CYBER PHYSICAL ANOMALY DETECTION FOR SMART HOMES: A SURVEY

Yasar Majib
School of Computer Science and Informatics
Cardiff University, UK
MajibY@cardiff.ac.uk

Omar Rana
School of Computer Science and Informatics
Cardiff University, UK
RanaOF@cardiff.ac.uk

Andre Asaturyan
Building Research Establishment
Hertfordshire, UK

Sharadha Kariyawasam
VortexIoT
Neath, UK

Behzad Momahed Heravi
VortexIoT
Neath, UK

Charith Perera
School of Computer Science and Informatics
Cardiff University, UK
PereraC@cardiff.ac.uk

November 12, 2023

ABSTRACT

Twenty-first-century human beings spend more than 90% of their time in indoor environments. The emergence of cyber systems in the physical world has a plethora of benefits towards optimising resources and improving living standards. However, because of significant vulnerabilities in cyber systems, connected physical spaces are exposed to privacy risks in addition to existing and novel security challenges. To mitigate these risks and challenges, researchers opt for anomaly detection techniques. Particularly in smart home environments, the anomaly detection techniques are either focused on network traffic (cyber phenomena) or environmental (physical phenomena) sensors' data. This paper reviewed anomaly detection techniques presented for smart home environments using cyber data and physical data in the past. We categorise anomalies as *known* and *unknown* in smart homes. We also compare publicly available datasets for anomaly detection in smart home environments. In the end, we discuss essential key considerations and provide a decision-making framework towards supporting the implementation of anomaly detection systems for smart homes.

***K*eywords** Internet of Things, Smart Homes, Anomaly Detection, CyberData, Physical Data

1 Introduction

Smart Homes are the merger of conventional Information Technology (IT) or cyber systems with the physical (real/tangible) world. There are three types of components in smart homes: i) cyber, components which have no interaction with physical space; ii) physical, components which have direct interaction with physical space but no connection with cyber-elements; and iii) cyber-physical, components that link cyberspace to physical space [1]. The primary role of automation in smart homes is played by sensors, actuators, control systems and the Internet of Things (IoT), a.k.a. smart devices. IoT devices are not just eyes and ears (audio/visual) anymore, as they evolved into other senses, e.g. temperature, humidity, occupancy, air quality and many more. The possibility of smart devices because of this evolution is limitless, such as a smart diaper proposed by [2], and Continuous Glucose Monitoring (CGM) devices [3].
Smart homes are primarily used for domestic living, where privacy and security are two major requisites. In the 21st century, humans spend more than 90% of their time in indoor environments [4]. The new emerging concept of smart homes brings a plethora of benefits towards optimising resources and improving living standards in smart homes coupled with privacy and security threats. Like every other system, smart homes face several cyber-physical threats. These threats can be detected in time if proper monitoring and defence mechanisms are in place. Knowledge of the systems' sources, motivations, attack vectors, targetable components, and outcomes can help secure the system in advance. Understanding human and system behaviour plays a vital role in detecting anomalies in the cyber-physical space. We reviewed both cyber and physical anomaly detection techniques for smart homes. We categorised the techniques into known and unknown categories and discussed anomaly detection in cyber and physical data. We compared publicly available datasets for anomaly detection in smart home environments. In the end, we discussed essential aspects and provided a decision-making framework to implement anomaly detection systems for smart homes. In order to improve the readability of this article, we have listed all the abbreviations we used in Table 1.

1.1 Contributions

- We reviewed anomaly detection techniques used by researchers in the past within smart home environments
- We categorised anomalies as (i) known and (i) unknown anomalies and explored how cyber and physical data could be used to detect anomalies.
- We compared how publicly available datasets are employed for the development, evaluation, and testing of anomaly detection techniques in smart home environments.
- We presented a decision-making framework for anomaly detection implementation in smart homes.

1.2 Research Questions

Following are the research questions we aim to address in this survey:

RQ1 What are the known anomalies and unknown anomalies in a smart home context?
Anomaly detection techniques in the smart homes context are based on two types. (i) Known anomalies in which data samples of anomalous activities are available, for example, DoS attacks, port scanning attacks, inactivity, and behaviour of people living with health conditions like hypertension, dementia, or Alzheimer's disease. (ii) Unknown anomalies in which data samples of anomalous activities are not available, i.e., only normal activities dataset is available.

RQ2 What are the key features and characteristics of anomaly detection techniques using cyber data in the smart home context?
Anomaly detection in the smart home context uses cyber data, that is, the cyber footprint of activities by humans or devices, mostly consisting of network traffic or system logs. There can be various ways to detect anomalies due to different types of attacks on heterogeneously configured smart home networks.

RQ3 What are the key features and characteristics of anomaly detection techniques using physical data in the smart home context?
Anomaly detection in the smart home context also uses physical data, which are the values of environmental sensors like temperature, humidity, illumination, motion, CO2, etc. and other smart devices (the physical phenomena). There are various combinations of sensors that can be employed to find anomalous activities in a smart home environment.

RQ4 What are the common publicly available datasets for smart home anomaly detection research?
To detect anomalies, datasets are required to understand the activities performed in smart homes and develop models to detect anomalies. A dataset may contain cyber, physical or both types of activities or anomalies data.

Table 1: List of abbreviations

Phrase	Acronym	Phrase	Acronym	Phrase	Acronym
Discriminant Analysis	DA	Ant Lion Optimization	ALO	Mean Absolute Error	MAE
Support Vector Machine	SVM	Self Organizing Maps	SOM	Principle Component Analysis	PCA
Logistic Regression	LR	Gaussian Mixture Models	GMM	Fuzzy Rule-Based System	FRBS
K- Nearest Neighbor	k-NN	Case-based reasoning	CBR	Spider Monkey Optimization	SMO
Decision Tree	DT	Markov Logic Network	MLN	Stacked-Deep Polynomial Network	SDPN
Probabilistic Neural Networks	PNN	Deterministic Finite Automation	DFA	Deep Learning	DL
Binary Class SVM	BSVM	Variation Autoencoder	VAE	Intrusion Detection System	IDS
One Class SVM	OSVM	Temporal Convolutional Network	TCN	User-to-Root	U2R
Naïve Bayes	NB	Wireless Sensor Network	WSN	Remote-to-Local	R2L
Bayes Network	BN	You Only Look Once	YOLO	Knowledge-Based Inference Engine	KBIE
Voted Perceptron	VP	Nonlinear AutoRegressive Network	NARX	Markov Logic Network Reasoner	MLNR
Simple Logistic Regression	SLR	Simple Moving Average	SMA	Dynamic Bayesian Network	DBN
Square Prediction Error	SPE	Mean Squared Percentage Error	MAPE	Salp Swarm Algorithm	SSA
Hidden Markov Model	HMM	Classification and Regression Trees	CART	Voyeurism	V
Hidden Semi-Markov Model	HSMM	Gradient Boosted Trees	GBT	Physical Harm	PH
Root Mean-Squared Error	RMSE	CatBoosting	CB	Hypertension	HPT
Mean Squared Error	MSE	Fuzzy Logic	FL	Physical Intrusion	PI
Optimal Clustering	OC	Physical Theft	PT	Industrial Control Systems	ICS
Genetic Algorithm	GA	Fault Detection	FTD	Network Traffic	NT
Grey Model	GM	Mechanical Exhaustion	ME	Property Loss	PL
Neurodegenerative Disorders	NDD	Mild Cognitive Impairment	MCI	Anomalous Energy Consumption	AEC
Anomalous Behaviour	AB	Fall Detection	FLD	Inactivity Detection	IAD
Accident	ACC	Energy Theft	ET	Intelligent Transportation System	ITS
Anomaly Detection	AD	Environmental Hazard	EH	Smart Homes	SH
Echo State Network	ESN	Back Propagation Through Time	BPTT	Real Time Recurrent Learning	RTRL
Electrocardiogram	ECG	Long Short-Term Memory	LSTM	Recurrent Neural Networks	RNN
Association Rule Mining	ARM	Convolutional Neural Networks	CNN	Property Theft	PT

1.3 Surveys Closer to Our Research

There is a minimal number of surveys previously conducted on anomaly detection in smart homes; most of these surveys focused on intrusion detection for cyber-attacks on IoT-based networks. Since smart homes are a sub-domain of IoT networks, it is relevant to incorporate work of the high-level domain. ElMenshawy and Helmy [5] surveyed Statistical, ML and NN-based anomaly detection techniques for IoT networks with a slight coverage of smart homes. Stellios et al. [6] did a comprehensive survey on IoT threats and discussed countermeasures for mitigating the threats. Their survey is focused on IoT-enabled cyber-attacks in ICS, smart power grids, ITS, healthcare and smart homes. Elrawy et al. [7] surveyed published research on IDS; their primary focus was on IoT-based smart environments, in which they briefly discussed smart homes as well. Their survey is also focused on cyber-based attacks only. Fahim and Sillitti [8] published a survey on anomaly detection in an IoT environment using physical sensors; the authors compared statistical with machine learning methods and also briefly touched smart homes as an *"intelligent inhabitant environment (IIE)"* since the participation of IoT devices is essential in a smart homes environment. They did not compare their review with similar work, which may reflect the limited unavailability of surveys on the topic. Asharf et al. [9] presented their survey focusing on intrusion detection in the IoT domain, briefly covering smart homes. They presented their work along with taxonomies and ML methods for intrusion detection.
Hammi et al. [10] summarised vulnerabilities, risks and countermeasures; the survey is focused on smart homes only. The authors discussed general mitigation but did not focus on anomaly detection techniques. Wang et al. [11] surveyed attacks and defences of smart home automation systems; their survey also lacked anomaly detection techniques. The study conducted by Yu et al. [12] discussed applications of deep learning, RNN and CNN in smart homes; the survey's main focus was anomaly detection, activity recognition and other security aspects. They discussed work published on cyber and physical anomaly detection; perhaps the cyber-related (network traffic/intrusion detection) work was comparatively minor compared to physical sensor-based applications. Another survey was conducted by DeMedeiros et al. [13] about AI-based anomaly detection systems recently. The authors of this survey touched on smart homes briefly, as the main focus of their work was related to IoTs and sensor networks in general. They consider all types of anomalies for their work, e.g. malicious attacks, faults and abrupt environmental changes. Keersmaeker et al. [14] surveyed both cyber and physical, publicly available datasets in the anomaly detection domain. The authors focus on protocol and other technical details of the datasets available. Table 2 provided an outlook about similar surveys related to anomaly detection in smart homes.

1.4 Paper Structure

The remainder of the paper is structured as follows. Section 2 discusses the adopted methodology to conduct this research; we have provided details of the survey method/protocol, eligibility criteria, and the risk of any possible

Table 2: Surveys Closer to Our Research

Paper	Year	SH	IoT	Threats	AD	ST	ML	NN	Cyber	Physical	Datasets
ElMenshawy and Helmy [5]	2018	◐	●	○	●	●	●	●	○	○	○
Stellios et al. [6]	2018	◐	●	●	○	○	○	○	●	●	○
Elrawy et al. [7]	2018	◐	●	●	◐	●	●	●	●	○	○
Fahim and Sillitti [8]	2019	◐	◐	○	●	●	●	●	○	●	○
Asharf et al. [9]	2020	◐	●	●	◐	○	●	●	●	○	●
Hammi et al. [10]	2021	●	◐	●	○	○	○	○	●	●	○
Wang et al. [15]	2022	●	●	●	○	○	○	○	●	●	○
Yu et al. [12]	2022	●	●	○	●	○	●	●	●	●	●
DeMedeiros et al. [13]	2023	◐	●	○	●	○	●	●	●	●	●
Keersmaeker et al. [14]	2023	◐	●	○	○	○	○	○	●	●	●
Our Survey	2023	●	●	●	●	●	●	●	●	●	●

Index: SH: Smart Home, AD: Anomaly Detection, ST: Statistical Techniques, ML: Machine Learning, and NN: Neural Networks

bias. Section 3 discusses smart homes, layers and components. Section 4 provides an overview of anomaly detection, approaches, and techniques with a taxonomy of presented anomaly detection techniques in the past. Sections 5 and 6 provide details of the surveyed work under known and unknown anomaly detection techniques in both cyber and physical data. In Section 7, we discuss and compare publicly available datasets for anomaly detection in smart homes. We have presented an anomaly detection design framework for consideration and systems design in Section 8. Section 9 provides future research directions, and Section 10 provides the conclusion of this survey.

2 Methodology

In this section, we discuss the scope of this paper and our adopted survey method. We don't consider areas like internet security firewalls, cryptography protocols, non-IoT networks, core-medical domain datasets, frameworks without anomaly detection techniques, and smart applications to ensure this work remains focused on the specific topic.

2.1 Survey Method

We used the Scopus tool to search for papers on anomaly detection in smart home environments as a starting point, followed by backward and forward snowballing techniques to reach out for relevant research papers. We used the Scopus search engine to find papers related to anomaly detection in smart homes using the search-term ("anomal* OR outlier OR abnormal*") AND "detection" AND ("dwelling" OR "smart home" OR smarthome* OR "home computing" OR "home automation" OR "smart house" OR "connected home" OR domotic"). We found 92 papers, of which 69 were from Scopus and 23 using forward and backwards snowballing. We then ran a filtering process based on the eligibility of papers and got 92 articles remaining. From this output, we filtered by full paper read and selected 58 articles to be added to the survey. Figure 1 shows the process of our selection criteria.

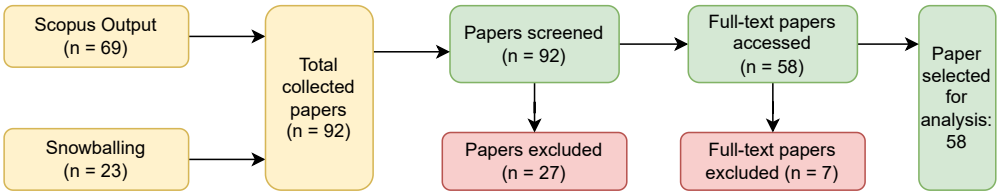

Figure 1: Search and Selection Method

2.1.1 Eligibility Criteria

Papers should present an anomaly detection technique using one or more publicly available or private datasets related to smart homes. Papers should not be surveys or lecture notes. Papers should be written in the English language. We also selected papers using forward and backward searches of papers that we collected from the initial keyword search engine.

2.1.2 Risk of Bias

Scopus is a well-known keyword search engine for papers of all disciplines; the only known bias about Scopus is towards literature in the English language [16]. Since our selection criteria are configured to filter out non-English literature, we see no issue using Scopus to search papers. Since this is not a systematic literature review, we may have ignored some important work in this domain.

3 Smart Homes: Overview

A smart home was defined by Housing Learning & Improvement Network (HLIN) as *"a dwelling incorporating a communications network that connects the key electrical appliances and services, and allows them to be remotely controlled, monitored or accessed"*, researched by Intertek as part of Department of Trade and Industry (DTI) smart homes Project [17]. It can be safe to infer that a smart home is a place where electric devices can command or control the cyber-physical space based on activities in the physical environment on top of automation rules. A smart home can be a set of smart devices for various parts of life, including Health, Comfort, Illumination, Food/Kitchen, Cleaning, Water Management, Physical Safety & Security, Metering, Sensing, Communication, Entertainment, Agility, Furniture, e-Pet systems and others [18]. Smart devices have penetrated the smart home market pervasively; off-the-shelf products are available for every aspect of life. There are various commercial and non-commercial/open-source home automation platforms available for smart homes; most of these platforms, e.g. Google Home, Apple HomeKit, Samsung SmartThings, AWS IoT, HomeAssistant, and OpenHAB, are based on a central hub for controlling off-the-shelf devices in the same network using either built-in support or via third-party APIs, with some exceptions e.g. cloud-based IFTTT service and AWS IoT for supporting both cloud and hub-based architectures. Other than a few non-commercial/open-source platforms, e.g. HomeAssistant and OpenHAB, most of the platforms are commercial [15].
Systems working within a smart home are vulnerable to errors, faults and failures [19]. There are three categories of threat sources, i.e., accidents (e.g. power surge/cuts, water damage or physical damage, etc.), environmental (natural causes, e.g. storms or floods) and adversarial (opportunist or targeted). With the evolution of IoTs, the threats landscape has expanded from the physical realm to cyber and cyber-physical domains [1]. Tracing the source of the threat is difficult in smart homes due to the transformation of cyber threats into physical incidents. Authors of [20] mentioned six significant types of cyber threats in their research, i.e., eavesdropping, masquerading, replay attacks, message modification, denial of service, and malicious codes. Threats were further categorised into three categories, i.e., unintentional, malfunction, and intentional in [21]. Unintentional threats can result from accidental changes, information leakage, actions based on unreliable information, or lack of planning. Malfunctions can result from internet failure, communication medium/channel failure, device failure, power failure, or damage resulting from third-party devices. Intentional threats are identity fraud, denial of service, manipulation or hijacking of the system. Prevention of cyber or physical threats requires data to develop an understanding or generate detection methods.

3.1 Smart Home Data Sources

Smart homes are equipped with various physical sensors connected to a central system (smart hubs) via various communication mediums. To interact with users for better user experiences, these smart hubs maintain a connection with cloud-based services where the user applications can connect to interact with the smart home devices. Figure 2 shows a holistic visualization of obtrusive and unobtrusive physical data sources, cyber-physical integration and cyber data sources used in smart home anomaly detection. In the following subsections, we discussed various data sources categorised in cyber and physical sections.

3.1.1 Cyber Data in Smart Homes

A significantly large number of IoT devices are connected to the internet across the globe. These devices can be compromised to become privacy and security threats. IoT devices can generate massive DDoS attacks [22]. Using network traffic, a lot of work has been done in anomaly detection in smart homes. Most of the work presented is based on detecting known attacks using known samples of attack scenarios along with normal traffic [23, 24, 25, 26, 27, 28, 29, 30]. There is a wide range of data sources for network traffic, e.g. network traffic monitoring using a device in the network, smart hub, network router, or system logs from devices. On top of various available data sources, many different mediums can provide different data types, e.g. ZigBee, WiFi, Bluetooth, etc.
The cyber sources are also responsible for traditional IT tasks, e.g. communication protocols, data storage, data processing, user interaction applications, etc. The format of cyber data for anomaly detection in smart homes depends on various factors, and we discuss these factors in the sub-sections. Figure 2 provides a visual brief of these factors.

- **Data Transport** Communication between devices is a critical task for automation. This activity is carried out on a communication medium by adopting a communication protocol supported by the medium and the environment, application, sensors, actuators and control devices. There are two data transmission modes, i.e., machine to machine (M2M) or IoT, where m2m is a connection between two or more devices, versus IoT is a connection between a device and an application [31]. A wide variety of communication protocols are available for the purpose, e.g. IP or IPv4, IPv6, Bluetooth, BLE, ZigBee, Z-Wave, LonTalk [32] and Message Queuing Telemetry Transport (MQTT) [33]. Proprietary protocol stacks are also being migrated towards IP, e.g. ZigBee and BLE.
- **Control Systems** are deployed in most scenarios. There are several Smart Hubs [10] off-the-shelf systems for monitoring and controlling devices in a smart home.
- **Data Storage** allows historical data to be stored and processed from time to time to provide better user experience, auditing, model generation and knowledge transfer [34, 35, 36, 37].
- **User Interaction Applications** enables interaction with users, allowing users to monitor ongoing activities and providing manual instruction to the systems.
- **Third-party Systems** are either third-party integration provided via Application Programmable Interfaces (APIs) or Web-hooks [38] e.g. IFTTT [39], etc. or independent protocols like MQTT protocols.

3.1.2 Physical Data in Smart Homes

Physical data is based on sensors transmitting values from the physical world (the physical phenomena). Sensors are the major data source from the physical world, whereas actuators and communication mediums may also contribute to the smart home context.

- **Sensors** are the core components in activity or anomaly detection in a smart home environment when using physical data. Sensing is not limited to one type of sensor; perhaps eyes, ears, olfactory, gustatory, tactile, and others are widely researched fundamental components in the smart homes context. The sensor's resolution is the most important factor in sensing; a higher resolution detects the smallest change in the physical environment. Sensors have been used for various purposes in smart homes and widely researched, e.g. environment monitoring using Temperature, Humidity, Air Quality (AQ), Carbon Dioxide (CO2), Air Velocity (AVi), occupancy monitoring using Passive Infra-Red (PIR) and Motion Sensors, energy monitoring using energy monitoring sensors, intrusion detection using Infra-Red (IR), water and gas leakage detection using Thermal sensors. Sensors transmit the readings to control devices, e.g., a temperature sensor provides temperature value in the room, or a motion sensor detects activity. The activities can be used to detect abnormal behaviour of smart home users. There are two categories of sensors: (i) **Obtrusive** sensors, which interfere with humans, provide in-depth information but require user interaction at some level, and have privacy concerns. Door contact, accelerometers, tags, body wearable, medical information related (blood pressure, glucometer, heart rate, etc.) sensors, cameras, and microphones as shown in Figure 2. (ii) **Unobtrusive** sensors are the in-situ sensors placed in the environment to monitor the overall place change. Motion (PIR/IR), pressure (for bed/sofa), temperature, humidity, CO2, and other sensors are Unobtrusive. Figure 2 contains a visualisation of different types of unobtrusive sensors as physical data sources for anomaly detection in smart homes.
  The most common sensors being used in this domain are motion sensors, i.e. IR or PIR, and contact sensors, e.g. door or window state (open/close) [40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57]. Other sensors used in this research area are microphones [58, 59], video cameras [60], thermal cameras [61], power meters are also commonly used for energy-based anomaly detection in smart homes [62, 63, 64, 22, 53, 65, 55, 56, 61], and environmental sensors like temperature, humidity, illumination, pressure, CO2, CO, LPG, smoke, flame, noise, air pressure [47, 50, 51, 66, 52, 53, 67, 55, 56, 61]. Some work also focuses on human biology-related sensors for anomaly detection in smart homes, mostly based on people suffering from diseases like dementia, hypertension, MCI, Alzheimer and NDD [68, 69]. Since the network traffic between IoT devices is encrypted, detecting anomalous activities from cyber data alone is not feasible. There are a large number of sensors being used for understanding the activities in smart homes.
- **Actuators** are devices designed to change the physical environment; in smart homes, the operations of actuators are commanded by a central system and, in some cases, manually. Actuators require an energy source, which can then be converted into hydraulic, pneumatic, electrical, thermal, magnetic, and mechanical energy. Depending on the system's configuration, the control devices may trigger actuators to change the physical environment, e.g., changing an HVAC system's temperature or turning on lights. In that case, the actuator may send confirmation or a regular update of its state.
- **Communication mediums** enables devices to connect and transfer data back and forth. A large number of communication medium exists; the choice of medium depends on the number of devices, required throughput,

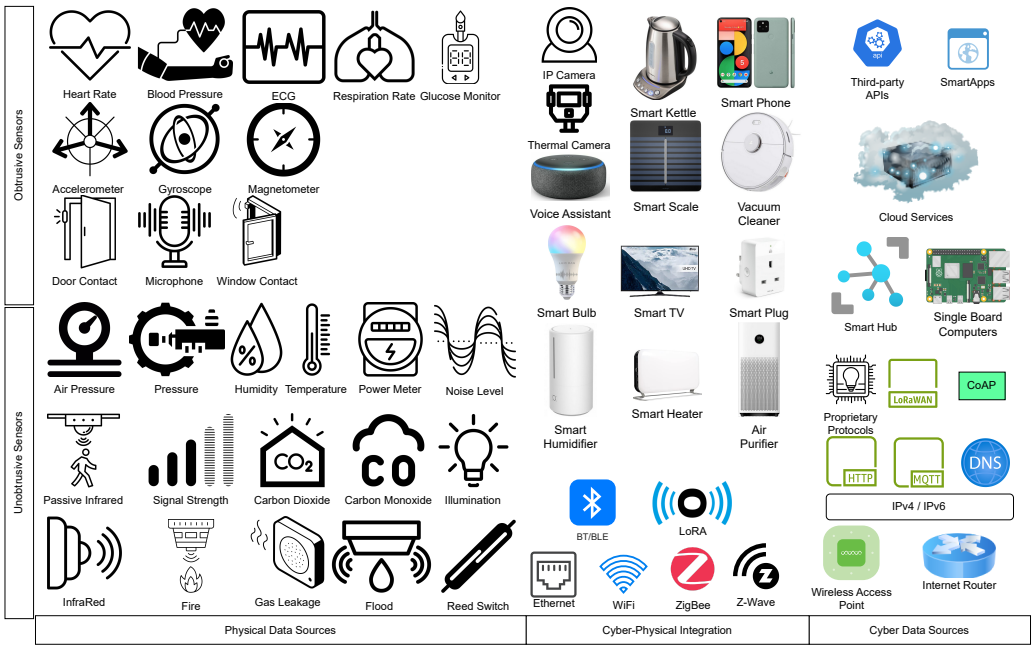

Figure 2: Cyber and Physical Data Sources for Anomaly Detection in Smart Home

type and size of vicinity. Hardware modules attached to sensors or actuators connect the physical devices to the cyber world. There are two categories of communication mediums: wired and wireless. Each has its benefits, limitations and hardware modules.

Wired mediums are less penetrable due to the requirement of physical interaction but less expandable than wireless. There are various wire types available for wired networks: Coaxial [70], RS-232 [71], RS-485 [72], Universal Serial Bus (USB) [73], Universal Powerline Bus (UPB) [74], and Ethernet [75]. Each type has its range, speed, application and bandwidth properties. Wireless mediums are more penetrateable, but they provide flexibility for adding devices compared to wired.

Wireless mediums are also critical due to the application requirements. There are several types for wireless communication open-source as well as proprietary, mainly used are RFID, NFC, Bluetooth, BLE, WiFi, ZigBee, Z-Wave, MiWi, Cellular (GSM/LTE/NB-IoT), DASH7, SigFox, LoRa, nWave, ISA100 [9] as well as BidCos [76] and others. On top of physical communication mediums, typologies are used to set up a communication system for smart homes. Topologies for the wired medium are Point-to-Point, Bus, Start, Mesh, and Hybrid. For wireless mediums, the typologies are Infrastructure Mode, Ad-Hoc Mode, and Mesh [77]. The cyber footprint of physical data depends on the communication medium, hardware module, and topology used for communication.

3.1.3 Cyber-Physical Integration

is a mechanism, a shared memory [78], that converts the physical world's interactions into the cyber world's digital form and vice versa. IoT devices provide physical data but can also be a source of cyber data as they are embedded with built-in processing units which can store/process/transmit data. Many off-the-shelf IoT devices extensively used in smart homes are embedded with cyber and physical components, e.g. IP cameras, voice assistants, and smart appliances, to provide physical interaction and generate network traffic. Figure 2 shows a glimpse of cyber-physical integration.

4 Anomaly Detection

An anomaly is a peculiar sample from all other known samples. There are three types of anomalies: point, contextual, and combined/collective anomalies [79, 13, 80] or perhaps global, local, or group anomalies, respectively, as per [81]. Contextual anomalies are the single entity of a data stream that varies from the context; contextual anomalies are mostly researched in time and space data [82, 83] e.g. temperature reading from a sensor can be too high or too low but still can be considered normal if the context supports it. A collective anomaly is a data sample peculiar from the complete data set; for example, an abnormal electrocardiogram (ECG) differs from all other normal ECGs in the training dataset. From

a smart home's perspective, an anomaly can result from an abnormal change in the physical environment, abnormal behaviour of a smart device, a sudden change in the sensor's value, or an accident or unauthorised intrusion. Anomaly detection is a process to detect anomalies in the available data; it has been an important research topic for centuries [84, 85, 86] evolved from single dimensional to high dimensional streaming data [87]. The anomalies can differ from other research areas for a smart home context.
We categorised anomaly detection approaches for the smart home context into (i) known anomaly detection, for example, detecting known cyber-attacks or human activities performed anomalously, and (ii) unknown anomaly detection, such as anomalous behaviour of humans, devices or systems in a smart home environment. Various approaches with different techniques were presented to address each category of anomaly detection in the past. We discuss the presented anomaly detection approaches and techniques in subsections. We discuss both anomaly detection categories in depth in Sections 5 for known anomalies and 6 for unknown anomalies. are Figure 3 shows a high-level view of anomaly detection categories and approaches.

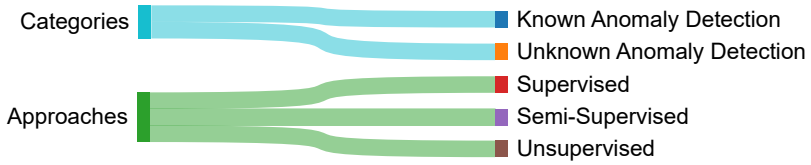

Figure 3: Anomaly Detection Categories and Approaches for Smart Homes

4.1 Anomaly Detection Approaches

The anomaly detection approach depends on the available data and knowledge on the subject area [88]. If abnormal data is unavailable and only normal data is available, then an unsupervised learning approach must be adopted. In this approach, a threshold point has to be determined based on learning from normal data. A sample is considered anomalous if the resulting output value of a given sample is greater than the threshold. A supervised learning approach can be adopted if both data types are available, e.g., normal and abnormal. This approach can identify anomalous samples by looking at the given abnormal data and thus does not require a threshold value. According to [89], there are only two approaches for anomaly detection in medicine, but in other domains, there can be a situation in some cases when the normal data is limited; in those cases, a semi-supervised approach can be adopted for anomaly detection [90].

4.2 Anomaly Detection Techniques

This section discusses anomaly detection techniques used in smart home contexts in the past. The presented techniques are mixed for known and unknown anomalies; we discuss both categories later in separate sections. There are many different ways presented to structure anomaly detection techniques; in Figure 4, we presented a taxonomy for anomaly detection techniques using previous research by [88, 80, 91, 92, 93].

4.2.1 Classical Machine Learning

Classical Machine Learning techniques are easily explainable using basic arithmetic. In the past, supervised learning techniques were heavily used in anomaly detection; now, the trends are towards unsupervised learning techniques [94]. Supervised, unsupervised, semi-supervised, ensemble and reinforcement learning approaches can be adopted using classical ML. Each technique can be implemented with various algorithms, as discussed below. We divided Classical ML techniques into three groups of approaches.

1. **Supervised Learning** used when data with labels are available; for example, if data contains samples with features of fruits, e.g. apple, banana or orange, each sample has a label for representative fruit. The machine will learn from these samples, and the output of a new given sample will be apple, banana or orange. Many algorithms can be utilised to train a machine for supervised learning. There are two sub-categories of supervised machine learning: i) Classification and ii) Regression.
   Classification algorithms can divide objects into categories. It requires a pre-categorised dataset to perform the machine learning process. For example, an email spam filter based on the classification method can predict an email is either "spam" or "not spam" if trained using a dataset containing spam and non-spam emails. Other clustering applications can be sentiment analysis, document or image classification; it is also applied for anomaly detection in IoT sensor data [95].

Regression methods look for correlations between variables (dependent or independent). There are several algorithms for supervised learning; the most popular algorithms are Logistic Regression, Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machines.

2. **Unsupervised Learning** is used when labelled data is unavailable or only one data type is present for learning. A threshold needed to be determined from model training data to predict output for a sample to be different from the training dataset based on threshold value. There are two major sub-categories of unsupervised learning: i) clustering and ii) dimension reduction.
   Clustering is a process of merging similar objects in a cluster; in contrast to classification, a pre-categorised dataset is not required in the clustering process. For example, a dataset contains samples of features for apples, bananas and oranges without labels. The clustering engine transforms the dataset into three separate clusters, e.g. 3; the result of a given sample can be the cluster number. Similar to classification, there are several algorithms for clustering; the most popular of them are K-Means Clustering, Mean Shift and DBSCAN. Traditionally, clustering was used for many applications like market segmentation (customer categorisation), image compression, and for analysing or labelling new data. Clustering is also being widely used for detecting anomalies in several applications, including IoT data [96].
   Dimension Reduction is the process of converting a high-dimensional dataset into a low-dimensional one while keeping it as meaningful or as good as the original one. Principle Component Analysis (PCA) is the most common dimension reduction technique and helps improve the performance of machine learning models [97].
3. **Semi-Supervised Learning** is based on a limited number of labelled samples of normal data; any new sample that is not classified as a normal sample is considered abnormal. For example, [28] used a semi-supervisor-based technique to detect anomalies in a smart home, implemented using DS2OS data.

4.2.2 Neural Networks

are based on complex sets of functions (linear and non-linear) and layers. There are two types of activation functions in neural networks: i) Activation functions (e.g. linear, tanh, ReLU, and sigmoid) are the regulators of nodes (neurons) for given inputs. A layer in neural networks is a collection of neurons; multiple layers can be in a neural network [98]. A few examples of layers are FullyConnected, CNN, RNN, LSTM, etc. Deep neural networks combined with RL show promising outcomes; for example, DeepMind can learn to play the Atari game without human supervision and can archive human-level performance [99]. Neural networks are also categorised under supervised, unsupervised and semi-supervised, but the below methods can be used by any approach, unlike classical machine learning. Further, into deep learning, we will discuss techniques used for anomaly detection related to IoT and smart homes.

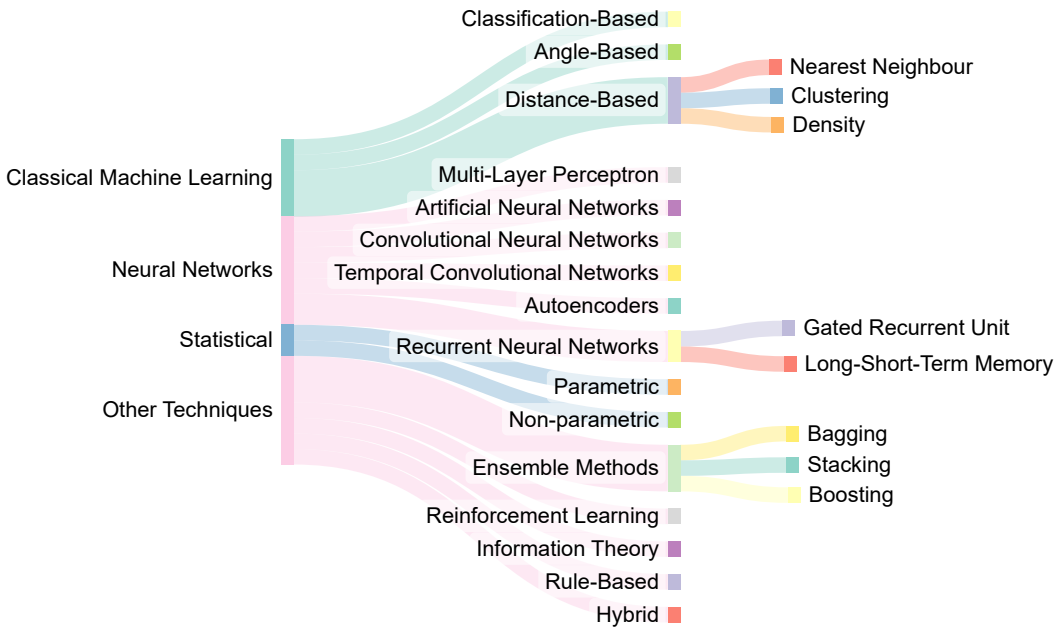

Figure 4: Taxonomy of Anomaly Detection Techniques

1. **Artificial Neural Networks (ANN)**, also known as Neural Networks, were built on the idea of the biological brain (human or animal). It is a collection of artificial neurons connected as a unit. ANN is being used in many domains, including anomaly detection. For example, [100] uses ANN to detect anomalies for a smart home using the N-BaIoT dataset.
2. **Multi Layer Perception (MLP)** is a feed-forward ANN based on three types of fully connected layers: i) input, ii) hidden, and iii) output. There can be many hidden layers, and the computations performed on these hidden layers are not visible to the user. MLP model consists of an input, an output layer, and hidden layers. All layers contain nodes connected to each node in the following layers. The loss function is defined in the output layer. MLP is used for anomaly detection in other domains. For example, [101] used MLP with other ML techniques to detect anomalies in Ultrasonic Sensor data collected using Arduino and NodeMCU, training accuracy of MLP was recorded at 99% against 100% accuracy of RF and GBC techniques.
3. **Convolutional Neural Networks (CNN)** are also ANN, mostly used for image-related work. CNN are an update of MLP, which was prone to over-fitting due to its fully connected layers. It is also known as Space Invariant Artificial Neural Networks (SIANN). CNN is widely used in academia for many applications, including image/video recognition systems, image segmentation/classification, object detection, audio/music analysis, time-series prediction, brain-computer interface and anomaly detection. For example, CNN was used by [102] for detecting anomalies in a smart environment using the WSU dataset.
4. **Recurrent Neural Networks (RNN)** are designed for data in a sequenced form, for example, time series, text sentences, or biological sequences. RNN has multiple sub-categories; the most used are **Long Short Term Memory (LSTM)** and **Gated Recurrent Unit (GRU)**. RNN is widely used in research, especially for time-series data. For example, [103] used the LSTM-RNN model and other techniques to detect anomalies in the temperature data stream in an IoT-based time-series data from the Intel Berkeley Research Laboratories (IBRL) sensor dataset. Authors of [104] published another example of RNN in which they applied LSTM, GRU and other techniques on various IoT-based datasets.
5. **Temporal Convolutional Networks** is an emerging technique for anomaly detection [28] [105], a variation of CNN to deal with sequential tasks; TCN is a combination of RNN and CNN.
6. **Autoencoder (AE)** is based on the idea of the same dimension of both input and output layers to reconstruct each input dimension in output after passing through the neural network. AE replicates input to the output, thus called replicator neural network. The number of units in the middle layer is fewer than in the edge layers, so the data is represented in a compressed/reduced form. There are five steps in an AE model: i) input (original data), ii) encoder, iii) reduced data (code), iv) decoder, and v) output (reconstructed data) [106]. An AE model can formed using any of the neural network functions and layers.

4.2.3 Statistical

techniques are based on the assumption that the probability of a normal data sample is high while an anomalous sample occurrence probability is very low [80]. Statistical techniques are also widely used for feature extraction in anomaly detection research, for example, [92] and [107]. Statistical models are divided into two types. First, the parametric technique assumes that the data is generated by a function with parametric distribution and observation [80]. Three types of parametric methods exist: Gaussian, Regression, and Mixture. The second technique is Non-Parametric, which is based on data instead of pre-defined parameters; there are two sub-categories of non-parametric styles: histograms and kernel function-based.

4.2.4 Other Techniques

Some other techniques have been used in anomaly detection in smart homes in the past; we categorise these techniques in this subsection.

1. **Reinforcement Learning** is a reward-based algorithm in which an agent learns in an environment without supervision and gets rewarded for the correct result to improve its learning model. Reinforcement learning is relatively new in the smart home anomaly detection domain but is getting popular; it is considered one of the future trends in anomaly detection by [94].
2. **Ensemble Learning** is the technique of grouping multiple algorithms as a single model; results show that ensemble learning can improve output performance and detect anomalies effectively. There are different ways to ensemble machine learning models, e.g. bagging, stacking, and boosting. Ensemble learning can be adopted for many applications; it is also being used for anomaly detection in various domains, especially in IoT sensors-based anomaly detection [108].

3. **Information Theory** based anomaly detection methods used theoretical framework to quantify and understand the concept of information and uncertainty of the data samples. Information theory is based on (i) **Entropy**: uncertainty or randomness of a random variable, (ii) **Information Gain**: quantifies the reduction in uncertainty or entropy when new information is observed, and (iii) **Kullback-Leibler Divergence (KLD)**: compares the distribution of normal data to the distribution of observed data, a higher value of KLD reflects potential anomaly [80].
4. **Rule-based** anomaly detection determines a given sample is anomalous based on specified rules. There has been some work done in this area by researchers in the smart home domain [66] [109] [48].
5. **Hybrid** learning approach leverages combining multiple techniques for an outcome; many examples of hybrid-based models in anomaly detection exist. Authors of [110] and [111] have used hybrid models to detect anomalies in their work.

4.3 Anomaly Detection Evaluation

To evaluate an anomaly detection technique for bench-marking or proving the algorithm's strength, it is necessary to have some defined criteria. Evaluation of anomaly detection models is based on two categories by default: labelled data or unlabelled data. We have presented a visualisation of evaluation metrics used for anomaly detection in the surveyed work in Figure 5.

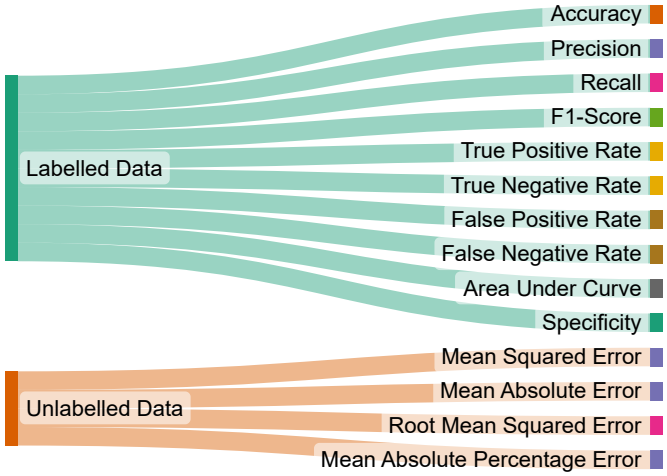

Figure 5: Evaluation Criteria for Anomaly Detection Techniques

4.3.1 Labelled data

means the data samples are associated with normal or anomalous categories/tags. Labelled data contains ground truth and makes it convenient to adopt supervised learning. Thus, performance metrics like Accuracy, Recall, Precision, F1-Score, etc., can be used to evaluate the technique with accurate results.

4.3.2 Unlabelled data

Provides data samples without predefined labels or categories, making it unclear whether the data points are normal or anomalous. Since the unlabelled data is more realistic, it is the only option to develop an anomaly detection technique in many real-world scenarios. Unlabelled data can be used for unsupervised and semi-supervised learning. One way is to visualise the data samples of graphs or use score/loss-based techniques using a set threshold for error. The techniques are tested based on a loss score (error) using MAE, MSE, RMSE or MAPE.

5 Known Anomalies Detection

In this section, we discuss the presented techniques for detecting known anomalies in a smart home context. These anomalies can be known cyber attacks, e.g. Recon, DoS, Replay, Spoofing, or known activities performed anomalously, e.g. anomalous electricity usage or a medicine box not returned to the right place after taking. In most cases, data

samples for abnormal activity are required to classify the abnormality of the activity. However, some work is presented without the availability of anomalous samples using unsupervised machine learning; the commonality in the proposed works is the availability of normal data.

5.1 Known Anomalies in Cyber Data

Cyber anomaly detection using network traffic is most commonly presented in the smart home domain as many datasets are available, with attack samples, publicly for bench-marking of normal/benign activity and known anomalies. Most of the presented work is either signature-based, anomaly-based, specification-based, hybrid or other form of IDS. DoS is the most common known threat in the research and to evaluate proposed methods. An overall view of cyber techniques, datasets and threats is presented in Figure 6. We used the STRIDE model [112], a well-known model to explore threats in the cyber domain, to analyse threats in this area of research. The proposed techniques for detecting

Table 3: Known Anomalies' Detection Techniques using Cyber Data

Ref	Year	AD	IDS	IPS	Dataset	Duration	Label	SH	IoT	LAN	ADA	ADT	Threats
[24]	2016		✓	✓	Private	7.5h	✓	✓	✓		SL	LR, SVM	D, T
[113]	2017	✓			Private	N-A			✓		USL	ANN	D
[27]	2018	✓	✓		Private	N-A	✓	✓	✓		ST	DFA	S, E
[23]	2019	✓	✓		Private	5w	✓	✓	✓	✓	SL	DT	I, S, T, D
[26]	2019	✓	✓		NSL-KDD	N-A	✓	✓	✓		SL	SMO, SDPN	D, I
[29]	2020	✓	✓		Multiple	N-A	✓		✓		ST	NICTER	D
[25]	2022	✓	✓		DS2OS	N-A	✓		✓	✓	EL	CB	D
[28]	2022	✓			DS2OS	N-A	✓		✓	✓	SSL	VAE, TCN	D
[30]	2022	✓			Multiple	N-A	✓	✓	✓	✓	USL	LOF, OSVM, IF	D

AD: Anomaly Detection, IDS: Intrusion Detection System, IPS: Intrusion Prevention System, Duration: [w: weeks, h: hours, N-A: No information Available], SH: Smart Homes, IoT: Generic IoT, LAN: Generic Network, ADA: Anomaly Detection Approach, ADT: Anomaly Detection Techniques, Threats: [I: Information Disclosure, S: Spoofing, T: Tempering, D: Denial of Service, E: Elevation of Privileges]

known anomalies in cyber datasets address addressing classification problems. A significant portion of the work is presented using supervised learning as the attack data samples are available by generating and capturing in a controlled laboratory environment. This area is overwhelmingly explored using classical machine learning techniques with a few other examples like NN, AE, statistical, ensemble and rule-based approaches. Looking at datasets for the classification of anomalies (attacks), we see that D2SOS is being used more than other publicly available datasets [28, 25]. On the contrary, most researchers perform anomaly detection on their private datasets, which are not available for other researchers to perform bench-marking most of the time. Most of the research in this category focuses on IDS-based anomaly detection with a few presented prevention techniques (IPS) on top of their proposed IDS [24]. There is also limited availability of real-world datasets for this research area; most of the datasets are captured for a limited period in a laboratory, as seen in Table 3, the most extended Duration of the dataset used is five weeks long by [23], in the papers we reviewed.

The datasets are mostly labelled with the corresponding situations; all the datasets contain normal traffic and attack samples. We found Mirai and other DoS/DDoS attacks most common in the available dataset; not only this, but the work with the private dataset also presented these attacks in their papers. Most of the work considered DoS threats with a few exceptions of tempering [23, 24] information disclosure [26, 23], spoofing [23, 27] and elevation of privileges [27]. Most papers focus on a single dataset, but [30, 29] presented their work using multiple datasets.

5.2 Known Anomalies in Physical Data

Physical anomaly detection covers the work related to environmental (in-situ) sensors, either obstructive (cameras, microphones, accelerates, and tags attached to user's body) or unobtrusive (motion sensors, pressure sensors placed on bed or sofa, temperature, humidity and other) sensors [43]. We considered energy consumption data as physical as well due to its nature. Physical sensor datasets are unavailable in large numbers and variety compared to network traffic datasets. Thus, the comparison and bench-marking are not performed on a regular level. A visualisation of physical data-based techniques, datasets and threats is presented in Figure 7.

Anomaly detection using a dataset with physical sensors is an established field of research. Still, when data from IoT devices comes in, it becomes challenging due to the heterogeneity of the data generated by these devices.

In contrast to cyber datasets, most of the datasets in this area of research are generated in real environments due to the necessity of human participation in the physical world. The share of approaches used in this subsection is pretty balanced. Nevertheless, supervised learning has a higher percentage of adaptation by researchers. Under supervised learning approaches most of the techniques are based on classical machine learning in which [58] uses GMM and HMM, [109] applied NB, [62, 22, 61] involved SVM making it most used technique in this category, [46] presented

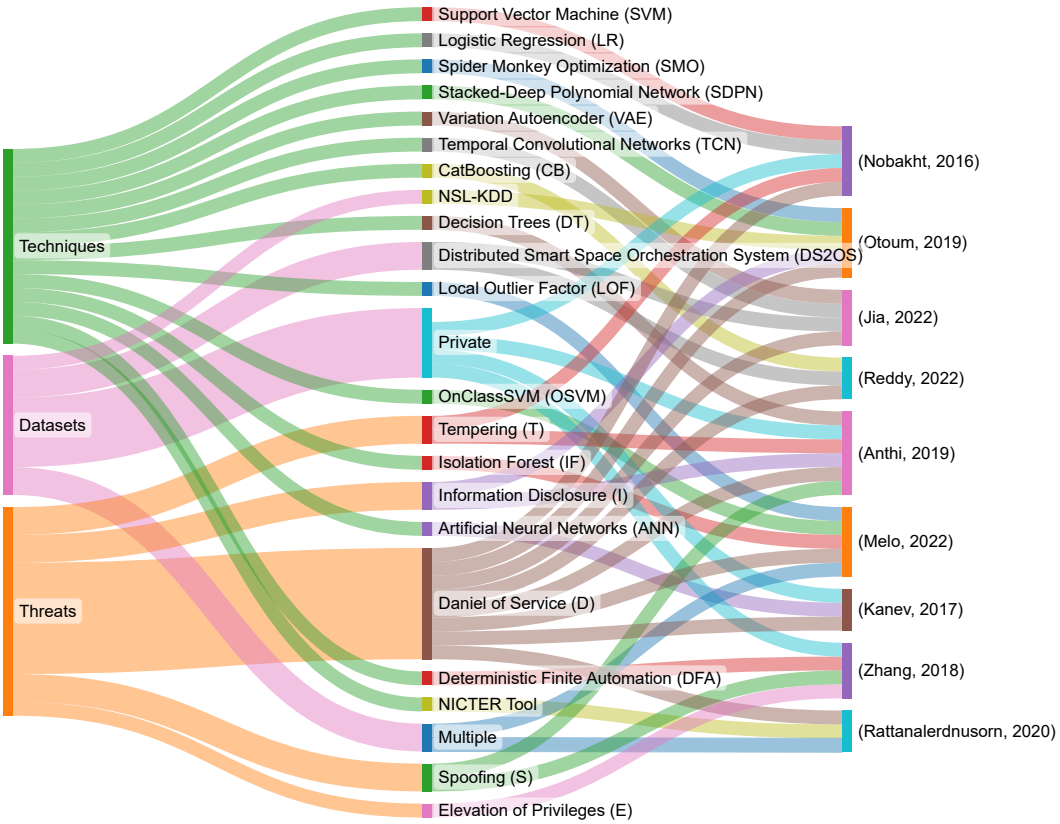

Figure 6: Known Anomaly Detection Techniques, Datasets and Threats for Cyber Data

their DBN based technique, [67] usage CART and [59] implemented their anomaly detection system using RF. In the

Table 4: Known Anomalies' Detection Techniques using Physical Sensors Data

Ref	Year	AD	IDS	Dataset	Duration	Label	Collection	SH	IoT	LAN	ADA	ADT	Threats
[40]	2007	✓		Private	658d		R	✓			ST	Probabilistic	IAD
[41]	2010	✓		Private	49d	✓	R	✓			IT	HD	AB, Alzheimer
[58]	2011	✓		Private	N-A	✓	R	✓			SL	GMM, HMM, SVM	PT, ACC
[109]	2012	✓		Private	N-A	✓	L	✓			SL	NB	D, S
[62]	2013	✓		CASAS	2m	✓	R	✓			SL	SVM	AEC
[69]	2014	✓		Private	N-A	✓	R	✓			H	CBR, FL	AB
[44]	2015	✓		MavHome	N-A	✓	R	✓			USL	DBSCAN	AB
[45]	2015	✓		Private	21d	✓	L	✓			H	MLNR	NDD
[46]	2015	✓		Private	N-A	✓	L	✓			SL	DBN	AB
[114]	2016	✓		Private	21d	✓	L	✓			H	MLN	AB
[49]	2017	✓		MavHome	N-A	✓	R	✓			USL	RNN	Dementia
[66]	2018	✓		Multiple	2h	✓	R	✓			RB	ARM, MLN	ACC, AB
[22]	2019	✓	✓	Private	2h	✓	L	✓	✓		SL	SVM	D
[115]	2019	✓		Multiple	N-A	✓	N-A	✓			SL	CNN, LSTM	Dementia
[65]	2019	✓		Private	4m		R	✓	✓		ST	SMA	ET
[67]	2019	✓		Private	1.5d	✓	R	✓			SL	CART	EH, PL
[59]	2020	✓		Multiple	61d	✓	N-A	✓	✓		SL	RF	I
[116]	2021	✓		CASAS	2-3m		R	✓	✓		SL	H2O	AB
[68]	2022	✓		Pima	NA		R	✓			SL	SVM	AB, Hypertension
[60]	2022	✓		Private	80h	✓	R	✓	✓		USL	YOLO	ACC, Fault
[117]	2022	✓		CASAS	2-3m		R	✓	✓		USL	NARX	AB, Alzheimer
[61]	2022	✓		Private	2w	✓	R	✓	✓		SL	SVM	PT, PI, V, I, PH, EH, T

AD: Anomaly Detection, IDS: Intrusion Detection System, Dataset: [N-A: Information not Available], Duration: [m: months, w: weeks, d: days, h: hours, and mi: minutes, N-A: Not Applicable], Collection: [N-A: Not Applicable, R: Real, L: Lab], SH: Smart Homes, IoT: Generic IoT, LAN: Generic Network, ADA: Anomaly Detection Approach, ADT: Anomaly Detection Techniques, Threats: [D: Denial of Service, S: Spoofing, I: Information Disclosure, T: Tempering, and see to Table 1 for others]

relatively recent work, there is a trend being settled for the exploration of neural networks in this area where techniques like RNN, CNN, LSTM, H2O, etc. are being adopted for anomaly detection [49, 115, 116]. An exception from the usual techniques, we found [60] used YOLO to detect anomalous activities like fall detection using video feeds from

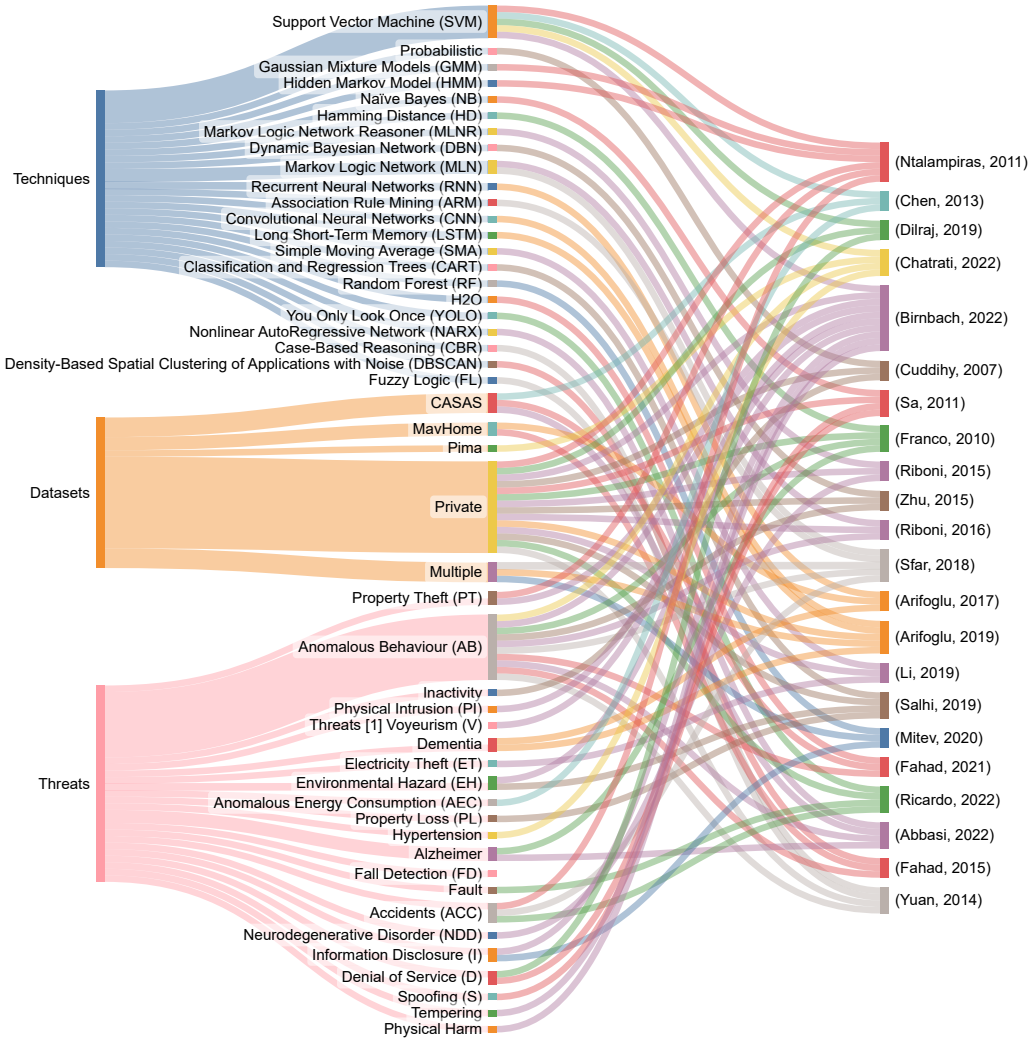

Figure 7: Known Anomaly Detection Techniques, Datasets and Threats for Physical Data

smart robots. Other than supervised approaches, there are some works presented using hybrid [45, 114, 48] for their FABER and its variants, statistical [40, 65], rule-based [66] and information theory-based [41] approaches in this area. A few papers used unsupervised learning to differentiate between the normality and abnormality of particular activities [44, 60, 117].

The usage trend of publicly available datasets in this area is similar to that of cyberspace, as a large percentage of work is performed on private datasets. There is many of research work done on publicly available datasets, e.g. [62, 116, 117] presented their work using the CASAS dataset, [44, 49] performed their technique using MavHome datasets, and [68] adopted Pima dataset for evaluating their presented technique. Some researchers present their work using multiple datasets [66, 115, 59] to evaluate their proposed techniques.

Various types of threats to the physical world can be determined using physical datasets; most of these are related to the physical activities of the subjects, especially older adults living alone, e.g. abnormal movements of people with dementia, hypertension, MCI, or Alzheimer's - can skip medicine, or take it twice, fall by accident, or other environmental hazards, in a smart home environment. External threats are faced by people living in smart homes, like physical theft, physical intrusion, voyeurism, physical harm, energy theft, and many others. Another noticeable aspect of this segment is that only one technique is presented for IDS in physical data by [22]; all other techniques only focus on anomaly detection.

6 Unknown Anomalies Detection

This section discusses the proposed techniques for detecting unknown anomalies in a smart home context. As discussed in the previous section, most of the work in known anomalies' detection focuses on known attacks or known anomalies. The unknown can be determined by behaviour, e.g. detecting anomalous behaviour of a device, system or human. Device or system behaviour can be categorised into two types, i.e., normal behaviour or abnormal, but human behaviour is complex and multi-dimensional.

6.1 Device or System behavior

System behaviour can be classified into two categories: a. Normal Behaviour, b. Abnormal behaviour.

- **Normal Behaviour**: means the system is running normally (as designed), and there is no warning or critical alert. Normal behaviour can be used to train unsupervised models or statistical analysis, and a threshold limit can be defined; the model can predict anomalous behaviour of the system.
- **Abnormal or Anomalous Behaviour**: means that the system is either showing an alert or not working as expected. Abnormal or anomalous behaviour can be in case of compromised [118], under attack [119], remotely controlled [120], attacking other devices as part of a botnet [121], etc. it is difficult or some times not practical to determine abnormal behaviour from system itself because if the system is already compromised the information is also control by adversary or malware like in case of [122] [123].

Various supervised and unsupervised approaches have been presented for system behaviour using cyber and physical observations, which will be discussed in the anomaly detection section below. Human behaviour can be classified in various ways because of intelligence, but system behaviour is binary. The cyber-physical behaviour of a smart home can be determined by analysing both behaviours but via external observations [124]. Let's compare system behaviour with human behaviour using the classification of human behaviour. System behaviour is always moral, overt in some cases, and covert in most cases, involuntary because the system performs actions as programmed (no choice of its own).

6.2 Human Behaviour

Human behaviour can be defined as a super-set of activities performed by an individual, as demonstrated in Figure 8, a timeline consisting of actions and activities of a user entering a smart home while watching TV in the living room. For example, a person wakes up at 8 AM, goes to the washroom for routine hygiene activities and then makes breakfast in the kitchen with coffee before leaving for work. These activities, if performed regularly, can establish the user's behaviour. Figure 8 represents the devices involved in the physical footprint on top of user actions, which can confirm each user's action on the floor map of a home. The user's physical actions (1-9) are shown in orange lines, and the cyber footprint is in blue. Activity number 10 is multiple actions with continuous cyber and physical footprints. Various activities are involved in this example; these activities collaboratively define the behaviour of an object.
It is one of the most important differences that behaviour represents a countless number of a person's performances. In contrast, action means something a person does to complete a specific objective or purpose. Also, the action can often be understood as a movement or gesture of the person towards other people or things. At the same time, the behaviour is a response or a gesture that the person usually does in any circumstance. As exemplified by [125], to illustrate the behaviour, we can say that Adam has terrible behaviour towards his sister, and for the action example, it would be that Adam's action fully scared me. According to [126], when moving from action to activity to behaviour, the degree of semantics increases to evaluate human behaviour with the help of technology. In contrast, if we head in the opposite direction from behaviour to action, the smaller the time-lapse the person will spend. Behavioural psychologists have classified human behaviour into three classes to understand and analyse human behaviour. Every human being can have common types of these classes.

- **Molecular and Moral Behavior** Molecular behavior occurs suddenly or without thinking [127]. For example, closing the eye without thinking when something is about to hit the eye. Moral behaviour occurs after the thinking process and is the opposite of molecular behaviour, for example, changing the direction of the road when there is a closed road [127].
- **Overt and Covert Behaviour** Overt behaviour is visible and occurs outside of a human being, for example, by pressing a key on the keyboard. Covert behaviour is not visible and occurs inside the human being, which is the opposite of overt behaviour [128], for example, thinking the word "stop".
- **Voluntary and Involuntary Behaviour** Voluntary behaviour depends on human want, for example, walking. Involuntary behaviour occurs naturally and without thinking, as opposed to voluntary behaviour, for example,

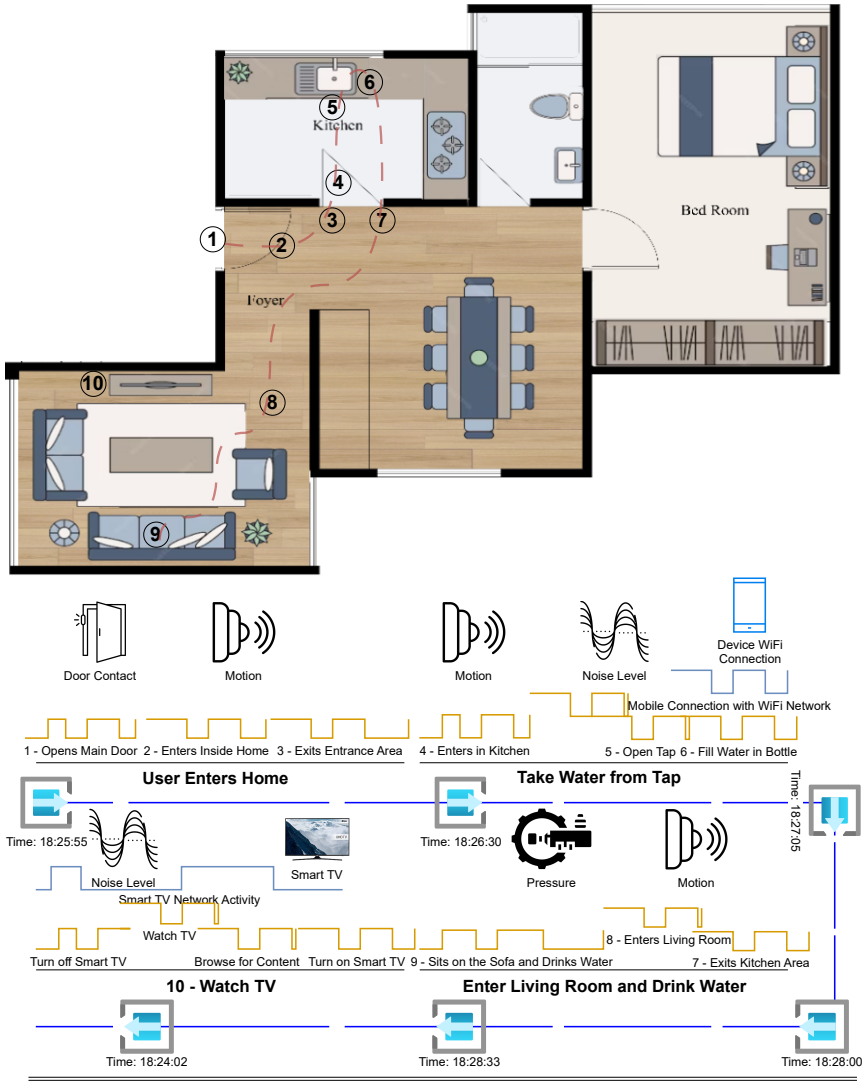

Figure 8: The relationship between behaviour, activity, and action in a person's daily routine.

breathing air. It is difficult but possible to control such behaviours through education, training, and experience [129].

Moreover, it can be said that a person's behaviour consists of several activities that a person performs. These activities consist of one or more actions in a pattern. Detecting anomalies in devices, systems, or human behaviour requires large datasets, as the activities can be performed in different sequences, circumstances, or time variations.

6.3 Unknown Anomalies in Cyber Data

Since the cyber domain is well established in known anomaly detection due to the availability of various attack samples in enormous quantities along with somewhat similar behaviour of network packets at the time of attacks, there is comparatively less work found for unknown anomaly detection in this area. Figure 9 provides an idea of this area's threats, datasets and techniques. The proposed techniques in this area are based on supervised learning except a rule-based technique proposed by [131]; this work is also distinct from others as the authors proposed a flow-based analysis scheme to save computational resources. Another exceptional work presented by [130] in which the authors proposed a simulation-based abnormal behaviour detection system for smart homes is exceptional to all other presented techniques in known and unknown areas. Table 5 contains a list of cyber-related unknown anomaly detection techniques. A couple of proposed methods are based on classical machine learning using LOF by [133] and a dual technique

Table 5: Unknown Anomalies' Detection Techniques using Cyber Data

Ref	Year	AD	IDS	Dataset	Duration	Label	smart homes	IoT	LAN	ADA	ADT	Threats
[130]	2016	✓	✓	Private	N-A	✓		✓			BSES	AB
[131]	2017		✓	Private	18mi	✓	✓	✓		RB	Flow Based	T
[132]	2019	✓	✓	Private	7d	✓	✓	✓		SL	GRU	D, I
[133]	2020	✓	✓	Private	12h	✓	✓	✓	✓	SL	LOF	I, D
[134]	2020	✓	✓	IoT-Botnet	7w	✓		✓		SL	DT, RF	D, I

AD: Anomaly Detection, IDS: Intrusion Detection System, Duration: [w: weeks, d: days, h: hours, mi: minutes, N-A: Not Applicable], SH: Smart Homes, IoT: Generic IoT, LAN: Generic Network, AP: Anomaly Detection Approach, ADT: Anomaly Detection Techniques, Threats: [I: Information Disclosure, S: Spoofing, T: Tempering, D: Denial of Service, E: Elevation of Privileges, and AB: Anomalous Behaviour]

comparison using DT and RF by [134]. The datasets used in this research segment were found to be from 18 minutes to 7 weeks long, whereas [130] was proposed without any dataset. Most researchers used their private dataset, except [134] presented their work using the IoT-Botnet dataset. The threats considered in this area mainly were Denial of Service and Information Disclosure with a couple of exceptions where [130] work was for abnormal behaviour and [131] was looking at Tempering.

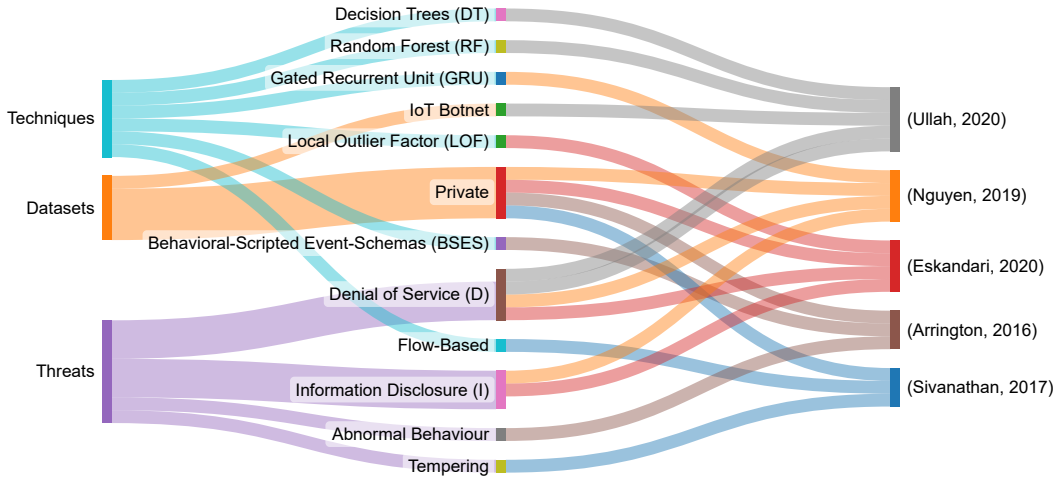

Figure 9: Unknown Anomaly Detection Techniques, Datasets and Threats for Cyber Data

6.4 Unknown Anomalies in Physical Data

Traditionally, cyber systems behaviour can be determined by analysing network traffic, but in IoT or smart home environments, it is not practical or feasible [135]. For example, if IoT is receiving temperature value from a cloud application using a TLS connection, the network traffic analysis will show only the one connection packet (encrypted) being received; let's say the packet contains a 4-byte float value. It will be impossible to determine the temperature value in the packet as it can be from 1.2E-38 to 3.4E+38. Now imagine that this value will instruct the heating system to increase or decrease the temperature in the room, and another actuator is configured to open the windows in case the temperature is above 40.50 degrees centigrade.

All communication in network traffic will describe this behaviour as normal because the cyber domain has limitations. It is pertinent to establish smart homes' secondary sensing mechanisms isolated from the network to understand better how a system's behaviour in the context of cyber-physical systems can be determined. So, physical sensing data is vital in achieving anomaly detection tasks in a smart home. We have noticed that research in unknown anomaly detection is more focused on physical sensing data than cyber data. Figure 10 provides an idea of this area's threats, datasets and techniques. Detecting unknown anomalies in physical datasets is tricky as the outcome of most of the presented work is looking for a shift in normal behaviour (training dataset) to be considered an anomaly. This shift can be novel behaviour of the human and system. Most of the studies presented under this segment use unsupervised machine learning or statistical approaches except only a couple of techniques we found are based on supervised learning [57, 50], both of these techniques use CASAS dataset for evaluation of the proposed work. The majority of work in this area is focused on abnormal behaviour detection, with only a few targeting Dementia [57, 42], whereas [42] also considered patients of Alzheimer in their work. Table 6 contains the list of work published under this segment. All the work in this segment only focused on anomaly detection without IDS. There is a noticeable aspect of unknown anomaly detection that the

Table 6: Unknown Anomalies' Detection Techniques using Cyber Data

Ref	Year	Dataset	Duration	Label	Collection	SH	IoT	LAN	ADA	ADT	Threats
[42]	2012	Private	1y		R	✓			USL	RNN	Dementia, Alzheimer, AB
[43]	2012	MavHome	2m		R	✓			USL	SOM	AB
[63]	2016	Private	1y		R	✓			ST	PCA, FRBS	AB
[47]	2016	Private	4m	✓	R	✓			ST	P	AB
[50]	2017	CASAS	N-A	✓	R	✓			SL	HMM, NB, SVM	AB
[51]	2018	Multiple	N-A	✓	N-A	✓	✓		ST	P	Fault
[136]	2018	Private	5ys	✓	L	✓				GM	AB
[52]	2019	Private	N-A		L	✓	✓		USL	AE, OSVM	AB
[54]	2019	MavHome	21d		R	✓			USL	MLP, LSTM	AB
[56]	2021	Private	3w	✓	R	✓	✓		SA	SA	AB, S, T, E, D
[57]	2022	CASAS	N-A	✓	R	✓	✓		SL	SVM	Dementia

Duration: [ys: years, y: year, m: month, w: weeks, d: days, h: hours, and mi: minutes], Collection: [N-A: Not Applicable, R: Real, L: Lab], SH: Smart Homes, IoT: Generic IoT, LAN: Generic Network, AP: Anomaly Detection Approach, ADT: Anomaly Detection Techniques, Threats: [S: Spoofing, T: Tempering, D: Denial of Service, E: Elevation of Privileges, AB: Anomalous Behaviour, FTD: Fault Detection]

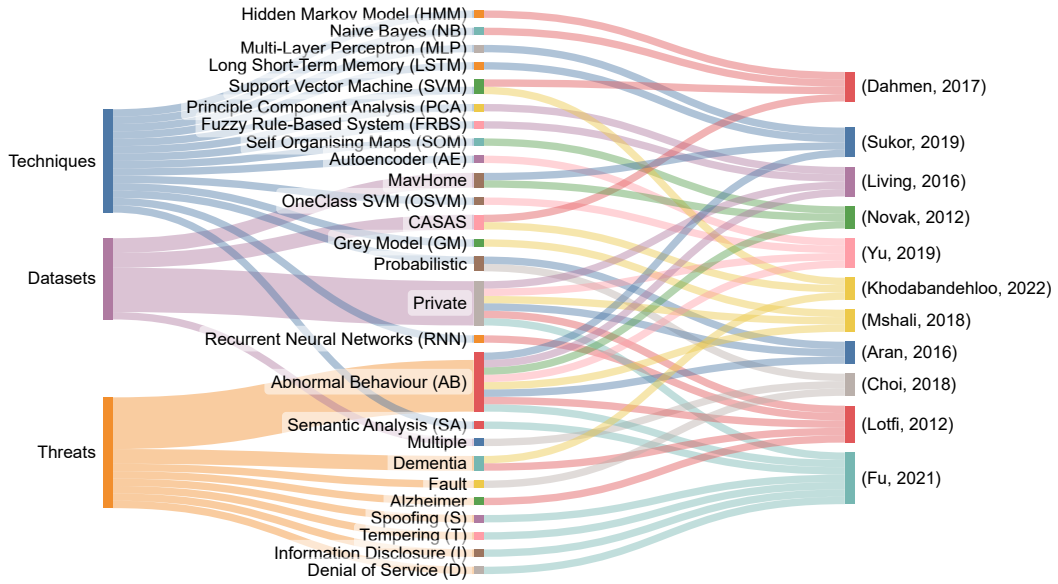

Figure 10: Unknown Anomaly Detection Techniques, Datasets and Threats for Physical Data

Duration of the datasets is much longer than other segments; the minimum dataset used by [54] for 21 days, and the maximum dataset collected over five years by [136]. Similar to known anomalies in physical data, this segment also has more real home data collection examples; only a couple of papers show laboratory-based data collection [136, 52] with only one paper with multiple datasets [51]. Another pertinent aspect is that all the work in this area focuses on sensors in smart homes, e.g. only motion, PIR, and environmental sensors with or without IoT data. Authors of [56] introduce a new type of technique in which the authors used semantic analysis to detect anomalous behaviour of people living in smart homes; their focus for the proposed work was not only anomaly detection but also look for spoofing, tempering, elevation of privileges and denial of service threats. An other another exceptional work was presented by [51] to detect faults in the system using a probabilistic technique. This section has a balanced mix of classical machine learning and neural network techniques. There is an example, under CML approaches, of using SOM by [43], an HMM, NB and SVM-based work by [50], [57] uses only SVM, and [52] using OSVM. For neural networks, we can see techniques from [42] using RNN, [52] presenting their comparison of AE with SVM, and [54] proposed MLP and LSTM for detecting unknown anomalies in smart homes.

7 Anomaly Detection Datasets

The accuracy of anomaly detection mostly depends on behaviours that are frequent and predictable [137], which reflects the importance of datasets. This section discusses publicly available datasets and smart home-related research focused on anomaly detection. Table 7 compares all discussed datasets based on type, research area, applications, release year, capture period, location, setup type and label.

- **CASAS** project [138] is widely used in smart home research. The first dataset was published in 2009, but the authors continuously added new datasets from various homes in this project. There are 66 datasets available in the project, with the last dataset published in 2021. The datasets contain data streams from various environmental sensors in smart homes.
- **NSL-KDD** dataset [139] is an upgrade of the KDD99 dataset [140] due to its poor evaluation properties for anomaly detection. The NSL-KDD dataset has a few improvements in KDD9,9, such as removing redundant and duplicate records in train and test sets.
- **MavHome** Project different datasets [137] from real smart homes, collected using motion sensors in both Home and Lab settings named MavPad and MavLab, respectively.
- **UNSW-NB15** dataset [141] was published by UNSW Canberra; it was in by Cyber Range Laboratory using the IXIA PerfectStrom tool. The dataset contains more than two million records in RAW traffic (PCAP files,) including normal and nine types of attack behaviour in the network.
- **IoTPOT** presents multiple datasets [142] based on Honeypot for detecting cyber threats in IoTs. The datasets contain network traffic in the existence of a honeypot in the network. This dataset is used in anomaly detection research in the general IoT domain and smart homes.
- **REFIT** [143] is an aggregated and appliance-level energy consumption dataset from 2013 to 2014 from 20 houses with 1 to 6 occupants. It was published in 2017. Data dataset was collected from houses in the Loughborough area in the U; these houses were constructed at different times between 1850 and 2005.
- **IoT-Sentinel** [144] is a dataset collected in a laboratory that contains packet capture of devices while setting up. There were 31 devices used in the dataset; 27 were different types, and four were divided into two types, which are the most applicable in a smart home. It is useful for fingerprinting common commercial IoT devices for a smart home.
- **N-BaIoT** [145] is a labelled attack dataset captured during botnet attacks on nine commercially available IoT devices affected by Mirai and BASHLITE. The devices used for this dataset are primarily used in a home, i.e., doorbell, webcam, baby monitor, and security camera. The dataset contains 7,062,606 instances and 115 different attributes; it can be used with both classification and clustering techniques for anomaly detection.
- **CSE-CIC-IDS2018** dataset [146] contains network captures of 7 types of attacks, including multiple DDoS attacks. The network infrastructure is based on 50 machines; the university has five departments with 420 devices and 30 servers. The main focus of releasing this dataset is anomaly detection in networks in general. Many researchers have used it in IDS and abnormal event detection in smart homes.
- **DS2OS** (Distributed Smart Space Orchestration System) dataset [147] captured in a virtual IoT network environment, it contains labelled normal and abnormal network packet captures from various IoT devices like light controllers, motion sensors, thermometers, washing machines, and smart doors etc. The dataset has more than 357K data points, of which 347K are normal and almost 10K are abnormal.
- **IoT host-based** datasets proposed by [148] were collected using an RPi, which emulates different IoT devices, e.g. multimedia devices or surveillance cameras. The dataset contains network traffic and hardware utilisation logs during normal and DDoS attack scenarios.
- **IoT Network Intrusion** dataset [149] is based on two commercially available IoT devices and usual network devices, e.g. smartphones and laptops. The dataset was captured using aircrack-ng with monitor mode enabled; all attacks were generated using Nmap software except the Mirai attack. There are 42 raw network packets available in the dataset, including different attacks on the devices and network and normal traffic.
- **BaT-IoT** is an upgrade to UNSW-NB15 dataset [150] with focus on IoT devices. This dataset was captured in the same lab, the Cyber Range Lab of UNSQ Canberra. The dataset contains network captures in PCAP format with 72 million records of various DoS and other network-based attacks on an IoT-based network.
- **IoT Analytics** dataset [151] contains network traffic captures from a living lab containing 28 IoT devices, including lights, cameras, motion sensors, smart plugs, health monitors and other appliances. The dataset is six months long and inflated synthetically by the authors.
- **MQTT-IoT-IDS2020** (Message Queuing Telemetry Transfer Internet of Things Intrusion Detection System 2020) [152] published in 2020, is a labelled dataset collected using a synthetic network containing twelve different sensors, an MQTT broker, a simulated camera, and an attacker. The dataset consists of five scenarios, including normal traffic and four attacks. This dataset is also considered to be used for smart home research due to the type of devices and protocol used for capturing it.

Table 7: Datasets for Smart Home Anomaly Detection Research

Ref	Name	Phy	NT	IoT	SH	Gen	Known	AD	IDS	Year	Captured	Setup	Label
[138]	CASAS	✓		✓	✓			✓	✓	2009	2009-21	L	
[139]	NSL-KDD		✓			✓	✓		✓	2009	2009	L	✓
[141]	UNSW-NB15		✓			✓	✓	✓	✓	2015	2015	L	✓
[142]	IoTPOT		✓	✓			✓	✓	✓	2015	2015-2022	L	✓
[143]	REFIT	✓			✓		✓	✓		2017	2013 - 2014	R	
[144]	IoT-Sentinel		✓	✓	✓		✓	✓	✓	2017	2016	L	✓
[145]	N-BaIoT		✓	✓	✓		✓	✓	✓	2018	N-A	R	✓
[146]	CSE-CIC-IDS-2018		✓			✓	✓	✓	✓	2018	2018	R	✓
[147]	DS2OS		✓	✓			✓	✓	✓	2018	2018	S	✓
[148]	IoT host-based datasets		✓	✓			✓	✓	✓	2018	2018	L	✓
[149]	IoT Network Intrusion		✓	✓	✓		✓	✓	✓	2019	N-A	L	✓
[150]	Bot-IoT		✓	✓			✓	✓	✓	2019	2019	L	✓
[151]	IoT Analytics		✓	✓	✓		✓	✓	✓	2019	2019	L	✓
[152]	MQTT-IoT-IDS2020		✓	✓	✓		✓	✓	✓	2020	N-A	S	✓
[153]	IoT-23		✓	✓	✓		✓	✓	✓	2020	2018-2019	L	✓
[154]	GHOST-IoT		✓		✓	✓	✓	✓	✓	2020	2019	L	✓
[155]	IoTID20		✓	✓			✓	✓	✓	2020	2020	L	✓
[156]	TON-IoT		✓	✓			✓	✓	✓	2020	2020	L	✓
[157]	MQTTset		✓	✓			✓	✓	✓	2020	2020	L	✓
[158]	CUBRE	✓	✓	✓	✓		✓	✓	✓	2023	2020	R	✓

Phy: Physical Sensors, NT: Network Traffic, SH: Smart Home, Gen: General Network, AD: anomaly detection, IDS: Intrusion Detection, Year: First Year of Release, Setup: [R: Real, L: Laboratory, S: Synthetic]

- **IoT-23** [153] is a labelled malware traffic dataset which was captured in a laboratory consisting of non-infected real IoT devices, i.e., Amazon Alexa, Philips Hue devices, and Somfy smart door lock. There are 20 malware and three normal traffic capture, connection log and summaries capture using Wireshark and Zeek. This dataset is also considered for smart homes due to the involved devices being mostly used in an indoor inhabitant environment.

- **GHOST-IoT** (Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control) dataset [154] is collected using various network interfaces. This dataset is one of its kind because of the variety of network interfaces adapted to capture the traffic. The dataset was captured focusing on the smart home network environment, which can be reflected by ZigBee, Bluetooth and RF interfaces.

- **IoTID20** is an intrusion detection-focused dataset generated from the previously discussed IoT Network Intrusion Detection dataset [155]. The authors of this dataset significantly improve it to detect attacks like DoS, Mirai, MiTM and other scanning attacks.

- **TON_IoT** is also published by UNSW Canberra, the focus of this dataset is Industrial IoT (IIoT) and Industry 4.0 [156], but some researchers have used it for anomaly detection and IDS for smart homes[159] [160].

- **MQTTset** is a network traffic dataset [157] focusing on MQTT packets from various IoT devices. The dataset contains packets from Temperature, Light, Humidity, Motion, CO2, Smoke, Fan speed, Door lock, and fan sensors. The dataset was captured from a two-room setup scenario where the broker, in the centre, is exposed to a malicious node. The frequency of packets ranges from fixed or periodic, 60 to 3600 seconds.

- **CUBRE** dataset is recently published by [158]; it contains data from all aspects of a smart home, i.e. network traffic, environmental sensors, smart devices, and energy consumption. The dataset was captured during October and November 2022 and published in 2023. The core novelty of this dataset is that it contains data from all types on a shared singular timeline, which can help researchers look at both cyber and physical anomalies in smart homes.

8 Considerations and System Design Framework

This section discusses considerations and provides a decision-making framework to design and implement an anomaly detection system for a smart home environment. Three main questions need to be considered before addressing the problem. This requires understanding all potential systems that presumably become part of the system. Figure 11 presents a graphical representation of our design framework.

8.1 Data Source

It is pertinent to consider data sources for anomaly detection in smart homes, as every activity is based on certain sensors and types of network activity. For example, [61] presented a comprehensive table of sensors and activities relation in their presented work on anomaly detection. We have discussed sensor usage for activities and anomalies in Sections 3.1.2 and 3.1.1 in detail.

8.2 Network Element for Anomaly Detection

Similar to antivirus and intrusion detection schemes, which perform host-based or network-based detection, smart homes-related anomalies in the presence of multiple smart devices with computational resources introduce novel network elements where anomalies can be detected in an active environment.

- **External Cloud Infrastructure (ECI)** As most model generation uses high-performance computing systems, the anomaly detection process can also be implemented on ECI, online and offline. In the online scenario, the required data streams should be sent to the ECI in almost real-time; in the case of offline, the data can be captured, stored and later passed through an anomaly detection engine.
- **Smart Hub Devices (SHD)** IoT Hubs or Smart Hubs are the central point for all devices in the network; all types of anomalies can be detected at this point. There are a couple of issues to be looked into, i.e., Smart Hubs are well-resourced, network availability is mandatory, and devices must be connected via a smart hub (some off-the-shelf IoT devices connect to the cloud service directly via encrypted sessions).
- **Directly connected IoT devices (DCD)** Devices on the same wireless network can sniff the traffic between IoT devices and cloud services (or smart hubs). There is a limited possibility of determining the data because most devices use webhooks via TLS for communicating with servers. Information like size, time and other context of packets can be monitored. These observations depend on network availability between the devices.
- **Cross Over Observation (COO)** Devices can observe each other in a local environment/network. This type of observation can be both cyber and physical as the devices are in the same wireless network, and network connectivity is required for the cyber perspective of this type of observation. Another major issue about cross-over observations is that the IoT devices are already resource-constrained, so having instances of anomaly detection for other devices is limited while running primary device operations in parallel.
- **Energy Consumption Monitoring (ECM)** Abnormal power consumption detection can also help look for anomalies in IoT devices but only applies to appliances or always-on power devices. IoT devices and sensors can be off-grid (battery-operated).
- **Secondary External Observation (SEO)** External low-cost sensors can be deployed in a smart home using a different communication medium to observe only physical channels. These sensors can be off-the-shelf micro-controllers, e.g., Arduino, ESP32, or single-board computers (SBC) like Raspberry Pi. In this case, the observation device must be physically presence in the environment. These devices can feed a second opinion on the physical environment and perform anomaly detection locally (on edge) to reduce overall response time in case of accidents.

8.3 Latency

Early alerts can reduce the resulting loss against attacks, accidents or environmental hazards. Particularly for independently living elderly people, it can be life-saving if anomalies can be detected earlier to allow caregivers time to respond [47]. An anomaly detection algorithm's low latency means that the time to alert is shorter in case an incident occurs [40]. That's why considering latency is one of the core factors in designing an anomaly detection system. There can also be a trade-off between latency and computational resources if the known implications are thoroughly accessed and mitigated by other means. Some researchers in this domain have discussed the time/latency factor in their work, like [61] mentioned their detection took five seconds for most of the events and [42] compares ESN, BPTT and RTRL based on training and testing times and concludes that ESN takes less time with better results. [114] results show that the number of temporal sequences increases execution time with the worst results. Authors of [161] proposed an LSTM-based model to detect anomalies in IoT systems with a latency of 532 milliseconds using a fog device. This area of research for anomaly detection in smart homes is not well explored in academia.

8.4 Computational Resources

Computational resources are mandatory for any task in the cyber systems; resources required for anomaly detection depend on the algorithm and the training dataset on which the model is trained. If the computational resources and

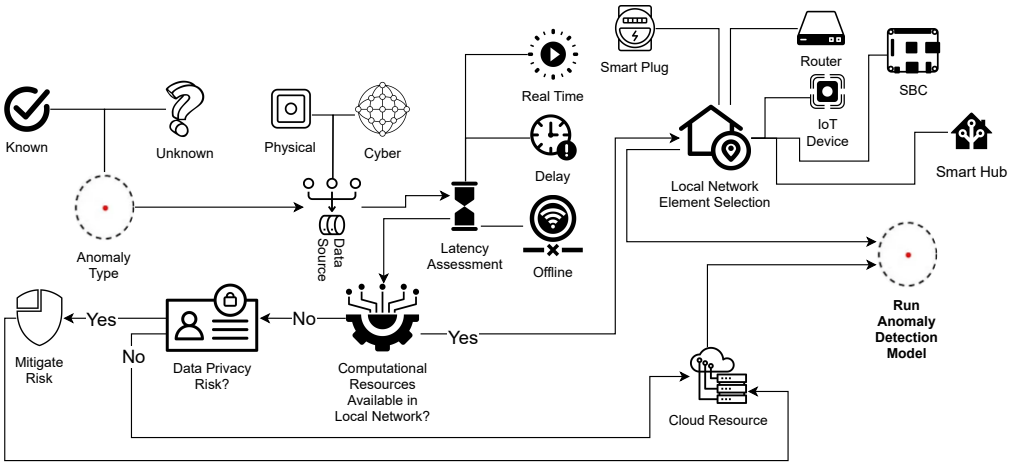

Figure 11: Design Framework for Anomaly Detection in Smart Homes

required data streams are known, then it will be easier to determine which network element can be assigned the task for anomaly detection; the element must have the required computational resources available, and data streams are accessible.

8.5 Data Privacy

One of the main concerns is data privacy regarding private lives or people. Data streams, including network or physical sensors. The data is required in both model generation and anomaly detection implementation processes. If the anomaly detection is done in a local network (inside a smart home), privacy may not be a serious concern. Still, if it is being done on an ECI, it should be covered under regulatory requirements. [61] uses a thermal camera with additional video effects to detect human presence without identification.

9 Future Research Directions

This section discusses some of the important future research directions for anomaly detection in smart homes, including real-time detection of anomalies using edge computing with privacy preservation, context-aware human behaviour modelling for anomaly detection to reduce false positives while increasing the overall accuracy, multimodal data fusion, deep learning or neural networks based techniques, and human feedback integration.

9.1 Real-Time Anomaly Detection

Three main deployment strategies are proposed for anomaly detection models in smart homes [162]. First is edge computing-based anomaly detection, where the model is directly installed on an IoT device in the network for real-time local processing. This requires the device to be equipped with adequate computational resources on board. In the edge-computing-based anomaly detection work proposed by [163], the authors compared computation vs data storage on edge and cloud, in which they have shown that the latency is minimum, i.e. 125.48 milliseconds when computation and storage are performed on the edge device with no communication cost, but the latency increases drastically to 6590.53 milliseconds when data is sent to the cloud with additional communication cost. The second strategy can be fog computing-based anomaly detection, where a central node within the smart home, e.g. smart hub, internet router or SBC, processes data from all devices that can be near real-time. The first and second strategies are friendly as smart home residents' data remain local, but on the other hand, computational resources to run anomaly detection models might be limited. The third strategy involves an externally managed (cloud or external third-party) central node for processing. This strategy brings the challenge of privacy preservation as the smart home residents' data has to leave the local network and be stored or processed at a third-party-owned system. The choice depends on IoT devices' resources, network connectivity, security, and design factors. Further research is essential to pinpoint the best approach for specific scenarios.

9.2 Privacy-preserving Anomaly Detection

Privacy-preserving anomaly detection aims to identify unusual patterns in data without compromising the sensitive information of the smart home's residents. There are three key factors to consider. First is anomaly detection, in which the goal is to spot anomalies like security breaches (cyber or physical), user behaviour, or system issues. The second factor is data privacy, which focuses on keeping residents' data under regulatory compliance and individual rights by ensuring that data is not exposed during anomaly detection. The third factor is to maintain a balance between data utility and privacy. A privacy-preserving solution proposed by [163] in which, while discussing the trade-offs between data utility and privacy, the authors presented an honest and curious model to detect anomalies for smart home security. A survey published by [164] states that privacy sensitivity is highest in IoT applications for smart homes when compared with smart grids and smart transportation, and the security goal should be confidentiality for the smart home context. Further research, e.g. data anonymisation for anomaly detection in smart homes, is needed to find the best balance between data privacy and utility.

9.3 Context-Aware Anomaly Detection

Context-aware anomaly detection in smart homes promises enhanced security and functionality by considering the environment and circumstances in addition to sensor or network traffic data. As stated by [165], context-aware systems are not for delivering all information to anyone at any time, but *"the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person"*. Context-awareness boosts detection accuracy while reducing false alarms and optimizes resource use. For example, an illumination or temperature sensor's data stream might differ between weekdays and weekends. This contextual information can be used to deploy two different models for detecting anomalies in light and temperature in each context. Most of the presented approaches in this area are based on cyber aspects, i.e. network traffic and application data [166]; however, authors of [161] presented a context-aware approach for anomaly detection in IoT systems. This area of research is also critical to be explored further.

9.4 Human-in-the-loop for Anomaly Detection

Humans can play a crucial role in the data mining process to implement effective eXplainable Artificial Intelligence (XAI) for anomaly detection [167]. A human user can either be a smart home resident or an analyst who can access the data and provide feedback. It will be invaluable for refining anomaly detection models. Other than that, real-world datasets with human activities from the smart home environments can also play a vital role in training new and better models and evaluating existing techniques. The datasets should also be considered to provide both cyber and physical data from the smart home so that a holistic understanding can be developed and accuracy can be enhanced. Capturing datasets should also consider nycthemeral cycle-based activities in the smart home; it can help design 24/7 monitoring solutions to detect abnormal behaviour in smart homes. Further research may explore novel methods to collect human activity datasets along with receiving and incorporating feedback from smart home residents to enhance accuracy and performance while reducing false positives.

9.5 Multimodal Data Fusion for Anomaly Detection

Data fusion techniques for cyber and physical data sources from smart homes may also be explored in the future as both datasets are of different types or structures, e.g. network traffic, wearable, environmental, or medical. Merging different data sources can be helpful to optimize and enhance the accuracy of anomaly detection models [164]. Despite some work done on data fusion in smart homes [168, 169] limited to fusing univariate data streams from physical sources into multivariate datasets, there is a need for research on fusing physical sources data streams with network traffic. Once the cyber-physical datasets from human activities in the smart home context are available at a reasonable level, bench-marking for such datasets can be done and compared with other techniques and domains. Followed by human feedback integration for refining the accuracy and reducing false positives. There is a need for further research to develop new architectures for anomaly detection models as well as training strategies focused on smart home data.

10 Conclusions

Smart home residents are exposed to new threats due to the additional attacks via cyberspace. There are various types of abnormal behaviours, both in humans and cyber systems. Recognising behaviour cyber-physical systems is the first step, which leads to determining the nature of the behaviour, normal or abnormal. Anomaly detection is being researched widely, but the context-awareness in anomaly detection needs to be addressed for each scenario. Activity and behaviour recognition is critical to detect anomalies in a smart home. We presented a detailed comparison between known and

unknown anomalies and detection techniques. Combining both cyber and physical data further opens ways to look for suitable methods for detecting anomalies in cyber-physical behaviour data in a better way. Anomaly detection can be performed at different stages of devices. Each stage has its limitations and pros/cons.

11 Acknowledgements

This work is partially supported by EPSRC PETRAS (EP/S035362/1), the Building Research Establishment, the GCHQ National Resilience Fellowship and Cardiff University.

References

[1] Abdulmalik Humayed, Jingqiang Lin, Fengjun Li, and Bo Luo. Cyber-Physical Systems Security-A Survey. *IEEE INTERNET OF THINGS JOURNAL*, 2017.

[2] Mohamed Y.E. Simik, Feng Chi, Randa S.I. Saleh, and Abdeldime M.S. Abdelgader. A design of smart diaper wet detector using wireless and computer. *Lecture Notes in Engineering and Computer Science*, 2220:685–689, 2015.

[3] P. Sanchez, S. Ghosh-Dastidar, K. S. Tweden, and F. R. Kaufman. Real-World Data from the First U.S. Commercial Users of an Implantable Continuous Glucose Sensor. *Diabetes Technology and Therapeutics*, 2019.

[4] Neil E. Klepeis, William C. Nelson, Wayne R. Ott, John P. Robinson, Andy M. Tsang, Paul Switzer, Joseph V. Behar, Stephen C. Hern, and William H. Engelmann. The National Human Activity Pattern Survey: A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, 2001.

[5] Dina ElMenshawy and Waleed Helmy. Detection techniques of data anomalies in IoT: A literature survey. *International Journal of Civil Engineering and Technology*, 2018.

[6] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez. A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. *IEEE Communications Surveys and Tutorials*, 2018.

[7] Mohamed Faisal Elrawy, Ali Ismail Awad, and Hesham F.A. Hamed. Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing*, 2018.

[8] Muhammad Fahim and Alberto Sillitti. Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review. *IEEE Access*, 2019.

[9] Javed Asharf, Nour Moustafa, Hasnat Khurshid, Essam Debie, Waqas Haider, and Abdul Wahab. A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions. *Electronics (Switzerland)*, 9(7), 2020.

[10] Badis Hammi, Sherali Zeadally, Rida Khatoun, and Jamel Nebhen. Survey on smart homes: Vulnerabilities, risks, and countermeasures. *Computers and Security*, 2022.

[11] Ruoying Wang, Kexin Nie, Yen Jung Chang, Xinwei Gong, Tie Wang, Yang Yang, and Bo Long. Deep Learning for Anomaly Detection. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020.

[12] Jiyeon Yu, Angelica de Antonio, and Elena Villalba-Mora. Deep Learning (CNN, RNN) Applications for Smart Homes: A Systematic Review. *Computers*, 2022.

[13] Kyle DeMedeiros, Abdeltawab Hendawi, and Marco Alvarez. A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks. *Sensors*, 23(3), 2023.

[14] François De Keersmaeker, Yinan Cao, Gorby Kabasele Ndonda, and Ramin Sadre. A Survey of Public IoT Datasets for Network Security Research. *IEEE Communications Surveys & Tutorials*, 2023.

[15] Zhibo Wang, Defang Liu, Yunan Sun, Xiaoyi Pang, Peng Sun, Feng Lin, John C.S. Lui, and Kui Ren. A Survey on IoT-Enabled Home Automation Systems: Attacks and Defenses. *IEEE Communications Surveys and Tutorials*, 2022.

[16] Raminta Pranckut. Web of Science (WoS) and Scopus : The Titans of Bibliographic Information in Today ' s Academic World. 2021.

[17] Nicola King. Smart Home – a Definition. *Health (San Francisco)*, 2003.

[18] Michael Schiefer. Smart Home Definition and Security Threats. *9th International Conference on IT Security Incident Management and IT Forensics*, 2015.

[19] Shixi Liu, Xiaojing Hu, and Jingming Wang. Hierarchical Modeling Fault-Error-Failure Dependencies for Cyber-Physical Systems. In Zhixiang Yin, Linqiang Pan, and Xianwen Fang, editors, *Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013*, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[20] Shafiq Ul Rehman and Selvakumar Manickam. A Study of Smart Home Environment and its Security Threats. *International Journal of Reliability, Quality and Safety Engineering*, 23(3), 2016.

[21] Malik Nadeem Anwar, Mohammad Nazir, and Khurram Mustafa. Security threats taxonomy: Smart-home perspective. *3rd International Conference on Advances in Computing, Communication and Automation, ICACCA*, 2018.

[22] M. Dilraj, K. Nimmy, and S. Sankaran. Towards Behavioral Profiling Based Anomaly Detection for Smart Homes. In *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2019.

[23] Eirini Anthi, Lowri Williams, Malgorzata Slowinska, George Theodorakopoulos, and Pete Burnap. A Supervised Intrusion Detection System for Smart Home IoT Devices. *IEEE Internet of Things Journal*, 2019.

[24] Mehdi Nobakht, Vijay Sivaraman, and Roksana Boreli. A host-based intrusion detection and mitigation framework for smart home IoT using OpenFlow. *11th International Conference on Availability, Reliability and Security*, '16.

[25] D.K.K. Reddy and H.S. Behera. *CatBoosting Approach for Anomaly Detection in IoT-Based Smart Home Environment*, volume 281. 2022.

[26] Yazan Otoum, Dandan Liu, and Amiya Nayak. DL-IDS : a deep learning – based intrusion detection framework for securing IoT. *Transactions on Emerging Telecommunications Technologies*, (September), 2019.

[27] Wei Zhang, Yan Meng, Yugeng Liu, Xiaokuan Zhang, Yinqian Zhang, and Haojin Zhu. Homonit: Monitoring smarthome apps from encryptedtraffic. *ACM Conference on Computer and Communications Security*, 2018.

[28] Yifei Jia, Yongliang Cheng, and Jun Shi. Semi-supervised Variational Temporal Convolutional Network for IoT Communication Multi-anomaly Detection. *ACM International Conference Proceeding Series*, 2022.

[29] Ekkachan Rattanalerdnusorn, Montida Pattaranantakul, Phithak Thaenkaew, and Chalee Vorakulpipat. IoTDePT: Detecting Security Threats and Pinpointing Anomalies in an IoT environment. *ACM International Conference Proceeding Series*, 2020.

[30] Pedro H A D De Melo, Rodrigo Sanches Miani, and Pedro Frosi Rosa. FamilyGuard : A Security Architecture for Anomaly Detection in Home Networks. *Sensors (Basel)*, 2022.

[31] Jan Höller, Vlasios Tsiatsis, Catherine Mulligan, Stamatis Karnouskos, Stefan Avesand, and David Boyle. M2M to IoT – An Architectural Overview. *From Machine-To-Machine to the Internet of Things*, 2014.

[32] Iván Froiz-Míguez, Tiago M. Fernández-Caramés, Paula Fraga-Lamas, and Luis Castedo. Design, implementation and practical evaluation of an iot home automation system for fog computing applications based on MQTT and ZigBee-WiFi sensor nodes. *Sensors (Switzerland)*, 2018.

[33] Biswajeeban Mishra and Attila Kertesz. The Use of MQTT in M2M and IoT Systems: A Survey. *IEEE Access*, 2020.

[34] Mehdi Bahrami. Cloud computing for emerging mobile cloud apps. In *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*. IEEE, 2015.

[35] Chunqiang Hu, Yuwen Pu, Feihong Yang, Ruifeng Zhao, Arwa Alrawais, and Tao Xiang. Secure and Efficient Data Collection and Storage of IoT in Smart Ocean. *IEEE Internet of Things Journal*, 7(10):9980–9994, 2020.

[36] Pallavi Srivastava and Navish Garg. Secure and optimized data storage for IoT through cloud framework. In *International Conference on Computing, Communication & Automation*, pages 720–723, 2015.

[37] Eko Sakti Pramukantoro, Widhi Yahya, Gabreil Arganata, Adhitya Bhawiyuga, and Achmad Basuki. Topic based IoT data storage framework for heterogeneous sensor data. In *2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA)*, 2017.

[38] Martin Thomson, Elio Damaggio, and Brian Raymor. Generic Event Delivery Using HTTP Push. RFC 8030, 12 2016.

[39] Seonyeong Heo, Seungbin Song, Jong Kim, and Hanjun Kim. RT-IFTTT: Real-Time IoT Framework with Trigger Condition-Aware Flexible Polling Intervals. In *2017 IEEE Real-Time Systems Symposium (RTSS)*, 2017.

[40] Paul Cuddihy, Jenny Weisenberg, Catherine Graichen, and Meena Ganesh. Algorithm to automatically detect abnormally long periods of inactivity in a home. *1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments*, 2007.

[41] C. Franco, J. Demongeot, C. Villemazet, and N. Vuillerme. Behavioral telemonitoring of the elderly at home: Detection of nycthemeral rhythms drifts from location data. *24th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010*, 2010.

[42] A. Lotfi, C. Langensiepen, S.M. Mahmoud, and M.J. Akhlaghinia. Smart homes for the elderly dementia sufferers: Identification and prediction of abnormal behaviour. *Journal of Ambient Intelligence and Humanized Computing*, 2012.

[43] M. Novák, M. Biňas, and F. Jakab. Unobtrusive anomaly detection in presence of elderly in a smart-home environment. In *Proceedings of 9th International Conference, ELEKTRO 2012*, 2012.

[44] Labiba Gillani Fahad and Muttukrishnan Rajarajan. Anomalies detection in smart-home activities. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015.

[45] Daniele Riboni, Claudio Bettini, and Gabriele Civitarese. Fine-grained Recognition of Abnormal Behaviors for Early Detection of Mild Cognitive Impairment. *IEEE International Conference on Pervasive Computing and Communications*, 2015.

[46] Chun Zhu, Weihua Sheng, and Meiqin Liu. Wearable Sensor-Based Behavioral Anomaly Detection in Smart Assisted Living Systems. *IEEE Transactions on Automation Science and Engineering*, 2015.

[47] Oya Aran B, Dairazalia Sanchez-Cortes, and Minh-tri Do. Anomaly Detection in Elderly Daily Behavior. 2016.

[48] Zaffar Haider Janjua, Daniele Riboni, and Claudio Bettini. Towards automatic induction of abnormal behavioral patterns for recognizing mild cognitive impairment. *Proceedings of the ACM Symposium on Applied Computing*, 2016.

[49] Damla Arifoglu and Abdelhamid Bouchachia. Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks. *Procedia Computer Science*, 2017.

[50] Jessamyn Dahmen, Brian L. Thomas, Diane J. Cook, and Xiaobo Wang. Activity learning as a foundation for security monitoring in smart homes. *Sensors (Switzerland)*, 2017.

[51] Jiwon Choi, Hayoung Jeoung, Jihun Kim, Youngjoo Ko, Wonup Jung, Hanjun Kim, and Jong Kim. Detecting and identifying faulty IoT devices in smart home with context extraction. *Proceedings - 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2018*, 2018.

[52] Y. Yu, C. Li, M.A. Jonas, C. Ma, F.H. Shezan, S. Shen, P. Gao, and Y. Tian. Detecting Abnormal Behaviors in Smart Home. In *IEEE 16th International Conference on Mobile Ad Hoc and Smart Systems Workshops*, 2019.

[53] Hemant Ghayvat, Muhammad Awais, Sharnil Pandya, Hao Ren, Saeed Akbarzadeh, Subhas Chandra Mukhopadhyay, Chen Chen, Prosanta Gope, Arpita Chouhan, and Wei Chen. Smart aging system: Uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. *Sensors (Switzerland)*, 19(4), 2019.

[54] A.S.A. Sukor, A. Zakaria, N.Abdul Rahim, L.M. Kamarudin, and H. Nishizaki. Abnormality detection approach using deeplearning models in smarthome environments. *ACM International Conference Proceeding Series*, 2019.

[55] Ryan Heartfield, George Loukas, Anatolij Bezemskij, and Emmanouil Panaousis. Self-Configurable Cyber-Physical Intrusion Detection for Smart Homes Using Reinforcement Learning. *IEEE Transactions on Information Forensics and Security*, 2021.

[56] Chenglong Fu, Qiang Zeng, and Xiaojiang Du. HAWatcher: Semantics-aware anomaly detection for appified smart homes. *Proceedings of the 30th USENIX Security Symposium*, pages 4223–4240, 2021.

[57] Elham Khodabandehloo, Abbas Alimohammadi, and Daniele Riboni. FreeSia: A Cyber-physical System for Cognitive Assessment through Frequency-domain Indoor Locomotion Analysis. *ACM Transactions on Cyber-Physical Systems*, 2022.

[58] Stavros Ntalampiras, Ilyas Potamitis, and Nikos Fakotakis. Probabilistic novelty detection for acoustic surveillance under real-world conditions. *IEEE Transactions on Multimedia*, 2011.

[59] Richard Mitev, Anna Pazii, Markus Miettinen, William Enck, and Ahmad Reza Sadeghi. LeakyPick: IoT Audio Spy Detector. *ACM International Conference Proceeding Series*, 2020.

[60] Glenn. Ricardo and Jason Halim. Towards Autonomous Robot Application and Human Pose Detection for Elders Monitoring. *6th International Conference on Information Technology, Information Systems and ElectricalEngineering*, 2022.

[61] Simon Birnbach, Simon Eberz, and Ivan Martinovic. Haunted House: Physical Smart Home Event Verification in the Presence of Compromised Sensors. *ACM Transactions on Internet of Things*, 3(3), 2022.

[62] Chao Chen, Diane J. Cook, and Aaron S. Crandall. The user side of sustainability: Modeling behavior and energy usage in the home. *Pervasive and Mobile Computing*, 2013.

[63] Sawsan M. Mahmoud, Ahmad Lotfi, and Caroline Langensiepen. USER ACTIVITIES OUTLIERS DETECTION; INTEGRATION OF STATISTICAL AND COMPUTATIONAL INTELLIGENCE TECHNIQUES. *Computational Intelligence*, 2016.

[64] Shubhangi Singh and Rajendra Singh Kushwah. Energy efficient approach for intrusion detection system for WSN by applying optimal clustering and genetic algorithm. *ACM International Conference Proceeding Series*, 2016.

[65] Weixian Li, Thillainathan Logenthiran, Van Tung Phan, and Wai Lok Woo. A novel smart energy theft system (SETS) for IoT-based smart home. *IEEE Internet of Things Journal*, 2019.

[66] H. Sfar, A. Bouzeghoub, and B. Raddaoui. Early anomaly detection in smart home: A causal association rule-based approach. *Artificial Intelligence In Medicine*, 2018.

[67] Lamine Salhi, Thomas Silverston, Taku Yamazaki, and Takumi Miyoshi. Early Detection System for Gas Leakage and Fire in Smart Home Using Machine Learning. *IEEE International Conference on Consumer Electronics*, 2019.

[68] Saiteja Prasad Chatrati, Gahangir Hossain, Ayush Goyal, Anupama Bhan, Sayantan Bhattacharya, Devottam Gaurav, and Sanju Mishra Tiwari. Smart home health monitoring system for predicting type 2 diabetes and hypertension. *Journal of King Saud University - Computer and Information Sciences*, 2022.

[69] Bingchuan Yuan and John Herbert. Context-aware hybrid reasoning framework for pervasive healthcare. *Personal and Ubiquitous Computing*, 18(4):865–881, 2014.

[70] W. T. Smith and W. L. Roberts. Design and Characteristics of Coaxial Cables for Community Antenna Television. *IEEE Transactions on Communication Technology*, 14(3):334–340, 1966.

[71] Xinghai Han and Xiangxin Kong. The designing of serial communication based on RS232. *Proceedings - 2010 1st ACIS International Symposium on Cryptography, and Network Security, Data Mining and Knowledge Discovery, E-Commerce and Its Applications, and Embedded Systems, CDEE 2010*, 2010.

[72] Su Xunwen, Wang Shaopmg, Zhu Dongmei, and Zhu Qishen. RS-485 serial port pseudo-full-duplex communication research and application. *Prognostics and System Health Management Conference*, 2010.

[73] V. Vijaya, Rama Valupadasu, B. Ramarao Chunduri, Ch Kranthi Rekha, and B. Sreedevi. FPGA implementation of RS232 to Universal serial bus converter. *IEEE Symposium on Computers and Informatics*, 2011.

[74] UPBTM Universal Powerline Bus. The UPB System Description, Powerline Control Systems, USA, 2007.

[75] IEEE. IEEE Standard for Ethernet, 1983.

[76] D Schwarz. The Current State of Security in Smart Home Systems. *SEC Consult Vulnerability Lab Vienna*, 2016.

[77] Allan Johnson. Wireless Concepts - Wireless Topologies, 2020.

[78] Béla Genge and Christos Siaterlis. Developing cyber-physical experimental capabilities for the security analysis of the future Smart Grid. *IEEE PES Innovative Smart Grid Technologies Conference Europe*, 2011.

[79] Yasar Majib, Mahmoud Barhamgi, Behzad Momahed Heravi, Sharadha Kariyawasam, and Charith Perera. Detecting anomalies within smart buildings using do-it-yourself internet of things. *Journal of Ambient Intelligence and Humanized Computing*, 2023.

[80] Vipin Kumar Varun Chandola, Arindam Banerjee. Anomaly detection. *ACM Computing Surveys*, 14(1):1–22, 2009.

[81] D. Samariya and A. Thakkar. A Comprehensive Survey of Anomaly Detection Algorithms. *Annals of Data Science*, 2021.

[82] Stan S., Philip C., and John B. Learning States and Rules for Time Series Anomaly Detection. 2004.

[83] Yufeng Kou, Chang-tien Lu, and Dechang Chen. Spatial Weighted Outlier Detection. 2003.

[84] F Y Edgeworth M.A. XLI. On discordant observations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1887.

[85] D. M. Hawkins. *Identification of Outliers*. 1980.

[86] Nadipuram R. Prasad, Salvador Almanza-Garcia, and Thomas T. Lu. Anomaly detection. *Computers, Materials and Continua*, 2009.

[87] Srikanth Thudumu, Philip Branch, Jiong Jin, and Jugdutt (Jack) Singh. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, 7(1), 2020.

[88] Victoria J Hodge and J I M Austin. A Survey of Outlier Detection Methodologies. (1969):85–126, 2004.

[89] Rahul C. Deo. Machine learning in medicine. *Circulation*, 2015.

[90] Amir Akcay, Samet. Atapour-Abarghouei and Toby P. Breckon. *GANomaly: Semi-supervised Anomaly Detection via Adversarial Training*. Springer International Publishing, 2022.

[91] Filipe Falcão, Anderson Santos, Tommaso Zoppi, Baldoino Fonseca, Andrea Bondavalli, Caio Barbosa Viera Silva, and Andrea Ceccarelli. Quantitative comparison of unsupervised anomaly detection algorithms for intrusion detection. *ACM Symposium on Applied Computing*, 2019.

[92] Mohammad Braei and Sebastian Wagner. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art. 2020.

[93] Yassine Himeur, Khalida Ghanem, Abdullah Alsalemi, Faycal Bensaali, and Abbes Amira. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, (November 2020), 2021.

[94] J.I.I. Araya and H. Rifà-Pous. Anomaly-based cyberattacks detection for smart homes: A systematic literature review. *Internet of Things (Netherlands)*, 22, 2023.

[95] Mahmudul Hasan, Md Milon Islam, Md Ishrak Islam Zarif, and M. M.A. Hashem. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things (Netherlands)*, 7:100059, 2019.

[96] Lincoln Best, Ernest Foo, and Hui Tian. *Utilising K-Means Clustering and Naive Bayes for IoT Anomaly Detection : A Hybrid Approach*. Springer International Publishing, 2022.

[97] John Carter, Spiros Mancoridis, and Erick Galinkin. Fast, lightweight IoT anomaly detection using feature pruning and PCA. *Proceedings of the ACM Symposium on Applied Computing*, 2022.

[98] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[99] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 2015.

[100] Fereshteh Abbasi, Marjan Naderan, and Seyed Enayatallah Alavi. Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset. *5th International Conference on Internet of Things and Applications (IoT)*, 2021.

[101] Amani Alzahrani, Tahani Baabdullah, and Danda B Rawat. *Attacks and Anomaly Detection in IoT Network*. Springer International Publishing, 2021.

[102] Naci Mert Ercan and Mustafa Sert. Anomaly Detection in Smart Home Environments using Convolutional Neural Network. *Proceedings - 23rd IEEE International Symposium on Multimedia, ISM 2021*, 2021.

[103] Ujjwal Sachdeva and Potukuchi Raghu Vamsi. Analysis of Deep Learning Models for Anomaly Detection in Time Series IoT Sensor Data. *ACM International Conference Proceeding Series*, pages 54–62, 2022.

[104] Imtiaz Ullah and Qusay H. Mahmoud. Design and Development of RNN Anomaly Detection Model for IoT Networks. *IEEE Access*, 10:62722–62750, 2022.

[105] Yongliang Cheng, Yan Xu, Hong Zhong, and Yi Liu. Leveraging Semisupervised Hierarchical Stacking Temporal Convolutional Network for Anomaly Detection in IoT Communication. *IEEE Internet of Things Journal*, 2021.

[106] Charu C. Aggarwal. *Teaching Deep Learners to Generalize*. 2018.

[107] Javed Ashraf, Marwa Keshk, Nour Moustafa, Mohamed Abdel-Basset, Hasnat Khurshid, Asim D. Bakhshi, and Reham R. Mostafa. IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities. *Sustainable Cities and Society*, 2021.

[108] Dukka Karun, Kumar Reddy, H S Behera, G M Sai Pratyusha, and Ravikiran Karri. *Ensemble Bagging Approach for IoT Sensor Based Anomaly Detection*. Springer Singapore, 2021.

[109] Muktikanta Sa and Amiya Kumar Rath. A simple agent based model for detecting abnormal event patterns in distributed wireless sensor networks. *ACM International Conference Proceeding Series*, pages 67–70, 2011.

[110] S. Saqaeeyan, H.H.S. Javadi, and H. Amirkhani. A novel probabilistic hybrid model to detect anomaly in smart homes. *CMES - Computer Modeling in Engineering and Sciences*, 2019.

[111] D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino. Kalis - A System for Knowledge-Driven Adaptable Intrusion Detection for the Internet of Things. *International Conference on Distributed Computing Systems*, 2017.

[112] Rafiullah Khan, Kieran Mclaughlin, David Laverty, and Sakir Sezer. STRIDE-based Threat Modeling for Cyber-Physical Systems. 2017.

[113] Anton Kanev, Aleksandr Nasteka, Catherine Bessonova, Denis Nevmerzhitsky, Aleksei Silaev, Aleksandr Efremov, and Kseniia Nikiforova. Anomaly detection in wireless sensor network of the 'smart home' system. *Conference of Open Innovation Association, FRUCT*, 2017-April:118–124, 2017.

[114] Daniele Riboni, Claudio Bettini, Gabriele Civitarese, Zaffar Haider Janjua, and Rim Helaoui. SmartFABER: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artificial Intelligence in Medicine*, 2016.

[115] Damla Arifoglu and Abdelhamid Bouchachia. Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. *Artificial Intelligence in Medicine*, (January 2018), 2019.

[116] Labiba Gillani Fahad and Syed Fahad Tahir. Activity recognition and anomaly detection in smart homes. *Neurocomputing*, 2021.

[117] Haniye Abbasi, Abdolreza Rasouli, and Kenari Mahboubeh. A Model for Identifying the Behavior of Alzheimer's Disease Patients in Smart Homes. *Wireless Personal Communications*, 2022.

[118] Jyoti Deogirikar. Security Attacks inIoT : A Survey. *International conference on IoT in Social, Mobile, Analytics and Cloud*, 2017.

[119] Hanan Mustapha and Ahmed M. Alghamdi. DDoS Attacks on the internet of things and their prevention methods. *ACM International Conference Proceeding Series*, 2018.

[120] Fan Dang, Zhenhua Li, Yunhao Liu, Ennan Zhai, Qi Alfred Chen, Tianyin Xu, Yan Chen, and Jingyu Yang. Understanding fileless attacks on linux-based IoT devices with HoneyCloud. *17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019.

[121] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, Jeffrey Voas, and Ieee Fellow. DDoS in the IoT. *Computer*, 2017.

[122] Sumayah Al-Rabiaah. The 'Stuxnet' Virus of 2010 As an Example of A 'APT' and Its 'Recent' Variances. *21st Saudi Computer Society National Computer Conference, NCC 2018*, 2018.

[123] D. Kushner. The Real Story of Stuxnet. *IEEE Spectrum*, pages 48–53, 2013.

[124] Jie Wan, Michael J. O'Grady, and Gregory M.P. O'Hare. Dynamic sensor event segmentation for real-time activity recognition in a smart home context. *Personal and Ubiquitous Computing*, 19(2):287–301, 2015.

[125] M Fishbein and I Ajzen. *Predicting and changing behavior: The reasoned action approach*. Taylor and Francis, 2011.

[126] B Rezaei, Y Christakis, B Ho, K Thomas, K Erb, S Ostadabbas, and S Patel. Target-specific action classification for automated assessment of human motor behavior from video. *Sensors*, 2019.

[127] William M Baum. Molar and molecular views of choice. *Behavioural Processes*, 2004.

[128] G E Schwartz and J D Higgins. Cardiac Activity Preparatory to Overt and Covert Behavior. *Science*, 9 1971.

[129] P Ekman and M O'Sullivan. From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception. *Behavioral sciences and the law*, 2006.

[130] Briana Arrington, Li Esa Barnett, Rahmira Rufus, and Albert Esterline. Behavioral modeling intrusion detection system (BMIDS) using internet of things (IoT) behavior-based anomaly detection via immunity-inspired algorithms. *2016 25th International Conference on Computer Communications and Networks, ICCCN 2016*, 2016.

[131] Arunan Sivanathan, Daniel Sherratt, Hassan Habibi Gharakheili, Vijay Sivaraman, and Arun Vishwanath. Low-cost flow-based security solutions for smart-home IoT devices. *IEEE International Conference on Advanced Networks and Telecommunications Systems*, 2017.

[132] T.D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.R. Sadeghi. DÏoT: A federated self-learning anomaly detection system for IoT. *International Conference on Distributed Computing Systems*, 2019.

[133] Mojtaba Eskandari, Zaffar Haider Janjua, Massimo Vecchio, and Fabio Antonelli. Passban IDS: An Intelligent Anomaly-Based Intrusion Detection System for IoT Edge Devices. *IEEE Internet of Things Journal*, 2020.

[134] Imtiaz Ullah and Qusay H. Mahmoud. A two-level flow-based anomalous activity detection system for IoT networks. *Electronics (Switzerland)*, 9(3), 2020.

[135] Huichen Lin and Neil W. Bergmann. IoT privacy and security challenges for smart home environments. *Information (Switzerland)*, 7(3), 2016.

[136] Haider Mshali, Tayeb Lemlouma, and Damien Magoni. Adaptive monitoring system for e-health smart homes. *Pervasive and Mobile Computing*, 2018.

[137] V. Jakkula and D. J. Cook. Anomaly detection using temporal data mining in a smart home environment. *Methods of Information in Medicine*, 47(1):70–75, 2008.

[138] Diane J. Cook, Aaron S. Crandall, Brian L. Thomas, and Narayanan C. Krishnan. CASAS: A smart home in a box. *Computer*, 2013.

[139] UNB. NSL-KDD Dataset, 2009.

[140] University of California (Irvine). KDD'99 Dataset, 1999.

[141] Nour Moustafa and Jill Slay. Intrusion Detection systems. *2015 Military Communications and Information Systems Conference (MilCIS)*, pages 1–6, 2015.

[142] Yin Minn Pa Pa, Shogo Suzuki, Katsunari Yoshioka, Tsutomu Matsumoto, Takahiro Kasama, and Christian Rossow. IoTPOT: Analysing the Rise of {IoT} Compromises. In *9th USENIX Workshop on Offensive Technologies*, Washington, D.C., 8 2015. USENIX Association.

[143] David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data*, 2017.

[144] Markus Miettinen, Samuel Marchal, Ibbad Hafeez, N. Asokan, Ahmad Reza Sadeghi, and Sasu Tarkoma. IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT. *Proceedings - International Conference on Distributed Computing Systems*, 2017.

[145] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici. N-BaIoT-Network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, 2018.

[146] UNB. CSE-CIC-IDS2018 Dataset, 2018.

[147] F X Aubet and M O Pahl. DS2OS traffic traces, 2018.

[148] Hugo Bezerra, Vitor da Costa, Victor Turrisi, Augusto Martins, Sylvio Ricardo Barbon, Miani Rodrigo, and Bruno Bogaz Zarpelão. Providing IoT host-based datasets for intrusion detection research. *Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg)*, 2018.

[149] Hyunjae Kang, Dong Hyun Ahn, Gyung Min Lee, Jeong Do Yoo, Kyung Ho Park, and Huy Kang Kim. IoT network intrusion dataset, 2019.

[150] Nickolaos Koroniotis, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems*, 2019.

[151] Arunan Sivanathan, Hassan Habibi Gharakheili, Franco Loi, Adam Radford, Chamith Wijenayake, Arun Vishwanath, and Vijay Sivaraman. Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. *IEEE Transactions on Mobile Computing*, 2019.

[152] Hanan Hindy, Ethan Bayne, Miroslav Bures, Robert Atkinson, Christos Tachtatzis, and Xavier Bellekens. Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study Dataset). *Lecture Notes in Networks and Systems*, 2021.

[153] Sebastian Garcia; Agustin Parmisano; Maria Jose Erquiaga. IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set], 2020.

[154] Marios Anagnostopoulos, Georgios Spathoulas, Brais Viaño, and Javier Augusto-Gonzalez. Tracing your smart-home devices conversations: A real world iot traffic data-set. *Sensors (Switzerland)*, 2020.

[155] Imtiaz Ullah and Qusay H. Mahmoud. *A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks*, volume 12109 LNAI. Springer International Publishing, 2020.

[156] Nour Moustafa. A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. *Sustainable Cities and Society*, 72(May):102994, 2021.

[157] Ivan Vaccari, Giovanni Chiola, Maurizio Aiello, Maurizio Mongelli, and Enrico Cambiaso. Mqttset, a new dataset for machine learning techniques on mqtt. *Sensors (Switzerland)*, 2020.

[158] Yasar Majib, Mohammad Alosaimi, Andre Asaturyan, and Charith Perera. Cyber-Physical Anomaly Detection in Smart Homes, 2023.

[159] Tahani Gazdar. A New IDS for Smart Home based on Machine Learning. *Proceedings - 2022 14th IEEE International Conference on Computational Intelligence and Communication Networks, CICN 2022*, pages 393–400, 2022.

[160] Pooja Anand, Yashwant Singh, Harvinder Singh, Mohammad Dahman Alshehri, and Sudeep Tanwar. SALT: transfer learning-based threat model for attack detection in smart home. *Scientific Reports*, 2022.

[161] Rozhin Yasaei, Felix Hernandez, and Mohammad Abdullah Al Faruque. IoT-CAD: Context-Aware Adaptive Anomaly Detection in IoT Systems through Sensor Association. *IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers, ICCAD*, 2020-Novem, 2020.

[162] Asif Rahim, Yanru Zhong, Tariq Ahmad, Sadique Ahmad, Paweł Pławiak, and Mohamed Hammad. Enhancing Smart Home Security: Anomaly Detection and Face Recognition in Smart Home IoT Devices Using Logit-Boosted CNN Models. *Sensors*, 23(15), 2023.

[163] Shagufta Mehnaz and Elisa Bertino. Privacy-preserving real-time anomaly detection using edge computing. *Proceedings - International Conference on Data Engineering*, 2020.

[164] Wenxiu Ding, Xuyang Jing, Zheng Yan, and Laurence T. Yang. A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Information Fusion*, 2019.

[165] Gerhard Fischer. Context-aware systems: The 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, pages 287–294, 2012.

[166] Rozhin Yasaei and Mohammad Abdullah Al Faruque. Context-Aware Adaptive Anomaly Detection in IoT Systems. In Sudeep Pasricha and Muhammad Shafique, editors, *Embedded Machine Learning for Cyber-Physical, IoT, and Edge Computing: Use Cases and Emerging Challenges*, pages 177–200. Springer Nature Switzerland, Cham, 2024.

[167] Grégory Smits, Marie-jeanne Lesot, Véronne Yepmo Tchaghe, and Olivier Pivert. PANDA:Human-in-the-Loop Anomaly Detection and Explanation. *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2022.

[168] Yu-liang Hsu, Po-huan Chou, Hsing-cheng Chang, Shyan-lung Lin, Shih-chin Yang, Heng-yi Su, Chih-chien Chang, Yuan-sheng Cheng, and Yu-chen Kuo. Design and Implementation of a Smart Home System Using Multisensor Data Fusion Technology. *Sensors*, 2017.

[169] Lan Zhang, Henry Leung, and Keith C C Chan. Information Fusion Based Smart Home Control System and Its Application. *IEEE Transactions on Consumer Electronics*, 2008.