail on integrating the Winter device into the DHOMUS platform.

Chapter 5

Data Driven Disaggregation Method for Electricity Based Energy Consumption for Smart Homes

This chapter presents the data-driven disaggregation method for electricity-based energy consumption in collaboration with the Italian national research agency ENEA as part of the DHOMUS Platform. This deals with the second objective of this thesis: Devise an approach that can disaggregate whole house energy consumption into specific energy sectors for feedback to smart homes users in the network. Related work and limitations are presented, followed by a brief description of the software architecture of the DHOMUS Platform. The data acquired from the smart home network and its hardware are discussed in detail. Afterward, we discuss the experimental results to evaluate the effectiveness of the proposed approach. Finally, experimental evidence of energy consumption saving based on proposed approach is presented.

5.1 Related Work

Energy monitoring is a fundamental aspect of the energy management system. It is, therefore, necessary to monitor the energy consumption of a building before taking practical initiatives to minimize energy consumption. Numerous studies have been conducted to understand consumers' energy consumption patterns comprehensively. Disaggregation of energy consumption is an optimal technique to gain deep insights into load monitoring. There are mainly two approaches to disaggregate energy consumption: distributed path monitoring and single point monitoring. In the literature, the first load monitoring method is called Intrusive Load Monitoring (ILM), while the other is called Non-intrusive Load Monitoring (NILM). The ILM approach is optimal due to its distributive nature but at a higher cost as the number of measurement devices increases. On the other hand, NILM is suboptimal, but the infrastructure cost is low.

A framework and collection of algorithms for implementing novel energy efficiency services using smart meters and smartphones are proposed in [119]. The equipment signature detection training procedure is greatly simplified signature detection is greatly simplified by using smartphones and a state-of-the-art filtering method. The proposed solution achieved a detection rate of 87 percent in a test with Eight simultaneous appliances. In another study [120], the authors discussed ILM and NILM approaches to energy disaggregation and the need to address security issues for such systems' rapid growth and adoption of such systems. In addition, it is claimed that the improvement of such systems may lead to the strong participation of energy consumers in energy-saving campaigns. Moreover, A pilot project for a Smart Homes Energy Management Monitoring System (SHEMS) based on Tridium's Niagara Framework using Fog (Edge) Cloud Computing with Non-Intrusive Appliance. Load Monitoring (NIALM) was developed as an IoT application in energy management [121]. The SHEMS prototype proposed uses an artificial neural network-based NIALM technique to non-invasively monitor relevant electrical appliances without the need for plug-in electricity meters (smart plugs), completing a two-tier NIALM approach. The SHEMS prototype is based on a compact, cognitive, embedded IoT controller that integrates IoT end devices such as sensors and meters and serves as a gateway for demand-side management (DSM) in smart homes.

Research in [122] presented that most datasets in the research are appropriate for pattern-based energy disaggregation (ED) techniques that require a lot of data. Subsequently, optimization-based ED methods have been devised that require information about the operating states of the devices. The development of repeatable state-of-the-art optimization-based ED algorithms are limited by the availability of standard datasets and acceptable assessment measures. Therefore, a dataset with several examples that reflect the various issues given by ED is provided. A NILM approach based on '0-1 sparse coding' was used [123] to disaggregate into one specific appliance from total electricity consumption. Data was collected from two households using smart meters, and the data granularity was 5 minutes. The results show that the "0- 1 sparse coding" method achieves 44.8% improved disaggregation accuracy compared to the standard sparse coding method. Authors in [50] discovered that there are concerns with load monitoring and management that need to be addressed, such as more accurate recognition and the requirement for a monitoring system that can monitor as many different types of devices as feasible. More work is needed to include NILM in the energy management of appliances. Finally, whether in homes, workplaces, or enterprises, electricity customers need to promote an energy management culture. In another study [48], the NILM approach was used for disaggregation based on load signatures composed of macroscopic and microscopic features, and then equipment classification was performed using publicly available datasets. Additionally, applications of energy disaggregation such as disaggregated energy calculations, accurate demand forecasting, and appliance anomalies are also presented.

In [124], the semi-supervised learning (SSL) approach is developed using a variety of signals from the unlabelled dataset to simultaneously learn the classification loss for the labeled dataset and the consistency training loss for the unlabelled dataset. The transformation that generates the samples for consistency learning is based on weighted versions of the DTW Barycentre Averaging method. Using data collected from an Internet-of-Things-based energy monitoring system in the context of a smart home, the method is evaluated and showed excellent results. This work in [125] is based on the values of active power as it examines the effectiveness of the load sharing method using the Rainforest Automation Energy Dataset (RAE) and Reference Energy Disaggregation Dataset (REDD) databases, which were collected using the Non-Intrusive Load Monitoring (NILM) measurement method. It explained how to assign labels based on the device combinations used, the device status (ON / OFF), and the selection of an appropriate temporal data frame. It also evaluated the effectiveness of known machine learning algorithms such as random forest, decision tree, and k-nearest neighbor (kNN). For the RAE and REDD databases, the results show that it is a very effective technique with low computational complexity, with an F1 score of over 95%.

It is found in [126] that providing disaggregated feedback leads to a 5% higher conservation effect than the usual (aggregated) input. Energy consumption can decrease by 10% to 15% during peak hours, which was mainly responsible for these conservation effects. Consumers who have appliance-level data can more accurately estimate how much energy different appliances consume. In addition, it was pointed out that the impact of smart meter roll-out on energy savings is significantly greater when appliance-level data is available. It is expected that device level feedback could increase consumer surplus for German households by about 570 to 600 million euros per year, based on a suitable statistical method. The authors in [127] believed that engaging consumers in demand-side management activities can lead to the achievement of energy management goals. An algorithm was applied to 14 households with different use cases for load-shifting activities. The results of these simulations showed an average decrease in theoretical flexibility of 53% instead of 66%, measured as the proportion of appliance cycles that are shifted compared to total cycles; in a single household, a maximum deviation of 29% is found. Finally, the monthly average shifted energy per dwelling drops by 32.5 percent, from 27 to 18 kWh.

Authors in [128] wanted to find what involvement the Italian residential sector may play in establishing load flexibility for Demand Response activities. A method of estimating the load profile of a housing cluster of 751 units based on experimental and statistical data was proposed. 14 housing archetypes were identified, and an algorithm was developed to categorize the sample units. After analyzing possible flexible loads for each archetype, a control technique was developed for implementing load time shifting. In this technique, both the power demand profile and the hourly electricity price are considered. Calculations reveal that a dwelling cluster in the Italian residential sector has an index of flexibility of 10.3%, as well as efficiency of 34%.Over the heating season (winter)for the weekends, the highest values have been recorded for flexibility purposes. The authors [129] suggested a technique for analyzing and designing a production, self-consumption, and storage system that serves a home user aggregate in order to optimize electric power demand. They achieved Peak Shaving of the electrical demand power curve by limiting the maximum power absorption from the grid and delivering the balance of the user's power demands via an electrical energy storage system charged from the solar plant during the daily overproduction period. The success of the Peak Shaving approach was assessed using a percentage parameter that depicts the number of grid power absorption peaks averted because of the storage system.

In the literature, two types of disaggregation methods are presented. The first is NILM, or energy disaggregation (ED), where whole-house energy consumption is collected with a single sensor. Energy consumption for each appliance is obtained with the help of supervised machine learning-based approaches. The second approach is ILM, where each device in the home is attached to a dedicated sensor. The techniques developed based on NILM (unsupervised or supervised) get complex due to the increased number and different types of devices [52, 82, 130].

Most of the NILM approaches in the literature are based on machine learning [51, 92, 131]. These approaches face problems like:

1. As the number of appliances grows, the amount of training data needed for feature extraction and model building grows exponentially.

2. Since each home is different in terms of its devices and how users use them, the training process must be done separately for each home or adjusted with the respective training data.

3. There are not enough unified load signatures to model how different types of appliances work.

4. The data collection and training processes must be done again to add new devices to the existing network, which makes it hard to use in the real world.

The disaggregation method proposed in our work is for a real-world project, an IOT-based smart homes platform built based on ENEA requirements. The data is acquired from the user using the online platform smart sim (SS) and the installed sensor (SH). The methods discussed in the literature were not well suited for the challenges with our real-world IOT-based smart homes platform.

In the literature, energy disaggregation is limited, which only provides appliance-level classification. Our proposed algorithm provides one step ahead of information, dividing the home into nine energy sectors. Users can access disaggregated information about the whole house for dedicated energy consumption sectors. The provided solution is a new method of its kind and does not face the limitations that have been faced by disaggregation methods reported in the literature. To our knowledge, no approach can disaggregate whole home energy consumption into different sectors. We put the proposed algorithm to the test in different homes with different numbers of devices and at different times of the year to see how well it works.

5.2 Software Architecture of the DHOMUS Platform

The smart home system architecture consists of one of the vertical applications of the smart district that can send a set of Key Performance Indicators (KPIs) to the Smart District Platform (SCP), which can be shared with the other application. There are two levels of application for the smart home system: home level and aggregator level. Specific functions and services have been assigned to the component associated with each level. The software architecture of the DHOMUS platform is shown in Figure 5.1.

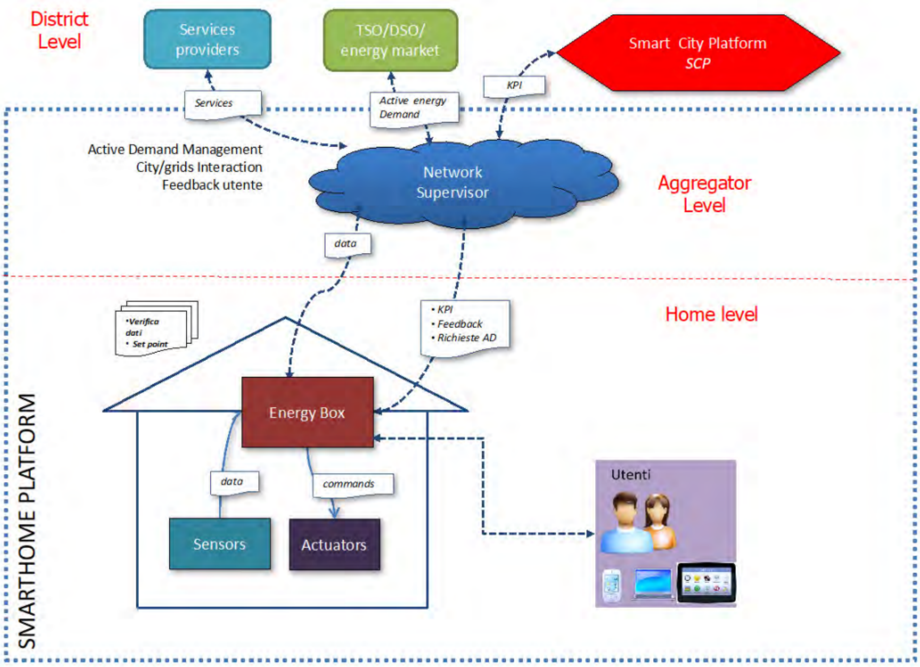


Figure 5.1: Software architecture of the DHOMUS platform

At the individual home level, an Energy Box (EB) collects data from the network of sensors installed in the home. Moreover, it enables the control of certain devices, smart plugs, smart switches, etc., and acts as a gateway for data collection and communication between the Home and Aggregator levels.

The Aggregator consists of an Information and Communication Technologies (ICT) platform that performs the functions of collection, aggregation, and analysis of data provided by the network of monitored homes to provide educational feedback to the user; it is also capable of providing the available data to external applications for further processing.

At the District level, the smart home system can communicate with third parties, both SCP and other stakeholders, to exchange information for home energy management or enable additional services useful to the end user, such as energy, security and assisted living.

5.3 Smart Home Network Overview

The smart home's architecture shown in Figure 5.2, is equipped with a range of smart devices such as smart plugs, sensors to measure consumption and comfort, and presence in the homes. All these devices are managed wirelessly, so no connection needs to be installed. The data from the sensors is collected, combined, and transmitted to the DHOMUS platform via the Energy Box, an electrical device connected to the internet.

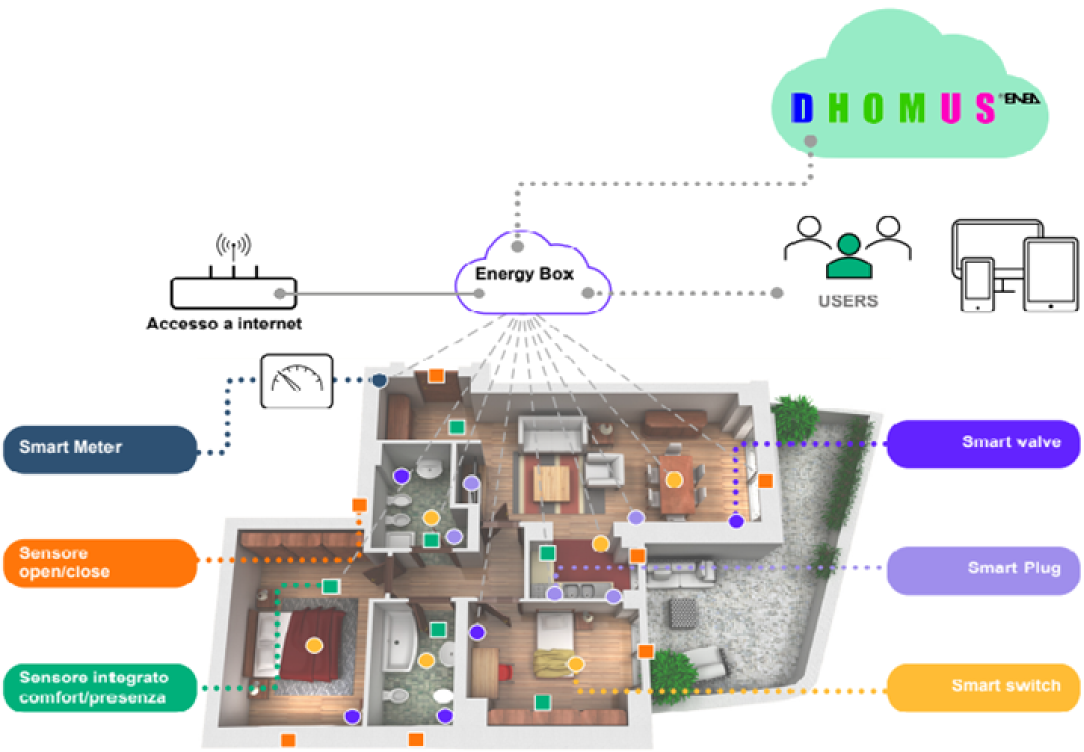


Figure 5.2: Smart Home

5.3.1 Data Acquisition Hardware

Table 5.1 shows the equipment used in the experimental demonstration in Rome. These are commercial sensors provided free of charge to participants in the study. The sensors can communicate via the wireless Z-Wave protocol with the Energy Box (EB), which consists of an Asus Mini PC - PN40 that acts as a gateway for smart homes. This box is made of a PC and is equipped with a USB dongle that acts as an antenna to receive signals from the sensors in the field, with which it communicates via the Z-Wave protocol.

Each sensor used in the experiment at the pilot homes has its own way of working and frequency of data collection. The collected data are sent in real-time to the cloud platform, where they are synchronized and put together. The verification of operation and data acquisition by the sensors installed in the pilot homes was done by checking the number of records received and stored, considering the mode of operation of each sensor. Table 5.1 summarize the measured quantities and the unit of measurement for each sensor. The point data acquisition interval and the expected number of records before the aggregation process are carried out at a minimum frequency of 15 minutes.

5.3.2 Data collection

The collected data is stored in the DHOMUS database and then used for data analysis. The dataset contains the energy consumption of smart homes with a data granularity of a quarter of an hour. The data we use contains information from 10 houses representing a particular home as an Energy Box (EB). Also, for the current scenario, we have six months of energy consumption data for (EB1, EB2, EB3, EB4, EB5, EB6, EB9, EB10, EB11, and EB12) for each smart home connected to the DHOMUS platform. However, each EB contains a different number of energy meters, smart plugs, and smart switches. The most important parameters of the data sets can be found in Table 5.2. Table 5.3 illustrates the smart homes for experimentation and sensors used for efficiently monitoring the maximum energy consumption. Table 5.4 presented the percentage energy meter data collection for six months (June to November 2021) for all smart households. The data presented has a granularity of a quarter of an hour, so for 24 hours, we have 96 observations when the sensors were working and sending 100% of the data correctly. We also counted the number of observations for all months to get some numbers for Table 5.4. The data collection

Table 5.1: Sensor kits for the experimental demonstration

Sensor	Make/Model	Measured magnitude	Data acquisition interval
Home Energy Meter	Aeotec/ ZW095-C	Instant Power (W)	The data acquisition reporting time according to the technical specification sheet ranges from 30 sec to 300 sec. However, using available parameters, we set the reporting interval to 60 sec for our experimentation
		Accumulated energy (kWh)	The sensor sends a report when it detects a change in Watts of 10%
Smart plug	Aeotec/switch 7	Instant Power (W) .	By default, there are no W thresholds for the sensor to send a report. A report is always sent every 600 seconds. However, in our case, we have set the reporting time to 60 seconds using configuration parameters available in technical specification sheet.
		Accumulated energy (kWh)	By default, there are no accumulated energy thresholds for the sensor to send a report. A report is always sent every 600 seconds. However, in our case, we have set the reporting time to 60 seconds using configuration parameters available in the technical specification sheet.
		Current (A)	By default, there are no thresholds of A for sending a report by the sensor. A report is always sent every 600 seconds. However, in our case, we have set the reporting time to 60 seconds using configuration parameters available in the technical specification sheet.
		Voltage (V)	By default, there are no V thresholds for sending a report by the sensor. A report is always sent every 600 seconds. However, in our case, we have set the reporting time to 60 seconds using configuration parameters available in the technical specification sheet.
Smart switch	Qubino /ES-22 QUBZMN-HBD1	Instant Power (W)	By default, the sensor sends a report when there is a variation of 10 of the committed power and a time interval of 300 seconds.

based on the numbers shows that the collection and recording systems for EB1 worked well in the first four months of the trial compared to the last two months. For EB2, data collection and recording are very poor as the system is switched off

Table 5.2: Important Parameters of the Datasets

Parameter	Definition	Units
home id	energy Box associated with the single house	None
Sensor	sensor associated with the appliance	None
Timestamp	date in datetime format year-month-day	hour:minutes:seconds
sum of energy of power	power measured by the integrated sensor on the quarter of an hour is therefore energy expressed in Wh	Wh
delta energy	energy detected by the sensor meter remains zero until a consumption threshold dependent on the sensor is exceeded, which can represented as Wh or kWh	Wh or kWh

Table 5.3: Appliances Information for Smart Homes Under Study

Smart Home	Energy Meter	Dryer	Computer TV	Air conditioner	House lighting	Dishwasher	Washing machine	Water heater and heat pump	Fridge	Coffee machine
EB1	1	1	4	3	1	1	1	1	x	x
EB2	1	1	x	2	x	1	1	1	x	x
EB3	1	x	1	x	x	1	1	x	x	x
EB4	1	x	x	1	x	1	1	x	1	x
EB5	1	x	1	x	x	x	1	x	1	1
EB6	1	1	x	1	x	1	1	x	1	x
EB9	1	1	x	2	x	1	1	1	x	x
EB10	1	x	x	x	x	1	1	x	1	x
EB11	1	x	x	2	x	1	1	x	1	x
EB12	1	x	x	x	x	1	1	x	1	x

for the first three months and hardly works for the next two months, i.e. September and October, but 76% of the data is recorded in November. Subsequently, both the collection and collection modules worked better for EB3, even in the month of July when 100% of the energy consumption data was collected. Also, for EB4, 100% data was collected in July and 98% in September, and satisfactory percentages for the other months. Also, for EB5, 100% and 96% of data were recorded in July and August respectively, compared to 67% in June. For EB6, the energy consumption is 0% in July, but 97% in August and September. For EB9, EB11, and EB12, the data for the month of September is 98% and for EB9, the data for the month of August is 97%. Moreover, for EB10, as for EB2, the data collection is very poor and

the reason for this type of reading is internet connection problems, so it is better not to consider the data of these houses for the algorithm. Finally, we only consider houses with a good amount of data collected from the smart meter.
Additionally, the involvement of the consumer is made possible by filling Smart Sim questionnaire (SSQ) [132] for their energy uses besides installed sensors. Moreover, SSQ is helpful for knowledge of unmonitored energy consumption when it is not possible to install sensors everywhere in the home.
Overall, 38 different energy consumption devices are identified in the Smart Sim Questionnaire and further subdivided into 9 sectors for better energy consumption disaggregation for unmonitored sectors.

Table 5.4: Data Collected from Energy Meter

Month	EB1	EB2	EB3	EB4	EB5	EB6	EB9	EB10	EB11	EB12
June	94%	0%	90%	73%	67%	0%	24%	9%	71%	70%
July	98%	0%	100%	100%	100%	56%	100%	26%	73%	56%
August	96%	0%	96%	96%	96%	97%	97%	25%	35%	97%
September	91%	29%	98%	91%	73%	97%	98%	98%	98%	98%
October	84%	37%	84%	84%	84%	84%	22%	84%	35%	83%
November	89%	76%	89%	81%	85%	88%	0%	14%	21%	14%

5.3.3 Data processing

Our algorithm requires monthly data, while the starting database had quarter-hour data. For this reason it was necessary to pre-process the data to obtain a separate CSV file for each month and for each house.
The data collected from the Smart Sim questionnaire (SSQ) were arranged in nine sectors according to the type of appliance. Table 5.5 presented the detail of each sector. For each month we summed up all the appliances' energy data and put in their specific sector.

5.4 Energy Consumption Disaggregation Algorithm

The main objective of this algorithm is to provide feedback to the consumer about the total energy consumption of the house by monitoring only a few appliances. The proposed smart homes energy consumption breakdown algorithm has been developed to analyze the energy consumption in detail. The flow chart of the proposed algorithm is shown in Figure 5.3, while the algorithm is shown in Algorithm 1.

Table 5.5: SS formation from SSQ

Energy Sector	Appliances
Lighting	Interior lighting , Outdoor lighting
EL kitchen	Microwave oven, Oven, Grill, Stove
Refrigeration	Fridge-freezer, Cockpit freezer, other fridge
Cooling	Mechanical ventilation, Fan, Portable dehumidifier, Cooling generator
EL heating	Generator for heating
ACS	ELElectric Domestic Hot Water
[Washing, cleaning, ironing, personal care]	Washing machine, Dryer, Dishwasher, Washer dryer, Vacuum cleaner, Electric broom, Iron without boiler, Iron with boiler, Hairdryer, Hair straightener
Computer/Tv	Desktop computer, Laptop, Modem, Inkjet printer, Laser printer
Other Uses	Heating auxiliaries, Cooling auxiliaries, ACS Production Auxiliaries

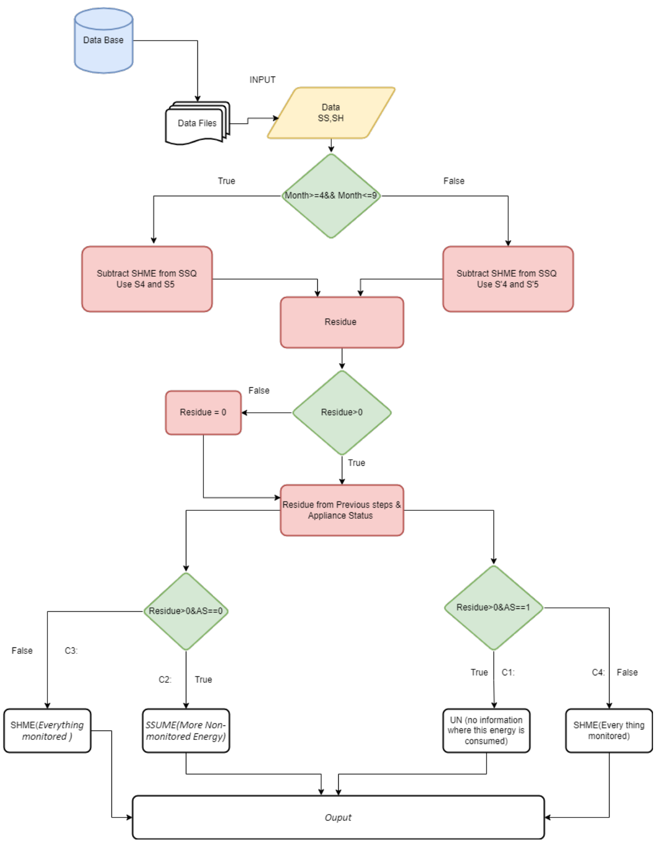


Figure 5.3: Energy Consumption Disaggregation Flow chart

Data:SH: is a csv file containing the monthly energy consumption data from the energy meters and smart plugs installed for a specific energy sector. The appliance status is also included in this file if the plug for the specified appliance is present AS =1 otherwise AS =0. For each month we have different SH files. The structure of the file is shown in Table 5.7.
SS: is a csv file containing the data from the Smart Sim questionnaire for all energy sectors of the sampled households. The structure of the file is shown in Table 5.7.

5.4.1 Equations and Parameters

The energy consumption of nine unmonitored sectors is represented by $S_1, S_2, ...S_9$ and N and M represent the number of devices, and their values could be different for each home. Moreover, the disaggregation can be calculated using all the sets of equations employed in the algorithm given below:

$$S_1 = \sum_{j=1}^{M} SS.Lighting(j) - \sum_{i=1}^{N} SH.Lighting(i) \tag{5.1}$$

$$S_2 = \sum_{j=1}^{M} SS_.kitchen(j) - \sum_{i=1}^{N} SH.kitchen(i) \tag{5.2}$$

$$S_3 = \sum_{j=1}^{M} SS_.Refrigeration(j) - \sum_{i=1}^{N} SH_.Refrigeration(i) \tag{5.3}$$

$$S_4 = \sum_{j=1}^{M} SS_.Cooling(j) - \sum_{i=1}^{N} SH_.Cooling(i) \tag{5.4}$$

$$S'_4 = \sum_{j=1}^{M} SS_.Cooling(j) \tag{5.5}$$

$$S_5 = \sum_{j=1}^{M} SS_.ELheating(j) - \sum_{i=1}^{N} SH_.heating(i) \tag{5.6}$$

$$S'_5 = \sum_{j=1}^{M} SS_.ELheating(j) - \sum_{i=1}^{N} SH_.heating - \sum_{i=1}^{N} SH_.Cooling(i) \tag{5.7}$$

$$S_6 = SS_.ACSEL - SH_.waterheather - SH_.heatpump \tag{5.8}$$

$$S_7 = \sum_{j=1}^{M} SS_{.}[Washing, cleaning, ironing, personalcare](j) - \sum_{i=1}^{N} SH_{.}[Washing, cleaning, ironing, personalcare](i) \tag{5.9}$$

$$S_8 = \sum_{j=1}^{M} SS_{.}Tv/Computer(j) - \sum_{i=1}^{N} SH_{.}Tv/Computer(i) \tag{5.10}$$

$$S_9 = SS_{.}OtherUses = SH_{.}OtherUses \tag{5.11}$$

Table 5.6: Algorithm Parameters

Parameters	**Explanation**
UN	Unknown Energy
SHME	Monitored Energy
SSUME	Unmonitored Energy information
R	Residue of SS - SH
P_k	Monitoring Plugs

Table 5.6 presents the parameters used in the algorithm for energy disaggregation.

5.4.2 Working of Algorithm

In the first step, the algorithm takes two input files SS and SH, and ensures that the files are from the same month. Then, in the second step, the differences are calculated using a series of equations from 1 to 11 for each of the sectors available in the house. This checks for seasonality, as the months of the year, affect the heating and cooling appliances and their energy consumption. The next, third step is the most important for this algorithm, which initially depends on two factors: the backlog of each unmonitored sector and the status of the appliance. If the backlog for a specific sector is negative, the device status does not contribute much and we set this specific unmonitored sector equal to zero because we have more monitored energy consumption for this specific sector, or we can say that everything is monitored and we could communicate this information to the user as SHME subdivision. Also, in the case where the residue of energy consumption from step two is positive for a specific unmonitored sector, the device status plays an important role in assigning the residue to the disaggregation subdivision UN and SSUME. If all

Algorithm 1 Energy Consumption Disaggregation

Step 1:INPUT dataset[SS,SH]
Step 2:Evaluate Differences SS from SH using equations (5.1) -(5.11)
Seasonality Checks [for S_4 and S'_4,S_5 and S'_5]
if $month >= 4$ & $month <= 9$ **then**
    Use S_4,S_5
else if S'_4,S'_5 **then**
end if
Step 3:Disaggregation Allocation to Sectors
[SHME, SSUME, UN]
Checks based upon difference from Step 2 and AS
C1= R $>$ 0 & AS==$\forall$ P_k : UN $\leftarrow$ R
C2= R $>$0 & AS==$\neg$ $\forall$ P_k : SSUME $\leftarrow$ R
C3= R $<$ 0 & AS==$\forall$ P_k : SHME
C4= R $<$ 0 & AS==$\neg$ $\forall$ P_k : SHME
Step 4:Output
Disaggregation=[SHME+SSUME+UN]

devices are monitored by sensors, e.g. for S1, and the energy consumption reported by SS is higher, which is not possible, we can assign this extra amount of energy to the UN subdivision. If not all devices of sector S1 are monitored, but only a few sensors are available, we can add the extra amount of energy consumption to the subdivision SSUME, i.e. there is a possibility that the extra energy is consumed by other devices of this sector. Similarly the whole process is repeated for all other energy sectors ($S_2, ...S_9$).

Finally, step 4 shows the total disaggregation of the house, i.e. the combination of the subdivisions SHME, SSUME, and UN. It can be concluded that with very little information about the monitored energy consumption and with the help of SS it is possible to get a complete knowledge of the energy consumption profile and provide feedback to the consumer. The algorithm is written in Python 3.8 and uses various libraries.

5.4.2.1 Example Scenario

To evaluate the functioning of the algorithm, an example is presented to explain the disaggregation of energy consumption. The smart house under consideration is equipped with a smart meter that is responsible for measuring the total energy consumption of the house. In addition, three [TV, dishwasher, washing machine] smart sockets are installed in the same house for specific measurements, which are aggregated for a specific energy consumption sector, as mentioned earlier. Furthermore, an important parameter of the disaggregation algorithm is the data provided by the consumer on SS for the same house and for a whole year.

Table 5.7: EB-3 Summarized Monitored and Unmonitored from dataset

SH	kWh		SS	
Meter	164.34	AS	Enerfy Sector	September
Meter FV	0	0	Lighting	5
Conditioner	0	0	EL kitchen	11
Switch Conditioner	0	0	Refrigeration	12
Dishwasher	9.16	1	Cooling	0
Fridge	0	0	EL heating	0
Fan heater	0	0	ACS EL	0
Coffee machine	0	0	Washing, cleaning, ironing, personal care	73
Vacuum	0	0	Computer / TV	31
Iron	0	0	Other Uses	6
TV	27.48	1		
Thermomix	0	0		
Microwave	0	0		
Washing machine	9.50	1		
Lamp	0	0		
Kitchen	0	0		
Dryer	0	0		
Water heater and heat pump	0	0		
Oven	0	0		

Table 5.7 illustrates the components from the datasets SH and SS for the month of September, that is stored in CSV files used as algorithm input. In this example, SH displays the energy consumption monitored by three separate plug-in devices. In addition, the device status is an essential algorithm parameter. If a certain device is present during the monitoring period, a 1 is assigned; otherwise, a 0 is assigned. The other data set file comprises the questionnaire data that reveals the energy usage of a certain sector.

Table 5.8: Energy consumption Disaggregation Monitored and Unmonitored

Final Disaggregation			
SSUME		SHME	
	kWh		kWh
Lighting	6.43	Conditioner	0
EL kitchen	14.15	Switch Conditioner	0
Refrigeration	15.44	Dishwasher	9.16
Cooling	0	Fridge	0
EL heating	0	Fan heater	0
ACS EL	0	Coffee machine	0
Washing, cleaning, ironing, personal care	69.92	Vacuum	0
Computer / TV	0	Iron	0
Other Uses	6.0	TV	27.48
UN	6.0	Thermomix	0
		Microwave	0
		Washing machine	9.50
		Lamp	0
		Kitchen	0
		Dryer	0
		Water heater and heat pump	0
		Oven	0
Total			164.34

The results of our method are shown in Table 5.8, which contains all of their specifics. In general, three primary components must be accounted for in the output: the first is SHME, which displays the total amount of energy consumed by three distinct appliances across two energy sectors. Both the washing machine and the dishwasher come from the S_7 sector, which is responsible for washing, cleaning, ironing, and personal care. In addition, the amount of energy used by a washing machine is 9.50 kWh, while the energy used by a dishwasher is 9.16 kWh. The

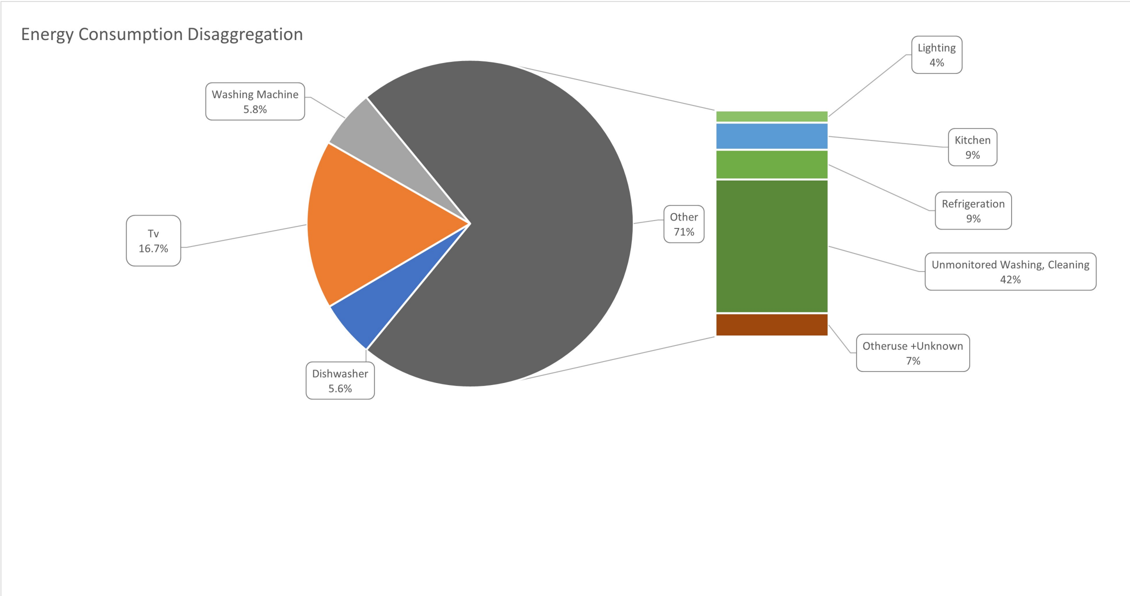


Figure 5.4: Energy Consumption Disaggregation

remaining information for this sector comes from SSUME, which indicated a total of 69.92 kWh.

Secondly, if there is more than one TV or computer in a given Home, the data must be combined to show only the total energy consumption of that computer/TV sector. However, in this Home, we only have one TV and it consumes 27.48 kWh of energy. In the SH files, the parameter "AS =1" confirms that this sector is shown with all plugs. So the algorithm showed 0 kWh in SSUME, which means that everything has been monitored.

Additionally, Because there are no plugs available for the other areas of energy consumption, all of the energy that is used must originate from an unmonitored portion of the output. The Refrigeration sector use15.44 kWh of energy, the Lighting sector used 6.43 kWh, and the Kitchen sector used 14.15 kWh of energy. In addition, the cooling and heating sector did not consume any energy as ACS did, and the value of 6 kWh is unknown, which is the same as the value for the other uses sector. Lastly, It can be observed that our algorithm provides us the information about 96.8 % of consumed whereas SSUME provides information on 68% of the energy consumed in the Home. 28% energy presented by SHME subdivision by means of sensors and 4% information comes from UN subdivision

Figure 5.4 shows a graphical representation of the energy use of EB-3 which further demonstrated SSUME 68% disaggregation sectors in more detail where 4%

energy consumed is by the lighting sector. The kitchen sector and refrigeration sector consumed 9% of energy followed by 47% for washing, cleaning, ironing, personal care sector and remaining 3% goes to another use sector.

5.5 Results and Analysis

The Algorithm is employed on different Homes for its performance and monthly feedback to the consumer for their energy consumption patterns. However, here in this section, only a few are presented.

Table 5.9: Energy Consumption Disaggregation EB-3 (Month of June)

Energy Sectors	**Algorithm Results**	**Actual Energy**	**Absolute Difference**
Lighting	3.9%	3.6%	0.29%
EL kitchen	8.6%	8.0%	0.64%
Refrigeration	9.4%	8.7%	0.70%
Cooling	0.0%	0.0%	0.00%
EL heating	0.0%	0.0%	0.00%
ACS EL	0.0%	0.0%	0.00%
Washing, cleaning, ironing, personal care	53.9%	52.9%	1.00%
Computer / TV	16.7%	22.5%	5.74%
Other Uses	3.7%	4.3%	0.6%
	monitored	28%	
	unmonitored	68.1%	
	unknown	3.8%	

Table 5.9 presented the energy consumption disaggregation algorithm output compared with the data acquired by the actual energy reported by the user. As discussed in the previous section, this EB-3 has three plugs beside an energy meter. The Lighting sector of this home consumed 3.9% of overall energy consumption, which has a marginal difference compared to figures provided by the user, which is 3.6%. Subsequently, the kitchen appliance reported 8.6% of consumption compared to 8.0% of data from the user values. Furthermore, in the case of the Refrigeration sector, the difference of comparison is 1.1 %, where the algorithm reported consumption of 9.4% of overall consumption. However, cooling, heating, and ACS have 0% consumption for September. The washing, cleaning, ironing, and personal care

sector reported 53.9% from the algorithm, whereas 52.9% is extracted from user information. The Computer/Tv sector energy consumption by the algorithm is 16.7% and 22.5% from the questionnaire, but an important factor needs to be discussed. The questionnaire also approximated values, and the gap between these values could be reduced by employing different optimization methods. The algorithm results from Table 5.9 provide phenomenal energy knowledge that can improve energy knowledge from 28% to approximately 97%.

Table 5.10: Energy Consumption Disaggregation EB-12 (Month of September)

Energy Sectors	Algorithm Results	Actual Energy	Absolute Difference
Lighting	4.6%	4.3%	0.24%
EL kitchen	16.1%	15.2%	0.84%
Refrigeration	27.4%	16.3%	11.09%
Cooling	2.3%	2.2%	0.12%
EL heating	0.0%	0.0%	0.0%
ACS EL	0.0%	0.0%	0.0%
Washing, cleaning, ironing, personal care	13.4%	40.2%	26.83%
Computer / TV	19.5%	18.5%	1.02%
Other Uses	1.6%	3.3%	1.68%
	monitored	41%	
	unmonitored	44%	
	unknown	15%	

The output of the Energy consumption disaggregation algorithm compared to the data collected through the questionnaire in Table 5.10 is presented for EB-12. This EB-12 has three plugs in addition to an energy meter. This home's lighting sector utilized 4.6 % of total energy usage, a slight variance from the 4.3 % stated by the user while filling out the questionnaire. Following that, the kitchen appliance indicated 16.1% usage vs 15.2% from the questionnaire. Furthermore, 27.44% of energy is consumed reported by the refrigeration sector of the algorithm, where the questionnaire results indicated the use of 16.3% of total consumption. Afterward, algorithm results for cooling are 2.3% compared to 2.2%. However, cooling, heating, and ACS all had zero percent usage for this home. There is a big difference between Washing, cleaning, ironing, and the personal care sector. finally, It can be discovered that our algorithm provides us the information about 85 % of consumed whereas unmonitored sectors SS provide information on 44% of the energy consumed in the Home. 41% energy is monitored by means of sensors but we have no information about 15%.

Table 5.11 presented the Energy consumption disaggregation comparison of home (EB-6) under experimentation with data collected from the consumer of the home. In the energy sector Lighting reported 6.9% consumption by means of the algorithm compared to 7.1% of information provided by the user. Likewise, the kitchen sector consumed 14.1% of the total energy according to our developed algorithm. Moreover, the results of all other sectors can be seen in the table. Lastly, 65% of energy consumption information is provided by the algorithm whereas only 12% of the information comes from sensors which showed the productivity of the algorithm.

Table 5.11: Energy Consumption Disaggregation EB-6 (Month of September)

Energy Sectors	**Algorithm Results**	**Actual Energy**	**Absolute Difference**
Lighting	6.86%	7.11%	0.25%
EL kitchen	14.13%	14.64%	0.52%
Refrigeration	7.14%	5.02%	2.12%
Cooling	8.53%	8.79%	0.25%
EL heating	0.32%	0.00%	0.32%
ACS EL	0.00%	0.00%	0.00%
Washing, cleaning, ironing, personal care	4.8%	38.5%	33.73%
Computer / TV	23.8%	24.7%	0.87%
Other Uses	1.47%	1.26%	
	monitored	12%	
	unmonitored	55%	
	unknown	33%	

Table 5.12: Energy Consumption Disaggregation EB-3 (Month of July)

Energy Sectors	Algorithm Results	Actual Energy	Absolute Difference
Lighting	2.22%	2.18%	0.04%
EL kitchen	4.89%	4.80%	0.09%
Refrigeration	5.78%	5.68%	0.11%
Cooling	36.93%	36.24%	0.69%
EL heating	0.00%	0.00%	0.00%
ACS EL	0.00%	0.00%	0.00%
Washing, cleaning, ironing, personal care	32.79%	33.19%	0.39%
Computer / TV	9.23%	13.97%	4.74%
Other Uses	3.66%	3.93%	0.27%
	monitored	12%	
	unmonitored	55%	
	unknown	33%	

Table 5.13: Energy Consumption Disaggregation EB-3 (Month of October)

Energy Sectors	Algorithm Results	Actual Energy	Absolute Difference
Lighting	5.2%	4.8%	0.39%
EL kitchen	8.2%	7.6%	0.62%
Refrigeration	9.73%	9%	0.73%
Cooling	0%	0%	0.00%
EL heating	0%	0%	0.00%
ACS EL	0%	0%	0.00%
Washing, cleaning, ironing, personal care	52%	52.5%	0.46%
Computer / TV	7.9%	22.1%	14.19%
Other Uses	3.2%	4.1%	0.94%
	monitored	20%	
	unmonitored	66%	
	unknown	14%	

Table 5.12 and 5.13 presents the energy consumption disaggregation of Home EB-3 for the month of July and October. The absolute difference between the energy acquired from the user (Actual energy) and the algorithm results(Disaggregated energy) shows greater variations for the computer/Tv sector for both months. However, small variations are noted for all other energy sectors.

The energy disaggregation comparison of five homes is demonstrated in Figure 5.5. Here, it can be seen that EB-6 has the lowest SHME compared to other EB under experimentation, whereas EB-3 has the highest SSUME and, in the case of EB-11, zero UN. Moreover, the pattern of all the homes is the same less SHME energy consumption and more SSUME, which seconds our motivation that from less SHME to more knowledge of overall energy consumption with the addition of SSUME and the subdivision SSUME is also further disaggregated as already shown as an example in Figure 5.4.

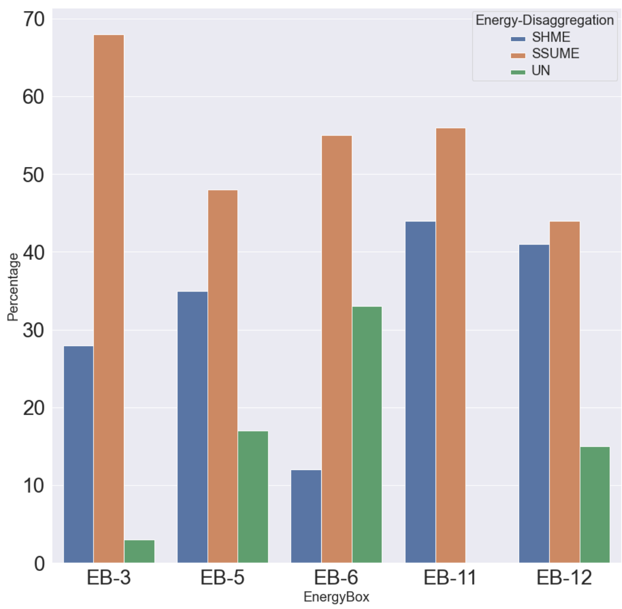


Figure 5.5: Energy Disaggregation Comparison

The results of three different homes have been presented and compared with consumers' information, showing excellent algorithm performance with slight divergence. The results of the algorithm for EB-3 provide us with phenomenal energy knowledge that we can improve energy knowledge from 28% to approximately 97% compared to four other homes under experimentation. Further work can be done to reduce the unknown sector values. By doing this, energy information will be enhanced, narrowing the energy consumption gap between actual and obtained consumption.

The algorithm has been applied to data collected from different homes with different numbers of devices. Furthermore, the algorithm was used in the same house in other months. Regarding the performance analysis, we added an absolute difference metric to each results table, which shows the error for each sector's disaggregated energy for each home.

Table 5.14: Performance of the proposed method with different homes

Month	HomeEB1	HomeEB3	HomeEB5	HomeEB6	HomeEB9
June	75%	83%	-	-	-
July	90%	95%	93%		94%
August	91%	93%	92%	85%	95%
September	72%	96%	83%	88%	90%
October	78%	86%	79%	-	-
November	58%	90%	96%	-	-

Table 5.14 shows how the algorithm disaggregates the energy for five homes over different months. Each home has a different number of appliances and a different number of users. For home EB3 in September, the algorithm gave the user 96% of the energy information. However, in the case of home EB1 in November, energy consumption disaggregation results show 58% of energy knowledge, which is also acceptable for user feedback applications. The comparison of the proposed method is concerned. To the author's knowledge, no dataset is available similar to ours. So it is not possible to apply the method for comparison. The other methods cannot handle our dataset because we have incorporated user feedback. Every home in the dataset is different (in terms of the number of users and devices installed). So we have provided a comparison of several homes where the energy consumption disaggregation ability of homes is shown.

5.6 Experimental Evidence of Energy Consumption Reduction

The energy consumption feedback based on the proposed disaggregation method is provided to all home users. In addition, data on electricity consumption has been collected during the proceeding years beside data used for feedback. So, it was possible to validate the proposed methods' findings regarding energy savings. Figure

5.6 visually represents the typical monthly consumption and the proportion of energy savings. Even though the greater incidence was found in single or two-component families, where the effects of the individual user's lifestyle changes and habits are more evident, the results suggest that the average savings were approximately 10% for each household. The results are encouraging, particularly when considering that most of the credit goes to a change in user behavior.

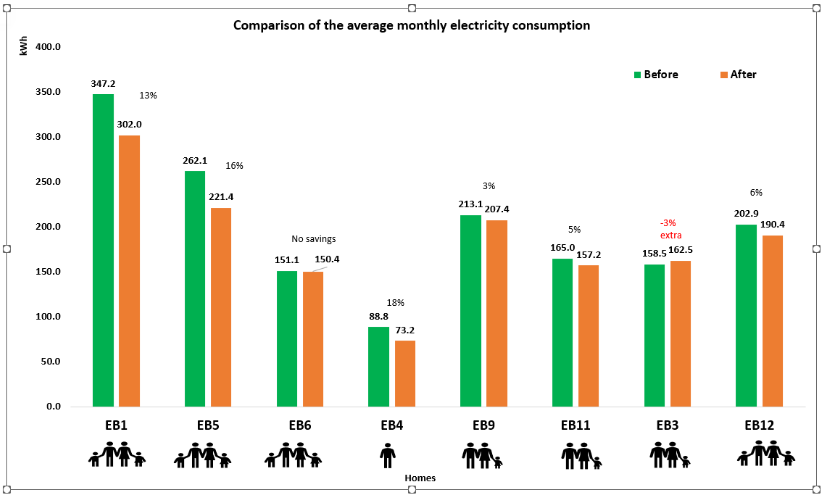


Figure 5.6: Comparison of the average monthly electricity consumption before and after the proposed method feedback.

5.7 Summary

Energy consumption disaggregation for smart homes and buildings is always exciting for researchers. In this chapter, we intend to provide a data-driven disaggregation algorithm to further explore the energy consumption monitoring field to give the user a better breakdown of their energy use. After the data collection, it was transformed so that our developed algorithm can accept and work on the provided data.

Then, we proposed the method that disaggregated the total energy consumption into a number of distinct sectors, as indicated by the equations described in section 5.4. Following that, for example, a scenario was presented to further explain the proposed algorithm. Our proposed method yields desirable outcomes for disaggregation of total energy usage, as illustrated by the graphs and tables given in section 5.5. Finally, experimental evidence of energy consumption saving based on proposed approach is presented in section 5.6, which shows the achievement of our second objective.

Chapter 6

Energy Consumption Patterns Detecting Technique for Household Appliances for Smart Home Platform

This chapter is the third contribution to the thesis, which is energy consumption patterns detecting technique for Household Appliances. The chapter is organized as follows. The related work is presented at the start, tracked by a Methodology that contains the description of the platform, and then information on Datasets and algorithms is explained in detail—followed by described results and a discussion of our proposed algorithm. Finally, the summary of the chapter is presented.

6.1 Related work

The diffusion of smart meters, low-cost sensors, and smart appliances has paved the way for novel energy management strategies including communication and interaction among users, devices, and the grid. Whereas greater resolution data is necessary for more precise energy analysis, raw consumption data by itself is not enough to provide details on the causes of energy demand and what can be done to reduce it.

Demand–response (DR) strategies can be used to boost the efficiency of smart grids. There are several DR initiatives in the literature aimed at lowering customers' bills and reducing grid loads. Authors in [5] devised a homomorphic encryptionbased alternate direction technique of multipliers approach to address cost-aware appliance scheduling optimization in a distributed way and scheduled home appliances without compromising users' privacy. They show that the suggested secure appliance scheduling for a flexible and efficient energy-consumption method, termed SAFE, significantly decreases power costs while protecting users' privacy through an intensive simulation research by using real-world datasets. Appliances' operating mode identification, which uses the cycle-clustering technique proposed in [88] is a smart home energy-management system primary approach based on sensed power consumption values, which enables DR by allowing users to employ energy-saving appliance operation modes. Cycles from an appliance single-usage profile, are extracted and reshaped into characteristics in the form of clusters of cycles. By using K-nearest neighbours, these attributes are then used to determine the operating mode during each occurrence. Identification of operation modes is deemed fundamental for several possible smart DR applications inside HEMS. Authors in [133] presented the K-means clustering technique to classify different modes of operations of home appliances. To validate the proposed method, a case study based on a refrigerator was presented. By combining an automatic switch-off system with load balancing and a scheduling algorithm, authors in [134] proposed a smart-home energy-management system that reduces energy waste. The load-balancing technique operates within defined limitations to keep the household's total energy consumption within a certain limit. The least slack time (LST) method is used to schedule appliances, by considering the user's comfort. The simulation results demonstrated a considerable reduction in residential energy consumption aided by an automatic switch-off mechanism thanks to the suggested LST-based energy-management system (EMS). Support vector machines were used in [135] to identify load characteristics of appliances in various operating modes, and a smart home energy-conservation system was constructed to test the approach and validate the proposed EMS. The authors of [136] applied an artificial neural network to identify usage patterns of some common household electric appliances from daily profiles of energy records obtained with 15-min granularity, and the proposed method showed effectiveness when applied to several daily load diagrams captured by a few residential meters. The iterative disaggregation approach based on consumption patterns presented in [137] is a method that merges the fuzzy C-means clustering technique, which provides an initial working condition, with subsequence searching dynamic time warping, which recovers single-source energy usage based on typical power-consumption patterns.

The results demonstrate that the suggested method properly disaggregates power usage and is suited for situations when many appliances are used at the same time. To decrease peak power and improve users' comfort, authors of [138] proposed an efficient home energy-management controller (EHEMC) based on the genetic harmony search algorithm. With real-time electricity pricing and critical peak pricing tariffs, they evaluated EHEMC for a single house and multiple dwellings. The study [139] provided a nonintrusive affinity propagation clustering algorithm based on factor graph model and the belief propagation theory, and the results revealed that the algorithm correctly recognizes the basic and combination classes of household appliances. This strategy laid the groundwork for power-management firms to efficiently and effectively allocate electricity. The study [140] proposed a novel method for assessing the energy consumption of various appliances, and it was utilized to develop a recommendation system that would advise renters on how to reduce their usage.

Clustering, association rule mining, and artificial neural networks were all used to perform data-mining tasks. Authors in [141] presented various unsupervised machine learning techniques to find energy-usage patterns in a smart home. Moreover, they provide solutions for smart metering systems that may help to: (1) raise energy awareness; (2) assist accurate use forecasts; and (3) give input for demand response systems in homes that provide users with timely energy saving advice. In [142], three data-analysis processes and a data-mining framework, i.e., data classification, cluster analysis and association rule mining, were proposed for efficient data analysis of buildings. A systematic approach for providing immediate feedback and recommendations to building occupants to assist them in taking appropriate action to minimize energy use was developed in [143]. For a specific building, the notion of a reference building (RB), i.e., an energy-efficient building, was established, and RB was created by using data-mining techniques that included clustering analysis and neural networks. The performance of a building was determined in comparison to a reference building, which provided feedbacks to suggest occupants the appropriate actions to enhance building energy performance (e.g., turning off lights or heating and air conditioning, etc.). In some other studies [144, 145], the nonintrusive load-monitoring (NILM) method was designed to identify the activation of a target appliance by analysing the recorded active power transient response and estimating its consumption in real time.

From the literature studies, it has been observed that most of work reported is related to overall load-level consumption information, cost-aware appliance scheduling, load balancing, and a scheduling algorithm based upon energy meter data. To the best of our knowledge, we have not found simple models aimed at extracting patterns such as the number and duration of operations, cycle disaggregation for appliances that have cyclic operation (e.g., washing machine, dishwasher), and energy consumption throughout various time periods, which can be easily integrated into IoT platforms dedicated to residential users.

The key contributions of chapter are as follows:

1. A pattern detecting method for different appliances (cyclic and non-cyclic) has been developed that is quite simple, computationally less complex, and can accurately disaggregate energy consumption cycles occurred in the close succession. Moreover, the algorithm developed herein could be applied to any device dataset.

2. The developed algorithms have been implemented in a code used to process the data acquired by a IoT platform developed by ENEA and currently operating in some dwellings, aimed at fostering user engagement and energy awareness by providing customised feedbacks to the users.

3. The DHOMUS platform uses open and interoperable protocols for data acquisition at the device and HEMS level and therefore, can be easily replicated in broader contexts.

6.2 DHOMUS Platform

Data Homes and Users (DHOMUS), is a platform developed by ENEA (`https://dhomus.smartenergycommunity.enea.it`, accessed on 1 December, 2022) that is dedicated to residential users for smart home applications. The layout of the smart home is depicted in Figure 6.1 [146]. The main objective of DHOMUS is to make users aware of their energy data, to let them understand how much energy they consume and why, to support reduction of both consumption and costs, thereby contributing to decrease their impact on the environment, to increase energy awareness, and to transform residential users into active subjects that contribute to the stability of the grid. The DHOMUS platform currently collects and analyses real-time data from smart homes located in Rome. They consist of 24 homes equipped with a kit of commercial sensors (their characteristics are described in Table 6.1), which monitor the electrical consumption of the home meter and of selected appliances, as well as the presence of people and indoor comfort. The management of all these devices is wireless, based on Z-Wave protocol, and the gateway is connected to the Internet through the energy box, which collects, integrates, and sends sensors' data to the cloud platform. As the electrical data acquisition time varies from sensor to sensor, postprocessing is performed to calculate their 15-min averages, similar to the method used for smart meters.

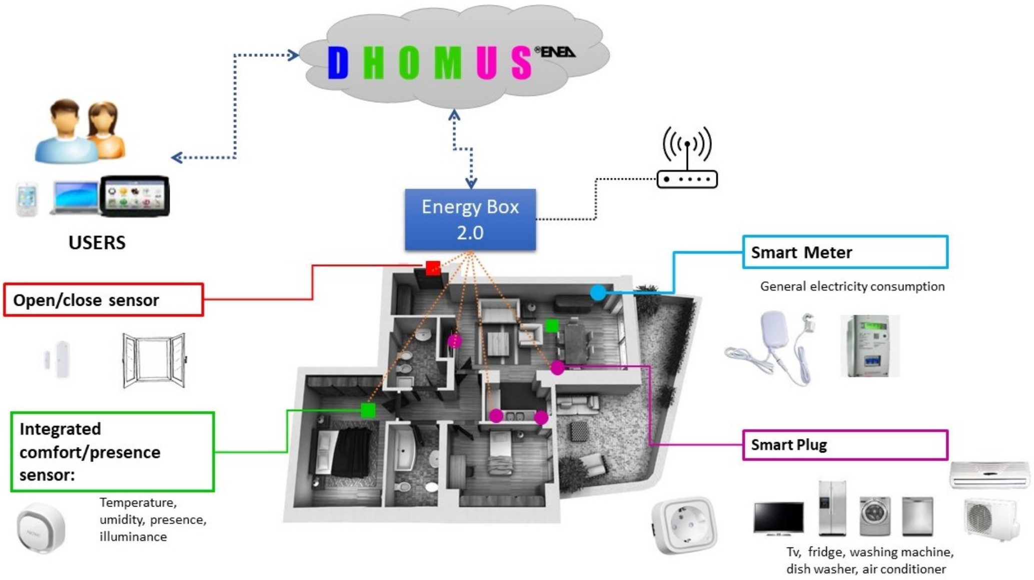


Figure 6.1: Smart home: representation of sensors.

Table 6.1: Sensors used in the experimentation.

Sensor	Manufacturer/Model	Measured Quantity
Home energy meter	Aeotec/ZW095-C	power, energy
Smart plug	Aeotec/switch 7	power, energy, current, voltage
Sensor and integrated or comfort-presence	Aeotec/Multisensory 6	temperature, illuminance, humidity, movement
Door/window contact	Aeotec/ Door window sensor	opening/closing

6.3 Datasets

The model has been calibrated and validated based on two datasets of electrical consumption of homes and single appliances monitored by DHOMUS platform, in the following Dataset A and Dataset B. Dataset A includes energy measurements of 58 appliances in 14 homes over 2.2 years, collected with a 15-min granularity. The start and end times of the measurements are not the same for all dwellings. Figure 6.2 depicts the available data for the appliances included in Dataset A. For most of the appliances, the data are recorded over the period from February 2018 to March 2020. Fridges, dishwashers, washing machines, and TVs are the most commonly monitored appliances.

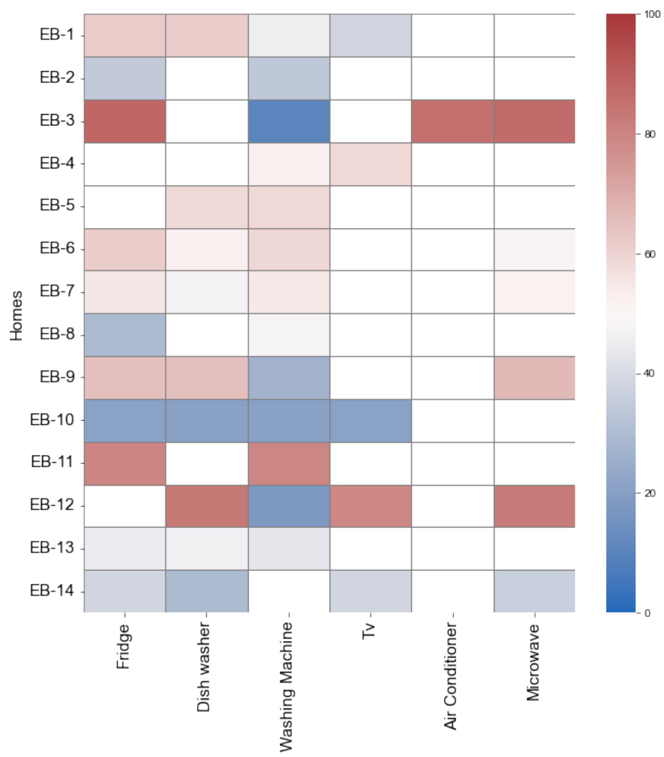


Figure 6.2: Percentage of available data for Dataset A

Dataset B includes measurements of 48 appliances from 10 homes from June 2021 to November 2021 and data were collected with a 15-min granularity. Figure 6.3 shows the available data for the appliances in dataset B. Dishwashers and washing machines have the most data available.

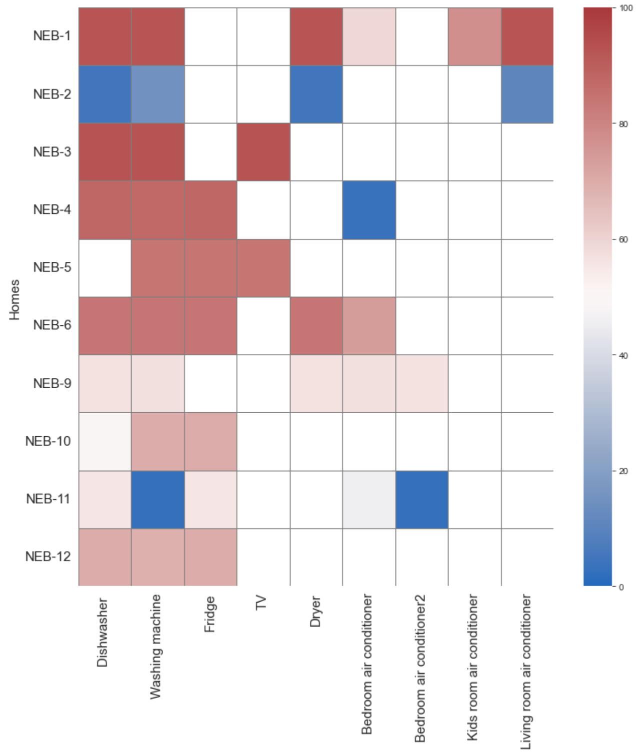


Figure 6.3: Percentage of available data for Dataset B

6.4 Methodology

6.4.1 Appliance Data Analysis Model

A model has been implemented in the MATLAB® environment to analyse the data collected by the data acquisition system interacting with ENEA's DHOMUS platform. The model performs data mining and a statistical analysis of the energy use of the dwelling energy meter and of single appliances, distinguishing between those that operate with cycles (e.g., washing machine, dishwasher, tumble dryer) and those that operate continuously (e.g., air conditioners, TV, etc.). The model can detect the operation mode (standby, operating, off, monitoring sensor disconnected/not working) of all domestic appliances, as well as the start, end, and length of operation. Moreover, a statistical analysis of electricity consumption is also performed for each operation, determining the total consumption, the average energy consumption on the reference time step, which is equal to 15 min, and the maximum and minimum energy. Additionally, consumption is split into time slots set by the ARERA Authority [147], as well as customised time slots between 8 a.m. and 8 p.m.

(with a minimum resolution of one hour). The model then carries out a monthly analysis of the operation and calculates the number of times the appliance is working, the duration of operation, and the total monthly energy, and then it compares these outputs with reference values, based on similar users or benchmarks derived from the technical literature, in order to make assessments on the actual use of the appliance. Additional outputs are calculated for domestic appliances that operate in a cyclic mode. The model is simple, fast, and it can distinguish nearby cycles in which one cycle runs immediately after another. It calculates the number of times the appliance has been used and the corresponding consumption; this information is used to provide households with feedback aimed at improving rational energy use. The algorithms are detailed as follows.

The flow chart in Figure 6.4 represents the structure of the numerical model. First, the algorithm imports monitoring data from .csv or .txt files exported from the data acquisition system. Data for various homes, different household appliances, and for an arbitrary time span could well be stored in the file. Input files have the following structure (taken from an example file), and the description of parameters is shown in Table 6.2:

```
home_id, sensor, date,sum_of_energy_of_power, delta_energy, last_value_switch
EnergyBox1, fridge plug,2019-01-01 00:30:00.0000,8.8,0,0
EnergyBox1, fridge plug,2019-01-01 00:45:00.0000,8.8,0,0
EnergyBox1, fridge plug,2019-01-01 01:00:00.0000,8.8,0,0
```

The quantities `sum_of_energy_of_power` and `delta_energy` provide similar information on the consumption of the appliance; therefore, the model uses the former for the energy calculations.

Module Smart Home Analyzer (SHAM) (i.e., module 1 in Figure 6.4) imports the monitoring data in a table format. Then, it sequentially calls the modules explained below. The call syntax is as follows:

```
smart_home_analyser_v2(B_id, B_name, elettrodom, energy_data).
```

The required inputs are `B_id` , the ID of the building which represents the energy box (e.g., EB-1 to EB-14 for dataset A, and NEB-1 to NEB-12 for dataset B), `B_name` , the acronym of the building used to plot the results and save the results, `elettrodom`, the name of the appliance (e.g., “dishwasher”), and `energy_data`, a table with the imported quarter-hour monitoring data. The code can recognize all possible appliances installed in the home.

Table 6.2: Input data for the model.

Parameters	Description
home_id	Name of the energy box
sensor	Name of the sensor associated with the appliance
date	Datetime
sum_of_energy_of_power	Energy provided in Wh on a quarter-hour basis
delta_energy	Energy counter, in Wh or kWh
last_value_switch	sensor parameter (not used)

The module detects the dwelling and the appliance provided as input, and if it cannot find the combination “house, appliance”, then it provides an error message and ends as shown in Figure 6.4.

The function “smart home” (module 2 in Figure 6.4) is automatically called if the dwelling and the appliance are found in the dataset. This function sorts the imported energy data over time and removes any duplicate; moreover, it verifies the unit of measurement of the energy quantities and, if necessary, converts from kWh to Wh. The module then produces a table-formatted array containing the input quantities sorted through time, checked, and with required units.

The function “appliances” (module 3 in Figure 6.4) automatically receives the following inputs from the previous function: `B_name` is the acronym of the dwelling (e.g., “EB-1”), `B_appl` and `appl_name` are the abbreviation and label of the appliance, respectively, and they are used to plot and save the results, `cyc_on` is a boolean variable that indicates whether the appliance has a cyclic operation or not, and `t_cdz_i` is the table array produced by the "smart home" function.

Furthermore, the “appliance” module performs data analysis and data mining. First, it provides a set of operating features that are used to calculate the appliance standby energy and to setup customised daytime periods. Additional parameters are adopted for the study of individual cycles if the appliance has a cyclic operation:

- Minimum cycle duration, i.e., shorter operations are neglected;
- Minimum cycle energy consumption, i.e., operations with a smaller consumption are neglected;

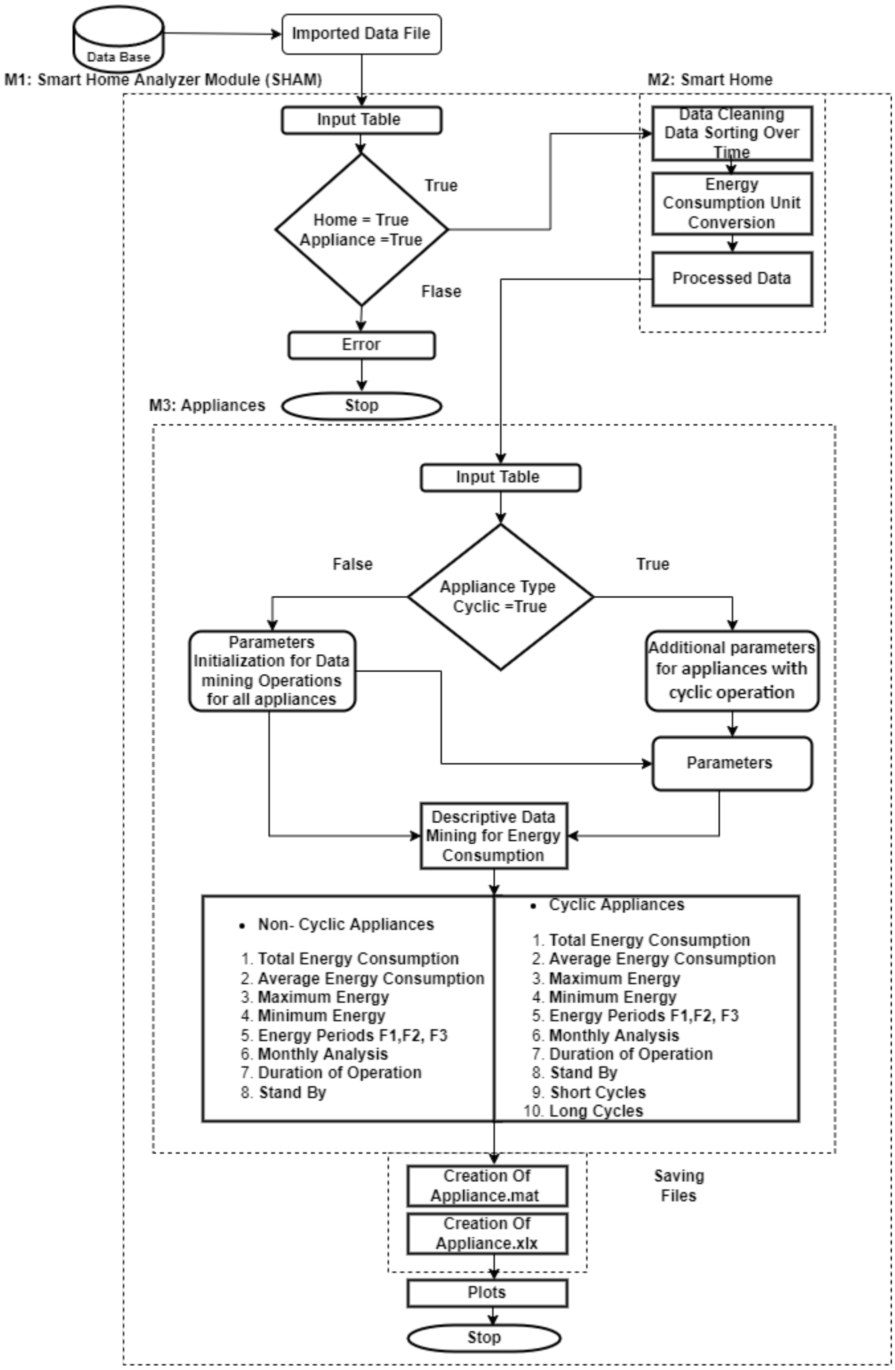


Figure 6.4: Flow chart of the numerical model.

- Parameters for the separation of nearby cycles;
- Threshold to skip the analysis for those months when insufficient data are available;

- Energy threshold to distinguish low consumption cycles;
- Time threshold for short cycles; and
- Time threshold for long cycles.

These parameters depend on the type of appliance, and therefore they require calibration.

The module creates a table with energy consumption and corresponding date-time, which starts on the first day of the first month and ends on the last day of the last month recorded in the imported dataset. This operation is useful to manage full months, especially to account for new sensors and for relying on a comparable data structure because sensors may be connected to different appliances in different moments.

6.4.2 Information Analysis

The model relies on consumption during standby mode in order to determine whether the appliance is operative or not. Standby energy (U) is calculated by the following equations,

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} E_{i\leqslant t} \tag{6.1}$$

$$F = \nabla \widehat{F}_n(t) \tag{6.2}$$

$$U = max(F), \tag{6.3}$$

where $\widehat{F}_n(t)$represents the empirical cumulative distribution of energy consumption $E_{i\leqslant t}$ function over time, and F represents the gradient of $\widehat{F}_n(t)$. The algorithm assumes that switching from standby to normal operation corresponds to the maximum energy gradient; therefore, the standby consumption is calculated as the energy that determines the maximum gradient of the cumulative distribution of measured consumption.

In actual operation, the appliance may perform several cycles in close succession (e.g., two washes, one immediately after the other for the washing machine). The model can separate cycles in close succession based on their energy-consumption profile. The disaggregation of cycles is computed with the following equations,

$$Y = \nabla \theta_A \tag{6.4}$$

$$\Delta Y = ZCI(Y), \tag{6.5}$$

Where θ_A is the energy consumption of aggregated cycles, Y is the gradient of consumption and ΔY is the position over time of the relative maxima of energy consumption in aggregated cycles. The ZCI(Y) function interpolates from the points it identifies near the zero crossings to linearly approximate the actual zero crossings. Maxima characterized by energy consumption above the value that corresponds to the percentile set as a parameter of the model (e.g., 85% of the consumption of the aggregated cycle) identify separate cycles and are used by the algorithm to disaggregate nearby cycles. Typically, energy consumption is higher during the first stage of the cycle (i.e., water heating) for washing machines and dishwashers.

As anticipated, the model neglects operations that last less than or that consume less than the thresholds set as input parameters in order to exclude those situations that may not represent real operating cycles. The corresponding data (spurious data) are stored separately in the results.

Then the model allocates energy consumption in time slots. In particular, hours between 8 a.m. and 8 p.m. can be grouped in a customized number of time slots provided as input. Moreover, the algorithm accounts for the standard time slots defined by the Italian Energy Authority (ARERA) as follows [147]:

- F1: from Monday to Friday, from 8 a.m. to 7 p.m. (except holidays);
- F2: from Monday to Friday from 7 a.m. to 8 a.m. and from 7 p.m. to 11 p.m., Saturday from 7 a.m. to 11 p.m., (except holidays); and
- F3: from Monday to Saturday, from 12 a.m. to 7 a.m. and from 11 p.m. to 12 a.m., Sundays and holidays.

For each individual period and cycle of operation, the following quantities are calculated: total energy consumption, average 15-min energy consumption, maximum and minimum consumption, and maximum/minimum ratio as an indicator of the variability of the consumption during operation.

The code then performs a monthly analysis, and calculates the following quantities:

a. Number of records in the dataset;

b. Number of NULL records, i.e., with sensor disconnected or inactive;

c. Number of records with zero consumption;

d. Number of records when the appliance is in standby mode;

e. Number of spurious records, i.e., characterized by either duration or consumption below the input thresholds;

f. Number of records when the appliance is operative;

g. Number of records when the appliance is operative during F1 time slot;

h. Number of records when the appliance is operative during F2 time slot;

i. Number of records when the appliance is operative during F3 time slot; and

j. Number of records when the appliance is operative during the i-th custom time slot, for all daytime slots set as input.

Similar quantities are calculated with reference to the monthly energy consumption. For household appliances with cyclic operation, the following additional quantities are calculated by differentiating cycles according to energy consumption and duration:

- Cycles with consumption above the minimum input value;
- Cycles with duration less than the input threshold S_{d1} (short cycles);
- Cycles with duration between boundaries S_{d1} and S_{d2} (medium cycles); and
- Cycles with duration greater than the input threshold S_{d2} (long cycles).

Moreover, for single cycles and for aggregated cycles (i.e., cycles before the application of the algorithm that distinguish cycles in close sequence) the following monthly quantities are calculated:

- Percentage of records in the month;
- Number of cycles recorded;
- Number of cycles extrapolated for every month, in the presence of missing data (i.e., the sensor returns NULL) and within the limits of acceptability set as input parameter (e.g., months with only 10% of records are discarded, because extrapolation over the month would not be reliable);

- Number of cycles extrapolated on annual basis, that comply similar condition on missing data as for monthly extrapolation;
- Average cycle duration;
- Average cycle consumption;
- Maximum consumption; and
- Minimum consumption.

Finally, the code performs a series of energy balance checks and then results are printed and plotted in graphs, and the relevant variables are saved in .xlsx and .mat files.

6.5 Results and Discussion

In this section, the analysis on the operation of a washing machine is used as an example to describe the results and the logic of the model. Indeed, this electrical appliance has many features that can be representative of other appliances, i.e., it is cyclic, and it can be used to illustrate the main features of the model. Similar remarks apply to other appliances.

6.5.1 Calibration of Parameters

As the algorithm can analyse all domestic appliances, the control parameters need to be tuned before deploying the model to the analysis of real data. Hence, Dataset A has been used for calibration of these parameters, which are then are applied to Dataset B in order to validate the model. Table 6.3 shows the range of parameters and the calibrated values obtained for washing machines. In detail, parameter `st_by_prc` is the reference percentile of energy used to evaluate the usability of standby energy. Parameter `min_dur` represents the minimum length of useful operative cycles, and the calibrated value (30 min) has been obtained from all cyclic appliances. Parameter `min_Wh` represents the minimum energy of useful operative cycles (i.e., 100 Wh for washing machines). All the other parameters are equal for cyclic appliances.

Table 6.3: Parameter tuning for washing machines.

Parameter	Calibrated Value	Tuning Range
`st_by_prc`	98	[90–99]
`min_dur`	30 min	[15 30 45]
`min_Wh`	100 Wh	[50 100 150 . . . 500]
`thr_en`	800 Wh	[500 600 700 800]
`thr_d1`	45 min	[15 30 45]
`thr_d2`	120 min	[90 105 . . . 500]

6.5.2 Algorithm Results

The energy analysis starts with the calculation of the standby consumption of the appliance. Figure 6.5 depicts the cumulative distribution of the energy consumption for a washing machine from which the standby energy is equal to 0.15 Wh. The model uses this value to distinguish among periods of operation and inactivity for the specific appliance.

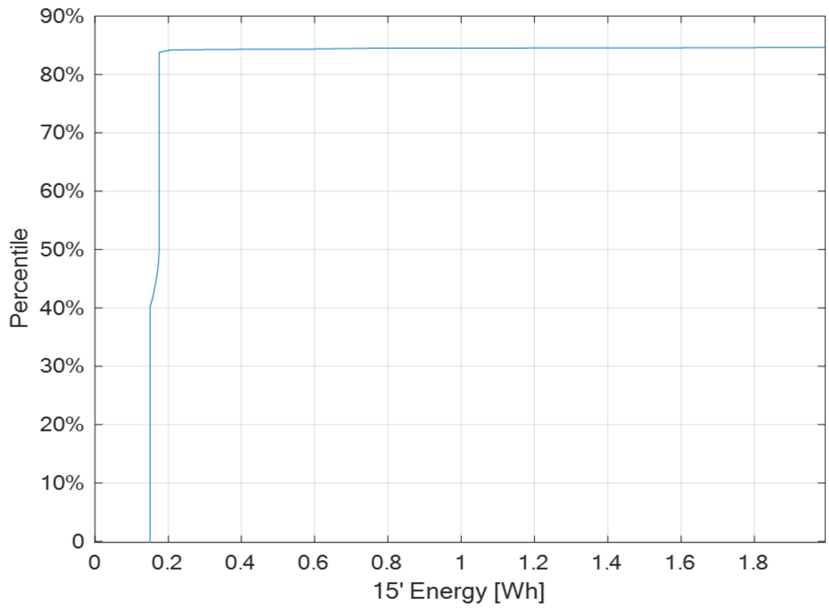


Figure 6.5: Cumulative distribution of the consumption.

The energy consumption profile in a typical cycle for washing machines is shown in Figure 6.6: an initial peak of consumption corresponding to water heating is followed by a longer phase (depending on the type of cycle set by the user as well as the type, model, and age of the appliance) with lower consumption.

The algorithm calculates the gradient of energy demand and finds consumption

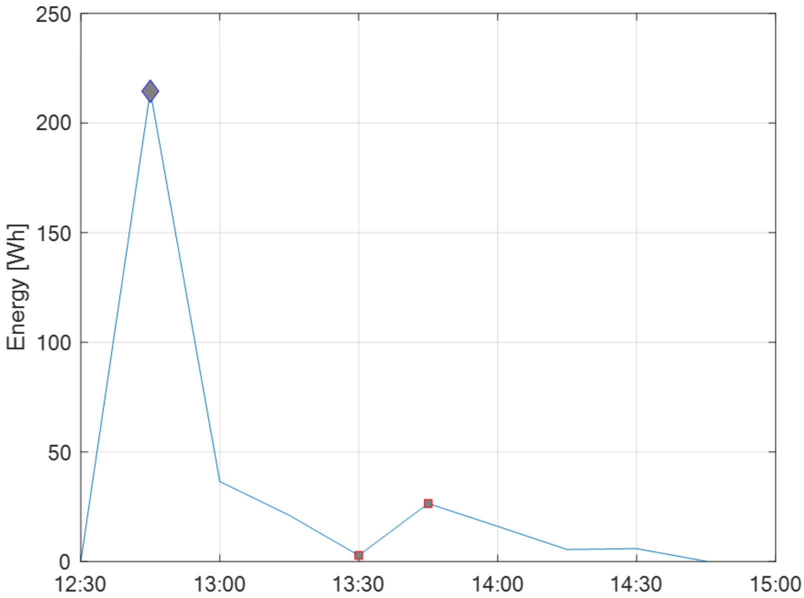


Figure 6.6: Automatic identification of the phases in the consumption profile of a single washing cycle of a washing machine

peaks, which correspond to relative maxima (i.e., zero gradient), for each aggregate cycle (multiple nearby cycles). The consumption of the peaks that separate individual nearby cycles (i.e., the first phase of operation) exceeds a specific percentile of the cycles' consumption, the value of which is selected among the parameters of the model. Once the various cycles have been recognized according to this methodology, the model determines the start and the end time (which corresponds to the lowest consumption before the peak of consumption of the next cycle), and duration of the single cycles. Tests have demonstrated that the algorithm is reliable in distinguishing nearby cycles, as shown in Figure 6.7. A finer sample period could produce better accuracy, but for the purposes of this investigation, the quarter-hour interval is appropriate.

Figure 6.8 illustrates a bubble chart with relevant information on cycle length and consumption of the same washing machine. The start hour during the day is displayed (abscissa) versus duration (ordinate), and the consumption is represented by the coloured scale of bubbles, whose size is proportional to the ratio between the highest and lowest quarter-hour consumption in the cycle. Larger bubble sizes in washing machines (and dishwashers) indicate washing cycles at a medium-high temperature.

Figure 6.9 depicts the subdivision of records (upper chart) and consumption

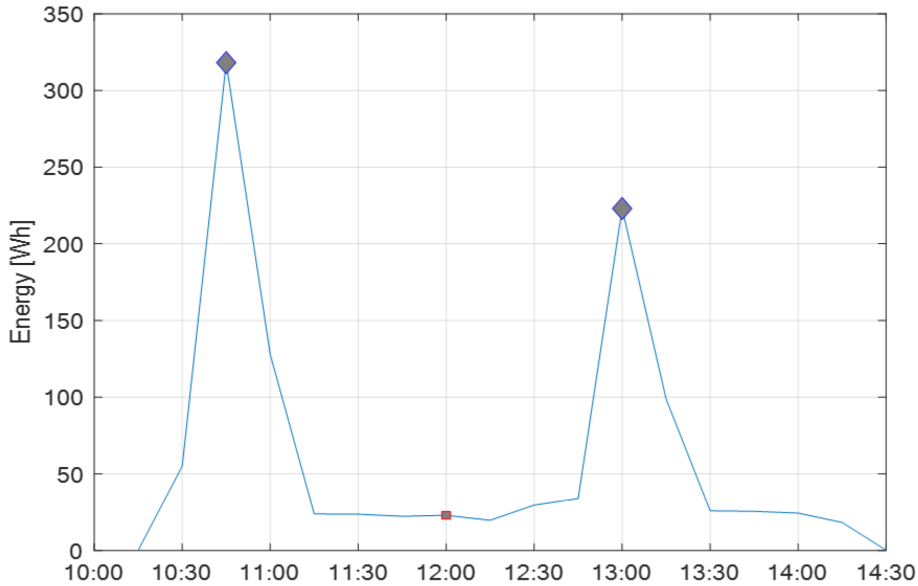


Figure 6.7: Automatic detection of two consecutive washing cycles.

(lower chart) based on the type of operation: appliance operative, standby mode, off, spurious data, and sensor not active (i.e., n/a label). Washing cycles correspond to a small fraction of the time but are associated with most of the consumption, unlike standby mode. Moreover, Figure 6.10 shows the subdivision of the records on a monthly basis, whereas Figure 6.11 illustrates the monthly (histogram on the left) and overall subdivision of consumption (pie chart on the right) for ARERA time slots. The upper graphs in Figure 6.12 show the monthly (histogram on the left) and overall (pie chart on the right) allocation of consumption according to the customized daytime slots, in the specific case three time slots (i.e., 8 a.m. to 12 p.m., 12 p.m. to 4 p.m., 4 p.m. to 8 p.m.), whereas the bottom graphs represent the number of operating hours (histogram on the left) and the overall operating hours (pie chart on the right).

6.5.3 Comparative Analysis

This section illustrates the comparison of appliances characterized by different energy classes. As described, the proposed algorithm is able to detect many features, and some of them have been selected in order to compare the performance of appliances: total number of cycles (recorded, extrapolated), percentage of consumption during peak hours, percentage of short cycles (recorded, extrapolated), percentage of long cycles (recorded, extrapolated), average energy consumption of short cycles

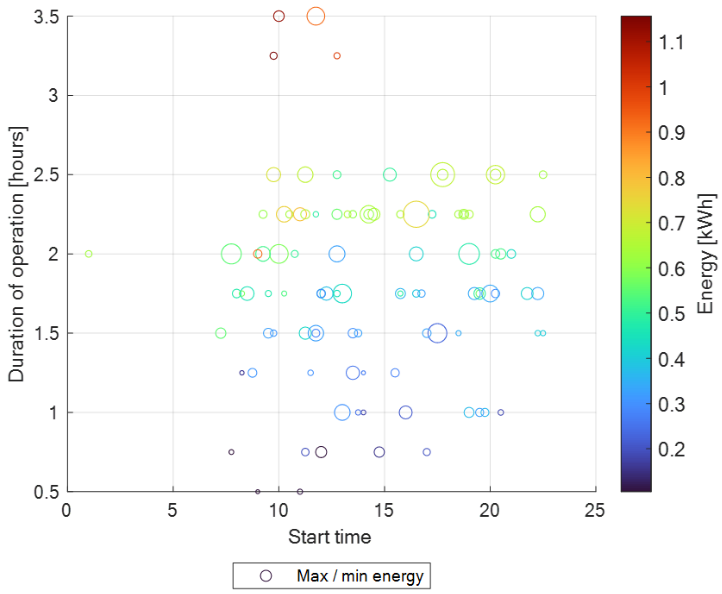
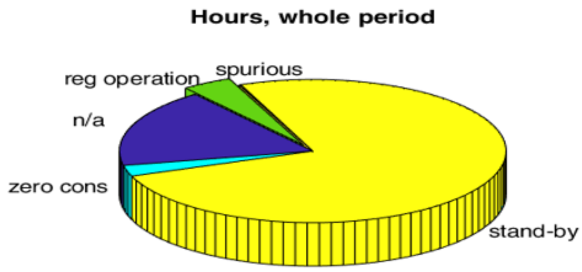


Figure 6.8: Duration, start time, and energy consumption of a washing machine's cycles.



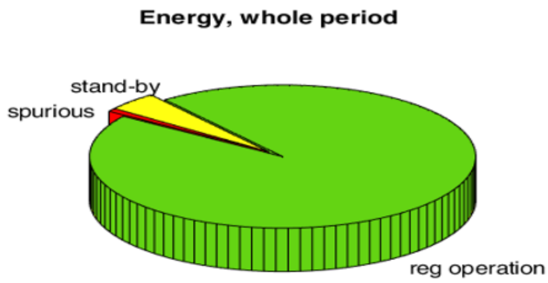


Figure 6.9: Subdivision of records (upper graph) and consumption (lower graph) based on the type of operation for a washing machine.

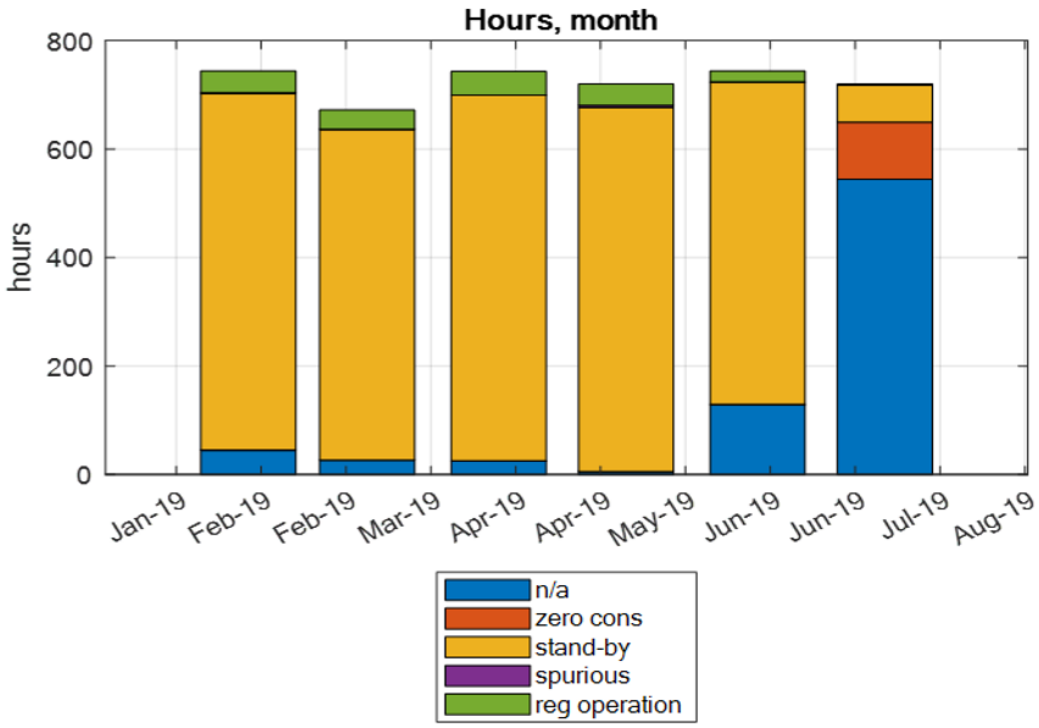


Figure 6.10: Monthly subdivision of records based on the type of operation for a washing machine.

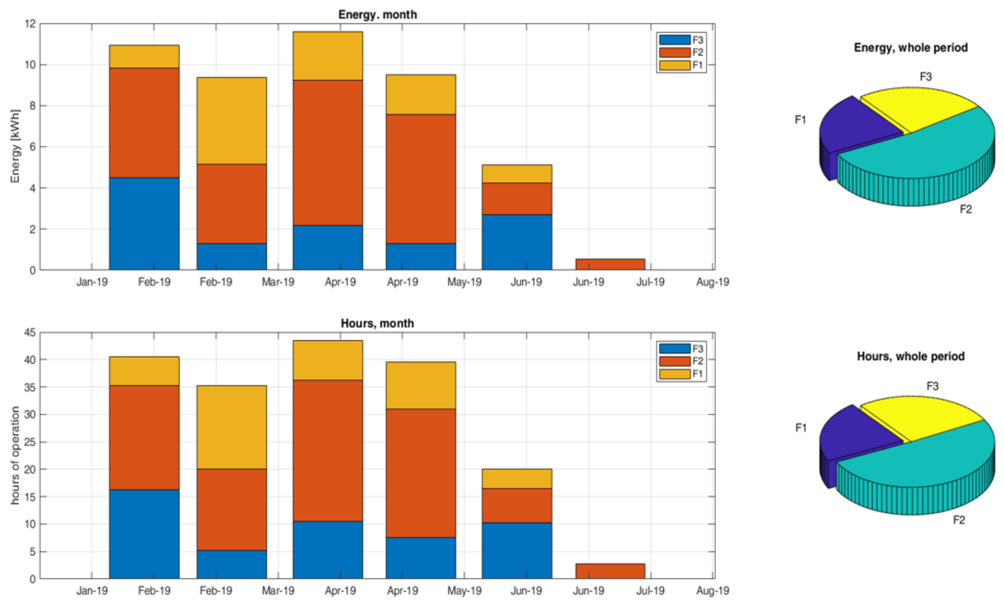


Figure 6.11: Monthly allocation of consumption in the three time slots defined by ARERA.

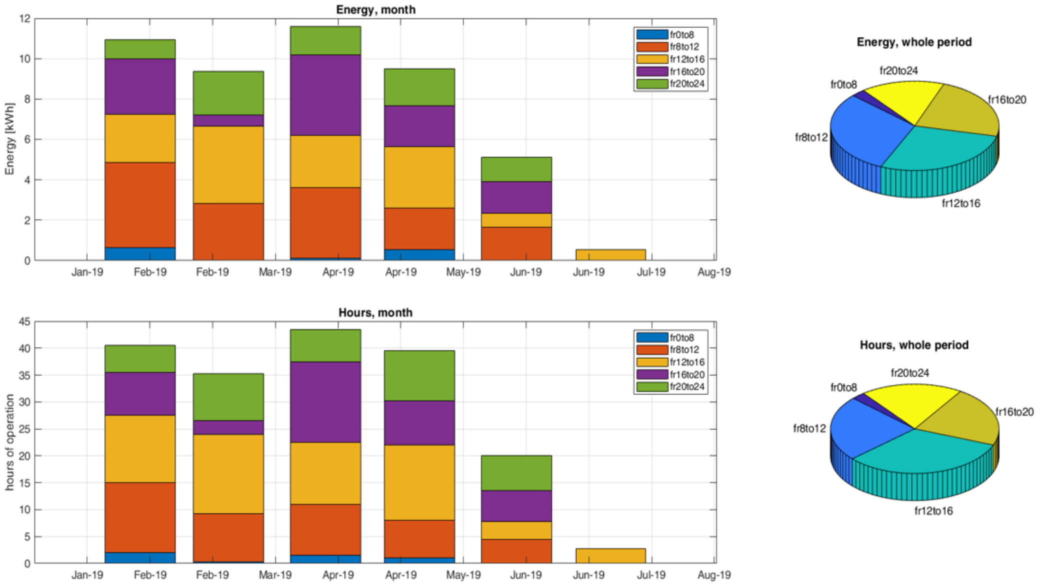


Figure 6.12: Monthly allocation of consumption according to the user-defined daytime slots.

and average energy consumption of long cycles.

Table 6.4 compares seven washing machines representing three energy classes (A, A++, and A+++) belonging to both datasets described in Section 6.3. Washing machines installed in homes EB-2 and NEB-4 are class A, those installed in homes EB-3, NEB-9, and NEB-10 are class A++, whereas homes EB-9 and NEB1 use washing machines of class A+++. Homes have a different number of residents, and they use the appliance in different manners. The model can determine the usage pattern of households in such different contexts. Furthermore, Table 6.4 presents the relevant quantities considered in the comparative analysis: the total number of cycles (Tn_θ) that is subdivided into short and long cycles, the energy consumption during peak hours (PEC), and the percentage of short cycles (P_{θ_s}) and long cycles (P_{θ_L}). The washing machine installed in home EB-3 carried out the largest number of cycles (Tn_θ) in both datasets, as confirmed by the energy consumption of both short cycles (AEC_{θ_s}) and long cycles (AEC_{θ_L}). As regards the energy consumption during peak hours, the washing machine in NEB-9 operated at 96% in this time slot, whereas the household of EB-2 showed a wiser behavior since he run his washing machine only for 28% of the time during peak hours. The percentage of short cycles for the washing machines of NEB-4 and NEB-1 is approximately the same even though the number of short cycles is 23 for the former and 52 for the latter, and the corresponding energy consumption during the analyzed period is 2.263 kWh and 3.264 kWh, respectively. Therefore, the average energy consumption in a single short cycle calculated for these appliances is consistent with their energy labels, i.e., class A and A+++ respectively.

Table 6.4: Comparative analysis on energy consumption for washing machines.

Home	EC	NoP	Tn_θ		PEC	P_{θ_s}		AEC_{θ_s}	P_{θ_L}		AEC_{θ_L}
			R_θ	Ex_θ	Peak Hours	PR_{θ_s}	PEx_{θ_s}	(kWh)	PR_{θ_L}	PEx_{θ_L}	(kWh)
EB-2	A	2	99	114	28%	6%	6%	1.140	54%	53%	8.523
NEB-4	A	2	31	36	45%	65%	64%	2.263	0	0	0
EB-3	A++	4	340	391	72%	7%	6%	2.062	42%	41%	18.295
NEB-9	A++	3	31	26	96%	9%	88%	3.759	0	0	0
NEB-10	A++	2	80	74	44%	86%	85%	3.662	0	0	0
EB-9	A+++	3	82	21	65%	12%	19%	1.492	4%	0	2.521
NEB-1	A+++	4	76	82	60%	64%	65%	3.264	1%	1%	1.558

Similarly, the results obtained with seven dishwashers monitored in the two datasets are reported from Figures 6.13–6.15. Figure 6.13 depicts the total number of cycles (recorded and extrapolated) in both datasets.

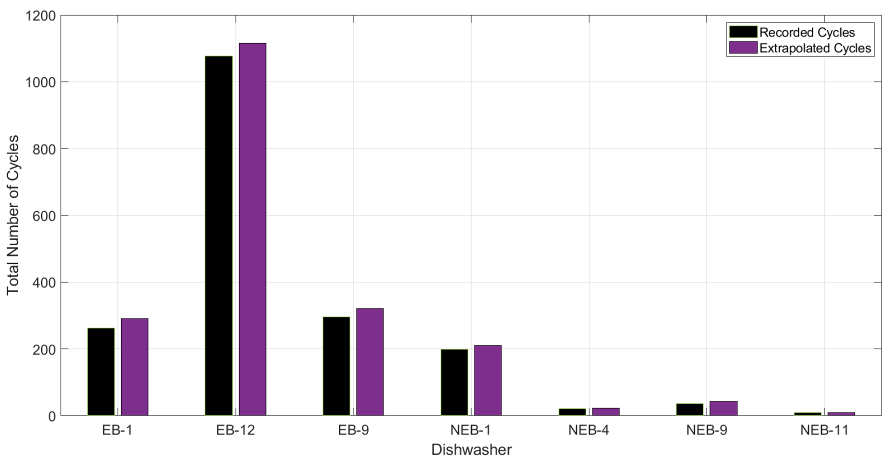


Figure 6.13: Energy recorded and extrapolated by the model for different dishwashers.

The model calculated the highest number of cycles for the dishwasher in EB-

12, while the total number of cycles for dishwashers in EB-1, EB-9 and in NEB-1 are similar. Figure 6.14 illustrates the percentage of short and long cycles for the dishwashers.

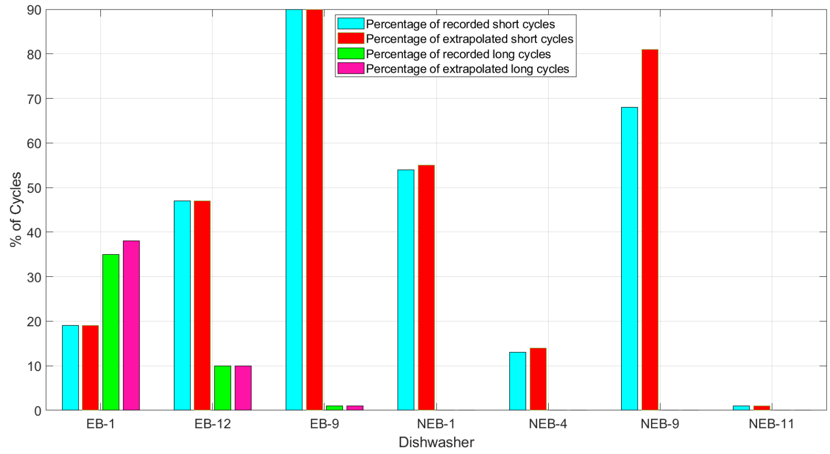


Figure 6.14: Percentage of short and long cycles for dishwashers.

The highest number of long cycles occurred in EB-1 while dishwashers in NEB-1, NEB-4, NEB-9, and NEB-11 operated only short cycles. Figure 6.15 depicts the total energy consumption of long (orange bar) and short (green bar) cycles during the analyzed period.

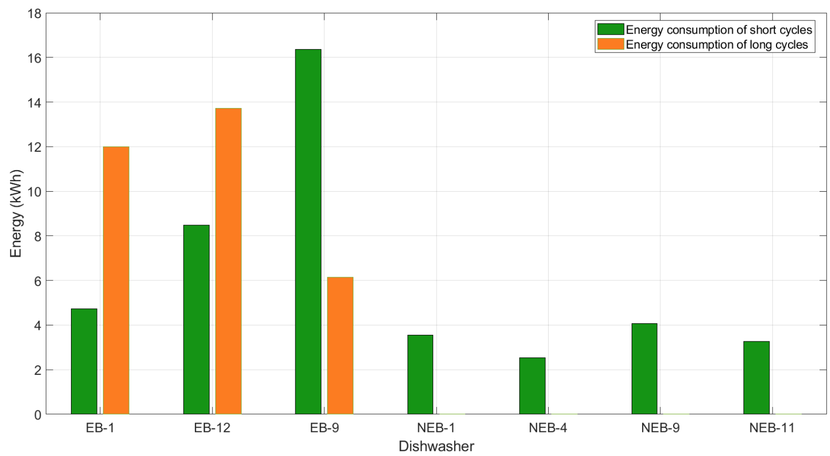


Figure 6.15: Energy consumption in short and long cycles for dishwashers.

The dishwasher in EB-12 consumed more energy during long cycles. By comparing the usage of appliances with their energy labels, it can be seen that even though dishwasher in EB-9 run fewer cycles than the dishwasher in EB-12 (see Figure 6.13), they consumed a similar amount of energy (see Figure 6.15). This outcome is consistent with the energy class of those appliances, i.e., class D for the former and class A++ for the latter.

In order to test the model on other datasets found in the literature, we applied it on the GreenD Energy database [148], containing detailed information on energy consumption obtained through a measurement campaign in households in Austria and Italy (December 2013 to November 2014). The results show that our implemented algorithms can extract all patterns for the available appliances in the dataset with the advantage of obtaining the same result with a much smaller amount of data, including disaggregation of energy consumption cycles occurred in close succession, which is one of the distinctive outcomes of the model, as shown in Figure 6.16.

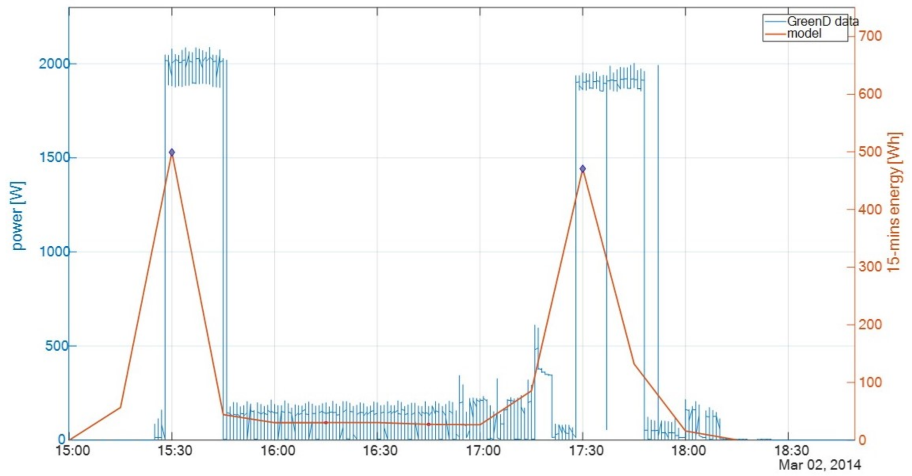


Figure 6.16: Comparison of energy consumption cycles occurred in close succession for the washing machine.

6.5.4 Feedbacks to the Consumer in DHOMUS Platform

The implemented model has been integrated in the DHOMUS platform in order to provide users with feedback on the electrical consumption and to compare their energy habits and patterns with other users participating in the experimentation, and with benchmarks as well. The platform provides a web interface for user feedback to encourage their virtuous and conscious behaviours. Moreover, a monthly report presented in Figure 6.17 contains summary of the calculations carried out by model and some suggestions to promote energy saving.

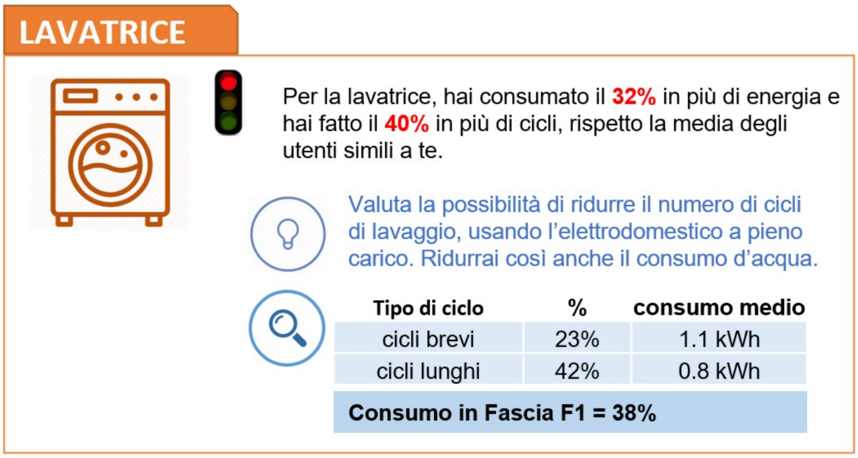


Figure 6.17: Feedback on the DHOMUS platform regarding the use of washing machine (in Italian).

In particular, the monthly report shows the detailed results obtained from the model and related to consumption during peak hours and to the shares of short cycles ("cicli brevi" in Italian) and of long cycles ("cicli lunghi" in Italian). Furthermore, tips are provided to the users, e.g., those reported in Figure 6.18 for the washing machine (infographics are in Italian since the DHOMUS platform is native for Italian users), which shows a comparison with similar users.

Figure 6.18: Section of the DHOMUS dashboard reporting the user's consumption vs. average consumption of other similar users (in Italian).

Moreover, tips are provided to assess the impact of virtuous behaviours, e.g., the reduction of the number of cycles by using the appliance at full load in order to reduce water consumption as well. On the other hand, Figure 6.18 shows the comparison of the average consumption with simular customers: washing machine, refrigerator, and dishwasher are compared in a bar graph on the left side of the illustration. The orange bar reflects the average consumption of other users, while the blue bar represents the average consumption of the specific user.

6.6 Summary

In this chapter, a method for detecting energy consumption patterns from household appliances' data for a smart home platform is developed. First, data sets are presented for the method's implementation, calibration, and validation. Then a detailed methodology of the proposed method is presented, followed by detailed working of the developed codes. Section 6.4.2 provides information on how different patterns are extracted from the data set using mathematical techniques. As the model is sensitive to parameters, parameters are finalized after fine-tuning, as provided in the 6.5.1 section. Algorithm results provided detailed statistics of the acquired features and energy consumption comparison for some appliances is provided which showed how algorithms behaved on different appliances. In addition, proposed model has been tested on the publicly available data set and produced excellent results. Lastly, Feedback based on our model findings is provided to the Consumer on DHOMUS Platform.

Chapter 7

Analysis of prediction on electricity demand of Smart Homes

This chapter presents the final objective of this dissertation, which is to compare different predictive models to select a good model for the smart home platform. To achieve this objective, we present an energy consumption forecasting approach for smart homes based on machine learning models. A comparative analysis is performed on two different homes data to compare the performance and determine a suitable forecasting model for smart home platform.

7.1 Related Work

As buildings are fully equipped with smart meters to monitor energy consumption at fine-grained intervals, research on energy consumption forecasting in buildings has become more critical over the past few years. In a study, [149], statistical and machine learning methodologies based on the evolutionary hybrid system were proposed through error series modeling for aggregated energy demand prediction. Experimental evaluation was carried out on a one-step-ahead scenario. The results were based on well-known metrics like mean squared error and were compared to existing statistical, hybrid, and machine learning methods. The proposed hybrid model showed statistically significant improvements.

In the domain of energy forecasting, ML-based system development has been emphasized [150]. Due to their relationship with demand, supply, environmental issues, and customer awareness, electricity load and energy consumption forecasts have received significant attention in this field. Supervised machine learning models were proposed to forecast Brazilian power electricity consumption for six different periods [151]. The models used were ARIMA, Random Walk, and different machine learning for all periods and results compared to benchmark models. Moreover, findings indicated that machine learning techniques, particularly Random Forest and Lasso Lars, provide more precise forecasts for all time horizons. Random Forest and Lasso Lars were able to track the trend and seasonality across many time ranges.

Authors in [152] proposed Multi-Layer Perceptron's enhanced deep learning capabilities for aggregated energy demand prediction. The day-ahead forecasting performance is obtained and compared with other machine learning models. Results showed that Multi-Layer Perceptron Outperformed Support Vector Machines, Gaussian Processes, Regression Trees, Ensemble Boosting, and Linear Regression. Moreover, a model for Multiple electric energy consumption prediction models was provided [153] for smart buildings using Transfer Learning and Long Short-Term Memory (TLL). Extensive tests were performed to compare the computing time and other performance indicators for numerous electric energy consumption forecasts for two smart buildings in South Korea. The experimental findings demonstrate that the proposed method can generate excellent results with less computational cost. Consequently, the suggested method may be efficiently implemented for intelligent energy management in smart buildings.

In addition, effective load forecasting techniques based on artificial neural networks have been developed [154]. Determining the optimal hyperparameters of neural networks is essential for accurate load forecasting, which is a complex and time-consuming task. The effectiveness of several artificial neural networks–based on building electric energy consumption forecasting models adopting various combinations of the hyperparameters was evaluated. Moreover, results reveal that neural networks with scaled exponential linear units and five hidden layers outperform other forecasting models on average. A study [155] introduced kCNN-LSTM, a deep learning model that exploited energy consumption data acquired at predetermined intervals to provide accurate building energy consumption predictions. The initiatory part of the algorithm uses k means clustering to perform cluster analysis to comprehend the energy consumption pattern/trend. In the second part, CNN is applied to extract complex features with non-linear interactions that affect energy consumption. In the last part, LSTM is employed to handle long-term dependencies by modeling temporal information in the time series data. The efficiency and applicability of kCNN-LSTM were proved using real-time building energy consumption data collected from a four-story IIT-Bombay building in India.

Another study [156] offered forecasting models based on a transversal set of Grey models, Nonlinear Grey Bernoulli models, and multi-layer feed-forward back-propagation networks that are more accurate than the Slovak Distribution Company's official load prediction. Models are evaluated using many performance metrics and report better results than officially available forecasting. A study [157] proposed a geographic and temporal ensemble forecasting framework for predicting short-term electricity usage. The Long Short-Term Memory Unit (LSTM) and Gated Recurrent Unit (GRU) are the two deep learning models that comprise the ensemble forecasting model. The model forecasts apartment, building, and floor power consumption hourly, daily, and weekly. Results demonstrated that the model learned the sequential behaviour of electric consumption by delivering superior performance with the lowest MAPE of 4.182 and 4.54 at the building and floor level prediction levels, respectively.

Authors in [158] proposed RABOLA (short for RAnager-Based Online Learning Approach) as a two-stage building-level short-term load forecasting model that facilitates practical and rapid pattern learning for unseen data. Furthermore, it was established through comprehensive comparison trials that the RABOLA model surpasses the prediction performance of cutting-edge stacking ensemble and deep learning techniques in terms of mean absolute percentage error and coefficient of variation of root-mean-square error. A study [159] proposed a new method for forecasting a building's energy demand for the following day utilizing the unique data collected from its digital twin model. Researchers examined naive approaches, linear regression, LSTM, and Prophet methods. They discovered that the Prophet model provided the most accurate prediction of energy consumption for the day ahead forecast. A unique hybrid AI-enabled forecasting model that combines single spectrum analysis (SSA) and parallels long, short-term memory (PLSTM) neural networks were proposed [160]. The SSA decomposition improved the performance of the PLSTM network. The model surpasses the state-of-the-art models at different points regarding predictive performance and computing efficiency.

A study [161] provides a model for estimating the energy consumption in Agartala, Tripura, India, which can reliably anticipate the demand for the next 24 hours and one week to 1 month. Several city-specific factors (temperature, humidity, air density, and pressure) have been analyzed to identify variables that could directly influence the pattern of electricity use. The combination of Random Forest and XGBoost machine learning models demonstrated excellent performance for energy consumption prediction. ARIMA-based models to forecast future power usage in Libya were proposed by authors in [162]. The ACF and PACF plots and the stationarity of the data was utilized to determine (p,d,q) values. Using the Mean Absolute Percentage Error (MAPE) to measure the accuracy of the prediction, the model was able to predict with an error of 4.332%.

Long short-term memory (LSTM) neural network, adaptive neuro-fuzzy inference system (ANFIS) with subtractive clustering (SC), ANFIS with fuzzy cmeans (FCM), and ANFIS with grid partition (GP) were applied for the one-day ahead short-term electricity energy consumption prediction [163] for Turkey. The findings demonstrated that forecasting short-term daily electrical energy consumption time series with the LSTM method yield accurate results compared to the other models. Another study [164] used machine learning models like SVM, EGB, and LGB for long-term load forecasting. These models were tested on London Smart Energy Meter data. All these models were trained to measure RMS error. Authors in [165] investigated ARIMA and TBATS with other models on datasets from home with photovoltaics and an energy management system. The TBATS model's reported mean absolute error of 73.62 Watts is better than the neural forecasting approach. A study [166] investigates gradient boosting machine learning models, particularly LightGBM, CatBoost, and XGBoost, to compare their performance on the Chicago office buildings dataset. When trained on the provided dataset, XGBoost outperforms LightGBM and CatBoost, according to the preliminary findings. The above studies provide valuable insight into how different statistical, machine learning, and deep learning models are productive for forecasting time series electrical energy consumption.

A comparative analysis of predictive models to identify a suitable forecasting model for the smart home platform is performed. For an hour ahead of energy consumption forecasting, We used machine learning based forecasting approach with a focus on building energy consumption forecasting. We consider calendar variables and historic consumption to train selected machine learning models such as Seasonal Auto regressive Integrated Moving Averages (SARIMA), Linear Regression (LR), Support Vector Machine (SVR), Random Forest (RF), Multilayer Perceptron (MLP), and Long Short-Term Memory (LSTM). We predict one hour ahead energy for two different houses located in Rome, Italy. We extensively compare the overall prediction performances of the selected ML models by considering different configuration settings based on grid search approach.

7.2 Smart Homes Electricity Energy Consumption Dataset

The data set used in this study has time series data about how much electricity smart homes use. The dataset includes electricity collected from 14 smart homes under experimentation, part of the DHOMUS [1] platform. The meter acquired punctual data every 5 minutes, and then the data was aggregated with a granularity of 15 minutes. For further forecasting analysis, the fifteen-minute granularity has been aggregated to one hour, and two homes have been chosen out of fourteen. Figure 7.1 shows an example of electrical energy consumption time series data for a home EB-3.

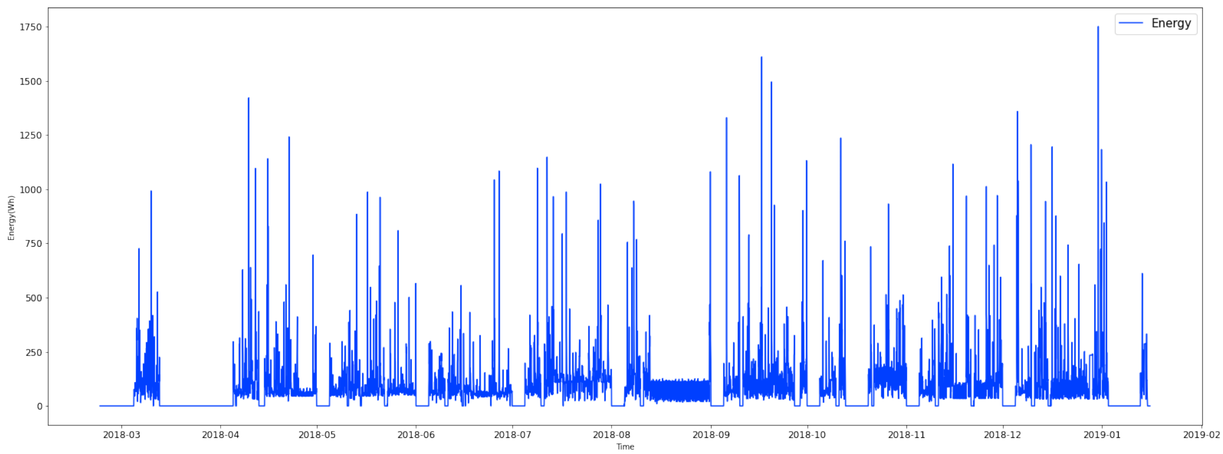


Figure 7.1: Electrical Energy consumption data for Home EB-3

7.3 Proposed Time Series Forecasting Work Flow

The overall workflow employed for the study of different forecasting models for smart home energy demand prediction is shown in Figure 7.2. The proposed framework consists of three main modules: data preprocessing, data partitioning, and model

[1](https://dhomus.smartenergycommunity.enea.it)

evaluation. All these modules are discussed independently in the following subsections.

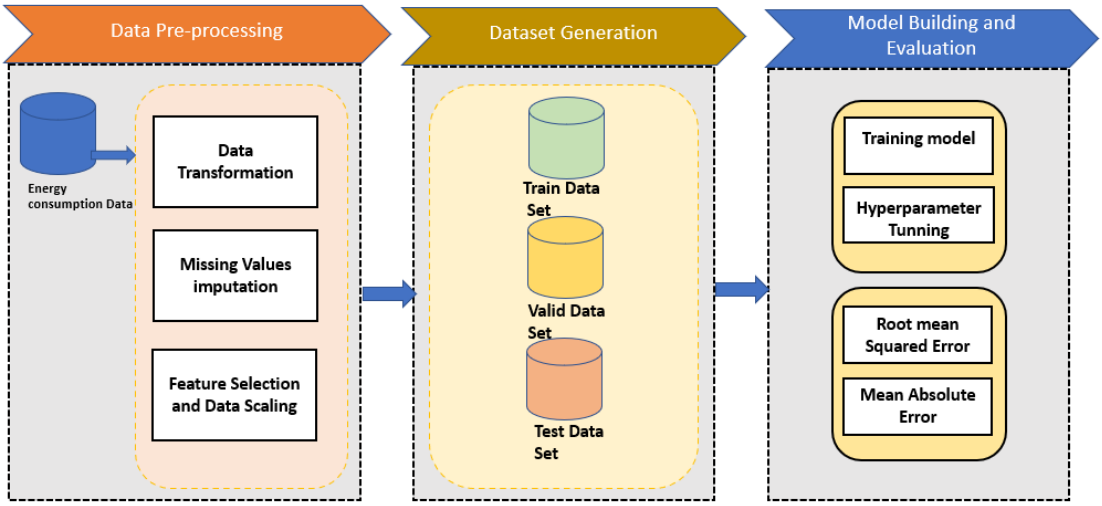


Figure 7.2: Workflow for Energy consumption forecasting

7.3.1 Data Preprocessing

This module is accountable for preparing the data for subsequent analysis and processing. It consists of three steps: data transformation, missing data handling, and normalization.

7.3.1.1 Data Transformation

The energy consumption data (every five minutes) is acquired through a smart meter installed in homes; afterward, an algorithm available on the DHOMUS platform aggregates it to 15 minutes. MySql extracted the energy data from the database in a .csv file. Further, for hourly energy consumption forecasting, acquired data is transformed into one-hour timestamps. Then the remaining preprocessing steps need to perform to prepare the data for the training and testing of forecasting algorithms.

7.3.1.2 Missing data handling

There are two general approaches to handling missing values in data. The first is to delete data samples with missing values, as most data mining algorithms cannot process such data. This strategy is effective only when the percentage of missing values is small. The second step is to replace missing data with inferred values using missing value imputation techniques [167]. In this work, just the deletion approach has been adopted to deal with the missing data, as our data has few missing values.

7.3.1.3 Feature Selections

Feature selection is an essential step in building forecasting models. It has a significant impact on the performance of forecasting models. The considered home energy consumption data is $N * M$ dimensional data, where N denotes the rows (collected energy consumption measurements) and M represents the columns (home id, sensor, timestamp, sum of energy power, delta energy) described in Table 7.1. As a feature selection step, in this study, we used the autoregressive features AR(2) of delta energy and the time feature (hours of the day) from the home energy consumption data to forecast the energy consumption. The time feature (hours of the day) is extracted from time and represents a number from 0 to 23.

Table 7.1: Important Feature of the Datasets

Parameter	Definition	Units
home id	Energy Box associated with the single house	None
Sensor	Sensor associated with the appliance	None
Timestamp	date in datetime format year-month-day	hour:minutes:seconds
sum of energy of power	power measured by the integrated sensor on the quarter of an hour is therefore energy expressed in Wh	Wh
delta energy	energy detected by the sensor meter remains zero until a consumption threshold dependent on the sensor is exceeded, which can represented as Wh or kWh	Wh or kWh

7.3.1.4 Data Scaling

After the selection of needed features, data scaling is frequently required to assure the correctness of predictive modeling, particularly when input variables have varying scales. The max-min normalisation (i.e, $s' = s - s_{min}/s_{max} - s_{min}$) and standardization of z-score (i.e, $s' = s - \mu/\sigma$) are two widely used approches for building energy consumption data sacaling. In this study, min-max normalization has been used while applying deep learning and machine learning models.

7.3.2 Data Partitioning

The data acquired from the smart energy meter contains consumption values between January 2018 to September 2020. However, for forecasting data, we are only taking data from January 2019 to December 2019 because the data quality is excellent during this span. The data acquired after the transformation step has hourly granularity, so the total sample points (hours) we have are 8758. The transformed dataset is then split into two parts training data and the second testing data with ratios of 80%(training points 7006) and 20%(testing points 1752), respectively.

7.3.3 Model Building and Evaluation

In literature, forecasting methods are divided into three classes : (i) classical time series forecasting methods, (ii) Machine learning forecasting methods, and (iii) Deep learning forecasting methods. In this study, we have used one classical forecasting method: Seasonal Auto-regressive Integrated Moving Average (SARIMA). As far as machine learning methods are concerned, we have used methods such as Linear Regression (LR), Support Vector Regression (SVR), and Random Forest(RF). We have employed Multiplayer Perception (MLP) and Long short-term memory (LSTM) from deep learning forecasting methods. The classical forecasting methods correlates energy consumption with relevant features like climate data and historical energy consumption; the main advantage of this kind of method is that while applying these methods, we get to know the nature of the data, that it is stationary or not and is there any seasonality lie in the series or not because classical methods are sensitive to the nature of data and it effects the prediction a lot [86]. However, The machine and deep learning methods expose the non-linear relationship between the input (historical data) and output (target consumption) by learning consumption patterns from past energy consumption data [168].The next section presented more details on the configuration and use of these adopted methods.

7.3.4 Applied Forecasting Models for Energy consumption

7.3.4.1 Classical Time Series Forecasting Models

In forecasting applications, the basic aim is the collection of time series data that consists of a set of observations collected in uniform time steps. The frequency of the recorded data may be every second, hour, day, week or month. It is a pre-processing step to develop a time-series data for forecasting.

7.3.4.1.1 Seasonal Autoregressive Integrated Moving Average

In time-series forecasting, autoregressive integrated moving average (ARIMA) has been used extensively. This technique is capable of handling the trend or a pattern in time-series data but incapable of dealing with seasonal components [169]. Hence, SARIMA was introduced that directly models the seasonal component of the time-series data. It is extended by adding seasonal terms that consist of three new hyper-parameters for the specification of auto regression (P), differencing (D), and moving average (Q).

$$ARIMA(p,d,q)(P,D,Q)_S \tag{7.1}$$

where (p,d,q) represents the non-seasonal term, $(P,D,Q)_S$ represents the seasonal term of the model and S denotes the season period number.

7.3.4.2 Machine Learning Forecasting Methods

7.3.4.2.1 Linear Regression

To exploit linear techniques for time-series forecasting and analysis is very common due to its insignificant computational cost. LR [86] explains the relationship between an independent variable x and a dependent variable y. The dependent variable y and the independent variable x are also called the forecast variable and the predictor variable; respectively LR methods try to draw a regression line that best fits across the forecast and predictor data. This relationship is defined in the following equation 7.2 as:

$$y_t = \alpha_0 + \alpha_1 X + \varepsilon_t, \quad t = 1, 2, \dots, n \tag{7.2}$$

Where $X = x_1, x_2, \dots, x_n$ represents the independent forecast variables,y_t are the dependent predictor variables, ϵ_t are the errors, and α represents the model regression coefficients. The Eq. 7.2 can also be written separately for each predictor variable as:

$$y_1 = \alpha_0 + \alpha_1 x_1 + \epsilon_1 \tag{7.3}$$

$$y_2 = \alpha_0 + \alpha_1 x_2 + \epsilon_2 \tag{7.4}$$

$$\vdots$$

$$y_n = \alpha_0 + \alpha_1 x_n + \epsilon_n \tag{7.5}$$

Combining all the above equations results in Eq. (7.6):

$$\sum y_n = n\alpha_0 + \alpha_1 \sum x_t + \sum \epsilon_t \tag{7.6}$$

In LR, the sum of errors i.e. $\sum \epsilon_t = 0$ is desired to achieve that reduces the Eq. (7.6) into Eq. (7.7) as:

$$\sum y_n = n\alpha_0 + \alpha_1 \sum x_t \tag{7.7}$$

Linear regression uses the least square method to decrease the squared distance from the regression line. Applying the least square to Eq. (7.6) and rearranging results in solving a quadratic problem of the form given in Eq. (7.8).

$$\sum \epsilon_t = ((\alpha_0 + \alpha_1 x_1 - y_1))^2 + ((\alpha_0 + \alpha_1 x_2 - y_2))^2 + \cdots + ((\alpha_0 + \alpha_1 x_n - y_n))^2 \tag{7.8}$$

Taking partial derivatives with respect to α_1 and setting the Eq. (7.8) to 0, gives:

$$\sum x_t y_t = \sum x_t \alpha_0 + \alpha_1 \sum (x_t)^2 \tag{7.9}$$

In LR, the following objective function given in Eq. (7.10) is minimized.

$$min \sum_{t=1}^{n} ((y_t - x_t \alpha_t))^2 \tag{7.10}$$

The regression coefficients α_0 and α_1 can be found by solving Eq. (7.10) and (7.11) given in the following section.

7.3.4.2.2 Support Vector Regression

The support vector machine (SVM) is a commonly used supervised ML algorithm. It can be applied to both classification and regression problems. When SVM is used in regression, also called support vector regression (SVR) [86], generates a model that does not consider the observations that are much similar to the target values. It tries to find a hyperplane in higher dimensional space that best maps to the input data. It allows the flexibility to choose a suitable error margin in the model. Unlike LR, SVR aims to minimize the coefficients instead of the objective function. The objective function of the SVR model is defined in Eq. (7.11), which denotes the absolute error. The accuracy of the SVR model can be tuned by changing the value ϵ.

$$\min \frac{1}{2}||\alpha||^2 \tag{7.11}$$

subject to constraint

$$|y_t - x_t\alpha_t| < \epsilon$$

7.3.4.2.3 Random Forest

Random Forest (RF) is also a good choice for the classification and regression problems in decision tree learning. It can model time series data for forecasting applications with better regression accuracy in comparison to other regression algorithms. An RF model constructs multiple décor-related decision trees and trains them simultaneously. To avoid biases and prediction variance, it estimates the output predictions by taking the average of the responses from each individual decision tree. The RF model is capable of yielding predictions in a probabilistic mode for both classification and regression problems. It also performs efficiently in exhibiting highly non-linear input and output mapping. Statistically, it is an advanced form of classification and regression trees (CART) with the advantage of dealing with non-linearities that for other algorithms such as neural networks may result in overfitting [170].

The fundamental concept of the RF technique is to make a forest of numerous weak performance classification regression trees by following particular rules and to estimate forecasting responses by majority voting of all trees in the forest. Eq. (7.12) explains an RF model in terms of the input vector and independently distributed random vector within k-th decision trees.

$$h(X, \theta_k), k = 1, 2, \ldots, n \tag{7.12}$$

The mapping between the input vector X and output vector Y is given in Eq. (7.13).

$$K(X, Y) = a_k I[h(X, \theta_k) = Y] - \max_{(j \neq Y)} a_k I[h(X, \theta_k)] = j] \tag{7.13}$$

Here, j and a_k denote the type of training set and average function, respectively. A higher value of mapping means more confidence in the accuracy of forecasting results. Eq. (7.14) explains the generalization error E^* of the RF model.

$$E^* = P_{X,Y}(K(X,Y) < 0) \tag{7.14}$$

The RF model follows the following two theorems in the case of a large number of decision trees in the forest.
Theorem 1. The generalization error converges to zero with more decision trees, for each θ_k as given in Eq. (7.15).

$$P_{X,Y}(P_\theta(h(X,\theta) = Y) - \max_{(j \neq Y)} P_\theta(h(X,\theta) = j) < 0) \to 0 \tag{7.15}$$

The theorem states that the increased amount of decision trees does not affect the generalization error to result in overfitting instead it reaches to some specific value.
Theorem 2. The generalization error of an RF model is upper bounded as given in Eq. (7.16)

$$E^* \leq \frac{\overline{\rho}(1 - s^2)}{s^2} \tag{7.16}$$

In Eq. (7.16), $\overline{\rho}$ and s represent the average correlation and the average strength of the tree, respectively. This theorem specifies that the fall in tree correlation and rise in strength of single tree results in reducing the upper bound on the generalization error consequent in effective controlling

7.3.4.3 Deep Learning Forecasting Models

7.3.4.3.1 Multilayer Perceptron (MLP)

The multi-layer perceptron (MLP) can also be exploited for time-series forecasting problems. After training, an MLP model can learn the complex non-linear mapping between input observations and target outputs with high accuracy by selecting the suitable number of layers and neurons at each layer. In the case of time-series data, the sequential output must be split into multiple feature vectors. The idea of training is to find those weights and biases that well describe the link between neurons. One of the most widely used supervised learning algorithm used for MLP based feedforward network is backpropagation. Neurons arranged in a single column are called a neural network layer. In an MLP, neurons are arranged in multiple layers consisting of three basic layers namely, input, hidden and output layer. The first and last layers are always the input and the output layer, respectively. There is single input and output layer, but the hidden layers can be more than one. The output of each layer is fed to the next one. All the neurons in one layer are connected to all the neurons in the preceding layer through the network parameters such as weights and biases. Except for the input layers nodes, all the neurons in an MLP employ a non-linear function also called the activation function. Two common activation functions are the predictive ability of a neural network lies in its multi-layered hierarchical structure [171].

7.3.4.3.2 Long Short-Term Memory (LSTM)

Lately, long short term memory (LSTM) has gained quite an attention in time-series forecasting tasks [172]. It is an extension of recurrent neural network (RNN). RNN are the neural networks where the previous output is fed as input to the current. The architecture of a simple RNN is shown in Figure 7.3.

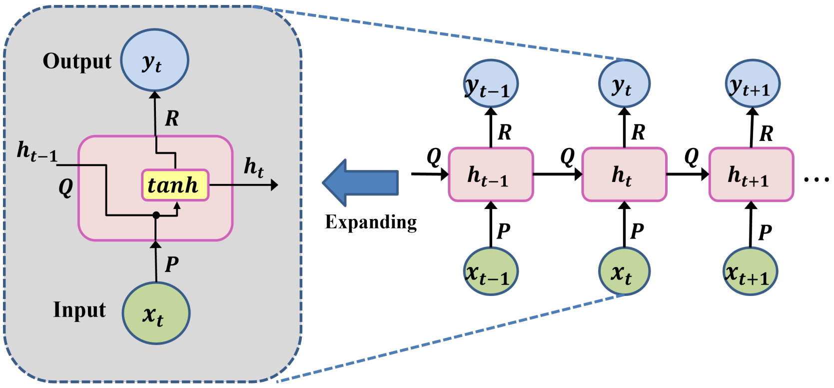


Figure 7.3: The architecture of recurrent neural network

The input to this neural network is x, hidden state is h acting as a memory element is the most important component of RNN. Since, RNN is evolved from MLP to specifically use for sequential data. Therefore, the subscripts $t-1$,t,$t+1$ are the previous, current and next time steps, respectively. The hyper-parameters of different layers are denoted by P,Q, and R. The uniqueness of this network lies in its ability to feed the current output of the hidden layer to the next hidden layer as input. Such looping behavior preserves the information from the previous step maintaining the data dependency and consequent in improved learning from the serial data. Nevertheless, the vanishing gradient problem still exists in RNN making it not a viable option for a long-sequenced dependent data. To mitigate the limitation of RNN, LSTM were proposed. The hidden state in LSTM is much more

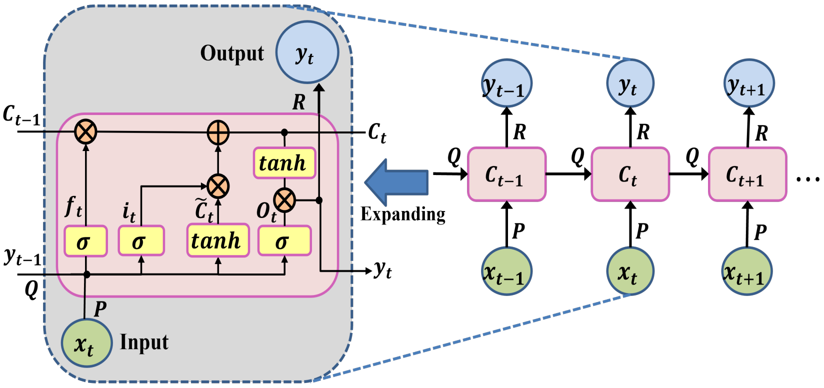


Figure 7.4: The architecture of long short term memory

complicated than RNN. The internal structure of LSTM is shown in Figure 7.4. The hidden state y_{t-1} in LSTM passes three gates i.e. input, output and forget gates that learn which data in the input sequence is important to keep or unnecessary to discard. This process makes it possible to keep only the relevant data required to make predictions. The stepwise procedure of LSTM followed in this neural network at time step t is given below:

- The previous hidden state y_{t-1} and the current input x_t is concatenated and fed into the forget gate to remove irrelevant data. The output of the forget gate f_t is calculated using Eq. (7.17).

$$f_t = \sigma(P_f x_t + Q_f y_{t-1}) \tag{7.17}$$

- In the next step, the possible values of x_t to add to the cell state C_t are identified. For this, the output of the input gatei_t and a candidate state $\tilde{C}_t$ should be updated.

$$i_t = \sigma(P_i x_t + Q_i y_{t-1}) \tag{7.18}$$

$$\tilde{C}_t = tanh(P_c x_t + U_c y_{t-1}) \tag{7.19}$$

$$C_t = (f_t * C_{t-1} + i_t * \tilde{C}_t) \tag{7.20}$$

$$O_t = \sigma(P_O x_t + Q_O y_{t-1}) \tag{7.21}$$

- Finally, a pointwise multiplication of the output gate O_t and the new cell state C_t results into the new hidden state y_t.

$$y_t = O_t * \tanh(C_t) \tag{7.22}$$

In the above equations, are the input weights, recurrent weight, and output weights, respectively. For the enhancement of network nonlinearity, two types of activation functions are used; sigmoid and hyperbolic tangent functions. The former activation function constrains the values between 0 and 1 while the latter between -1 and 1.

The architecture of the LSTM is designed in a way that resolves the vanishing gradient problem. Thus, it outperforms traditional RNN in capturing and extracting past information to forecast future values for time-series data.

7.4 Experimental Evaluation

After the discussion of the employed forecasting, we present the empirical evaluation of the models on home energy consumption data.The design of evaluation can be seen in next section followed by evaluation metrics used. Finally, we have the experiments and results of our empirical analysis on the smart home dataset.

7.4.1 Design of the Evaluation

To predict the electricity energy demand from time series energy consumption data, in this work, we have used electricity energy data as discussed in Section 7.3. The dataset includes electricity collected from 14 different smart homes under experimentation which is part of the DHOMUS platform. However, for forecasting we used two homes under study and we have utilized the energy usage measures which are collected from from January 2019 to December 2019.

7.4.2 Evaluation Metrics

Several statistical methods exist for evaluating the learning model's performance based on the difference between the actual and forecast value. The following defines the performance evaluation metrics utilized in this study.

i. Root mean squared error: It is the standard deviation of the differences between the true value and the predicted value shown in Eq (7.23).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y_i} - y_i)^2}{n}} \tag{7.23}$$

ii. Mean absolute error: It evaluates the absolute difference between the actual value and the predicted value as shown in equation (7.24).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y_i} - y_i| \tag{7.24}$$

iii. Correlation Coefficient (R): It measures the linear dependencies between the actual y_t and predicted value $\hat{y_t}$. Equation 7.25 presents the formula to calculate R.

$$R = \frac{n\sum y_t.\hat{y_t} - (\sum y_t)(\sum \hat{y_t})}{\sqrt{n(\sum y_t^2) - (\sum y_t)^2}\sqrt{n(\sum \hat{y_t^2}) - (\sum \hat{y_t})^2}} \tag{7.25}$$

7.4.3 Results

In this section, we present experimental results of predictive models with their exploratory settings on home energy consumption data. We have performed the implementation of all the selected forecasting models in the Python programming language by utilizing $scikit - learn^7$ machine learning library.

7.4.3.1 Naive Regressor Results

The Naive predictor(NP) is the initial model used as a baseline model. The Naive method ensures that the desired hours of energy consumption will be the same as the previous day.

$$p(t) = (t - \Delta t) \tag{7.26}$$

where $\Delta t = 24h$. This method is only useful to make raw forecasting. In our work, we make predictions for smart home energy consumption(two homes EB-3, ENEA-4) using a granularity of 1h, so in our case, $\Delta t = 1h$.The Table 7.2 shows the prediction errors for NP with $\Delta t = 1h$.

Table 7.2: Naive predictor forecasting errors

Home	**Model**	Δt	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	NP	1H	0.23	0.10	0.65
ENEA-4	NP	1H	0.34	0.16	0.51

7.4.3.2 SARIMA Forecasting Results

The first attempt is to evaluate the applicability of an autoregressive model. Before using this approach, it is critical to ensure that a series is not statistically significant during the first lags. The Ljung Box test is employed for this purpose, followed by a stationary analysis using the Dickey-Fuller(DF) test [173].

Table 7.3: Dickey-Fuller Test Statistic EB3

Test Statistic	-21.804699
p-value	0
#Lags Used	37
Number of Observations Used	8732
Critical Value (1%)	-3.87324311

Table 7.3 shows the results of the DF Test for EB3 home. Since the value of p <=0.05: rejects the null hypothesis (H0), the data do not have a unit root and are stationary. In addition, the test run prints the value of the test statistic of -21.804699. The more negative this statistic is, the more likely we will reject the null hypothesis (we have a stationary dataset). As part of the output, we get a lookup table to help determine the DF statistic. We can see that our statistical value of -21.804699 is less than the value of -3.434 to 1%. Results suggest that we can reject the null hypothesis with a significance level of less than 1%. Rejecting the null hypothesis means that the process has no unitary root and, in turn, the time series is stationary or has no time-dependent structure. Ljung-Box test for persistence/autocorrelation is adopted and the outcome revealed that the p-values for the first 10 lags are greater than 0.05, and the residues are distributed independently, so it is not persistent. In this scenario, it's not a good choice to use it for time-series forecasting.

Table 7.4 shows the results obtained by the SARIMA model. The correlation coefficient has a -0.03 value which is not significant in the case of data from EB3. However, the correlation coefficient of predicted and actual for ENEA 4 home is 0.02, which is not good. Thus, the SARIMA model is unsuitable for forecasting our energy prediction data. The evaluation matrix of the Naive method has better results compared to SARIMA model.

Table 7.4: SARIMA forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.31	0.14	-0.03
ENEA-4	HC	0.37	0.24	0.02

7.4.3.3 LR Forecasting Results

The LR model attempts to simulate the relationship between a predictor variable and one or more forecast variables [174] (details of the model shown in sec.7.3.4.2.1). Because of their ease of use, LR models have been described in the literature to predict the electricity demand of homes. We employed two distinct LR model versions in this work. The sole distinction is which features are used for training the model. The first version of the LR model includes historic consumption (HC) with AR(2) values. In contrast, the second version includes more sophisticated features that are time information in addition to HC (hour of the day). Table 7.5 illustrates the forecasting errors determined using two distinct linear regression models for homes EB-3 and ENEA 4.

Table 7.5: Linear regression forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.21	0.12	0.66
EB-3	HC & time (hour)	0.21	0.11	0.67
ENEA-4	HC	0.31	0.18	0.48
ENEA-4	HC & time (hour)	0.30	0.17	0.52

The results in Table 7.5 demonstrate that the LR model has produced better forecasting results than the naive predictor. The prediction errors show that adding the time feature (hour of the day) does not significantly improve forecasting accuracy. From Figure 7.5 to Figure 7.8 shows the comparison of actual and predicted energy consumption values of the LR model for both the used versions and Homes.

7.4.3.4 SVR Forecasting Results

The SVR is a transformation of a Support Vector Machine (SVM) using the same rules as the SVM. This variant is used to solve the regression problems [174] (details of the model shown in sec.7.3.4.2.2).In this work, We employed two alternative

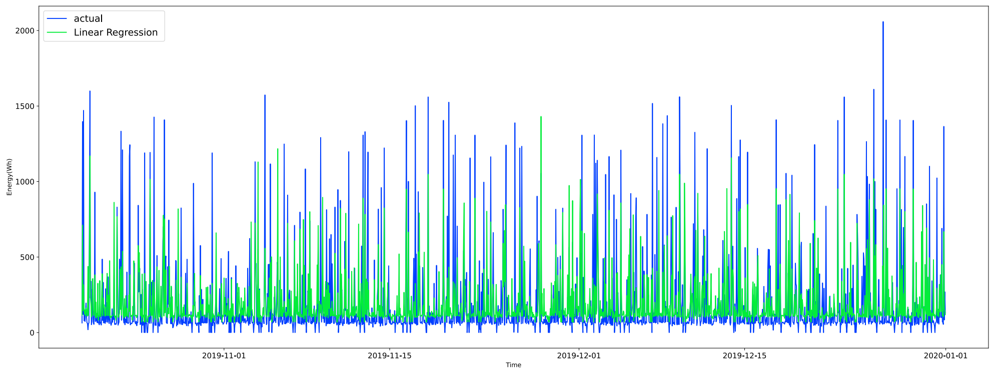


Figure 7.5: Actual and predicted energy consumption forecasting for EB3 home with HC feature.

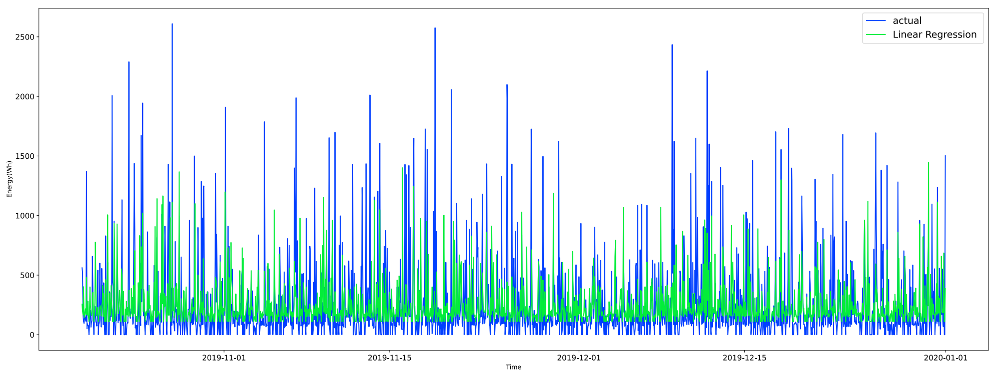


Figure 7.6: Actual and predicted energy consumption forecasting for ENEA 4 home with HC feature.

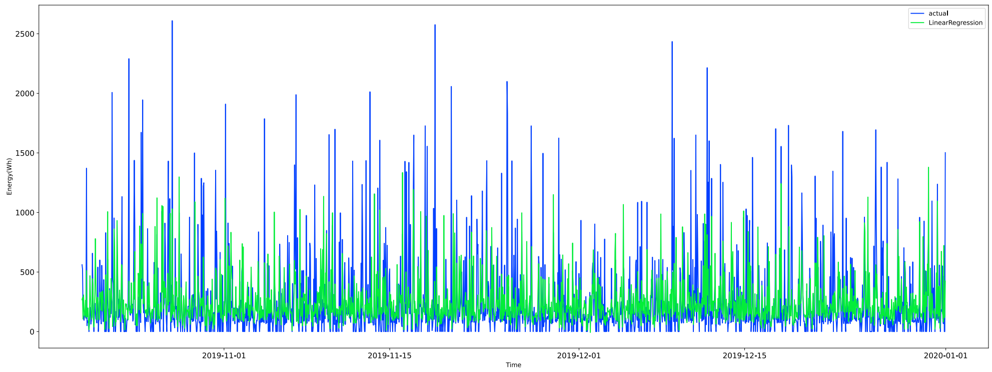


Figure 7.7: Actual and predicted energy consumption forecasting for EB3 home with HC and time(hour) feature.

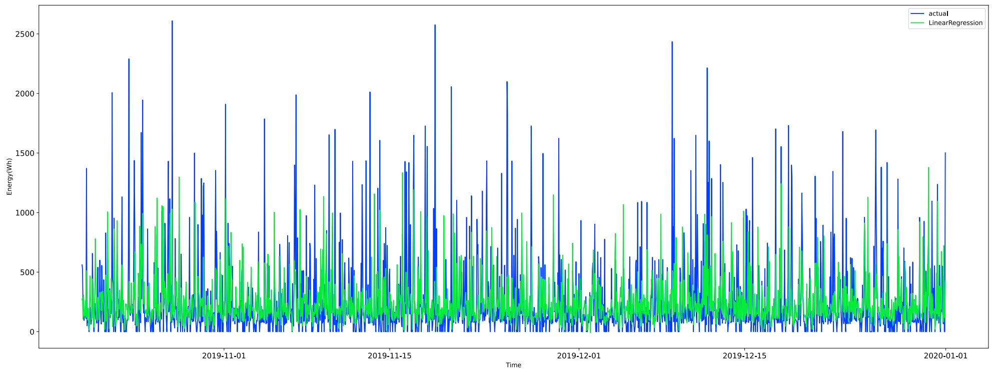


Figure 7.8: Actual and predicted energy consumption forecasting for ENEA 4 home with HC and time(hour) feature.

versions of the SVR model, the same as the LR model. In the initial version, we trained the model with HC and AR (2) values. We included time information in the second edition (hour of the day) to train the model. The SVR model's performance is incredibly reliant on the model's setup. We utilized the grid search approach to get the best designs for the SVR. It is a typical method for tweaking hyper parameters. It aims to carefully create and analyze a model for each conceivable combination of algorithm parameters presented in a grid.Table 7.6 shows the configurations that were determined using the grid search technique. Table 7.7 illustrates the forecasting errors determined using two distinct SVR models for homes EB-3 and ENEA 4.

Table 7.6: Parameter configurations of SVR model

Parameters	**Value**
Kernel	linear
gamma	0.8
C	4
epsilon	0.001

Table 7.7: SVR forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.22	0.10	0.66
EB-3	HC & time (hour)	0.22	0.10	0.66
ENEA-4	HC	0.32	0.16	0.48
ENEA-4	HC & time (hour)	0.31	0.15	0.50

The findings in Table 7.7 show that the SVR model performs better than the naïve predictor in forecasting. However, the MAE of SVR is better than LR in almost all the cases, but RMSE and R are not better than LR. We can see an improvement in predicting accuracy, especially when we train the models using HC and time features (hour of the day) in the case of Home ENEA 4. Figures 7.9 to 7.12 compare the SVR model's actual and anticipated energy consumption figures for both versions utilized.

7.4.3.5 RF Forecasting Results

RF is an ensemble model that consists of a bagging framework and an independent decision tree. In ensemble learning, models combine the output from various models.

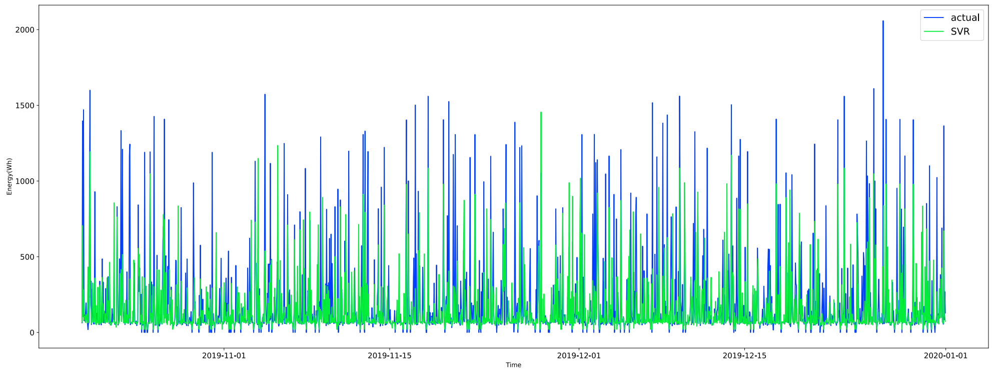


Figure 7.9: Actual and predicted energy consumption forecasting for EB3 home with HC.

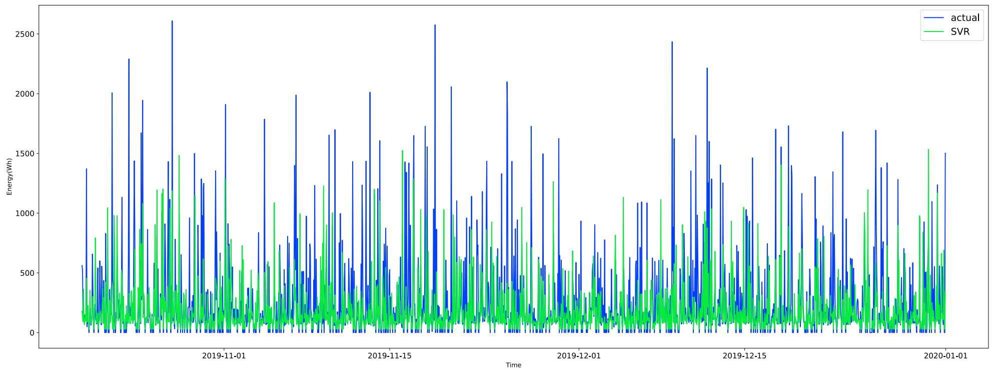


Figure 7.10: Actual and predicted energy consumption forecasting for ENEA 4 home with HC .

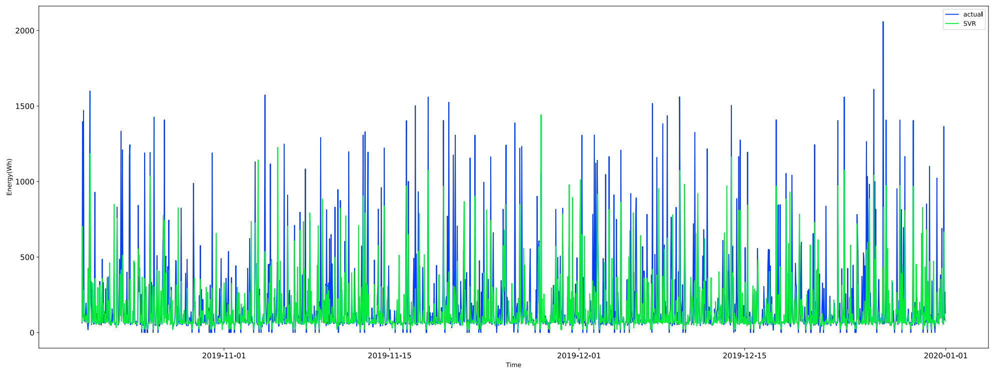


Figure 7.11: Actual and predicted energy consumption forecasting for EB3 home with HC and time(hour) feature.

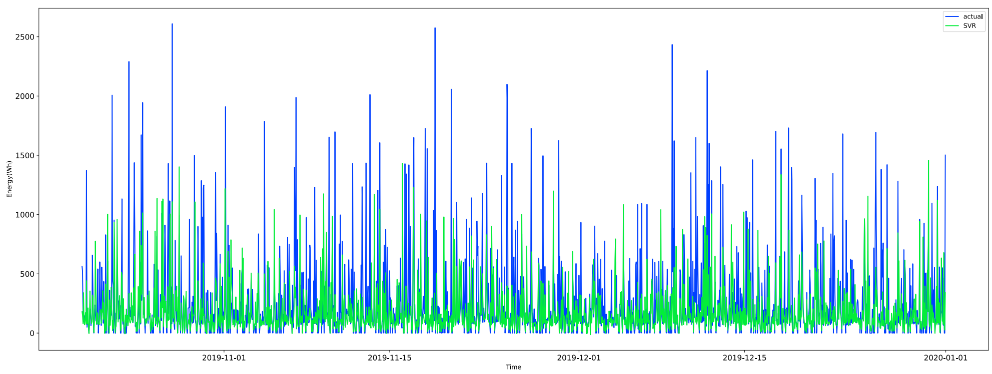


Figure 7.12: Actual and predicted energy consumption forecasting for ENEA 4 home with HC and time(hour) feature.

An ensemble model produces better results than a single model(details of the model are given in the early section 7.3.4.2.3). We used two different versions of the RF model to evaluate the application of this model for our datasets. We trained the model with HC and AR (2) values in the first version. Second, we introduced time information (hour of the day) to train the model. The performance of the RF model is hugely dependent on the model's configuration. We used the grid search method to get the most acceptable parameters for the Rf. It is a common way of fine-tuning hyperparameters. Table 7.8 displays the RF parameters chosen using the grid search approach. Only two hyperparameters are tuned: the number of trees and the maximum depth.

Table 7.8: Parameter configurations of RF model

Parameters	**Value**
n_estimators	1120
max_depth	7

Table 7.9: RF forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.20	0.10	0.69
EB-3	HC & time (hour)	0.19	0.09	0.74
ENEA-4	HC	0.29	0.16	0.55
ENEA-4	HC & time (hour)	0.27	0.15	0.62

Table 7.9 reveals that RF outperformed the other models in predicting accuracy. We can observe that incorporating time information with the HC improves predicting errors significantly. Figures 7.13 to 7.16 compare the model's actual and expected energy consumption values for both versions.

7.4.3.6 MLP Forecasting Results

The (MLP) is a type of feedforward artificial neural network in which input layers, hidden layers, and output layers are entirely linked, more details of the model are shown in section 7.3.4.3.1. Like previous models used in this work, we are using two different sets of features for the training of the model first one is HC with AR (2) values followed by time information (hours of the day). MLP is also sensitive to its

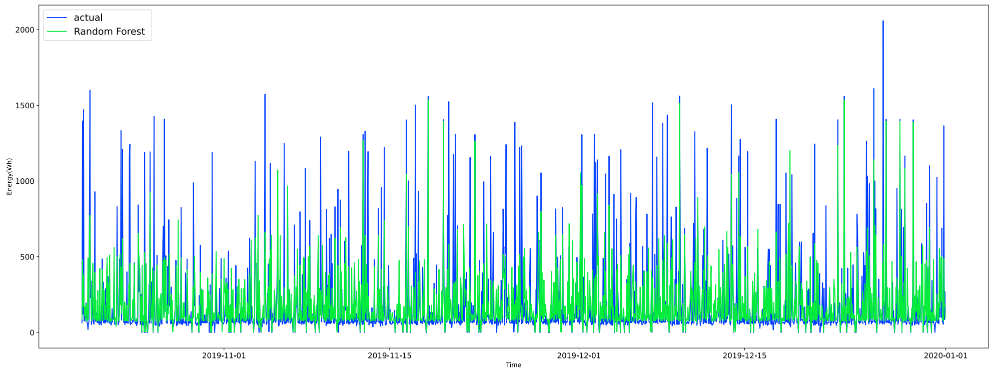


Figure 7.13: Actual and predicted energy consumption forecasting for EB3 home with HC.

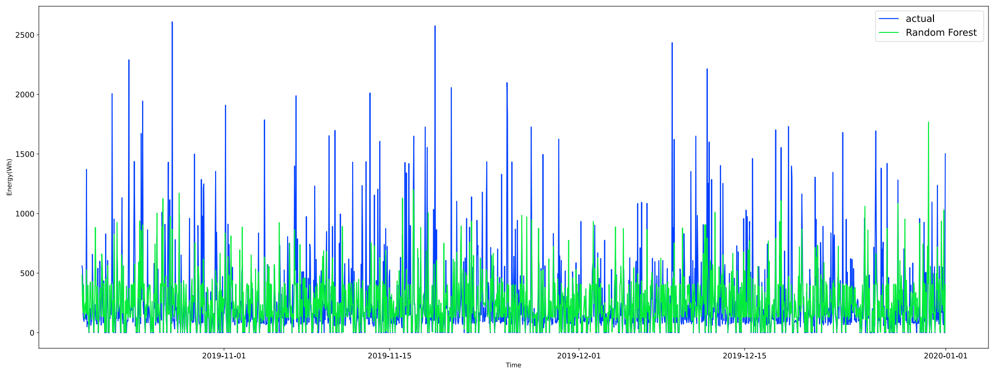


Figure 7.14: Actual and predicted energy consumption forecasting for ENEA 4 home with HC.

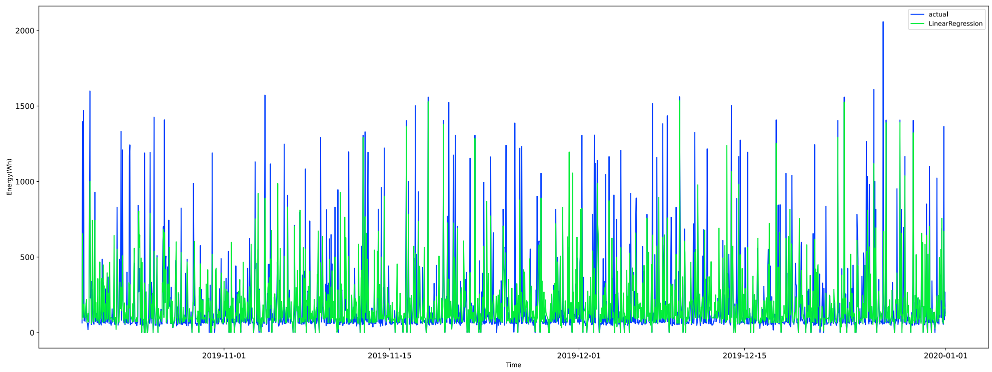


Figure 7.15: Actual and predicted energy consumption forecasting for EB3 home with HC and time(hour) feature.

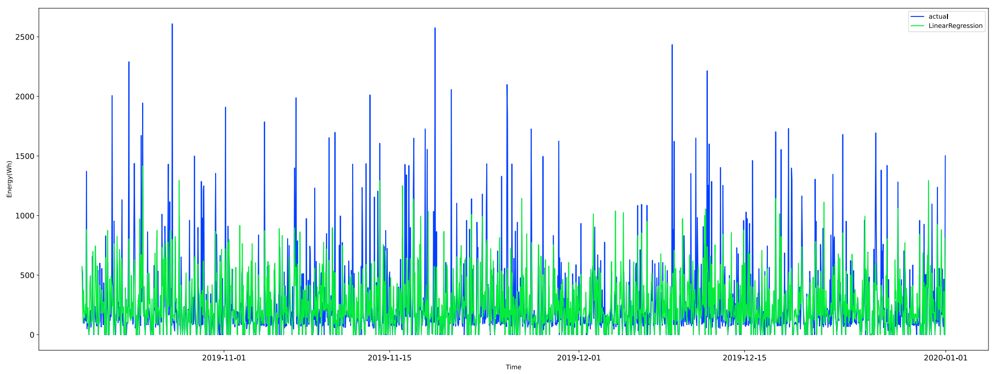


Figure 7.16: Actual and predicted energy consumption forecasting for ENEA 4 home with HC and time(hour) feature.

hyperparameters. Table 7.10 shows the list of the best parameters attained using grid search.

Table 7.10: Parameter configurations of MLP model

Parameters	**Value**
Hidden Layers	2
Activation Function	Relu
Optimizer	Adam
Batch size	16
Epochs	200

Table 7.11: MLP forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.21	0.10	0.73
EB-3	HC & time (hour)	0.20	0.09	0.75
ENEA-4	HC	0.29	0.14	0.61
ENEA-4	HC & time (hour)	0.29	0.13	0.63

Table 7.11 reveals that MLP performed well in terms of R values compared to other applied models. However, in some cases, MAE and RMSE are also improved. Figures 7.17 to 7.20 compare the model's actual and expected energy consumption values for both feature sets and homes under experimentation.

7.4.3.7 LSTM Forecasting Results

Apart from single data points, the LSTM provides feedback connections, which means it can comprehend the complete data sequence. The details of the model are already discussed in the earlier LSTM section. This work used the same two feature sets for the other model training. LSTM is also sensitive to its hyperparameters. Table 7.12 shows the list of the best parameters attained using grid search.

The findings in Table 7.13 show that the LSTM model performs better than the LR and SVR but below than RF predictor in forecasting. However, the MAE of LSTM is better than RF for home ENEA4. We can see no improvement in predicting accuracy, especially when we train the models using HC and time features (hour of the day) in the case of Home ENEA 4.Figures 7.21 to 7.24 compare the model's actual and expected energy consumption values for feature sets and homes.

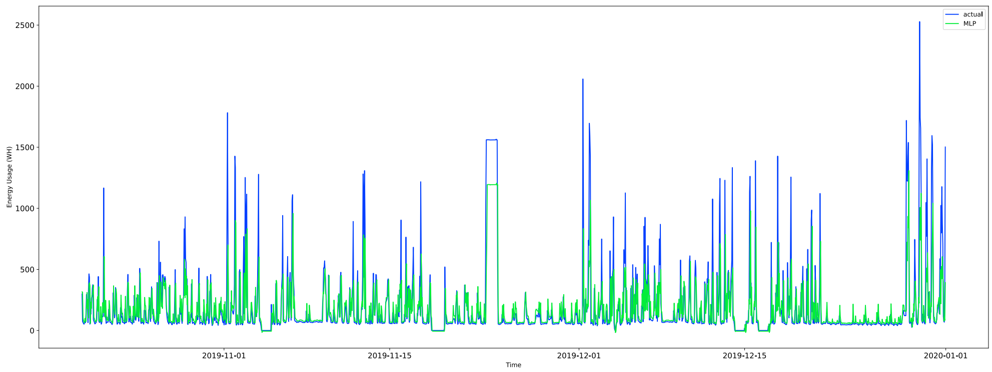


Figure 7.17: Actual and predicted energy consumption forecasting for EB3 home with HC.

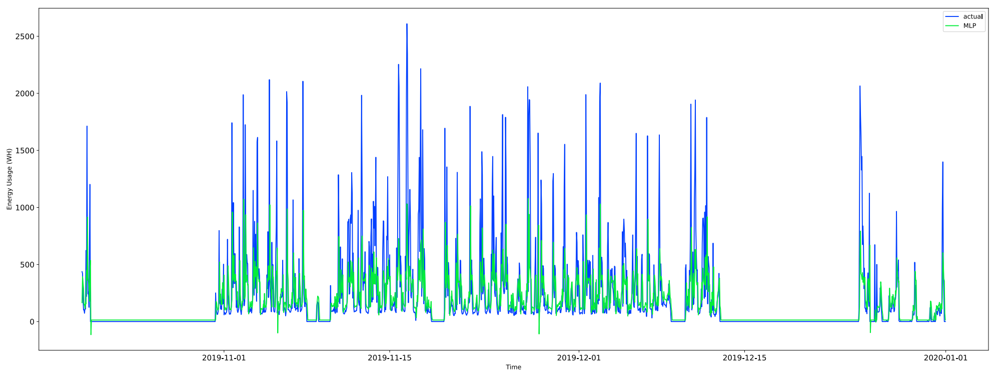


Figure 7.18: Actual and predicted energy consumption forecasting for ENEA 4 home with HC.

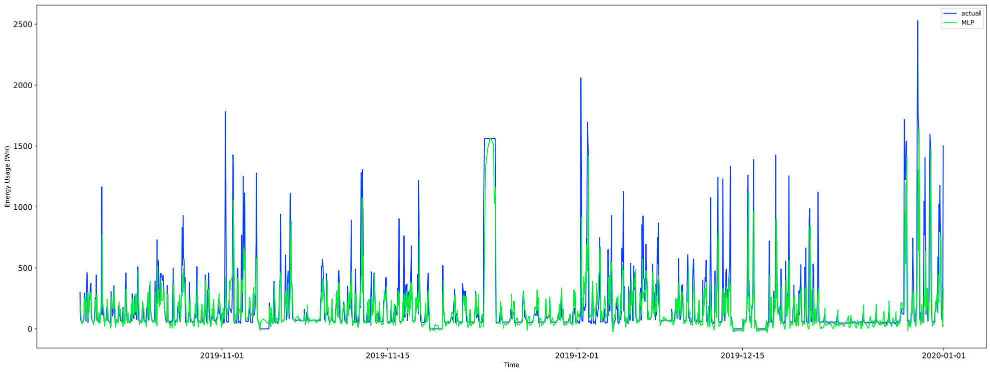


Figure 7.19: Actual and predicted energy consumption forecasting for EB3 home with HC and time(hour) feature.

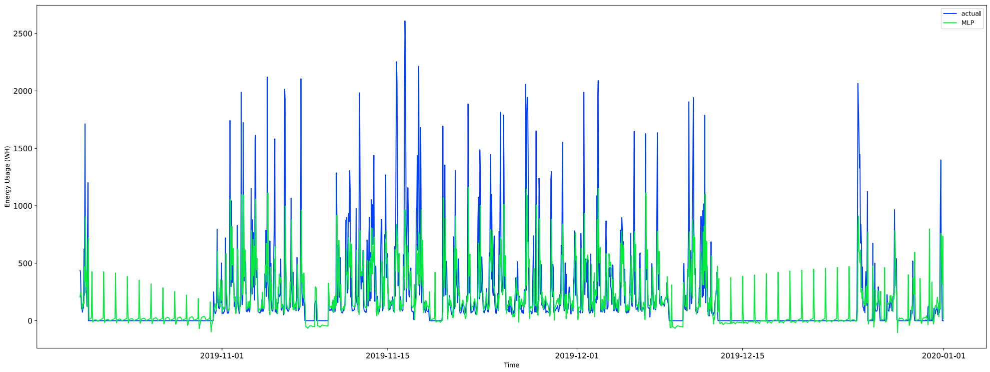


Figure 7.20: Actual and predicted energy consumption forecasting for ENEA 4 home with HC and time(hour) feature.

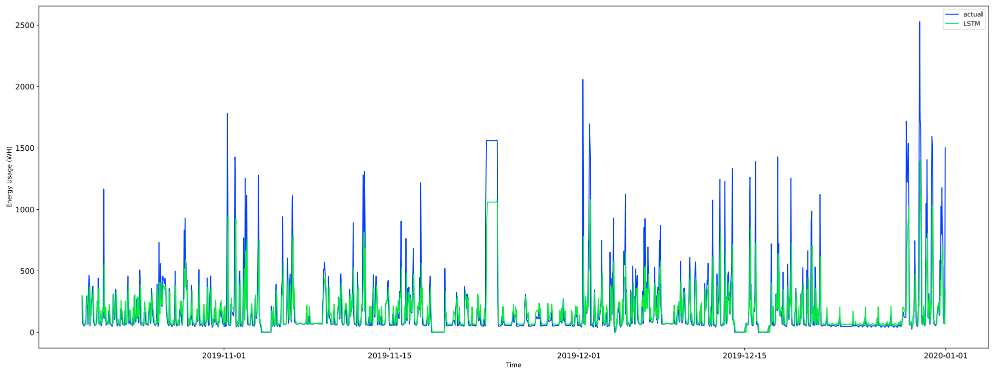


Figure 7.21: Actual and predicted energy consumption forecasting for EB3 home with HC.

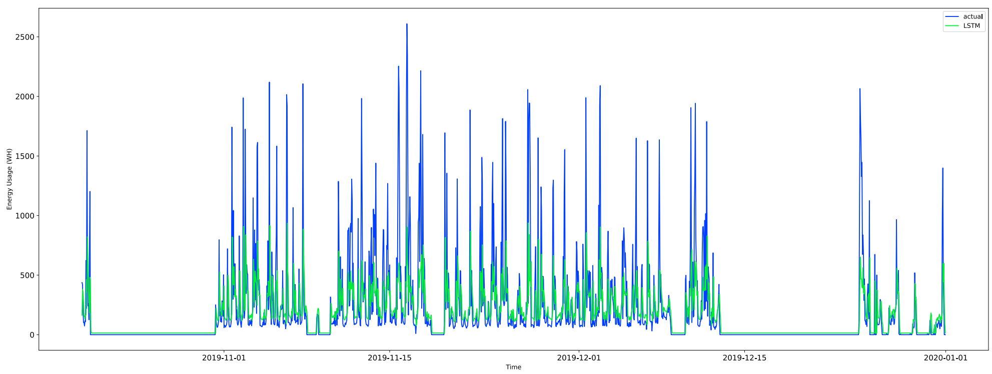


Figure 7.22: Actual and predicted energy consumption forecasting for ENEA 4 home with HC.

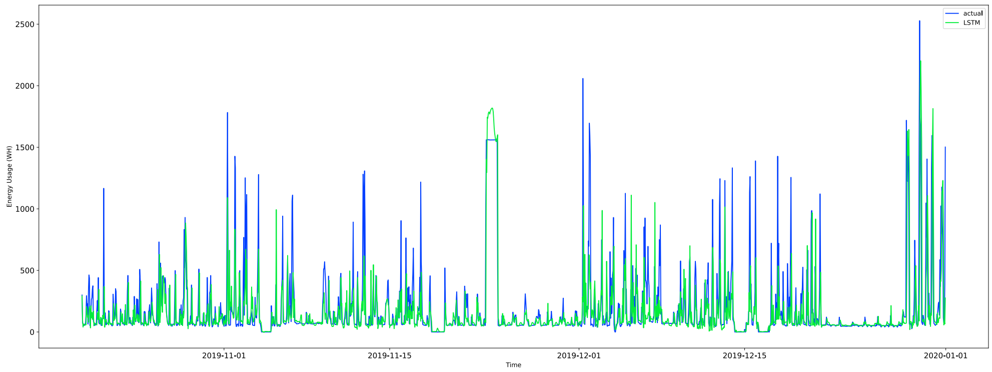


Figure 7.23: Actual and predicted energy consumption forecasting for EB3 home with HC and time(hour) feature.

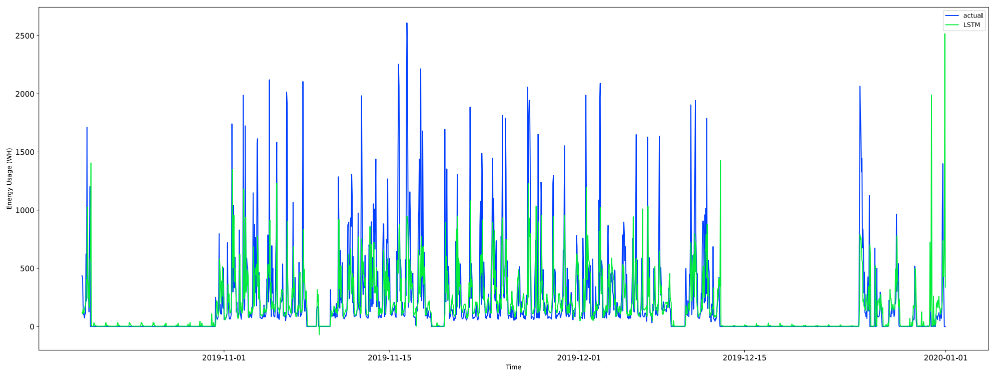


Figure 7.24: Actual and predicted energy consumption forecasting for ENEA 4 home with HC and time(hour) feature.

Table 7.12: Parameter configurations of LSTM model

Parameters	**Value**
Hidden Layers	2
Activation Function	Relu
Optimizer	Adam
Batch size	8
Epochs	200

Table 7.13: LSTM forecasting errors.

Home	**Input Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
EB-3	HC	0.21	0.10	0.72
EB-3	HC & time (hour)	0.20	0.09	0.72
ENEA-4	HC	0.29	0.13	0.64
ENEA-4	HC & time (hour)	0.30	0.13	0.58

7.5 Analysis and Discussion

The comparative analysis aims to evaluate which predictive model generated the best energy demand prediction for the homes under study. The very first model used is the SARIMA model. After the statistical test, it was found that this model is unsuitable for our actual historical data. Afterward, In empirical testing, most forecasting models enhance their prediction accuracy outcomes when models are trained to use HC and time information (hour of the day) as a feature set. Specifically, incorporating time information in addition to HC improves LR and RF predicting outcomes greatly. Moreover, MLP and LSTM for home ENEA 4 showed suboptimal results when hourly time information features were used. The findings in Table 7.14 show that RF outperformed its equivalent models in the case of EB3, and LSTM shows the best results for ENEA 4.

Table 7.14: Performance comparison of forecasting methods

.

Model	**Home**	**Features**	**RMSE** (kWh)	**MAE** (kWh)	**R**
Naïve	EB3	HC	0.23	0.10	0.65
	ENEA4	HC	0.34	0.16	0.51
SARIMA	EB3	HC	0.31	0.23	-0.03
	ENEA4	HC	0.37	0.24	0.02
LR	EB3	HC	0.21	0.12	0.66
	EB3	HC+Time(H)	0.21	0.11	0.67
	ENEA4	HC	0.31	0.18	0.48
	ENEA4	HC+Time(H)	0.30	0.17	0.52
SVR	EB3	HC	0.22	0.10	0.66
	EB3	HC+Time(H)	0.22	0.10	0.66
	ENEA4	HC	0.32	0.16	0.48
	ENEA4	HC+Time(H)	0.31	0.15	0.50
RF	EB3	HC	0.20	0.10	0.69
	EB3	HC+Time(H)	0.19	0.09	0.74
	ENEA4	HC	0.29	0.16	0.55
	ENEA4	HC+Time(H)	0.27	0.15	0.62
MLP	EB3	HC	0.21	0.10	0.73
	EB3	HC+Time(H)	0.20	0.09	0.75
	ENEA4	HC	0.29	0.14	0.61
	ENEA4	HC+Time(H)	0.29	0.13	0.63
LSTM	EB3	HC	0.21	0.10	0.72
	EB3	HC+Time(H)	0.20	0.09	0.72
	ENEA4	HC	0.29	0.13	0.64
	ENEA4	HC+Time(H)	0.30	0.13	0.58

7.6 Summary

In this chapter, comparative analysis is presented with different classical statistical (SARIMA) machine learning (LR, SVR, RF)and deep learning (MLP, LSTM) approaches for energy consumption demand prediction. Initially, we have data from 14 different homes, but for experimentation, we have used two homes' data(EB3, ENEA4). Other methods aim to find the best forecasting method for home users to provide them with their future energy consumption based on their own real historical data. To find a better model among selected models, we tested the models on two homes with different energy consumption patterns. The prediction models are sensitive to the input features used for the prediction for this purpose; two different feature sets are used first one is based on historical consumption with two lags, AR(2), and the second one is AR(2) with time information(hour of the day).

Furthermore, finding the best hyperparameters for a model is an important task; thanks to the grid search method, we adopted the process to find the best configuration of the machine and deep learning models. While conducting the experiments, it is observed that involving time information in the feature set has enhanced the performance of the employed models. According to our findings, RF is the best model for our home's energy demand forecasting, but MLP and LSTM provide good results in a few instances.

Chapter 8

Conclusion and Future Work

This chapter summarizes the dissertation with closing remarks based on important contributions of this research study and some directions toward future work.

8.1 Conclusion

This doctoral thesis contributes toward achieving the targets of ENEA and the university of Bergamo collaboration for data aggregation from sensors and then integrating that with the DHOMUS platform for data visualization and development of methods to provide the user of smart homes with better energy consumption knowledge. More explicitly, the accomplishments of this dissertation are:

- We have developed a multi module Winter device and used its environmental module to acquire environmental parameters (Temperature, Humidity, Pressure). Further DHOMUS platform cloud service makes it possible to aggregate, visualize, and analyze data streams in the cloud.

- We have developed a data-driven disaggregation method for electricity-based energy consumption for smart homes. The proposed algorithm disaggregates home energy consumption into different sectors. The data acquired for the method has two datasets: data from our deployed sensors and the second one with the active participation of users by filling out a questionnaire regarding their energy uses. The proposed method has provided efficient energy consumption disaggregation into nine different sectors, and these acquired results are provided to users of the smart home. Having energy knowledge about each energy sector(Heating, cooling, Kitchen, Lighting, etc.) can help users think more precisely about their energy usage. Finally, experimental evidence of energy consumption saving based on proposed approach is presented in section 5.6, which shows the achievement of our second objective.

- We devised a method to detect energy consumption patterns from household appliance data for a smart home platform. The proposed model is quite simple, computationally less complex, and capable of detecting patterns such as the number of times the appliance is on, off, on standby, start time, end time, and duration period for which the appliance remains operational. The method can also handle both cyclic and non-cyclic appliances. Using descriptive data mining techniques for cyclic appliances provides information on the total number of cycles and disaggregation of aggregate, short, medium, and large cycles. Moreover, the proposed model distinguishes energy consumption computation over customized periods and disaggregation aggregated cycles into the individual cycle. Afterward, the model is applied to all appliances. Dataset A is used for the training propose, and parameter tuning has been done because the cyclic appliance is sensitive to the minimum duration threshold and minimum energy consumed by the meaningful cycle. Dataset B is used for validation, and the model can detect all required patterns with high efficiency. The algorithm is robust to the time as both datasets have different starting and end timestamps. Furthermore, results of two cyclic appliances (washing machine, dishwasher) are presented that elaborate standby cumulative distribution of energy consumption, the shape of the single washing cycle, detection of close and consecutive washing cycles, the energy consumption of washing cycles, as well as monthly statistics. Energy consumption comparison of different class washing machines and dishwashers is presented. However, results depend on the amount of data acquired, but the model can still provide all patterns. In addition, proposed model has been tested on the publicly available data set and produced excellent results. Finally, the developed model has been integrated into the DHOMUS platform to offer users feedback on their electricity usage and enable them to compare their energy habits and patterns with those of other experiment participants.

- We have performed a comparative analysis of machine learning and deep learning methods in order to identify a suitable forecasting model for home energy consumption forecasting. Initially, we had data from 14 homes, but two homes were selected after data visualization with better data. We have applied several methods starting from classical statistical prediction methods. However, after a few statistical tests, it was found that the nature of the data was not aligned with these models, so we moved to machine learning models (LR, SVR, RF). We proposed two feature sets as input to the machine learning models and wanted to see the behavior of two feature sets. The first feature is the historical consumption with two lags AR(2) that we got during the statistical test, and the second one is AR(2) with time information(hour of the day). After parameter tuning, these models showed outstanding results for the first feature set. However, the result with the second feature set enhanced the prediction with less error computed by evaluation metrics (RMSE, MAE, R). The results from RF surpassed its other counterparts for both feature sets. Lately, we have adopted some deep learning models (MLP, LSTM) with both feature sets, and hyperparameter tuning results show some improvement for a few evaluation metrics. However, overall RF model has the best results among all used models.

8.2 Future Work

Our developed Winter device has been used as an IoT node for environmental monitoring (Chapter 4). However, in the future, the diversity of embedded sensors and the enhanced possibilities offered by the extension connection enable Winter to be used for a wide range of applications, including activity tracking, home habilitation help, physiological monitoring, and quality of life evaluation. The work regarding the development of algorithms for disaggregation of overall home electricity energy consumption for smart homes (Chapter 5) and energy consumption pattern detecting techniques for household appliances (Chapter 6) could be extended for smart city platforms. Moreover, the comparative analysis of prediction on electricity demand for smart homes (Chapter 7) for our datasets acquired from the platform could be extended to different forecasting approaches and horizons.

Bibliography

[1] Peck Yean Gan and ZhiDong Li. "An econometric study on long-term energy outlook and the implications of renewable energy utilization in Malaysia". In: *Energy Policy* 36.2 (2008), pp. 890–899.

[2] K Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. "Is disaggregation the holy grail of energy efficiency? The case of electricity". In: *Energy policy* 52 (2013), pp. 213–234.

[3] Pan Li, Hao Wang, and Baosen Zhang. "A distributed online pricing strategy for demand response programs". In: *IEEE Transactions on Smart Grid* 10.1 (2017), pp. 350–360.

[4] Jiajia Yang, Junhua Zhao, Fushuan Wen, Weicong Kong, and Zhaoyang Dong. "Mining the big data of residential appliances in the smart grid environment". In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE. 2016, pp. 1–5.

[5] Sai Mounika Errapotu, Jingyi Wang, Yanmin Gong, Jin-Hee Cho, Miao Pan, and Zhu Han. "SAFE: Secure appliance scheduling for flexible and efficient energy consumption for smart home IoT". In: *IEEE Internet of Things Journal* 5.6 (2018), pp. 4380–4391.

[6] Remo Lutolf. "Smart home concept and the integration of energy meters into a home based system". In: *Seventh international conference on metering apparatus and tariffs for electricity supply 1992*. IET. 1992, pp. 277–278.

[7] Davit Marikyan, Savvas Papagiannidis, and Eleftherios Alamanos. "A systematic review of the smart home literature: A user perspective". In: *Technological Forecasting and Social Change* 138 (2019), pp. 139–154.

[8] Frances K Aldrich. "Smart homes: past, present and future". In: *Inside the smart home*. Springer, 2003, pp. 17–39.

[9] Liyanage C De Silva, Chamin Morikawa, and Iskandar M Petra. “State of the art of smart homes”. In: *Engineering Applications of Artificial Intelligence* 25.7 (2012), pp. 1313–1321.

[10] Kara Saul-Rinaldi, Robin LeBaron, and Julie Caracino. “Making sense of the Smart Home”. In: *National Home performance council* (2013).

[11] Nazmiye Balta-Ozkan, Benjamin Boteler, and Oscar Amerighi. “European smart home market development: Public views on technical and economic aspects across the United Kingdom, Germany and Italy”. In: *Energy Research & Social Science* 3 (2014), pp. 65–77.

[12] Tom Hargreaves and Charlie Wilson. “Introduction: Smart homes and their users”. In: *Smart homes and their users*. Springer, 2017, pp. 1–14.

[13] Yolande Strengers and Larissa Nicholls. “Convenience and energy consumption in the smart home of the future: Industry visions from Australia and beyond”. In: *Energy Research & Social Science* 32 (2017), pp. 86–93.

[14] Jungwoo Shin, Yuri Park, and Daeho Lee. “Who will be smart home users? An analysis of adoption and diffusion of smart homes”. In: *Technological Forecasting and Social Change* 134 (2018), pp. 246–253.

[15] Kirsten Gram-Hanssen and Sarah J Darby. ““Home is where the smart is”? Evaluating smart home research and approaches against the concept of home”. In: *Energy Research & Social Science* 37 (2018), pp. 94–101.

[16] Benjamin K Sovacool and Dylan D Furszyfer Del Rio. “Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies”. In: *Renewable and sustainable energy reviews* 120 (2020), p. 109663.

[17] T Serrenho and P Bertoldi. “Smart home and appliances: State of the art”. In: *Energy, communications, protocols, standards. Brussels: JRC technical reports* (2019), pp. 2–36.

[18] Carles Gomez, Stefano Chessa, Anthony Fleury, George Roussos, and Davy Preuveneers. “Internet of Things for enabling smart environments: A technology-centric perspective”. In: *Journal of Ambient Intelligence and Smart Environments* 11.1 (2019), pp. 23–43.

[19] Rafik Belloum and Charles Consel. “End-user development of activity-supporting services for smart homes”. In: *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE. 2020, pp. 1–2.

[20] Alexandra Schieweck, Erik Uhde, Tunga Salthammer, Lea C Salthammer, Lidia Morawska, Mandana Mazaheri, and Prashant Kumar. “Smart homes and the control of indoor air quality”. In: *Renewable and Sustainable Energy Reviews* 94 (2018), pp. 705–718.

[21] Gonçalo Marques and Rui Pitarma. “mHealth: indoor environmental quality measuring system for enhanced health and well-being based on internet of things”. In: *Journal of Sensor and Actuator Networks* 8.3 (2019), p. 43.

[22] Ewa Rudnicka, Paulina Napierała, Agnieszka Podfigurna, Błażej Męczekalski, Roman Smolarczyk, and Monika Grymowicz. “The World Health Organization (WHO) approach to healthy ageing”. In: *Maturitas* 139 (2020), pp. 6–11.

[23] Laura Vadillo Moreno, Maria Luisa Martin Ruiz, Javier Malagón Hernández, Miguel Ángel Valero Duboy, and Maria Lindén. “The role of smart homes in intelligent homecare and healthcare environments”. In: *Ambient Assisted Living and Enhanced Living Environments*. Elsevier, 2017, pp. 345–394.

[24] Omar H Salman, Zahraa Taha, Muntadher Q Alsabah, Yaseein S Hussein, Ahmed S Mohammed, and Mohammed Aal-Nouman. “A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work”. In: *Computer Methods and Programs in Biomedicine* 209 (2021), p. 106357.

[25] George Myrizakis and Euripides GM Petrakis. “iHome: smart home management as a service in the cloud and the fog”. In: *International Conference on Advanced Information Networking and Applications*. Springer. 2019, pp. 1181–1192.

[26] Marco Dell’Isola, Giorgio Ficco, Laura Canale, Boris Igor Palella, and Giovanni Puglisi. “An IoT integrated tool to enhance user awareness on energy consumption in residential buildings”. In: *Atmosphere* 10.12 (2019), p. 743.

[27] Hitoshi Ikezawa and Masakatsu Imafuku. "Convenience survey of IoT house equipment for a smart life". In: *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE. 2020, pp. 290–294.

[28] Ad van Berlo Ilse Bierhoff. "Towards an inclusive future". In: East Sussex Press, 2020. Chap. Smart home environment.

[29] Joseph Bugeja, Paul Davidsson, and Andreas Jacobsson. "Functional classification and quantitative analysis of smart connected home devices". In: *2018 Global Internet of Things Summit (GIoTS)*. IEEE. 2018, pp. 1–6.

[30] Bing Dong, Vishnu Prakash, Fan Feng, and Zheng O'Neill. "A review of smart building sensing system for better indoor environment control". In: *Energy and Buildings* 199 (2019), pp. 29–46.

[31] Stefan Seer, Norbert Brändle, and Carlo Ratti. "Kinects and human kinetics: A new approach for studying pedestrian behavior". In: *Transportation research part C: emerging technologies* 48 (2014), pp. 212–228.

[32] Yuvraj Agarwal, Bharathan Balaji, Rajesh Gupta, Jacob Lyles, Michael Wei, and Thomas Weng. "Occupancy-driven energy management for smart building automation". In: *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*. 2010, pp. 1–6.

[33] Timothy W Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. "Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors". In: *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. 2012, pp. 309–322.

[34] Emilio Sardini and Mauro Serpelloni. "Self-powered wireless sensor for air temperature and velocity measurements with energy harvesting capability". In: *IEEE Transactions on instrumentation and measurement* 60.5 (2010), pp. 1838–1844.

[35] Sandhyalaxmi G Navada, Chandrashekara S Adiga, Savitha G Kini, et al. "A study on daylight integration with thermal comfort for energy conservation in a general office". In: *Int. J. Electr. Energy* 1.1 (2013), pp. 18–22.

[36] Nabil Nassif. "A robust CO2-based demand-controlled ventilation control strategy for multi-zone HVAC systems". In: *Energy and buildings* 45 (2012), pp. 72–81.

[37] Aakash C Rai, Prashant Kumar, Francesco Pilla, Andreas N Skouloudis, Silvana Di Sabatino, Carlo Ratti, Ansar Yasar, and David Rickerby. "End-user perspective of low-cost sensors for outdoor air pollution monitoring". In: *Science of The Total Environment* 607 (2017), pp. 691–705.

[38] Ayşe Berika Varol. "Compilation of Data Link Protocols: Bluetooth Low Energy (BLE), ZigBee and Z-Wave". In: *2019 4th International Conference on Computer Science and Engineering (UBMK)*. IEEE. 2019, pp. 85–90.

[39] Wael Ayoub, Mohamad Mroue, Fabienne Nouvel, Abed Ellatif Samhat, and Jean-Christophe Prévotet. "Towards ip over lpwans technologies: Lorawan, dash7, nb-iot". In: *2018 sixth international conference on digital information, networking, and wireless communications (dinwc)*. IEEE. 2018, pp. 43–47.

[40] Antonio Cilfone, Luca Davoli, Laura Belli, and Gianluigi Ferrari. "Wireless mesh networking: An IoT-oriented perspective survey on relevant technologies". In: *Future internet* 11.4 (2019), p. 99.

[41] Salim Jibrin Danbatta and Asaf Varol. "Comparison of Zigbee, Z-Wave, Wi-Fi, and bluetooth wireless technologies used in home automation". In: *2019 7th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE. 2019, pp. 1–5.

[42] Byomakesh Mahapatra, Sanket Kumar Panda, Ashok Kumar Turuk, Bibhudatta Sahoo, and Sarat Kumar Patra. "LW-AKA: A Security Protocol for Integrated RFID and IoT Based Smart Home Security System". In: *2019 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)*. IEEE. 2019, pp. 331–336.

[43] Eyhab Al-Masri, Karan Raj Kalyanam, John Batts, Jonathan Kim, Sharanjit Singh, Tammy Vo, and Charlotte Yan. "Investigating Messaging Protocols for the Internet of Things (IoT)". In: *IEEE Access* 8 (2020), pp. 94880–94911. DOI: 10.1109/ACCESS.2020.2993363.

[44] Mauro AA da Cruz, Joel JPC Rodrigues, Ellen S Paradello, Pascal Lorenz, Petar Solic, and Victor Hugo C Albuquerque. "A proposal for bridging the message queuing telemetry transport protocol to HTTP on IoT solutions". In: *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE. 2018, pp. 1–5.

[45] Pin Nie. "An open standard for instant messaging: eXtensible Messaging and Presence Protocol (XMPP)". In: *TKK T-110.5190 Seminar on Internetworking, Helsinki University of Technology, Finland*. Citeseer. 2006, pp. 1–6.

[46] Woochul Kang, Krasimira Kapitanova, and Sang Hyuk Son. "RDDS: A real-time data distribution service for cyber-physical systems". In: *IEEE Transactions on Industrial Informatics* 8.2 (2012), pp. 393–405.

[47] OASIS Standard. "Oasis advanced message queuing protocol (amqp) version 1.0". In: *International Journal of Aerospace Engineering Hindawi www. hindawi. com* 2018 (2012).

[48] Behzad Najafi, Sadaf Moaveninejad, and Fabio Rinaldi. "Data analytics for energy disaggregation: Methods and applications". In: *Big data application in power systems*. Elsevier, 2018, pp. 377–408.

[49] EJ Aladesanmi and KA Folly. "Overview of non-intrusive load monitoring and identification techniques". In: *IFAC-PapersOnLine* 48.30 (2015), pp. 415–420.

[50] I Abubakar, SN Khalid, MW Mustafa, Hussain Shareef, and M Mustapha. "Application of load monitoring in appliances' energy management–A review". In: *Renewable and Sustainable Energy Reviews* 67 (2017), pp. 235–245.

[51] George William Hart. "Nonintrusive appliance load monitoring". In: *Proceedings of the IEEE* 80.12 (1992), pp. 1870–1891.

[52] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey". In: *Sensors* 12.12 (2012), pp. 16838–16866.

[53] MN Valero Pérez. "A non-intrusive appliance load monitoring system for identifying kitchen activities". In: *Department of Automation and System Technology, Aalto University-School of Electrical Engineering, Espoo, Finland* (2011).

[54] Jian Liang, Simon KK Ng, Gail Kendall, and John WM Cheng. "Load signature study—Part II: Disaggregation framework, simulation, and applications". In: *IEEE Transactions on Power Delivery* 25.2 (2009), pp. 561–569.

[55] Yung Fei Wong, Y Ahmet Şekercioğlu, Tom Drummond, and Voon Siong Wong. “Recent approaches to non-intrusive load monitoring techniques in residential settings”. In: *2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG)*. IEEE. 2013, pp. 73–79.

[56] Hsueh-Hsien Chang, Ching-Lung Lin, and Jin-Kwei Lee. “Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms”. In: *The 2010 14th International Conference on Computer Supported Cooperative Work in Design*. IEEE. 2010, pp. 27–32.

[57] Yi-Cheng Chen, Yu-Lun Ko, Wen-Chih Peng, and Wang-Chien Lee. “Mining appliance usage patterns in smart home environment”. In: *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part I 17*. Springer. 2013, pp. 99–110.

[58] Mithra Moezzi. “The predicament of efficiency”. In: *Proceedings of the 1998 Summer Study on Energy Efficiency in Buildings*. Vol. 4. Citeseer. 1998.

[59] Robert Hendron and Mark Eastment. *Development of an energy-savings calculation methodology for residential miscellaneous electric loads*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2006.

[60] Noah Horowitz, Chris Calwell, and Susanne Foster. “Opportunities and recommendations for reducing the energy consumption of consumer electronics products”. In: *Proceedings of the 2005 IEEE International Symposium on Electronics and the Environment, 2005.* IEEE. 2005, pp. 135–139.

[61] Steven Firth, Kevin Lomas, A Wright, and R Wall. “Identifying trends in the use of domestic appliances from household electricity consumption measurements”. In: *Energy and buildings* 40.5 (2008), pp. 926–936.

[62] Paolo Bertoldi, Andrea Ricci, and Anibal de Almeida. *Energy Efficiency in househould appliances and lighting*. Springer Science & Business Media, 2012.

[63] Ke Ding, Jonathan M Gilligan, and George M Hornberger. “Avoiding" day-zero": a testbed for evaluating integrated food-energy-water management in Cape Town, South Africa”. In: *2019 Winter Simulation Conference (WSC)*. IEEE. 2019, pp. 866–877.

[64] Mark Eastment and Robert Hendron. *Method for evaluating energy use of dishwashers, clothes washers, and clothes dryers*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2006.

[65] Keith W Hipel and A Ian McLeod. *Time series modelling of water resources and environmental systems*. Elsevier, 1994.

[66] Ratnadip Adhikari and Ramesh K Agrawal. "An introductory study on time series modeling and forecasting". In: *arXiv preprint arXiv:1302.6613* (2013).

[67] Kenneth C Land and Stephen H Schneider. *Forecasting in the social and natural sciences*. Springer Science & Business Media, 2012.

[68] Nan Wei, Changjun Li, Xiaolong Peng, Fanhua Zeng, and Xinqian Lu. "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review". In: *Journal of Petroleum Science and Engineering* 181 (2019), p. 106187.

[69] Deyslen Mariano-Hernández, Luis Hernández-Callejo, Felix Santos Garcıa, Oscar Duque-Perez, and Angel L Zorita-Lamadrid. "A review of energy consumption forecasting in smart buildings: methods, input variables, forecasting horizon and metrics". In: *Applied Sciences* 10.23 (2020), p. 8323.

[70] Balram Avittathur and Debabrata Ghosh. "Forecasting and demand planning". In: *Excellence in Supply Chain Management*. Routledge, 2020, pp. 19–34.

[71] Arunesh Kumar Singh, S Khatoon, Md Muazzam, DK Chaturvedi, et al. "Load forecasting techniques and methodologies: A review". In: *2012 2nd International Conference on Power, Control and Embedded Systems*. IEEE. 2012, pp. 1–10.

[72] Nagraj Raju Balakrishnan, Barry Render, Ralph Stair, and Chuck Munson. "Forecasting Models". In: *Managerial Decision Modeling*. De Gruyter, 2017, pp. 647–730.

[73] James W Taylor. "A comparison of univariate time series methods for forecasting intraday arrivals at a call center". In: *Management Science* 54.2 (2008), pp. 253–265.

[74] Cristina Gabriela Butca, George Suciu, Adelina Ochian, Octavian Fratu, and Simona Halunga. "Wearable sensors and cloud platform for monitoring environmental parameters in e-health applications". In: *2014 11th international symposium on electronics and telecommunications (ISETC)*. IEEE. 2014, pp. 1–4.

[75] Ming Jin, Han Zou, Kevin Weekly, Ruoxi Jia, Alexandre M Bayen, and Costas J Spanos. "Environmental sensing by wearable device for indoor activity and location estimation". In: *IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2014, pp. 5369–5375.

[76] James Dieffenderfer et al. "Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease". In: *IEEE journal of biomedical and health informatics* 20.5 (2016), pp. 1251–1264.

[77] Nghi Dam, Andrew Ricketts, Benjamin Catlett, and Justin Henriques. "Wearable sensors for analyzing personal exposure to air pollution". In: *2017 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE. 2017, pp. 1–4.

[78] Patrick Locatelli and Dario Alimonti. "Differentiating essential tremor and Parkinson's disease using a wearable sensor—A pilot study". In: *2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*. IEEE. 2017, pp. 213–218.

[79] Sam Moayedi, Ahmad Almaghrebi, Jan Haase, Hiroaki Nishi, Gerhard Zucker, Nasser Aljuhaishi, and Mahmoud Alahmad. "Energy Optimization Technologies in Smart Homes". In: *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2020, pp. 1974–1979.

[80] Patricia Franco, Jose Manuel Martinez, Young-Chon Kim, and Mohamed A Ahmed. "IoT based approach for load monitoring and activity recognition in smart homes". In: *IEEE Access* 9 (2021), pp. 45325–45339.

[81] Mohamed Nait Meziane, Philippe Ravier, Guy Lamarque, Karim Abed-Meraim, Jean-Charles Le Bunetel, and Yves Raingeaud. "Modeling and estimation of transient current signals". In: *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE. 2015, pp. 1960–1964.

[82] Michael Zeifman and Kurt Roth. "Nonintrusive appliance load monitoring: Review and outlook". In: *IEEE transactions on Consumer Electronics* 57.1 (2011), pp. 76–84.

[83] Gu-yuan Lin, Shih-chiang Lee, Jane Yung-jen Hsu, and Wan-rong Jih. "Applying power meters for appliance recognition on the electric panel". In: *2010 5th IEEE Conference on Industrial Electronics and Applications*. IEEE. 2010, pp. 2254–2259.

[84] Antonio G Ruzzelli, Clement Nicolas, Anthony Schoofs, and Gregory MP O'Hare. "Real-time recognition and profiling of appliances through a single electricity sensor". In: *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*. IEEE. 2010, pp. 1–9.

[85] Tehseen Zia, Dietmar Bruckner, and Adeel Zaidi. "A hidden Markov model based procedure for identifying household electric loads". In: *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2011, pp. 3218–3223.

[86] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. "Review of smart meter data analytics: Applications, methodologies, and challenges". In: *IEEE Transactions on Smart Grid* 10.3 (2018), pp. 3125–3148.

[87] Daniel Jorde and Hans-Arno Jacobsen. "Event detection for energy consumption monitoring". In: *IEEE Transactions on Sustainable Computing* 6.4 (2020), pp. 703–709.

[88] Abdelkareem Jaradat, Hanan Lutfiyya, and Anwar Haque. "Appliance Operation Modes Identification Using Cycles Clustering". In: *arXiv preprint arXiv:2101.10472* (2021).

[89] Sami Rollins and Nilanjan Banerjee. "Using rule mining to understand appliance energy consumption patterns". In: *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE. 2014, pp. 29–37.

[90] Kwok Tai Chui, Kim Fung Tsang, Shu Hung Chung, and Lam Fat Yeung. "Appliance signature identification solution using K-means clustering". In: *IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2013, pp. 8420–8425.

[91] Youngwook Kim, Seongbae Kong, Rakkyung Ko, and Sung-Kwan Joo. "Electrical event identification technique for monitoring home appliance load using load signatures". In: *2014 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. 2014, pp. 296–297.

[92] D Srinivasan, WS Ng, and AC Liew. "Neural-network-based signature recognition for harmonic source identification". In: *IEEE transactions on power delivery* 21.1 (2005), pp. 398–405.

[93] Leijiao Ge, Wenlong Liao, Shouxiang Wang, Birgitte Bak-Jensen, and Jayakrishnan Radhakrishna Pillai. "Modeling Daily Load Profiles of Distribution Network for Scenario Generation Using Flow-Based Generative Network". In: *IEEE Access* 8 (2020), pp. 77587–77597. DOI: `10.1109/ACCESS.2020.2989350`.

[94] David Fischer, Andreas Härtl, and Bernhard Wille-Haussmann. "Model for electric load profiles with high time resolution for German households". In: *Energy and Buildings* 92 (Apr. 2015), pp. 170–179.

[95] "Constructing load profiles for household electricity and hot water from time-use data—Modelling approach and validation". In: *Energy and Buildings* 41.7 (2009), pp. 753–768. ISSN: 0378-7788.

[96] Tri Kurniawan Wijaya, Matteo Vasirani, Samuel Humeau, and Karl Aberer. "Cluster-based aggregate forecasting for residential electricity demand using smart meter data". In: *2015 IEEE International Conference on Big Data (Big Data)*. 2015, pp. 879–887. DOI: `10.1109/BigData.2015.7363836`.

[97] "A novel approach for load profiling in smart power grids using smart meter data". In: *Electric Power Systems Research* 165 (2018), pp. 191–198.

[98] "Assessing the relevance of load profiling information in electrical load forecasting based on neural network models". In: *International Journal of Electrical Power & Energy Systems* 40.1 (2012), pp. 85–93.

[99] "Electricity load profiles in Europe: The importance of household segmentation". In: *Energy Research & Social Science* 3 (2014), pp. 30–45. ISSN: 2214-6296.

[100] "Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management". In: *Energy* 180 (2019), pp. 665–677. ISSN: 0360-5442.

[101] "The North East Scotland Energy Monitoring Project: Exploring relationships between household occupants and energy usage". In: *Energy and Buildings* 75 (2014), pp. 493–503. ISSN: 0378-7788.

[102] "Smart Meter Data Analytics for Optimal Customer Selection in Demand Response Programs". In: *Energy Procedia* 107 (2017). 3rd International Conference on Energy and Environment Research, ICEER 2016, 7-11 September 2016, Barcelona, Spain, pp. 49–59. ISSN: 1876-6102.

[103] "Big data driven smart energy management: From big data to big insights". In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 215–225. ISSN: 1364-0321.

[104] "Understanding household energy consumption behavior: The contribution of energy big data analytics". In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 810–819. ISSN: 1364-0321.

[105] "Reducing Household Electricity Consumption through Demand Side Management: The Role of Home Appliance Scheduling and Peak Load Reduction". In: *Energy Procedia* 72 (2015). International Scientific Conference "Environmental and Climate Technologies, CONECT 2014, pp. 222–229. ISSN: 1876-6102.

[106] "Residential peak electricity demand response—Highlights of some behavioural issues". In: *Renewable and Sustainable Energy Reviews* 25 (2013), pp. 71–77. ISSN: 1364-0321.

[107] "Reducing household electricity consumption during evening peak demand times: Evidence from a field experiment". In: *Energy Policy* 144 (2020), p. 111657. ISSN: 0301-4215.

[108] "Move it! How an electric contest motivates households to shift their load profile". In: *Energy Economics* 68 (2017), pp. 255–270. ISSN: 0140-9883.

[109] Tracey Crosbie, John Broderick, Michael Short, Richard Charlesworth, and Muneeb Dawood. "Demand Response Technology Readiness Levels for Energy Management in Blocks of Buildings". In: *Buildings* 8.2 (2018). ISSN: 2075-5309. URL: https://www.mdpi.com/2075-5309/8/2/13.

[110] Filipe Rodrigues, Carlos Cardeira, and João Mamiel Ferreira Calado. “The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal”. In: *Energy Procedia* 62 (2014), pp. 220–229.

[111] Peter Lusis, Kaveh Rajab Khalilpour, Lachlan Andrew, and Ariel Liebman. “Short-term residential load forecasting: Impact of calendar effects and forecast granularity”. In: *Applied energy* 205 (2017), pp. 654–669.

[112] Hengfang Deng, David Fannon, and Matthew J Eckelman. “Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata”. In: *Energy and Buildings* 163 (2018), pp. 34–43.

[113] Federico Divina, Miguel Garcia Torres, Francisco A Goméz Vela, and Jose Luis Vazquez Noguera. “A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings”. In: *Energies* 12.10 (2019), p. 1934.

[114] Zeyu Wang, Yueren Wang, Ruochen Zeng, Ravi S Srinivasan, and Sherry Ahrentzen. “Random Forest based hourly building energy prediction”. In: *Energy and Buildings* 171 (2018), pp. 11–25.

[115] Md Abdulla Al Mamun and Mehmet Rasit Yuce. “Sensors and systems for wearable environmental monitoring toward IoT-enabled applications: A review”. In: *IEEE Sensors Journal* 19.18 (2019), pp. 7771–7788.

[116] Ashley Marie Polhemus et al. “Human-centered design strategies for device selection in mhealth programs: development of a novel framework and case study”. In: *JMIR mHealth and uHealth* 8.5 (2020), e16043.

[117] Patrick Locatelli, Asad Hussain, Andrea Pedrana, Matteo Pezzoli, Gianluca Traversi, and Valerio Re. “Winter: A Novel Low Power Modular Platform for Wearable and IoT Applications”. In: *International Conference on Applications in Electronics Pervading Industry, Environment and Society*. Springer. 2020, pp. 62–68.

[118] Andrea Pedrana, Daniele Comotti, Patrick Locatelli, and Gianluca Traversi. “Development of a telemedicine-oriented gait analysis system based on inertial sensors”. In: *2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*. IEEE. 2018, pp. 1–4.

[119] Markus Weiss, Adrian Helfenstein, Friedemann Mattern, and Thorsten Staake. "Leveraging smart meter data to recognize home appliances". In: *2012 IEEE International Conference on Pervasive Computing and Communications*. IEEE. 2012, pp. 190–197.

[120] Marco Danilo Burbano Acuña. "Intrusive and non-intrusive load monitoring (a survey)". In: *Latin American Journal of Computing* 2.1 (2015).

[121] Yung-Yao Chen, Ming-Hung Chen, Che-Ming Chang, Fu-Sheng Chang, and Yu-Hsiu Lin. "A smart home energy management system using two-stage non-intrusive appliance load monitoring over fog-cloud analytics based on Tridium's Niagara framework for residential demand-side management". In: *Sensors* 21.8 (2021), p. 2883.

[122] Oladayo S Ajani, Abhishek Kumar, Rammohan Mallipeddi, Swagatam Das, and Ponnuthurai Nagaratnam Suganthan. "Benchmarking Optimization-Based Energy Disaggregation Algorithms". In: *Energies* 15.5 (2022), p. 1600.

[123] Kanae Matsui, Yoshiki Yamagata, and Hiroaki Nishi. "Disaggregation of Electric Appliance's Consumption Using Collected Data by Smart Metering System". In: *Energy Procedia* 75 (2015), pp. 2940–2945.

[124] Vanh Khuyen Nguyen, Wei Emma Zhang, and Adnan Mahmood. "Semi-supervised intrusive appliance load monitoring in smart energy monitoring system". In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17.3 (2021), pp. 1–20.

[125] Dimas Augusto Mendes Lemes, Thales Wulfert Cabral, Gustavo Fraidenraich, Luis Geraldo Pedroso Meloni, Eduardo Rodrigues De Lima, and Fernando Bauer Neto. "Load disaggregation based on time window for HEMS application". In: *IEEE Access* 9 (2021), pp. 70746–70757.

[126] Andreas Gerster, Mark Andreas Andor, Lorenz Goette, et al. *Disaggregate Consumption Feedback and Energy Conservation*. SFB 823, 2021.

[127] Francesco Mancini, Jacopo Cimaglia, Gianluigi Lo Basso, and Sabrina Romano. "Implementation and Simulation of Real Load Shifting Scenarios Based on a Flexibility Price Market Strategy—The Italian Residential Sector as a Case Study". In: *Energies* 14.11 (2021), p. 3080.

[128] Francesco Mancini, Sabrina Romano, Gianluigi Lo Basso, Jacopo Cimaglia, and Livio de Santoli. "How the italian residential sector could contribute to load flexibility in demand response activities: A methodology for residential clustering and developing a flexibility strategy". In: *Energies* 13.13 (2020), p. 3359.

[129] Jacopo Cimaglia, Sabrina Romano, and Francesco Mancini. "Methodology and Simulation of Electrical Grid Peak Shaving Strategy based on Photovoltaic and Storage Optimization: an Italian Residential Sector Case Study". In: *E3S Web of Conferences*. Vol. 312. 2021, p. 01005.

[130] Lucas Pereira and Nuno Nunes. "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review". In: *Wiley Interdisciplinary Reviews: data mining and knowledge discovery* 8.6 (2018), e1265.

[131] Men-Shen Tsai and Yu-Hsiu Lin. "Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation". In: *Applied Energy* 96 (2012), pp. 55–73.

[132] Francesco Mancini, Gianluigi Lo Basso, and Livio De Santoli. "Energy use in residential buildings: Characterisation for identifying flexible loads by means of a questionnaire survey". In: *Energies* 12.11 (2019), p. 2055.

[133] Zeyu Wang and Ravi S Srinivasan. "Classification of household appliance operation cycles: A case-study approach". In: *Energies* 8.9 (2015), pp. 10522–10536.

[134] Bhagya Nathali Silva, Murad Khan, and Kijun Han. "Load balancing integrated least slack time-based appliance scheduling for smart home energy management". In: *Sensors* 18.3 (2018), p. 685.

[135] Ming-Tang Chen and Che-Min Lin. "Standby power management of a smart home appliance by using energy saving system with active loading feature identification". In: *IEEE Transactions on Consumer Electronics* 65.1 (2018), pp. 11–17.

[136] Alberto Prudenzi. "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel". In: *2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 02CH37309)*. Vol. 2. IEEE. 2002, pp. 941–946.

[137] Huijuan Wang and Wenrong Yang. "An iterative load disaggregation approach based on appliance consumption pattern". In: *Applied Sciences* 8.4 (2018), p. 542.

[138] Hafiz Majid Hussain, Nadeem Javaid, Sohail Iqbal, Qadeer Ul Hasan, Khursheed Aurangzeb, and Musaed Alhussein. "An efficient demand side management system with a new optimized home energy management controller in smart grid". In: *Energies* 11.1 (2018), p. 190.

[139] Shengli Du, Mingchao Li, Shuai Han, Jonathan Shi, and Heng Li. "Multipattern data mining and recognition of primary electric appliances from single non-intrusive load monitoring data". In: *Energies* 12.6 (2019), p. 992.

[140] "Development of building energy saving advisory: A data mining approach". In: *Energy and Buildings* 172 (2018), pp. 139–151. ISSN: 0378-7788. DOI: `https://doi.org/10.1016/j.enbuild.2018.04.052`.

[141] Krzysztof Gajowniczek and Tomasz Ząbkowski. "Data Mining Techniques for Detecting Household Characteristics Based on Smart Meter Data". In: *Energies* 8.7 (2015), pp. 7407–7427.

[142] Zhun Yu, Benjamin C. M. Fung, and Fariborz Haghighat. "Extracting knowledge from building-related data — A data mining framework". In: *Building Simulation* 6.2 (June 2013), pp. 207–222.

[143] Milad Ashouri, Benjamin C.M. Fung, Fariborz Haghighat, and Hiroshi Yoshino. "Systematic approach to provide building occupants with feedback to reduce energy consumption". In: *Energy* 194 (Mar. 2020), p. 116813.

[144] Christos Athanasiadis, Dimitrios Doukas, Theofilos Papadopoulos, and Antonios Chrysopoulos. "A scalable real-time non-intrusive load monitoring system for the estimation of household appliance power consumption". In: *Energies* 14.3 (2021), p. 767.

[145] Christos L Athanasiadis, Theofilos A Papadopoulos, and Dimitrios I Doukas. "Real-time non-intrusive load monitoring: A light-weight and scalable approach". In: *Energy and Buildings* 253 (2021), p. 111523.

[146] P Clerici Maestosi. *Sustainable Urban Transitions*. Santarcangelo di Romagna, Italy: Maggioli Publisher: 2021.

[147]] Aggiornamento delle Fasce Orarie Con decorrenza 1 gennaio. *ARERA. (n.d.). Retrieved*. https://www.arera.it/it/docs/06/181-06.html. [August 2, 2022]. 2007.

[148] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D'Alessandro, and Andrea M Tonello. "GREEND: An energy consumption dataset of households in Italy and Austria". In: *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE. 2014, pp. 511–516.

[149] Diogo M. F. Izidio, Paulo S. G. de Mattos Neto, Luciano Barbosa, João F. L. de Oliveira, Manoel Henrique da Nóbrega Marinho, and Guilherme Ferretti Rissi. "Evolutionary Hybrid System for Energy Consumption Forecasting for Smart Meters". In: *Energies* 14.7 (2021).

[150] "A review on machine learning forecasting growth trends and their real-time applications in different energy systems". In: *Sustainable Cities and Society* 54 (2020), p. 102010. ISSN: 2210-6707.

[151] "Machine learning models for forecasting power electricity consumption using a high dimensional dataset". In: *Expert Systems with Applications* 187 (2022), p. 115917. ISSN: 0957-4174.

[152] Nikolaos G. Paterakis, Elena Mocanu, Madeleine Gibescu, Bart Stappers, and Walter van Alst. "Deep learning versus traditional machine learning methods for aggregated energy demand prediction". In: *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 2017, pp. 1–6. DOI: 10.1109/ISGTEurope.2017.8260289.

[153] Tuong Le, Minh Thanh Vo, Tung Kieu, Eenjun Hwang, Seungmin Rho, and Sung Wook Baik. "Multiple Electric Energy Consumption Forecasting Using a Cluster-Based Strategy for Transfer Learning in Smart Building". In: *Sensors* 20.9 (2020).

[154] Jihoon Moon, Sungwoo Park, Seungmin Rho, and Eenjun Hwang. "A comparative analysis of artificial neural network architectures for building energy consumption forecasting". In: *International Journal of Distributed Sensor Networks* 15.9 (2019), p. 1550147719877616.

[155] Nivethitha Somu, Gauthama Raman MR, and Krithi Ramamritham. "A deep learning framework for building energy consumption forecast". In: *Renewable and Sustainable Energy Reviews* 137 (2021), p. 110591.

[156] Michal Pavlicko, Mária Vojteková, and Ol'ga Blažeková. "Forecasting of Electrical Energy Consumption in Slovakia". In: *Mathematics* 10.4 (2022), p. 577.

[157] Anam-Nawaz Khan, Naeem Iqbal, Atif Rizwan, Rashid Ahmad, and Do-Hyeun Kim. "An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings". In: *Energies* 14.11 (2021), p. 3020.

[158] Jihoon Moon, Sungwoo Park, Seungmin Rho, and Eenjun Hwang. "Robust building energy consumption forecasting using an online learning approach with R ranger". In: *Journal of Building Engineering* 47 (2022), p. 103851.

[159] Joanna Henzel, Łukasz Wróbel, Marcin Fice, and Marek Sikora. "Energy Consumption Forecasting for the Digital-Twin Model of the Building". In: *Energies* 15.12 (2022), p. 4318.

[160] Ning Jin, Fan Yang, Yuchang Mo, Yongkang Zeng, Xiaokang Zhou, Ke Yan, and Xiang Ma. "Highly accurate energy consumption forecasting model based on parallel LSTM neural networks". In: *Advanced Engineering Informatics* 51 (2022), p. 101442.

[161] Rita Banik, Priyanath Das, Srimanta Ray, and Ankur Biswas. "Prediction of electrical energy consumption based on machine learning technique". In: *Electrical Engineering* 103.2 (2021), pp. 909–920.

[162] Meftah Elsaraiti, Gama Ali, Hmeda Musbah, Adel Merabet, and Timothy Little. "Time series analysis of electricity consumption forecasting using ARIMA model". In: *2021 IEEE Green Technologies Conference (GreenTech)*. IEEE. 2021, pp. 259–262.

[163] Mehmet Bilgili, Niyazi Arslan, ALİİHSAN ŞEKERTEKİN, and ABDULKADİR YAŞAR. "Application of long short-term memory (LSTM) neural network based on deeplearning for electricity energy consumption forecasting". In: *Turkish Journal of Electrical Engineering and Computer Sciences* 30.1 (2022), pp. 140–157.

[164] Palanki Amitasree, Guntha Raghu Vamshi, VS Kirthika Devi, et al. "Electricity consumption forecasting using machine learning". In: *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE. 2021, pp. 1–8.

[165] Arpad Gellert, Ugo Fiore, Adrian Florea, Radu Chis, and Francesco Palmieri. "Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods". In: *Sustainable Cities and Society* 76 (2022), p. 103426.

[166] Abnash Bassi, Anika Shenoy, Arjun Sharma, Hanna Sigurdson, Connor Glossop, and Jonathan H. Chan. "Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models". In: New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450390125.

[167] Brandon K Vaughn. *Data analysis using regression and multilevel/hierarchical models*. 2008.

[168] Duc-Hoc Tran, Duc-Long Luong, and Jui-Sheng Chou. "Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings". In: *Energy* 191 (2020), p. 116552.

[169] Tingting Fang and Risto Lahdelma. "Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system". In: *Applied energy* 179 (2016), pp. 544–552.

[170] Guo-Feng Fan, Meng Yu, Song-Qiao Dong, Yi-Hsuan Yeh, and Wei-Chiang Hong. "Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling". In: *Utilities Policy* 73 (2021), p. 101294.

[171] Taleb Zarei and Reza Behyad. "Predicting the water production of a solar seawater greenhouse desalination unit using multi-layer perceptron model". In: *Solar Energy* 177 (2019), pp. 595–603.

[172] Xuanyi Song, Yuetian Liu, Liang Xue, Jun Wang, Jingzhe Zhang, Junqiang Wang, Long Jiang, and Ziyan Cheng. "Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model". In: *Journal of Petroleum Science and Engineering* 186 (2020), p. 106682.

[173] Palanisamy Manigandan, MD Shabbir Alam, Majed Alharthi, Uzma Khan, Kuppusamy Alagirisamy, Duraisamy Pachiyappan, and Abdul Rehman. "Forecasting natural gas production and consumption in United States-evidence from SARIMA and SARIMAX models". In: *Energies* 14.19 (2021), p. 6021.

[174] Jui-Sheng Chou and Duc-Son Tran. "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders". In: *Energy* 165 (2018), pp. 709–726.