

Machine Learning on the Edge for Sustainable IoT Networks: A Systematic Literature Review^{*}

Luisa Schuhmacher^{ID,a}, Jimmy Fernandez Landivar^{ID,a}, Ihsane Gryech^{ID,a}, Hazem Sallouha^{ID,a}, Michele Rossi^{ID,b}, Sofie Pollin^{ID,a,c}

^a *WaveCoRE, Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10 postbus 2440, Leuven, 3001, Belgium*

^b *Department of Information Engineering, University of Padova, Via Gradenigo 6/b, Padua, 35131, Italy*

^c *IMEC, Kapeldreef 75, Leuven, 3001, Belgium*

Abstract

The Internet of Things (IoT) has become integral to modern technology, enhancing daily life and industrial processes through seamless connectivity. However, the rapid expansion of IoT systems presents significant sustainability challenges, such as high energy consumption and inefficient resource management. Addressing these issues is critical for the long-term viability of IoT networks. Machine learning (ML), with its proven success across various domains, offers promising solutions for optimizing IoT operations. ML algorithms can learn directly from raw data, uncovering hidden patterns and optimizing processes in dynamic environments. Executing ML at the edge of IoT networks can further enhance sustainability by reducing bandwidth usage, enabling real-time decision-making, and improving data privacy. Additionally, testing ML models on actual hardware is essential to ensure satisfactory performance under real-world conditions, as it captures the complexities and constraints of real-world IoT deployments. Combining ML at the edge and actual hardware testing, therefore, increases the reliability of ML models to effectively improve the sustainability of IoT systems. The present systematic literature review explores how ML can be utilized to enhance the sustainability of IoT networks, examining current methodologies, benefits, challenges, and future opportunities. Through our analysis,

^{*}The present work has received funding from the European Union's Horizon 2020 Marie Skłodowska Curie Innovative Training Network Greenedge (GA. No. 953775).

we aim to provide insights that will drive future innovations in making IoT networks more sustainable.

Keywords: Internet of Things, Machine Learning, Sustainability, Energy Efficiency, Edge Computing, Hardware, Testbeds, Systematic Literature Review

1. Introduction

The Internet of Things (IoT) has become a cornerstone of modern technology, permeating various aspects of daily life and industrial processes. By connecting a myriad of devices, from household appliances to industrial machinery, IoT networks facilitate real-time data exchange and automation, leading to increased efficiency and convenience [1]. However, the exponential growth of IoT systems presents significant sustainability challenges. The vast number of connected IoT devices, in sum, contributes to substantial energy consumption, which in turn affects carbon emissions and environmental sustainability [2]. Furthermore, managing resources such as bandwidth and computational power in IoT networks often leads to inefficiencies and increased operational costs [3]. Addressing these sustainability and operational cost issues is paramount to ensuring the environmental and economic viability of IoT networks in the long term. To this end, optimization on the software side regarding the allocation of computational resources and IoT network operation becomes a crucial area to investigate.

Machine Learning (ML), a subset of Artificial Intelligence, has demonstrated remarkable success across various domains, including healthcare, finance, and transportation [4]. By utilizing sophisticated algorithms and vast amounts of data, ML can uncover patterns, make predictions, and optimize processes in ways that were previously unimaginable. In healthcare, for instance, ML algorithms have improved diagnostic accuracy and personalized treatment plans [5]. In finance, ML models enhance fraud detection and risk management [6]. Similarly, in transportation, ML has enabled the development of autonomous vehicles and optimized logistics [7, 8]. The proven capabilities of ML in these fields suggest its potential to address complex challenges, such as those found in IoT networks.

Compared to traditional statistical methods, ML offers several distinct advantages that are particularly beneficial for IoT networks. While statistical methods rely on predefined models and assumptions about data distribution, ML algorithms can learn directly from raw data. This makes them

more flexible and adaptive to complex and dynamic environments [9]. This adaptability is crucial for IoT systems, which generate diverse and often intermittent and non-stationary data. Additionally, ML techniques can uncover hidden patterns and relationships within the data that statistical methods might miss. More innovative solutions for optimizing IoT operations can be achieved through this [10]. However, ML usually requires a lot of data which implies collecting and transferring vast amounts of data over the network [11].

One promising development, therefore, is the execution of ML on the edge of IoT networks. Edge computing involves processing data near the source of data generation rather than in centralized cloud servers [12]. This approach offers several benefits for sustainability [13]. Firstly, edge computing significantly lowers bandwidth usage and associated energy consumption by reducing the need to transmit large data volumes to the cloud. Secondly, edge computing enhances the responsiveness and efficiency of IoT systems by enabling real-time data processing and decision-making. This is particularly important for applications requiring immediate actions, such as predictive maintenance and energy management. Moreover, processing data locally at the edge can enhance data privacy and security, further bolstering the overall robustness of IoT networks.

The importance of testing machine learning models on hardware at the edge cannot be stressed enough. While useful for initial model development, simulated environments often fail to capture the complexities and constraints of real-world IoT deployments [14]. Actual hardware testing ensures that ML models perform satisfactorily under the specific conditions they will encounter, such as limited computational resources, power constraints, and varying environmental conditions. It also allows for the fine-tuning of models to achieve the best trade-off between accuracy and resource usage. Moreover, testing on actual hardware helps identify potential issues related to hardware-software integration, which is crucial for the seamless operation of IoT networks. This hands-on approach is also essential for validating the effectiveness of ML models in improving the sustainability of IoT systems.

Linking ML with IoT networks presents a promising avenue for enhancing sustainability. Specifically, ML can be employed to optimize energy usage by predicting and adjusting the power consumption of IoT devices, thereby reducing overall energy demand [15, 16]. Resource allocation can be improved through ML algorithms that dynamically manage network resources based on real-time data, minimizing waste and maximizing efficiency [17]. Addi-

tionally, ML can aid in predictive maintenance, identifying potential failures in IoT systems before they occur, thus reducing downtime and the need for resource-intensive repairs [18]. By integrating ML into IoT networks, especially through edge computing and rigorous testing on actual hardware, we can not only enhance their operational efficiency but also significantly mitigate their environmental impact.

In this context, a few related surveys exist. Some provide a broader perspective on deploying machine learning in resource-constrained IoT environments through edge computing to mitigate network congestion, latency, and privacy concerns [19, 20]. However, these are conventional surveys rather than systematic literature reviews (SLRs). Others are more specialized and closely aligned with our work, even adopting a systematic approach, such as [21]. In this work, the authors present a mapping of Artificial Intelligence (AI)-based solutions to achieve energy sustainability together with better Quality of Service (QoS) in the different layers of IoT networks. Although this paper is significantly related to our field of interest, it distinguishes itself in three major ways from our systematic review: Firstly, next to energy efficiency, selected papers further needed to improve the QoS. Secondly, the authors considered AI broadly, which includes ML but also other AI methods such as Swarm Intelligence. Lastly, they focused on optimizing IoT networks and on the methodology to pursue it while we further investigate actual hardware implementations and the specifications of the therefore established IoT networks. Trends and methodologies for energy management in IoT networks were investigated in [22]. Though this is related, the authors focus on energy management, while we broaden it to sustainability in general. Moreover, they did not specifically look into ML methods, which is a key research focus of this paper. Further, a comparative study of existing algorithms to enhance the energy efficiency of IoT networks was conducted in [23]. Unlike our systematic literature review, they also include non-ML methods and investigate simulators instead of actual hardware implementations.

While the related surveys and reviews mentioned provide relevant information, they do not specifically address the problem of evaluating simulation-based conclusions on for instance energy consumption with hardware-based results. Simulation models can be very intricate, some modelling the real world in much more detail than others; however, most of them are based on simplifying assumptions. In general, it is impossible to define a complete model. To advocate performance verification on actual hardware and

show different ways of doing it, we present a systematically collected range of research evaluated on actual hardware. This research can start small by presenting a prototype, contain a network implementation in a testbed environment, or utilize a full-fledged implementation in an IoT network. To be specific, this systematic literature review is centred on the intersection of machine learning and IoT sustainability evaluated on real hardware on the edge. Thus, we offer a comprehensive analysis of current research, methodologies, and applications. We will explore how ML algorithms are being utilized to create more sustainable IoT networks, identify the benefits and challenges associated with these approaches, and highlight emerging trends and future opportunities. Through this analysis, we seek to provide valuable insights and a foundational understanding that will drive future innovations in making IoT networks more sustainable. Specifically, the main contributions of this paper are as follows:

- Conduction of a systematic review using the PRISMA methodology to ensure a structured, transparent, and replicable process.
- Adherence to a predetermined protocol to maintain objectivity and impartiality throughout the review process.
- Extraction and summary of papers from the literature that proposes ML models to improve the sustainability of IoT networks and validate their proposed methods by actual hardware implementation.
- Elucidation of different aspects of the network to improve sustainability in, using ML.
- Specification of the IoT networks regarding the hardware used and application implemented to validate sustainability improvements.
- Investigation of which ML methods have been used and compared with each other, and which showed the best performance for a given setup.
- List metrics used to evaluate the improvement of the network's sustainability.
- Identification of gaps in the literature and suggestions for future work to advance the sustainability of IoT networks.

The remainder of the paper is organized as follows. Section 2 covers the materials and methods, including the research questions, data sources, and inclusion criteria. Section 3 provides a summary of the review findings and a detailed comparison of the methods used in the surveyed literature. A discussion about the identified gaps, directions for future research, and practical implications is provided in Section 4. Finally, Section 5 presents the conclusions.

2. Materials and methods

This work presents a Systematic Literature Review (SLR), employing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology (Fig. 1) [24], and thus ensuring a structured, transparent, and replicable process. The study enhances rigour and reproducibility by analyzing indexed research from SCOPUS over the past decade (2013 – 2023), facilitating validation by other researchers. Explicit inclusion and exclusion criteria are established to minimize selection bias, guaranteeing comprehensive consideration of pertinent research. Adherence to a predetermined protocol maintains objectivity and impartiality throughout the review process, distinguishing this systematic review from traditional approaches.

The SLR procedure encompasses three primary phases: (i) Identification, (ii) Screening, and (iii) Eligibility. Upon formulating the research questions, the Identification phase establishes the search strategy, specifying the selection of data sources and extraction methods for gathering pertinent papers. In the screening and eligibility phases, inclusion and exclusion criteria, aligned with the review's specific requirements and scope, are delineated. During these phases, papers are filtered based on titles and abstracts (screening) and full-text (eligibility). Fig. 2 provides a comprehensive overview of the outcomes at each stage of the PRISMA process. Subsequently, responses to the research questions are synthesized, and the challenges, opportunities, and limitations are underscored. These processes are elaborated on in the following sections.

2.1. Research Questions

ML plays a crucial role in enhancing energy efficiency within the IoT framework. The deployment of ML algorithms at the edge of IoT networks can bring forth numerous energy-saving strategies [13, 27, 28]. ML can empower real-time decision-making at the edge and diminish the constant ne-

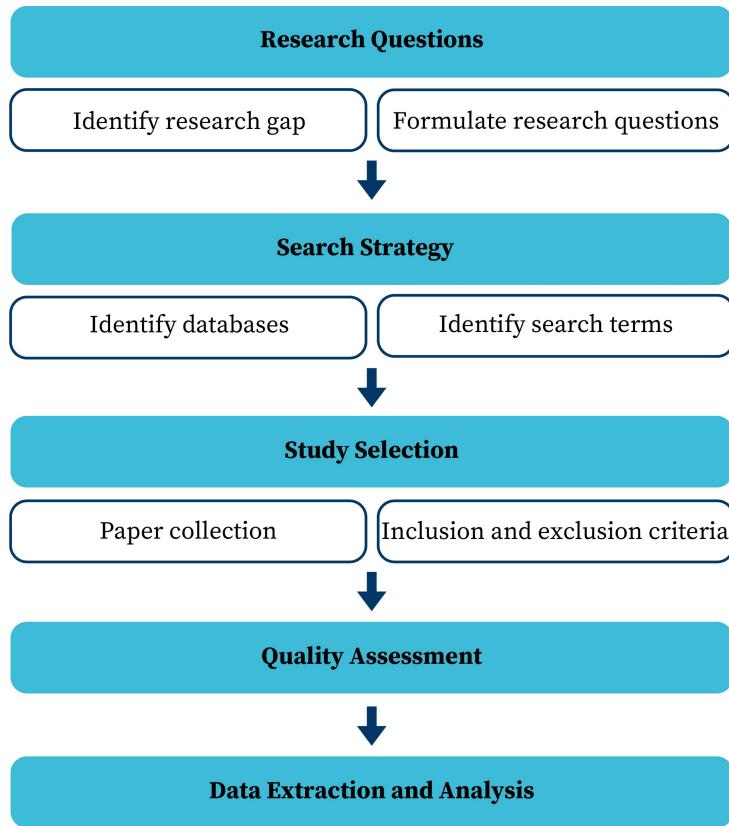


Figure 1: PRISMA Methodology followed in the systematic review process [25, 26]

cessity for data transmission to centralized servers. This localized processing capacity allows for direct filtering, compression, and analysis of data on IoT devices. It thereby reduces the quantity of information transmitted over the network and subsequently curbs energy consumption [29, 30]. ML's contribution could extend to predictive maintenance, enabling early detection of equipment failures or anomalies, thus averting unnecessary energy usage and downtime. Additionally, adaptive power management and personalized energy profiles, guided by ML models, can guarantee that IoT devices operate in the most energy-efficient states based on usage patterns and individual preferences. In conclusion, integrating ML at the edge of IoT networks has the potential to be a vital tool in advancing sustainable practices. This integration can optimize resource allocation, reduce data transfer, and promote intelligent, energy-conscious decision-making at the edge. However, there is

no strict path to enhancing sustainability, leaving the option of green-washing research. Due to the hype about sustainability, it occurs that authors state that their conducted research improves sustainability while no credible metric is being used for assessment, or not all components of the method are being examined [31].

To comprehend the current landscape in this context and to avoid including potentially green-washed research, our systematic review aims to identify critical research questions and seek relevant answers via a thorough investigation. The three research questions formulated for this systematic review are:

- **RQ1:** How can ML contribute to enhancing energy efficiency and sustainability at the edge of IoT networks?
- **RQ2:** What IoT network specifications were used to test the proposed ML methods?
- **RQ3:** Which ML tools are employed at the edge to enhance the energy efficiency and sustainability of IoT networks?

These inquiries have been formulated to fulfill the primary objective of this paper: conducting a systematic literature review on the application of machine learning methods in enhancing energy efficiency and sustainability within the IoT. Additionally, this paper seeks to offer a comprehensive synthesis of the ML-driven contributions to the sustainability of IoT devices, including the tools and models employed for this purpose. Furthermore, the paper provides insights into the current state of research in the field, highlighting key challenges and potential practical implications.

2.2. Search Strategy

To investigate the research questions, we leveraged the SCOPUS dataset [32], as it is recognized as the world's largest abstracting and indexing database, known for its continuous daily updates [33]. The search for relevant publications was initiated on August 8th, 2023.

Keywords and queries were formulated according to the requirements of the Scopus scientific database using the research questions. The following research query was employed: **ALL ("IOT") OR ALL ("Internet of Things") AND ALL ("energy efficiency") OR ALL ("sustainability") AND ALL ("machine learning") OR ALL ("deep learning"**

) AND ALL ("edge computing") AND PUBYEAR >2012 AND PUBYEAR <2024 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOC-TYPE , "cp")).

The search time frame covered a decade of publications, spanning January 2013 to August 2023. The initial search found 5,827 documents. These documents were added to Rayyan, a smart research collaboration platform designed to streamline the process of literature reviews and systematic reviews [34]. On this platform, 40 duplicates were identified, and only 7 were verified as true duplicates before being removed. The remaining 5820 studies were further processed according to the inclusion and exclusion criteria (Fig. 2).

2.3. Study Selection and Data Extraction

The collected papers from the initial search were screened according to the preset inclusion and exclusion criteria (Table 1). The paper selection process consisted of two phases. First, based on the inclusion and exclusion criteria, the papers were independently screened by two researchers through title and abstract screening (L.S & I.G). The papers selected in this phase were then assessed through full-text screening (L.S. & J.F.). The authors cross-checked the selection results and resolved any disagreement on the selection decisions. All disagreements in either the first or the second phase were resolved by consensus, and a third researcher (S.P.) was consulted to finalize the decision. The process is depicted in Fig. 2.

The inclusion and exclusion criteria were formulated by the authors to select relevant papers effectively. The documents underwent analysis to investigate diverse ML approaches related to the energy efficiency and sustainability of IoT networks.

Both journal and conference papers written in English and within the scope of the research questions were included. Commercial papers, letters to the editor, ebooks, books, posters, and PhD dissertations were excluded. Papers were further excluded if

- they were review papers,
- they solely focused on the consideration of energy efficiency within ML models,

Inclusion Criteria	Exclusion Criteria
- Within scope of research questions.	- Does not include edge computing. - Proposed method does not positively impact energy efficiency or sustainability of IoT networks, or is not discussed. - Only ML-specific hardware optimization. - Makes application on the edge more energy-efficient, but not the network itself.
- Application of ML on the edge.	- No direct application on the edge, solely research on energy-efficient ML methods for edge computing (e.g., optimized federated learning algorithm).
- Actively improves energy efficiency or sustainability.	- Only meets resource constraints (e.g., computational constraints by specific hardware, or financial constraints).
- Explicit usage of an IoT network.	- Usage of technologies that can be used for IoT networks but not embedded in an IoT network (e.g., UAVs, 5G, Mobile Edge Computing...).
- Evaluation of ML on an edge device.	- Only simulation-based evaluation, no real hardware implementation.
- Within the time frame of publication: August 2013 until July 2023.	- Outside the time frame of publication: August 2013 until July 2023.
- Is a journal or conference paper.	- Review, book chapter, book, short survey, editorial, retracted, conference review, data paper, note.
- Written in English.	- Other languages.

Table 1: Inclusion and Exclusion Criteria for Paper Selection

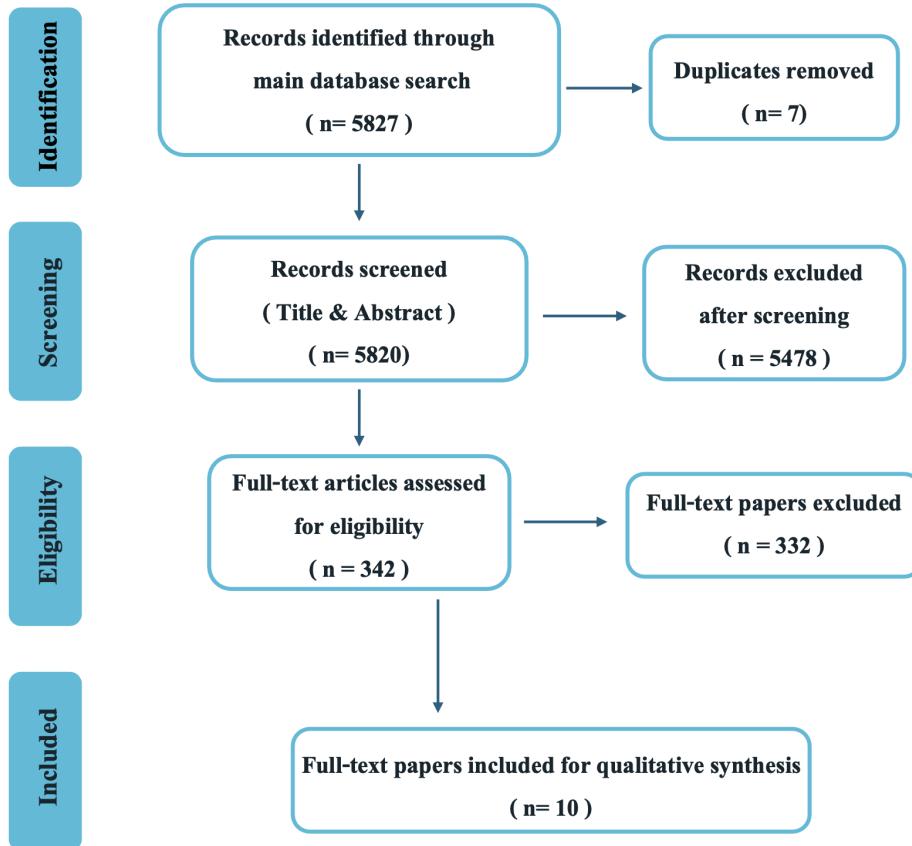


Figure 2: Flow Diagram for the selection of the literature reviewed.

- the energy efficiency was present in the hardware but not integrated into an IoT network,
- the technologies used were applicable to IoT but not integrated into an IoT network (e.g., UAVs, 5G, 6G, Mobile Edge Computing, etc.), or
- performance was not evaluated on hardware.

The authors deemed that the aim of the review was the integration of edge ML, IoT networks and energy efficiency with the goal of enhancing sustainability.

Relevant data was extracted from the selected publications for further analysis:

- Author list,
- Titles and abstracts,
- Year of publication,
- Associated database,
- IoT network application, IoT aspect optimized,
- ML model(s) and metric used, sustainability gain achieved,
- IoT network specifications: communication technology, number and type of sensors used, hardware deployed

2.4. Risk of Bias and Quality Assessment

This systematic literature review adhered to the PRISMA guidelines for screening and selecting relevant literature. However, certain limitations must be acknowledged, primarily related to potential biases. The choice of keywords for the initial query search may introduce bias, and the subjectivity in defining eligibility criteria poses an additional risk. The reliance on a single database, Scopus, may limit the comprehensiveness of the literature search. Despite these considerations, the authors followed the best possible criteria in line with PRISMA guidelines. Independent selection processes and disagreement resolution techniques were employed to enhance transparency and objectivity. Additionally, a quality assessment system, established through consensus among authors, ensured the inclusion of high-quality publications with substantial contributions. This system was based on a checklist of the following criteria:

- Are the methods used clearly defined and applied?
- Are the methods applied successfully and correctly?
- Are accuracy values and efficiency/confidence levels reported?
- Do the contributions outweigh the limitations of the study?

2.5. Characteristics of selected papers

The search process (Fig. 2) yielded a total of 5827 articles. After removing duplicates, 5820 papers remained. Of these, 5478 studies were excluded during the title and abstract screening for not meeting the inclusion criteria. Among the 342 studies that underwent full-text screening, 332 were found to not meet the full inclusion criteria and, therefore, further excluded. Ultimately, 10 studies were selected for the current review, as summarized in the following sections. The selection process took eleven months to complete.

The final selection for the systematic literature review comprised 9 journal papers [35, 36, 37, 38, 39, 40, 41, 42, 43] and one conference paper [44]. There was one paper each from China [43], Italy [42], the USA [35], France [37], Portugal [39], Canada [41], Australia [40], and India [38], and 2 papers from South Korea [36, 44]. These publications span the period from 2018 to 2023, with one publication in 2018 [44] and 2021 [39], two in 2019 [37, 42] and three each in 2022 [36, 38, 40] and 2023 [35, 41, 43].

3. Findings

To answer our research questions, we look into how the literature approached making IoT networks more sustainable. The layout of this section is shown in Fig. 3. We first elucidate different aspects of an IoT network upon which sustainability can be improved. We differentiate between communication- and computation-focused approaches. Afterward, as we are interested in actual hardware implementations, we delve into the different IoT networks implemented in the selected papers, their specifications, and which kind of IoT application they execute. We finish by examining the different ML methods used as well as metrics leveraged to measure the sustainability of an IoT network.

3.1. Communication- vs. computation-focused energy efficiency

To reduce the energy consumption of IoT networks, one can focus on the communication side of the network or on the computation level of each node in the network, respectively. Table 2 gives an overview of the network aspects in which sustainability was improved in the included papers, if its focus is on the communication or computation part, and how many papers looked into that specific aspect. It can be seen that the range of network design parameters that were researched is broad, tackling computation and communication subsystems.

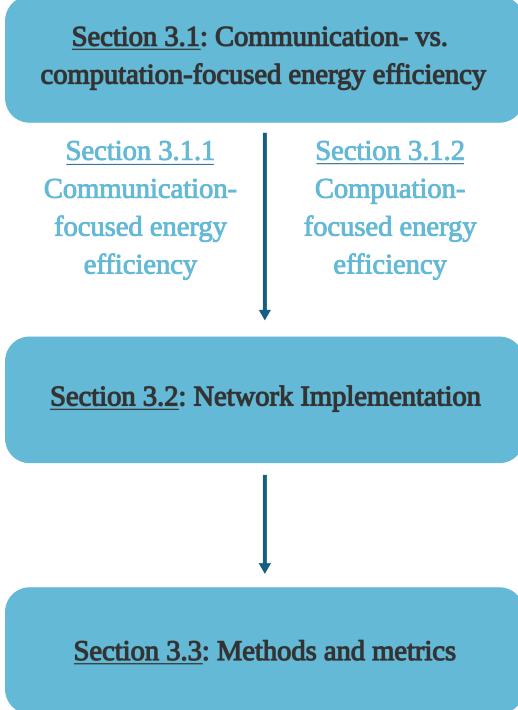


Figure 3: Layout of the findings section.

Network aspect	Focus	Nr. of papers	Ref.
Transmission period control	Communication	1	[36]
Data compression	Communication	2	[37, 38]
Sleep Scheduling	Communication	1	[40]
Communication protocol	Communication	2	[39, 42]
Transmission selection	Communication	1	[44]
Dynamic Voltage and Frequency Scaling	Computation	1	[41]
Computation offloading	Computation	2	[43, 44]
Co-scheduling of computational resources	Computation	1	[35]

Table 2: Aspect of the network tackled to enhance sustainability, if this aspect focuses on the communication part of the network or on computation and how many papers dealt with this aspect.

3.1.1. Communication-focused energy efficiency

We can further differentiate between network operation optimization and data transmission reduction when delving into communication-focused approaches. In the following, we start with covering papers that work on optimizing the network operation and then move on to those dealing with reducing the amount of data transmitted.

Network operation optimization. We find three aspects covered in the included papers, aiming to optimize the network operation: dynamically adjusting the data transmission period, tuning the duty cycle for sleep scheduling, and selecting the communication protocol.

The authors in [36] suggested increasing the sleep duration between each data transmission to transmit less often and, therefore, decreasing the energy consumption. This, however, entails that some data does not get transmitted at all. They dealt with this problem by introducing data imputation at the server. For this, they developed a Deep Neural Network (DNN) consisting of a Bidirectional Long Short-Term Memory (Bi-LSTM), a Convolutional Neural Network (CNN) and a simple one-layer Artificial Neural Network (ANN). This model takes the transmitted data and the data transmission period as input and, based on this, predicts the imputation accuracy. Using this result, they formulated an optimization problem to minimize the energy consumption of the IoT sensors and maximize the imputation accuracy. With their method, they achieved an energy consumption reduction of 18.23% on the device.

The work in [40] is dedicated to improving sleep scheduling. In particular, a range of ML methods were used to predict downlink packets so that the nodes only wake up when they expect data arrival. In this paper, different regression methods were compared, including Random Forest (RF), Gradient Boosting (GB), Extra Trees (ET), Histogram-based Gradient Boosting (HB), Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM). Finally, the LSTM model was implemented since it yielded the highest prediction accuracy. The authors of this work also compared their method to two power-saving modes, in two different scenarios: one using edge computing and one taking the computation of their application to the cloud. Their results are stated in Table 3.

Another field of interest to improve the sustainability of a network's communication is the choice of the communication protocol. Two papers proposed a solution in this regard. The work in [39] suggested self-configurable

Power saving mode	Scenario	Energy efficiency improvement
PSM	Edge computing	37%
	Cloud computing	46%
APSM	Edge computing	6%
	Cloud computing	26%

Table 3: Energy efficiency improvement results from [40] for the normal power save mode (PSM) and the adaptive power save mode (APSM), executed in an edge computing scenario and in a cloud computing scenario.

IoT nodes that choose the communication protocol and transmission power based on energy consumption and signal quality prediction. The protocols to choose from were ESP-Now [45], Bluetooth Low Energy (BLE) [46], LoRa [47], and ZigBee [48]. To determine the best prediction method, the work compared Linear Regression (LR), Decision Tree (DT), RF, MLP and Support Vector Machine (SVM). As the RF yielded the best prediction results, this was the model deployed in the network. Their method achieved up to 68% energy consumption reduction with only 7% of loss in network quality. While this approach used existing protocols and switched between them efficiently, [42] introduced a by-design energy-efficient MAC protocol. They used Reinforcement Learning (RL) to adjust the duty cycle of the node. In this way, they prolonged the node’s lifetime by 4.5 times compared to the conventional CSMA-CA MAC with a duty cycle at 60% and even by 26 times compared to the same conventional protocol with a duty cycle at 100%.

Data transmission reduction. Two different approaches can be identified to reduce data transmissions. One approach is compressing the data before sending it to reduce the amount of transmitted bits and, therefore, save energy. A second approach involves selecting which data to transmit instead of blindly transmitting everything.

For the first approach, the authors of [38] used an LSTM-based Autoencoder (AE) to learn to compress and decompress the data to be sent. They leveraged optimization methods to find the optimal compression size, and in addition, they formulated another optimization problem to find the optimal data transmission speed. This problem was then solved with a conventional optimization method. They compared their method against existing approaches and demonstrated that they achieved lower energy consumption;

however, no exact numbers were provided. The authors of [37] used traditional ML methods, namely Linear-Curve Fitting (LCF) and Quadratic-Curve Fitting (QCF), as part of the compression algorithm. Precisely, they only compressed unpredictable data points. Those data points predicted by the preceding ML methods are discarded and predicted using the same method by the receiver. Using this method, they achieved an increase of 27% of the device’s lifetime after four hours.

The second approach is pursued in [44]. The authors proposed to use a smart camera to train a DNN to determine if a captured image is of interest or not, and to only transmit the image if it is classified as being of interest. This method decreased the energy consumption by 41%. The specific DNN architecture is, however, not specified in the paper.

3.1.2. Computation-focused energy efficiency

Three aspects can be identified to enhance energy efficiency from the computational side. The work in [41] proposed the use of Deep Reinforcement Learning (DRL) to learn the optimal dynamic voltage and frequency scaling for local computation as well as the optimal data distribution for data offloading. It compared energy savings on three different IoT devices to two different Linux governors, which determines the Central Processing Unit (CPU) utilization. The results are given in Table 4.

IoT device	Governor	Energy saving
Linux laptop	Ondemand	3.25% - 5.78%
	Conservative	9.9% - 10.35%
Nano 2 GB	Ondemand	4.07% - 7.72%
	Conservative	7.13% - 9.89%

Table 4: Energy saving results from [41] for different IoT devices and CPU utilization governors. They also tested on a Raspberry Pi as an IoT device but stated that the energy savings there were negligible.

While [44] offloaded data to reduce the computational load on the edge device, the research in [43] offloaded parts of DNNs to divide computation between a local device and an edge server. It trained an RL model and a DRL model to find the optimal DNN model partition point. The authors stated that with the RL model, the energy consumption of the local device was reduced by 13.9% whereas, with the DRL model, this result was improved to

a reduction of 41.8%. However, the work mentioned that the edge server’s energy consumption increased but did not elaborate further.

The offloading methods dealt with executing certain computations on a mobile device or on a server. Another way of optimizing computational load is to co-schedule computational resources, i.e., efficiently use computational resources on heterogeneous hardware. To achieve this, [35] leveraged DRL to co-schedule the use of CPU and Graphics Processing Unit (GPU). They trained two models: one basic model to jointly optimize latency and throughput and one energy-aware model that optimizes these two factors and energy consumption. They compared their models to CPU-only usage, GPU-only usage, and a round-robin way of distributing computational load between CPU and GPU. Compared to CPU-only, their basic model improved average energy consumption per task by 42.24%, whereas their energy-aware model further increased this number by 4.09%. They did not state numbers for the comparisons against GPU-only and round-robin. However, from their results, it can be seen that both the energy-aware model and the basic model provided higher energy efficiency than the other baselines.

3.2. Network implementation

So far, we discussed how the literature has improved the sustainability of IoT networks. In this subsection, we will elaborate on the IoT network implementation in terms of scope, including network specifications and their IoT application. Fig. 4 visualizes the different IoT applications. We find that three papers implemented full networks consisting of more than one point-to-point connection, while seven papers provided prototypes with no more than two connected devices to support their findings. We will now elaborate on the number of nodes, type of sensors, communication technology and hardware used, as well as the use the network served for. An overview of the communication technologies used is shown in Fig. 5.

3.2.1. Full network implementation

Only a few works provided the implementation of a full network containing multiple nodes, summarized in Table 5. The biggest scope in the considered works was the network in [39], which consisted of 35 nodes and one gateway in an urban environment, spanning an area of 36ha. This network transmitted dummy data, generic to many applications for IoT networks. In this work, the authors switched between communication technologies, including four in total, namely ESP-Now [45], BLE [46], LoRa [47], and ZigBee [48].

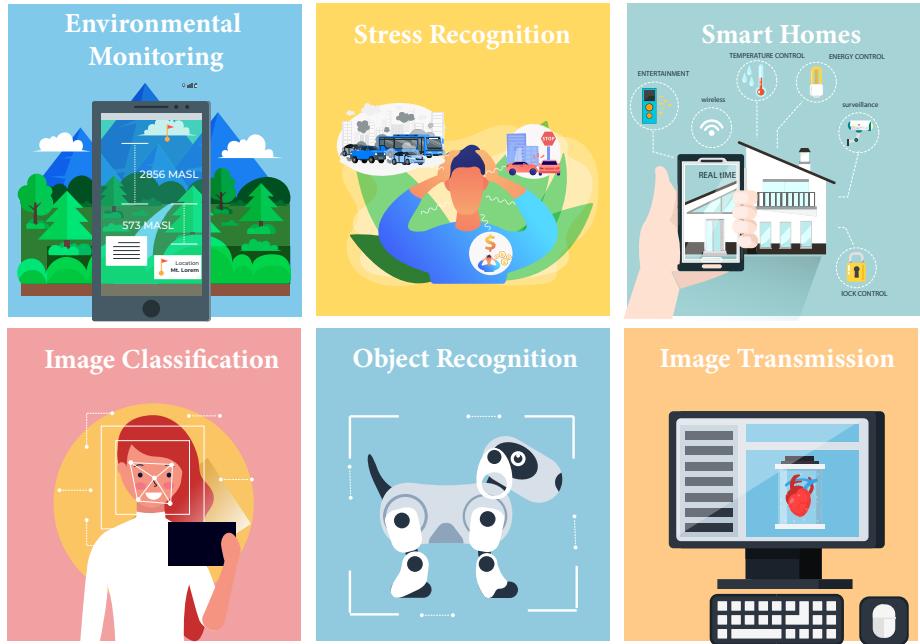


Figure 4: IoT applications in the included papers.

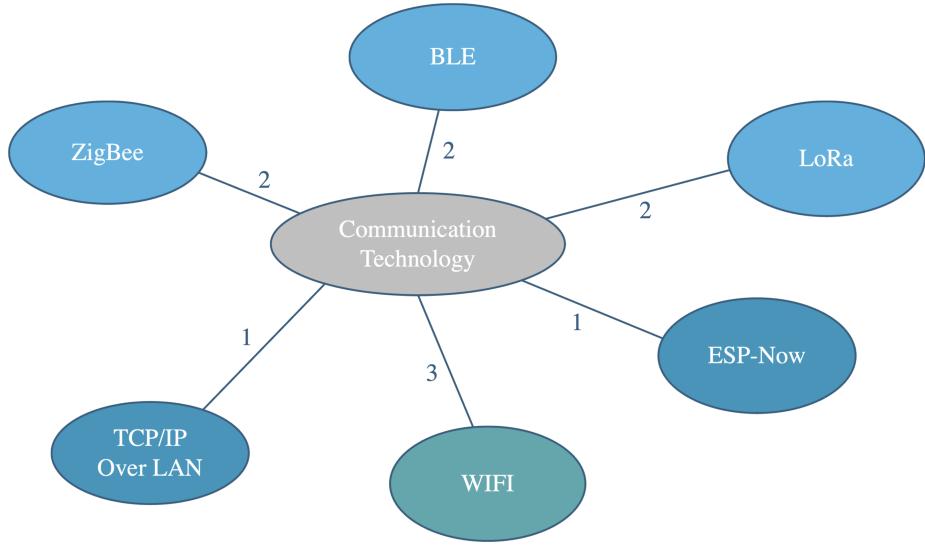


Figure 5: Different communication technologies and their usage count in the selected papers.

Application	Nr. nodes	Comm. techn.	Hardware	Ref.
Dummy data transmission	35	ESP-Now [45], BLE [46], LoRa [47], ZigBee [48]	ESP32-WROOM-32 [49]	[39]
Dummy data transmission	7	ZigBee [48]	TelosB [50]	[42]
Dummy data transmission	4	WiFi	Cypress CYW43907 [51], Raspberry Pi 4 [52]	[40]

Table 5: Implemented full IoT networks in the included papers. No actual sensed data has been used, and therefore no sensor deployed.

For their nodes, they used Espressif ESP31-WROOM-32 nodes [49]. The second largest IoT network in the included work, implemented in [42], also used ZigBee for communication. Seven TelosB nodes [50] were implemented in their work, which similarly transmitted dummy data. The network in [40] featured four Cypress CYW43907 nodes [51] and four Raspberry Pi edge devices. It was set in a residential area and transmitted, like the other networks, dummy data, but in their network, WiFi was being used for communication.

3.2.2. Prototype implementation

Most of the considered papers in the literature tested their methods on prototypes, with a maximum of two devices communicating with each other. We give an overview in Table 6. Using WiFi, [43] used two Raspberry Pi 4's [52] as mobile devices and two edge servers to test their model for image classification. An existing dataset was used instead of images captured by the mobile devices. The authors in [37] tackled a very specific IoT task. They provided a prototype of a wearable to classify the stress rate of car drivers. More specifically, they used the Polar M600 wearable [58] and transmitted via BLE [46]. They first recorded signals from a driver, collected a data set like this, and then deployed it on the wearable to test their method. The results in [41] verified the efficacy of their method for the IoT application of image classification. In [41], the authors used the well-known MNIST dataset [65] as dummy data for that and deployed it on a set of three different IoT devices. Those devices started small with a Raspberry Pi 4 [52], getting more powerful over a Jetson Nano 2 GB [61] to a Linux laptop with Intel

Application	Nr. nodes	Type of sensors	Comm. techn.	Hardware	Ref.
Energy consumption data transmission in smart home	6	Smart meter (energy consumption)	LoRa [47]	DDS238-4W [53], RFM95W-868S2 [54], Raspberry Pi 3 [55], Drigano LG01-SIOT [56]	[38]
Image classification	2	-	WiFi	Raspberry Pi 4 [52]	[43]
Environmental sensor data transmission	1	Temperature, humidity, dust	WiFi	Arduino Wemos D1 mini [57], Raspberry Pi 4 [52]	[36]
Stress rate classification in car drivers	1	-	BLE [46]	Polar M600 wearable [58]	[37]
Image transmission	1	Image sensor	BLE [46]	OV7725 [59], Nordic nRF528 40 [60], Raspberry Pi 3 [55]	[44]
Image classification	1	-	TCP/IP over LAN	Raspberry Pi 4 [52], Jetson Nano 2 GB [61], Linux laptop with Intel CPU [62], Lenovo Legion 5i laptop [63]	[41]
Facial expression classification and facial landmarks detection	1	-	-	Google Pixel 2 XL [64]	[35]

Table 6: Implemented prototypes in the included papers. If no sensor is stated, no actual sensed data has been used. Further, for the work of [35], the communication technology is unimportant, so none is given.

CPU [62]. All devices sent their data to a Lenovo Legion 5i laptop [63], which served as an edge server. They only specified that they use TCP/IP over LAN for communication technology. The work in [35] did not entail communication. Their method was solely based on computation on the local device. Therefore, sending data for testing was not needed. As IoT device, they used the Google Pixel 2 XL [64], and the used application was facial expression classification and facial landmarks detection.

In [35, 37, 39, 40, 41, 42, 43], no sensors were deployed, and hence, no actual sensor readings were being used. While some of the works provided prototypes, they did use sensors. The authors of [38] deployed DDS238-4W single-phase smart meters [53] to measure the energy consumption of up to 12 appliances each in six households. The smart meter transmitted its readings using LoRa [47] to the Raspberry Pi 3 edge device [55], which had an RFM95W-868S2 LoRa node [54] attached to it. This node then further transmitted to a Drigano LG01-SIOT LoRa Gateway [56]. These readings were conducted every 20 seconds for 240 hours. In [36], the Arduino Wemos D1 mini sensor [57] gathered information about temperature, humidity, and dust and transmitted it as an environmental monitoring application via WiFi to a Raspberry Pi 4B edge device [52]. The sensors were implemented in an office, and their experiment spanned 6 days. Sensor readings were performed at an interval of 10 seconds. It should be emphasized that both works tested their method on actual hardware in a real setting, coming very close to an actual application setting and, hence, trustworthy performance results. Last but not least, [44] used an OV7725 image sensor [59] to capture images. Connected to a Nordic nRF52840 node [60], these images were being transmitted through BLE [46] to a Raspberry Pi 3 gateway [55]. Information about the prototype’s location, the experiment’s time span, or amount of pictures taken were not given in the paper.

3.3. Methods and metrics

Finally, we examine the ML methods and metrics that are being used in the included papers. For the ML methods part, given the recent trend in using Deep Learning (DL), we specifically focus on the comparison between traditional ML methods’ and DL methods’ usage. From Table 7, we can see that a variety of traditional ML methods have been implemented, while only a few different DL methods are considered. However, each DL method, except the CNN, has been implemented in three papers, whereas the traditional methods are mainly used in one paper each. Only RF and RL have been

ML method	Nr. papers	Ref.
DT	1	[39]
ET	1	[40]
GB	1	[40]
HB	1	[40]
LCF	1	[37]
LR	1	[39]
QCF	1	[40]
RF	2	[40, 39]
RL	2	[42, 43]
SVM	1	[39]
ANN	3	[36, 40, 39]
CNN	1	[36]
DRL	3	[41, 35, 43]
LSTM	3	[36, 38, 40]

Table 7: Overview of implemented ML methods in the included papers and how many papers used the corresponding method. The first part covers the traditional ML methods used, while the second part lists the DL methods used. In "ANN", we summarize classical MLPs and single-layer ANNs. Note that [44] only stated that they use a DNN but did not specify the architecture, so it is not listed here.

Evaluation metric	Nr. papers	References
Percentage energy consumption reduction	7	[36, 38, 41, 39, 44, 35, 43]
Percentage energy efficiency improvement	1	[38]
Percentage device's lifetime increase	2	[37, 42]

Table 8: Overview of metrics used to evaluate the sustainability enhancement of an IoT network.

used in two papers. Further, [40] implemented RF, GB, ET, HB, an MLP and an LSTM to compare their performance. The LSTM model is the only method implemented in real devices, which emerged as the best model in their comparison. Also, [39] implemented different methods to compare, being LR, DT, RF, an MLP, and a SVM. In their work, however, the traditional ML method RF exhibited the best performance and beat the MLP, a DL model. In [43], the authors compared traditional RL against DRL and found that with DRL, more energy could be saved. Considering this, the papers where the best-achieving model was a traditional ML model amount to three, while DL models were implemented in the other seven papers. Note, however, that only three papers compared different ML methods.

As for the evaluation metrics used, we can differentiate three metrics to indicate the sustainability of a network. Table 8 shows that seven out of ten papers evaluated their methods by how much the normal energy consumption is being reduced. In [38], the authors presented their method's performance by the percentage of energy efficiency improvement instead, being the only paper from the selected ones that uses this metric. In [37] and [42], the authors used the lifetime of the IoT devices as an indicator and stated how much it increased using their proposed method.

As most papers used the percentage in energy consumption reduction to evaluate the performance of their proposed models, we further detail their results. An individual comparison between the papers' implemented ML methods and the gain in energy consumption reduction thereby achieved is shown in Table 9. The highest energy consumption reduction can be found in [39] with going up to 68%. In their work, the authors leveraged an RF to adapt the communication protocol. This energy consumption reduction, however, comes at the price of a 7% loss in network quality. The next

Method	Gain	Reference
DRL	46.33%	[35]
Bi-LSTM + CNN + ANN	18.23%	[36]
LSTM-based AE	-	[38]
RF	Up to 68% (with 7% loss in network quality)	[39]
DRL	Up to 10.35%	[41]
DRL	41.8% locally, but increase in edge server energy consumption	[43]
DNN	41%	[44]

Table 9: Comparison of implemented ML techniques and their achieved energy consumption reduction. [38] did not state exact numbers in their paper. Note that the gains of [39] and [43] come at the loss of network quality and energy consumption increase at the server’s side, respectively.

highest gain can be found in [35], with an energy consumption reduction of 46.33%. Their approach encompasses a DRL model to efficiently co-schedule the usage of CPU and GPU. Followed closely regarding energy consumption reduction are the works in [43] and [44]. Both advocate computational offloading to decrease energy consumption. In [43], the authors train a DRL model to partition the used DNN model and offload the second partition to the edge server, while in [44], data is offloaded only if it is classified by a DNN to be of interest. It is further stated in [43] that this offloading increases the energy consumption of the edge server, a statement which also holds for [44]. Both do not further elaborate on this drawback. Less energy consumption reduction is achieved in [36, 38, 41]. Their approaches tackle transmission period control, data compression, and dynamic voltage and frequency scaling, respectively. These results indicate that optimizing the usage and smartly combining existing resources, such as switching between communication protocols or optimizing the usage of CPU and GPU, yield higher energy consumption reduction potential. Less energy consumption reduction seems achievable by reducing the amount of transmitted data or going to very specific hardware instances, like voltage and frequency scaling. Note, however, that there is a mixture of applications, hardware and communication technologies in the papers discussed, so no definitive statement can be made.

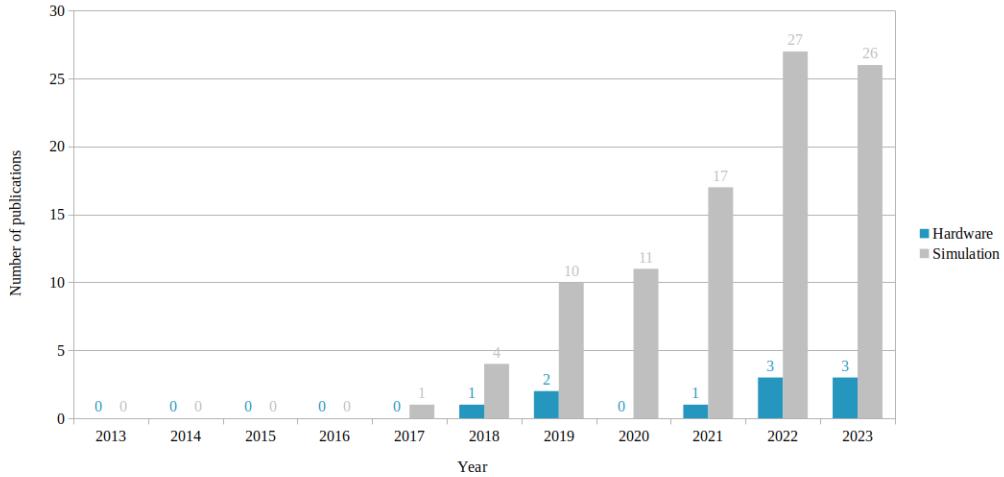


Figure 6: Distribution of selected papers over the included years: in cyan, final papers with hardware evaluation; in grey, papers that were evaluated with simulations but fulfilled all other inclusion criteria.

4. Discussion

In this systematic literature review, we have explored state-of-the-art works that used ML to enhance the sustainability of IoT networks. Of the 342 papers left after the abstract screening, 332 were excluded, and only 10 were included. Out of the 332 excluded papers, 96 papers fulfilled all criteria except actual hardware evaluation. This shows that almost tenfold of research evaluates their work on edge ML for sustainable IoT with simulations rather than hardware. From the publication years of the included papers, we noticed that all fall in the second half of the selected period. Further, if we omit the very restrictive inclusion criterion of hardware-based evaluation, this trend still remains. The distribution for both is shown in Fig. 6. This can be attributed to the fact that sustainability enhancement as a major goal in research, as well as a rapid increase in the development of DL methods, has only evolved in recent years. For both to further arrive at niche areas like IoT networks could explain the lack of work before 2017.

Among the included papers, a wide range of IoT applications (Fig. 4) and aspects were investigated and optimized. In [66], the authors proposed to maximize the harvested energy of each IoT device while satisfying its data rate requirements. They did so utilizing an online DRL algorithm, which they ran distributed on each device, only using local channel state informa-

tion. Going for the application of Industrial Internet of Things (IIoT), [67] designed a DRL model using dual-attention for resource allocation. Another specific IoT field is the Internet of Medical Things (IoMT). The authors in [68] employed in this setting a DRL model as well and trained it to allocate channels so that Age of Information (AoI) and energy consumption are being optimized. Broadening again to IoT in general, a different approach for reducing energy consumption has been conducted in [69]. They claimed that some of the collected data is untrustworthy due to harsh environments where sensor nodes are deployed and can be discarded before transmission. For that, they trained an SVM and further enhanced it with online learning, so that the model stays up-to-date upon employment. Focusing as well on the collected data, [70] suggests fusing data to reduce redundant information transmission. They, therefore, proposed an Extreme Learning Machine (ELM) and optimized it by a regular optimization algorithm.

While simulations are a good starting point and give valuable insights, validating performance on actual hardware is crucial, as it might vary greatly from simulation performance [71]. From the included papers, which also performed hardware testing, we have seen various aspects of an IoT network being taken and investigated to improve energy efficiency. We split those aspects into communication-focused and computation-focused ones. While focusing on the communication part of the network is intuitive, with the trend of putting more computation on the edge of IoT networks [72], improving computational efficiency is of utmost importance as well. For the former one, we identified that transmission period control [36], data compression [37, 38], sleep scheduling [40], the communication protocol itself [39, 42], and transmission selection [44] have been taken to improve sustainability. For the latter one, we saw that dynamic voltage and frequency scaling [41], computation offloading [43, 44], and co-scheduling of computational resources [35] have been deemed to have room for improvement via ML. We have seen that more research has been conducted to enhance sustainability from the communication perspective [36, 37, 38, 40, 39, 42, 44] rather than the computational perspective [41, 43, 44, 35]. Though this variety of network aspects showcases creativity in how to make IoT networks more sustainable, it also demonstrates that tackling a different area for improving sustainability is favored over improving existing areas. This is emphasized by the fact that baseline comparison happens mainly to the normal network operation [35, 37, 44, 40, 41, 42, 39]. Only [36, 38, 43] compare their methodology to existing works. Building on and outperforming existing work, like done so

often in the ML community as challenges with leaderboards and leading to rapid performance improvement [73], seems to not be implemented yet.

Different approaches have been taken to improve energy efficiency, and we have seen a diversity of ML methods, partially intertwined, to realize them. Those techniques range from classical and well-known ML methods like DTs and SVMs [39] to less popular methods like HB [40] and to booming DL methods like DRL [35, 41, 43] and LSTMs [36, 38, 40]. A trend towards DL models could be identified, whereas DRL [35, 41, 43] and LSTMs [36, 38, 40] took the lead in mainly used models. It should be highlighted that traditional ML methods were investigated [37, 39, 40, 42, 43] and in one instance, they outperformed the nowadays hyped DL models [39]. However, this statement has to be viewed with caution, as not necessarily the best fit DL model or ML model has been implemented for comparison.

We further investigated the implemented IoT networks and discovered that various hardware and IoT applications have been used to verify the proposed method's effectiveness. The IoT applications utilized dealt with either simply data transmission [36, 37, 38, 39, 42, 40, 44], or was an image-based application, like object recognition [43, 35] or image classification [41]. To implement these applications, three papers leveraged a full IoT network in a for their purpose adequate environment, ranging from a total of 4 nodes to a total number of 35 nodes [39, 40, 42]. They, however, only used dummy data to transmit instead of real sensor readings. The other seven papers provided prototypes to test their proposed method [35, 36, 37, 38, 41, 43, 44]. Three papers started with simulations and then backed up their findings with prototype implementations [38, 42, 43]. Communication technologies used were ESP-Now [39], BLE [39, 37, 44], LoRa [39, 38], ZigBee [39, 42], and WiFi [40, 43, 36] (Fig. 5). The data transmitted by the networks were either dummy sensor readings [39, 42, 40], images/sensed data loaded on the node beforehand [43, 37, 41, 35], energy consumption readings from a smart meter [38], environmental sensor readings [36], or images captured by an image sensor [44]. Distinguishing themselves from the rest are the last three works [36, 38, 44]. They very closely replicate actual applications by using sensor readings on the device instead of dummy data and being set in realistic environments. In total, we can see that a multitude of different IoT applications and hardware were considered. We emphasize at this point again the value of executing ML models on actual hardware, further strengthening trustworthiness by the usage of sensor readings on device, and evaluating their performance in this way rather than only through simulations. However,

broadening the evaluation to several different IoT networks and showing that the proposed method gives the expected performance in different scenarios and environments would further increase credibility for deployment by other researchers or industry.

As for the performance metrics, all of them are concerned with energy, be it just reducing its consumption [35, 36, 39, 40, 41, 43, 44], improving efficiency [38], or measuring its impact by an increase in a device’s lifetime [37, 42]. Despite them being all related, having different metrics makes it difficult to compare methods from different papers. But even if the same metric was being used throughout the literature, differences like IoT application, communication technology, hardware, or network size, among others, make it, in any case, practically impossible to compare findings from different papers [74].

4.1. Identified gaps

From our findings, we identify the following gaps in the current literature:

- Lack of evaluation in real IoT networks. Though 342 papers were selected after the full-text screening to match all inclusion criteria, excluding the implementation of actual hardware, the majority carried out simulations to evaluate the effectiveness of their model. Although simulations provide important insights and can be a first step when proposing a method to enhance an IoT network’s sustainability, testing the method on actual hardware is necessary. There are two main reasons for this. First, simulation results often greatly differ from performance on real hardware [71]. Second, implementing and executing ML models on resource-constrained hardware is a hurdle that might need to be overcome with modifications of the model, which then may lead to poorer performance than before. Further, only three offered a complete IoT network implementation from the ten included papers. The other seven papers provided prototypes consisting of two connected devices. Testing in a bigger IoT network consisting of multiple nodes and implemented in an environment where it would be implemented in practice is expected to uncover unseen issues and performance variations. However, the full network implementations only transmit dummy data, neglecting the effect the usage of actual sensor readings could have, i.e., their temporal and spatial dynamics.

- Lack of implementation of computation models in real hardware, along with a quantitative assessment of their energy and latency requirements. This involves the identification of the tradeoffs between the energy savings and other performance improvements (e.g., related to decision-making for network or protocol control) that are enabled by the use of ML and the additional energy and latency burden due to its execution. This includes the investigation of ML models in isolation and, subsequently, their integration into the whole IoT ecosystem, jointly considering communication and scheduling aspects.
- Lack of available datasets and code. In the spirit of open science, making datasets and code accessible would greatly help advance research further. Other researchers can build on existing code or work with existing datasets, making it additionally meaningful to compare methods against each other on the same dataset. The implemented work would be reproducible, aiding in actual implementation in the industry. Despite this, *none of the included papers* provided the data or code open-source, except those that use existing datasets that have been open-source before already, like MNIST. Falling in this category would also be to further make prototypes / implemented networks accessible. Collaboration between researchers could be enhanced and existing IoT infrastructure leveraged.
- Lack of anomaly management. All methods assume normal network behavior. This is not necessarily the case in real-life implementations. Anomalies in the environment can impact communication, which could lead to drastic quality drops for those methods, increasing energy efficiency through compromising network quality. Also, methods focusing on computational efficiency might lead to unreasonable performance when hardware anomalies in the computational unit of the sensors occur.
- Lack of comparison between different models and algorithms to enhance sustainability. Most papers propose an ML method or DL model architecture and then compare it to the normal network operation or standard energy-saving methods. To support the superiority of the proposed method, comparing it to other (ML) methods is recommended. Ablation studies to ensure optimal hyperparameter settings, to which ML methods are very sensitive, would also be appreciated to convince

that the methods described are chosen in the best way possible for the given task.

4.2. Directions for future research

Taking the findings from our systematic literature review and the identified gaps into account, we suggest the following direction for future research:

- Implementing more advanced DL models. While DRL and LSTMs are powerful models from the DL family, in recent years, other mechanisms have risen in the booming fields of image processing and natural language processing. To name the prominent attention mechanism, these kinds of recent advances in DL have not been applied to the field of sustainable IoT yet. Though originating from image processing or natural language processing, those new methods are being successfully applied in many other research areas, making them promising to yield good performance for sustainable IoT. Along these lines, modern energy-efficient ML designs, such as ELMs [75], reservoir neural networks [76] or even spiking neural networks and neuromorphic hardware [77, 78] should be considered in addition to classical convolutional, recurrent networks or transformer architectures.
- The energy and latency cost of ML and DL models is still insufficiently explored. Analytical, semi-analytical or empirical models to estimate the energy consumption of neural network architectures would be highly valuable to implement optimization-based methods for two main reasons: to come up with highly energy efficient ML architectures for IoT via network architecture search, and to optimize ML models at runtime, e.g., via split computing or early exiting [79], in an effort to continuously minimizing their energy footprint. Therefore, the design of a metric to measure the energy cost of ML and DL models independent of the used hardware, which will be accepted as the metric to use for energy cost in the community, is of utmost importance to compare and assess research as well.
- Comprehensive frameworks, with solid validation on hardware, including neural network models and communication protocols, are still missing where the two are jointly optimized. This is the natural evolution of current network technology, which works following rigid protocol

rules, into future “intelligent” networks with self-optimization and self-assessment capabilities. This objective is being discussed extensively [80], and while simulation-based research can be found [81], methods validated on hardware are still lacking in the IoT domain.

- Open access to testbeds. Many more papers evaluate their methods on simulations rather than hardware due to a lack of financial means or expertise for building intricate testbeds for evaluation. Moreover, building new testbeds in each research group contradicts the sustainability principle, as many chips and hardware would be needed. Another crucial point is the time and effort it takes to build a testbed, which could be invested in further research instead. Therefore, we advocate using existing testbeds by making their usage also possible outside the research group implementing the testbed. This can be achieved either by collaboration or by taking it a step further, by making testbeds accessible and their collected datasets open access to the whole research community [82].
- Consideration of the existence of anomalies. To support the methods’ robustness, showcasing that existing methods also work in the presence of anomalies is a field of interest. In case methods cease to work properly in the presence of anomalies, anomaly detection, and removal should be investigated. Anomalies can not only hinder the effectiveness of proposed methods to enhance sustainability, but they also increase energy consumption. Detecting and removing them is a gap yet to be filled.
- Verifying existing methods evaluated by simulations on actual hardware. As we already stated, the amount of research based on solely simulations outnumbers actual hardware implementations. Instead of pushing forward for new methods, verifying existing ones and making them therefore applicable in real IoT networks would carry a high value.

4.3. Limitations

The findings of this review should be considered in light of some limitations. Although the data source covered many scientific databases, they did not encompass all available literature, limiting the generalizability of the

findings. The review specifically focused on actual hardware implementations, excluding many approaches to optimizing an IoT network's sustainability that were evaluated solely through simulations. Though the limitations of simulation-based results were discussed and their exclusion motivated, it must be noted that many simulations follow real-world-like assumptions, and their findings are meaningful and insightful for application on hardware. Collecting papers and reviewing them based on the quality of the simulations would be another critical contribution. Additionally, relevant works may have emerged during and beyond the specified data collection period, which spanned from August 2013 to July 2023. Finally, despite meticulous data extraction and analysis, the potential for bias remains.

5. Conclusion

Both IoT networks and ML have experienced an explosion in usage in recent years. IoT networks are getting deployed in many different environments for monitoring and automation, while ML aids in optimizing routines and finding solutions to problems traditional methods are not capable of solving. IoT networks, however, come at the expense of the environment. They consist of nodes using chips that must be manufactured and later disposed of. Further, their operation is energy-consuming. In this systematic literature review, we have looked into the application of ML on the edge of IoT networks to increase their sustainability. We could differentiate between communication-focused and computation-focused approaches, where ML has been used to optimize the network's communication or the computation on the IoT nodes. We have found that a variety of different aspects of an IoT network have been tackled to optimize by using ML, as well as numerous different kinds of IoT networks have been implemented to validate the proposed ML methodology. Based on this, we have identified gaps and proposed directions for future research.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to reformulate the abstract and introduction for a better reading flow. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] H. Prabhu, J. Joseph, K. K. Koushik, J. Karthikeyan, M. Deeksha, IoT and its potential for transforming industries, International Journal of Networks and Systems (2024).
- [2] R. Prakash, D. Singh, Green Internet of Things (G-IOT) for sustainable environment, EPRA International Journal of Multidisciplinary Research (2023).
- [3] K. Raghavendar, I. Batra, A. Malik, A robust resource allocation model for optimizing data skew and consumption rate in cloud-based IoT environments, Decision Analytics Journal (2023).
- [4] R. I. Mukhamediev, Y. Popova, Y. Kuchin, E. Zaitseva, A. Kalimoldayev, A. Symagulov, V. Levashenko, F. Abdoldina, V. Gopejenko, K. Yakunin, et al., Review of Artificial Intelligence and Machine Learning technologies: Classification, restrictions, opportunities and challenges, Mathematics (2022).
- [5] K. Venigandla, Integrating RPA with AI and ML for enhanced diagnostic accuracy in healthcare, Power System Technology (2022).
- [6] S. Tatineni, A. Mustyala, Enhancing financial security: Data Science's role in risk management and fraud detection, ESP International Journal of Advancements in Computational Technology (2024).
- [7] S.-H. Chung, Applications of smart technologies in logistics and transport: A review, Transportation Research Part E: Logistics and Transportation Review (2021).
- [8] H. Khayyam, B. Javadi, M. Jalili, R. N. Jazar, Artificial Intelligence and Internet of Things for autonomous vehicles, Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems (2020).
- [9] T. Averbuch, K. Sullivan, A. Sauer, M. A. Mamas, A. A. Voors, C. P. Gale, M. Metra, N. Ravindra, H. G. Van Spall, Applications of Artificial Intelligence and Machine Learning in heart failure, European Heart Journal-Digital Health (2022).

- [10] M. Alkhayyal, A. Mostafa, Recent developments in AI and ML for IoT: A systematic literature review on LoRaWAN energy efficiency and performance optimization, *Sensors* (2024).
- [11] L. Zhou, S. Pan, J. Wang, A. V. Vasilakos, Machine Learning on big data: Opportunities and challenges, *Neurocomputing* (2017).
- [12] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, X. Yang, A survey on the Edge Computing for the Internet of Things, *IEEE Access* (2017).
- [13] Z. Ning, X. Kong, F. Xia, W. Hou, X. Wang, Green and sustainable Cloud of Things: Enabling collaborative Edge Computing, *IEEE Communications Magazine* (2018).
- [14] R. Almutairi, G. Bergami, G. Morgan, Advancements and challenges in IoT simulators: A comprehensive review, *Sensors* (2024).
- [15] N. Tekin, A. Acar, A. Aris, A. S. Uluagac, V. C. Gungor, Energy consumption of on-device Machine Learning models for IoT intrusion detection, *Internet of Things* (2023).
- [16] L. Schuhmacher, S. Pollin, H. Sallouha, ecoBLE: A low-computation energy consumption prediction framework for Bluetooth Low Energy, *International Conference on Embedded Wireless Systems and Networks* (2023).
- [17] H. Djigal, J. Xu, L. Liu, Y. Zhang, Machine and Deep Learning for resource allocation in multi-access Edge Computing: A survey, *IEEE Communications Surveys & Tutorials* (2022).
- [18] M. A. Sami, T. A. Khan, Forecasting failure rate of IoT devices: A Deep Learning way to predictive maintenance, *Computers and Electrical Engineering* (2023).
- [19] M. Merenda, C. Porcaro, D. Iero, Edge Machine Learning for AI-enabled IoT devices: A review, *Sensors* (2020).
- [20] M. G. S. Murshed, C. Murphy, D. Hou, N. Khan, G. Ananthanarayanan, F. Hussain, Machine Learning at the network edge: A survey, *ACM Computing Surveys* (2021).

- [21] N. Charef, A. Ben Mnaouer, M. Aloqaily, O. Bouachir, M. Guizani, Artificial Intelligence implication on energy sustainability in Internet of Things: A survey, *Information Processing & Management* (2023).
- [22] D. Wang, D. Zhong, A. Souri, Energy management solutions in the Internet of Things applications: Technical analysis and new research directions, *Cognitive Systems Research* (2021).
- [23] A. B. Guiloufi, S. El khediri, N. Nasri, A. Kachouri, A comparative study of energy efficient algorithms for IoT applications based on WSNs, *Multimedia Tools and Applications* (2023).
- [24] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *British Medical Journal* (2021).
- [25] I. Gryech, C. Asaad, M. Ghogho, A. Kobbane, Applications of Machine Learning & Internet of Things for outdoor air pollution monitoring and prediction: A systematic literature review, *Engineering Applications of Artificial Intelligence* (2024).
- [26] I. Gryech, E. Vinogradov, A. Saboor, P. S. Bithas, P. T. Mathiopoulos, S. Pollin, A systematic literature review on the role of UAV-enabled communications in advancing the UN's sustainable development goals, *Frontiers in Communications and Networks* (2024).
- [27] D. Li, M. Lan, Y. Hu, Energy-saving service management technology of Internet of Things using Edge Computing and Deep Learning, *Complex & Intelligent Systems* (2022).
- [28] X. Chen, J. Zhang, B. Lin, Z. Chen, K. Wolter, G. Min, Energy-efficient offloading for DNN-based smart IoT systems in cloud-edge environments, *IEEE Transactions on Parallel and Distributed Systems* (2021).
- [29] S. Bebortta, A. K. Singh, B. Pati, D. Senapati, A robust energy optimization and data reduction scheme for IoT based indoor environments using local processing framework, *Journal of Network and Systems Management* (2021).

- [30] Y. Cai, A. Genovese, V. Piuri, F. Scotti, M. Siegel, IoT-based architectures for sensing and local data processing in ambient intelligence: Research and industrial trends, IEEE International Instrumentation and Measurement Technology Conference (2019).
- [31] L. Gatti, P. Seele, L. Rademacher, Grey zone in—greenwash out. a review of greenwashing research and implications for the voluntary-mandatory transition of CSR, International Journal of Corporate Social Responsibility (2019).
- [32] Scopus. 2024 elsevier b.v.,
<https://www.scopus.com>. [online] accessed on: September 2024.
- [33] M. Schotten, M. El Aisati, W. Meester, S. Steiginga, C. Ross, A brief history of scopus: The world's largest abstract and citation database of scientific literature, Research Analytics (2017).
- [34] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyan — a web and mobile app for systematic reviews, Systematic Reviews (2016).
- [35] Z. Xu, D. Yang, C. Yin, J. Tang, Y. Wang, G. Xue, A co-scheduling framework for DNN models on mobile and edge devices with heterogeneous hardware, IEEE Transactions on Mobile Computing (2023).
- [36] J. Han, G. H. Lee, J. Lee, T. Y. Kim, J. K. Choi, A novel Deep-Learning-based robust data transmission period control framework in IoT Edge Computing system, IEEE Internet of Things Journal (2022).
- [37] J. Azar, A. Makhoul, M. Barhamgi, R. Couturier, An energy efficient IoT data compression approach for edge Machine Learning, Future Generation Computer Systems (2019).
- [38] P. Kumari, R. Mishra, H. P. Gupta, T. Dutta, S. K. Das, An energy efficient smart metering system using Edge Computing in LoRa network, IEEE Transactions on Sustainable Computing (2022).
- [39] A. F. X. Glória, P. J. A. Sebastião, Autonomous configuration of communication systems for IoT smart nodes supported by Machine Learning, IEEE Access (2021).

- [40] J. Sheth, C. Miremadi, A. Dezfouli, B. Dezfouli, EAPS: Edge-assisted predictive sleep scheduling for 802.11 IoT stations, IEEE Systems Journal (2022).
- [41] S. K. Panda, M. Lin, T. Zhou, Energy-efficient computation offloading with DVFS using Deep Reinforcement Learning for time-critical IoT applications in Edge Computing, IEEE Internet of Things Journal (2023).
- [42] C. Savaglio, P. Pace, G. Aloisio, A. Liotta, G. Fortino, Lightweight Reinforcement Learning for energy efficient communications in Wireless Sensor Networks, IEEE Access (2019).
- [43] Y. Xiao, L. Xiao, K. Wan, H. Yang, Y. Zhang, Y. Wu, Y. Zhang, Reinforcement Learning based energy-efficient collaborative inference for mobile Edge Computing, IEEE Transactions on Communications (2023).
- [44] J. Lim, J. Seo, Y. Baek, CamThings: IoT camera with energy-efficient communication by Edge Computing based on Deep Learning, International Telecommunication Networks and Applications Conference (2018).
- [45] 2021 Espressif Systems (Shanghai) Co., Ltd. ESP-Now overview.
<https://www.espressif.com/en/solutions/low-power-solutions/esp-now>. [online] accessed on: September 2024.
- [46] 2017 Bluetooth SIG, Inc. Bluetooth Low Energy.
<https://web.archive.org/web/20170310111443/https://www.bluetooth.com/what-is-bluetooth-technology/how-it-works/low-energy>. [online] accessed on: April 2025.
- [47] 2024 LoRa Alliance. what is LoRaWAN?
<https://lora-alliance.org/about-lorawan>. [online] accessed on: September 2024.
- [48] 2015 ZigBee Alliance. ZigBee specifications.
<https://zigbeealliance.org/wp-content/uploads/2019/11/docs-05-3474-21-0csg-zigbee-specification.pdf>. [online] accessed on: September 2024.
- [49] 2023 Espressif Systems (Shanghai) Co., Ltd. ESP32-WROOM-32 datasheet.

- https://www.espressif.com/sites/default/files/documentation/esp32-wroom-32_datasheet_en.pdf. [online] accessed on: July 2024.
- [50] CrossbowTechnology, Inc. TelosB mote platform.
https://www.willow.co.uk/telosb_datasheet.pdf. [online] accessed on: July 2024.
- [51] Cypress Semiconductor. CYW43907: IEEE 802.11 a/b/g/n SoC with an embedded applications processor,
<http://www.cypress.com/file/298236/download>. [online] accessed on: September 2024.
- [52] Raspberry Pi Ltd. Raspberry Pi 4 model B,
<https://datasheets.raspberrypi.com/rpi4/raspberry-pi-4-product-brief.pdf>. [online] accessed on: September 2024.
- [53] 2016 Engelec Power Technology Co., Ltd. EEDDS238-4W single phase DIN rail WiFi energy meter,
<https://www.engelecgroupp.com/eeddss238-4w-single-phase-din-rail-wifi-energy-meter-pd49263996.html>. [online] accessed on: September 2024.
- [54] Hoperef Electronic. RFM95/96/97/98(W) - low power long range transceiver module v1.0,
https://www.mouser.be/datasheet/2/975/1463993415rfm95_96_97_98w-1858106.pdf. [online] accessed on: September 2024.
- [55] Raspberry Pi Ltd. Raspberry Pi 3 model B,
<https://www.raspberrypi.com/products/raspberry-pi-3-model-b>. [online] accessed on: September 2024.
- [56] Dragino Technology Co., Ltd. Open source LoRa WiFi gateway LG01,
https://www.dragino.com/downloads/downloads/datasheet/en/datasheet_lg01.pdf. [online] accessed on: September 2024.
- [57] 2021-2024, wemos.cc. LOLIN D1 mini,
https://www.wemos.cc/en/latest/d1/d1_mini.html. [online] accessed on: September 2024.

- [58] Polar. Polar M600,
https://support.polar.com/e_manuals/m600/wear-os/polar-m600-user-manual-english/manual.pdf. [online] accessed on: September 2024.
- [59] 2024 Omnipixel. (End of life) CMOS VGA (640x480) image sensor with OmniPixel 2 technology,
<https://www.ovt.com/products/ov7725>. [online] accessed on: September 2024.
- [60] 2024 Nordic Semiconductor. nRF52840,
<https://www.nordicsemi.com/products/nrf52840>. [online] accessed on: September 2024.
- [61] Nvidia. Jetson nano 2GB developer kit,
<https://developer.nvidia.com/embedded/jetson-nano-2gb-developer-kit>. [online] accessed on: September 2024.
- [62] Intel. Intel core i3-6006U processor specification,
<https://ark.intel.com/content/www/us/en/ark/products/91157/intel-corei3-6006u-processor-3m-cache-2-00-ghz.html>. [online] accessed on: September 2024.
- [63] Lenovo. Legion 5i 17,
<https://www.lenovo.com/in/en/p/laptops/legion-laptops/legion-5-series/legion-5i-17/88gmy501441?orgref=https%253a%252f%252f>. [online] accessed on: September 2024.
- [64] Wikipedia. Pixel 2,
https://en.wikipedia.org/wiki/pixel_2. [online] accessed on: September 2024.
- [65] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE (1998).
- [66] Y. Kim, B. Jung, Y. Song, Online Learning for joint energy harvesting and information decoding optimization in IoT-enabled smart city, IEEE Internet of Things Journal (2023).

- [67] Y. Wang, F. Shang, J. Lei, X. Zhu, H. Qin, J. Wen, Dual-attention assisted Deep Reinforcement Learning algorithm for energy-efficient resource allocation in Industrial Internet of Things, Future Generation Computer Systems (2023).
- [68] K. Wei, L. Zhang, S. Wang, Intelligent channel allocation for Age of Information optimization in Internet of Medical Things, Wireless Communications and Mobile Computing (2021).
- [69] T. Wang, H. Ke, X. Zheng, K. Wang, A. Sangaiah, A. Liu, Big data cleaning based on mobile Edge Computing in industrial sensor-cloud, IEEE Transactions on Industrial Informatics (2020).
- [70] L. Cao, Y. Cai, Y. Yue, S. Cai, B. Hang, A novel data fusion strategy based on Extreme Learning Machine optimized by bat algorithm for mobile heterogeneous Wireless Sensor Networks, IEEE Access (2020).
- [71] J. Caba, F. Rincón, J. Dondo, J. Barba, M. Abaldea, J. C. López, Testing framework for in-hardware verification of the hardware modules generated using HLS, International Symposium on Power and Timing Modeling, Optimization and Simulation (2018).
- [72] A. Gudnavar, K. Naregal, Edge Computing in Internet of Things (IoT): Enhancing IoT ecosystems through distributed intelligence, Advancement of IoT in Blockchain Technology and its Applications (2023).
- [73] D. Donoho, Data science at the singularity, Harvard Data Science Review (2024).
- [74] A. Glória, F. Cercas, N. Souto, Comparison of communication protocols for low cost Internet of Things devices, South Eastern European Design Automation, Computer Engineering, Computer Networks and Social Media Conference (2017).
- [75] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme Learning Machine: Theory and applications, Neurocomputing (2006).
- [76] M. Lukoševičius, H. Jaeger, Reservoir Computing approaches to Recurrent Neural Network training, Computer Science Review (2009).

- [77] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, A. Maida, Deep Learning in Spiking Neural Networks, *Neural Networks* (2019).
- [78] A. R. Young, M. E. Dean, J. S. Plank, G. S. Rose, A review of spiking neuromorphic hardware communication systems, *IEEE Access* (2019).
- [79] Y. Matsubara, M. Levorato, F. Restuccia, Split Computing and Early Exiting for Deep Learning applications: Survey and research challenges, *ACM Computing Surveys* (2022).
- [80] B. Rong, 6G: The next horizon: From connected people and things to connected intelligence, *IEEE Wireless Communications* (2021).
- [81] A. Biason, C. Pielli, M. Rossi, A. Zanella, D. Zordan, M. Kelly, M. Zorzi, EC-CENTRIC: An energy- and context-centric perspective on IoT systems and protocol design, *IEEE Access* (2017).
- [82] F. Minucci, R. M. Noguera Oishi, H. Xiong, D. Verbruggen, C. Thys, R. Hersyandika, R. Beerten, A. Colpaert, V. Ranjbar, S. Pollin, Building a real-time physical layer labeled data logging facility for 6G research, *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks* (2024).