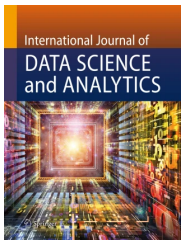


International Journal of Data Science and Analytics

Publishing model

Hybrid

Submit your manuscript

Editorial board

Aims and scope

Journal updates

Overview

The International Journal of Data Science and Analytics is a pioneering journal in data science and analytics, publishing original and applied research outcomes.

Focuses on fundamental and applied research outcomes in data and analytics theories, technologies and applications.

Promotes new scientific and technological approaches for strategic value creation in data-rich applications.

Encourages transdisciplinary and cross-domain collaborations.

Strives to bring together researchers, industry practitioners, and potential users of data science and analytics.

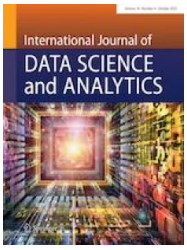
Addresses challenges ranging from data capture, creation, storage, retrieval, sharing, analysis, optimization, and visualization.

Editor-in-Chief
João Gama

Impact factor
2.4 (2022)

Submission to first decision (median)
17 days

Downloads
257,345 (2022)



Latest issue

October 2023
Volume 16, Issue 4

Special Issue: Theoretical and Practical Data Science and Analytics: DSAA 2022

View all volumes and issues →

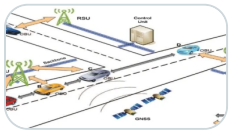
Latest articles

Robust and optimal alignment of high-dimensional data using maximum likelihood estimation through a random sample consensus framework

Regular Paper | Open access | 16 January 2024

Part of 1 collection:
Applications

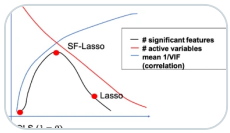
Anonymity and security improvements in heterogeneous connected vehicle networks



Regular Paper | 16 January 2024

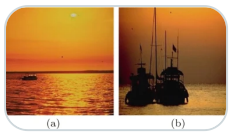
Part of 1 collection:
Innovative Hardware and Architectures for Ubiquitous Data Science

A stable model for maximizing the number of significant features



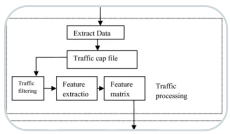
Regular Paper | Open access | 16 January 2024

Robust multi-label feature learning-based dual space



Regular Paper | Open access | 13 January 2024

Enhancing smart home device identification in WiFi environments for futuristic smart networks-based IoT



Regular Paper | 07 January 2024

Part of 1 collection:
Innovative Hardware and Architectures for Ubiquitous Data Science

View all articles →

This journal has 119 open access articles →

Journal updates

CfP: Innovative Hardware and Architectures for Ubiquitous Data Science

Submission Deadline: 10 September 2023

Guest Editors: Dr. Faheem Khan, Dr. Umme Laila, Dr. Muhammad Adnan Khan.

CfP: CCF BigData conference Journal Track on 'Data Science in China'

CfP: Learning from Temporal Data

Submission Deadline: 17 November 2023

Guest Editors: João Mendes-Moreira, Joydeep Chandra, Albert Bifet

CfP: DSAA'2023 Journal Track on Theoretical and Practical Data Science and Analytics

Submission Deadline: April 15th, 2023

Guest Editors: Feida Zhu, Bin Yang, Wei Wei

View all updates →

Journal information

Electronic ISSN	Print ISSN
2364-4168	2364-415X

Abstracted and indexed in

Baidu

CLOCKSS

CNKI

CNPIEC

DBLP

Dimensions

EBSCO Discovery Service

EI Compendex

Emerging Sources Citation Index

Google Scholar

INSPEC
Japanese Science and Technology Agency (JST)
Naver
OCLC WorldCat Discovery Service
Portico
ProQuest-ExLibris Primo
ProQuest-ExLibris Summon
SCImago
SCOPUS
TD Net Discovery Service
UGC-CARE List (India)
WTI AG
Wanfang

For authors

Submission guidelines →
Language editing services ↗
Ethics and disclosures →
How to publish with us →
Open Access fees and funding →

Contact the journal →

Collections and calls for papers →

Language quality checker

Get your manuscript edited for free →

Use our pre-submission checklist →

Avoid common mistakes on your manuscript.

This journal's calls for papers →

Collections this journal is participating in.

Sign up for alerts →

Get notified when new articles are published.

Explore

Articles →

Volumes and issues →

Collections →

International Journal of Data Science and Analytics

Publishing model: **Hybrid**

Back to Overview

Editors

Cervero, Richard Chbeir, Taotao Cai, Olawande Daramola, Wei Du, Souad Ghazouani, Zhimin Gao, Jialin Song, Michele Starnini, Christoph Weisser, Qinfen Wang, Warwick Graco, Marouene Kachroudi, Maram Hasanain, Ahmad Hashemi, Dengzhao Hong, Panagiotis Karampelas, Kaicheng Yang, Jie Zhou, Mucahid Kutlu, Yanchi Liu, Elio Mansour, Patrick Cheong-Iao Pang, Cheng Peng, Brian Schwartz, Sana Sellami, Gautam Kishore Shahi, Mattia Samory, Zeinab Noorian, Kevin Labille, Sabri Allani, Lucile Sautot, Guansong Pang, Francesca Pratesi, Sai Kumar Popuri, Shengsheng Qian, Paula Raissa, Enayat Rajabi, Aghiles Salah, Shamima Mithun, Sha Lu, Zhigang Lu, Golshan Madraki, Queen Nguyen, Mihir Narayan, Chuanren Liu, Dominique Laurent, Ki Yong Lee, Rina Kumari, Yawen Zheng, Shengming Zhang, Ping Zhang, Zhilin Zhao, Taoufik Yeferni, Ming Yin, Sihong Xie, Hua Yuan, Wenhui Zeng, Atikur Khan, Marwa Herzi, Azza Harbaoui, Abderrazek Jemai, Siyu Huang, Xiaodi Huang, Fujiao Ju, Songlei Jian, Mingzhe Wang, Pengyang Wang, Venu Madhav Tammali, Yorgos Tsitsikas, Damiano Spina, Chang Su, Masood Ghayoomi, Amir Mohammad Fathollahi-Fard, Antonio Ferrara, Ahmed Dridi, Alessandra Teresa Cignarella, Oliver Chi, Yuyue Chen, Veronika Batzdorfer, Michael Bewong, Riadh Bouslimi, Shun Cao, Mete Celik, Mohammad Alian Nejadi, Salim Amri, Jeff Ansah, Liesbeth Allein, Ahmed Hamed, Samira Zad, Alireza Rezvanian, Ali Ayadi, Wojtek Buczynski, Murat Alper Basaran, Hu Cao, Ronald Denaux, Pierpaolo D'Urso, Faezeh Ensan, Hamidreza Esmalifalak, Kirill Fedyanin, Mohamed Ferah, Nadia Soudani, Ruofei Shen, Iraklis Varlamis, Lennart van de Guchte, Chenxu WANG, Jianwu Wang, Levin Wiebelt, Valerio Grossi, Mayank Jobanputra, Adel Jebali, Xiuyi Jia, Mourad Khayati, Ashraf Kamal, Li Yang, Snezana Lawrence, Alina Lazar, Majd Latah, Hao Liu, Mahmood Neshati, Md Sarowar Morshed, Athanasios Salamanis, Pablo Sánchez Pérez, Frederic Andres, Jose María Alvarez Rodríguez, Bernd Amann, Manoj Apte, Sabeur Aridhi, Dennis Assenmacher, Gamal Attiya, Ayoub Bagheri, Ritwik Banerjee, Sattam Al-Matarneh, Özgür Akgün, Belhassen Akrout, Malak Abdullah, Svetlana Abramova, Raksha Agarwal, Mahima Agumbe Suresh, Mohammad Akbari, Yonghua Cen, Vitor Cerqueira, Ricardo Cerri, Wei Cao, Sahil Chelaramani, Alycia Carey, Maaz Amjad, Ruichu Cai, David Camacho, Paula

Branco, Mongi Boulehmi, David Biesner, Alejandro Bellogín Kouki, Lamjed Ben Said, Manel Ben Sassi, Nadav Beno, Arup Baruah, Cesare Bernardis, Paidamoyo Chapfuwa, Dimitris Chatzopoulos, Oliver Chi, Davide Ciucci, Etienne Côme, Giuliano Cornacchia, Luciano Costa, Dipankar Das, Ravi Doddavaram, Sana Fakhfakh, Ahmed Eassa, Baris Erman, Jacques Fize, Francky Fouedjio, Fabíola Fernandes, Fethi Ghazouani, Georgios Siolas, Yu Song, Yu Song, Mikolaj Stanek, Julia Maria Struß, Stan Szpakowicz, Jaya Sreevalsan-Nair, Mohamed Amine Ferrag, Francesca Spezzano, Yanjie Fu, Kenichi Fukui, Guojun Gan, Lilia Georgieva, Dilip Kumar Sharma, Manik Sharma, Pavel Shcherbakov, Lifeng Shen, Tsugawa Sho, Anu Shrestha, Eduarda Silva, Nicolas Turenne, Ayad Turky, Antonela Tommasel, Panagiotis Traganitis, Ha Xuan Tran, Damian Trilling, ANUSUA TRIVEDI, Qiang Tang, Dan Taninecz Miller, Mingfei Teng, Souza Tharsis, Friedhelm Victor, Priyesh Vijayan, Marco Viviani, Nguyen Vo, Guifeng Wang, Yue Wang, Yifang Wei, Ziqi Wei, Edgar Weippl, Jinjin Guo, Bin Wu, Gramoz Goranci, Cyril Goutte, Navneet Goyal, Junbeom Hur, Dongmin Hyun, Gary Holness, Wissem Inoubli, HARIPRIYA HARIKUMAR, Loni Hagen, Marwan Hassani, Harry Halpin, Shuchu Han, Teruaki Hayashi, Shoujin Wang, Delia Irazu Hernandez Farias, Chaker Katar, Ravindra Khattree, Krishna Kant, Chandrika Kamath, Armin Kirchknopf, Valery Kirzner, Gerhard Klassen, Benjamin Kille, Samaneh Khoshrou, Xintao Wu, ZHANGKAI WU, Jing Yuan, Dazheng Zhang, Haijun Zhang, Huaiwen Zhang, Mouna Jouini, Chao Yan, Depeng Xu, Fengli Xu, Hongzuo Xu, Jianfei Yu, Xiangyu Zhao, Yu-Dong Zhang, Kaiwen Zhang, Weijia Zhang, Hao Zhong, Arkaitz Zubiaga, Chaoyuan Zuo, Supriya Kumar De, Nayeon Lee, Mark Last, Carson Leung, dichao Li, Zhizhong Liu, Hao Liao, K.V.L.V. Narayanachari, Anh Nguyen, Nuno Moniz, TAHIR MUNIR, Hannah Nithya, Tomonobu Ozaki, Reynier Ortega, Felipe Ortega Soto, Son Mai, Malamati Louta, Vinay Madanbhavi Shashidhar, Xufang Luo, Kumara Swamy M., Massimiliano Luca, Qingxin Meng, Mazin Abed Mohammed, Dana Mckay, David McMeekin, Alejandro Martin, Inês Martins, Albert Ali Salah, Carlos Rojas, Arnaud Sallaberry, Zghal Sami, Esteban Rissola, Neila Rjaibi, Elkosantini Sabeur, Diego Saez-Trumper, Morteza Poyan Rad, D M MOTIUR Rahaman, Joao Resende, Nikos Pelekis, Subba Reddy Oota, Jose E. Ramirez-

Marquez, Diana Portela, Isabel Praça, Daniela Perrotta, Vu Viet Hoang Pham, Danae Pla Karidi, Marco Polignano, Claudio Pomo, Guanqiu Qi, Ana Nogueira, Xiao Pan, Siti Aisyah Panatik, Dimitra Pappa, Pedro Pereira Rodrigues, Maria Pedroto, Paritosh Parmar, Lucia C. Passaro, Arindam Paul, Javier Sánchez Junquera, Mouna Selmi, Kalyani Selvarajah, Apurbalal Senapati, Raquel Sebastiao, Akrem Sellami, Mokhtar Sellami, Shaden Shaar

For authors

Submission guidelines →

Language editing services ↗

Ethics and disclosures →

How to publish with us →

Open Access fees and funding →

Contact the journal →

Collections and calls for papers →

Language quality checker

Get your manuscript edited for free →

Use our pre-submission checklist →

Avoid common mistakes on your manuscript.

This journal's calls for papers →

Collections this journal is participating in.

Sign up for alerts →

Get notified when new articles are published.

Explore

Articles →

Volumes and issues →

Collections →

International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-023-00489-3

REGULAR PAPER

Enhancing smart home device identification in WiFi environments for futuristic smart networks-based IoT

Hassan Falah Fakhruldeen[1] · Mohamed J. Saadh[2] · Samiullah Khan[3] · Nur Agus Salim[4] · Naveed Jhamat[5] · Ghulam Mustafa[5]

Received: 25 September 2023 / Accepted: 28 November 2023

Abstract
The recognition of smart home devices within WiFi environments stands as a pivotal focus within contemporary Internet of Things (IoT) security, especially in the context of Futuristic Smart Networks-based IoT. The inherent encryption feature of the 802.11 protocol in WiFi settings renders conventional identification methods, reliant on plaintext traffic patterns, ineffective for IoT devices. Through an examination of the 802.11 protocol, distinctive traits within data frames of various smart home devices are revealed. Building on these insights, this research selects attributes like frame length, frame arrival time, duration, and frame sequence number as salient traffic characteristics. Leveraging an enhanced decision tree CART algorithm, the study achieves robust device identification for smart home devices operating within WiFi environments. Experimental outcomes affirm the method's efficacy by accurately discerning device models, achieving an impressive identification accuracy of 91.3%.

Keywords Device identification · WiFi environments · Futuristic smart networks · IoT Security · Smart home devices · Traffic characteristics

✉ Samiullah Khan
samikhan@aup.edu.pk

Hassan Falah Fakhruldeen
hassan.fakhruldeen@gmail.com

Mohamed J. Saadh
msaadeh@meu.edu.jo

Nur Agus Salim
nuragussalim@uwgm.ac.id

Naveed Jhamat
naveed.jhamat@pugc.edu.pk

Ghulam Mustafa
gmustafa@pugc.edu.pk

1 Computer Techniques Engineering Department, Faculty of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq10011

2 Faculty of Pharmacy, Middle East University, Amman 11831, Jordan

3 Department of Mathematics, Statistics & Computer Science, The University of Agriculture Peshawar, Peshawar, KP, Pakistan

4 Universitas Widya Gama Mahakam Samarinda, Samarinda, Indonesia

5 Department of Information Technology, University of Punjab, Gujranwala Campus, Gujranwala, Pakistan

1 Introduction

The Internet of Things (IoT) has become an important component of the new generation of information technology, with widespread applications in various fields such as smart homes, intelligent healthcare, and smart cities. The quantity and variety of IoT devices have grown exponentially due to increasing market demand. It is projected that by the year 2025, there will be over 75 billion IoT devices connected to the Internet [1], and this number is expected to reach 125 billion by 2030 [2]. Presently, due to the advantages of wide wireless network coverage, strong mobility, and low construction costs, a large number of IoT devices are connected via WiFi networks. Especially in the case of smart home devices, connecting to WiFi allows for quick information exchange between devices, enhancing convenience in daily life.

Most of the research on IoT device identification is based on extracting traffic features at the TCP/IP layers. Author [3] extracted 23 features including protocols, packet sizes, IP addresses, and port numbers from different network layers. Author [4] identified devices by extracting 67 characteristics such as TTL and TCP window size from packet sequences. These methods utilize privileged access to router and network

Published online: 07 January 2024

protocol headers, enabling the extraction of a wide range of protocol feature fields. However, in unfamiliar WiFi environments, the captured traffic consists of encrypted link layer data based on the 802.11 protocol. Among the fastest growing segments of the communications sector today is 802.11g WLAN technology. Naturally, it does this without a network cable and offers constant network access. Workers from home or those who work remotely can build up networks without worrying about how to put cables through homes that were not intended to accommodate network equipment. A collection of specifications known as 802.11 covers computer communications over wireless local area networks (WLANs) operating in the 2.4, 3.6, and 5 GHz bands. For big enterprise wireless systems and household wireless access points, the most popular standards are 802.11a, b, and g. With data transmission rates of up to 54 Mbps, 802.11a is faster than 802.11b. A privacy mechanism called WEP is defined in 802.11 to safeguard connection data that is sent via WLAN. This speaks to the intention of giving wireless LAN users a privacy service like what a traditional LAN's built-in physical security offers. The RC4 symmetric stream cipher with 40-bit and 104-bit encryption keys is used for the WEP encryption. The 802.11 standard does not specify 104-bit encryption keys. However, several wireless AP manufacturers do support them.

Extracting specific fields of IP layers and above is challenging in such cases. Moreover, stability and interference resistance of data transmission in WiFi environments are weak, leading to significant difficulties in device identification. Currently, research on device identification in WiFi environments is relatively limited. Author [5] analyzed the duration field in 802.11 traffic for device identification author [6] used encrypted WiFi traffic's destination address, network name, frame size, and MAC protocol fields as implicit identifiers for wireless devices. Author [7] differentiated devices through temporal analysis of 802.11 probe request frames. Author [8] evaluated features like transmission rate, frame size, media access time, transmission time, and inter-frame arrival time in 802.11 traffic, with transmission time and inter-frame arrival time proving most effective. The author [9] utilized similar hashing algorithms to generate device fingerprints for 802.11 management frames. Author [10] transformed traffic from IoT devices in wireless networks into traffic grayscale images for identification experiments. Given the limited information obtainable from WiFi-based traffic, effectively identifying IoT devices in this environment remains a challenging aspect that requires further research.

Most IoT device identification algorithms primarily employ machine learning techniques, constructing classification models based on feature attributes extracted from network traffic. Author [3] utilized a random forest algorithm to identify devices using a $23 \times N$ feature matrix. The author [11] generated feature vectors with features like TCP window size and payload length and employed a gradient boosting tree algorithm for identification. The author [12] used the J48 decision tree algorithm to identify 23 IoT devices. Author [13] compared identification algorithms including random forest, k-nearest neighbors (KNN), decision tree (DT), and support vector machine (SVM), concluding that random forest and decision tree algorithms excel in recognition rate and speed, respectively. The author [14] used a naive Bayesian optimization algorithm and clustering algorithm for IoT device and non-IoT device identification.

Author [15] employed KNN for identifying 22 IoT devices based on features like average packet length, average arrival time, and packet length. The author [16] utilized an SVM algorithm for IoT device identification using various protocol features and device-specific attributes. Based on existing research, this paper further investigates link layer traffic features of the 802.11 standard under IoT environments by monitoring smart home device traffic under WiFi and using machine learning algorithms for device identification.

However, as the heterogeneous smart home devices connected to WiFi networks rapidly increase, the resulting wireless encrypted traffic becomes increasingly complex and chaotic. This leads to the growth of network complexity and scale, making the management of smart home devices more intricate. Additionally, since smart home devices often transmit sensory information and user privacy data, they are susceptible to attacks, thereby posing a serious threat to the security and privacy of IoT systems. Therefore, achieving smart home device identification in WiFi environments is currently a prominent area of research. Device identification methods can associate information about IoT devices, users, functionalities, and data within the network space. This aids in situational analysis, event tracing, and targeted control of these devices, thereby holding significant importance in addressing the management and security issues of smart home devices.

The Internet of Things (IoT) has become increasingly significant in the dynamic field of information technology, as it has made its presence felt in several sectors such as smart homes, healthcare, and urban environments. Given the anticipated increase in the quantity of Internet of Things (IoT) devices, it is imperative to explore effective methods for device identification. The principal objective of this study is to tackle the issue of discerning Internet of Things (IoT) devices within wireless fidelity (WiFi) networks, with a specific emphasis on the complex smart home setting.

The work makes a valuable contribution by:

1. The objective of this study is to explore and provide reliable methods for accurately identifying Internet of Things (IoT) devices within WiFi networks.
2. This paper aims to examine the intricacies and security concerns associated with smart home setups.

3. Facilitating situational analysis and monitoring of events.
4. Improving the administration and security of smart home devices.

2 Related work

This extensive literature review explores a wide range of advanced subjects encompassing smart buildings, the Internet of Things (IoT), and related disciplines. The study conducted by Ma et al. [17] focuses on the improvement of job engagement in smart offices as a means to enhance staff productivity. In their study, Huseien and Shah [18] conducted a thorough examination of the incorporation of 5G technology to enhance energy management and smart infrastructures, with a specific focus on the context of Singapore. In this study, Khalil et al. [19] present an innovative approach to nonintrusive occupant identification through the use of ultrasonic sensors. The primary objective of this method is to enhance energy efficiency in smart buildings. In their study, Malkawi et al. [20] provide a novel Internet of Things (IoT) architecture that aims to improve data-driven operations and experimentation in the context of smart buildings.

In the realm of healthcare, Gowda et al. [21] proposes a novel integration of the Internet of Things (IoT) and fog computing as a means to transform and enhance the delivery of high-quality industrial healthcare services. In their study, Nauman et al. [22] explore the integration of cognitive intelligence into post-5G networks, with a specific focus on enhancing network efficiency at the MAC layer. The authors Wirtz et al. [23] provide a valuable contribution by developing a comprehensive public Internet of Things (IoT) infrastructure specifically designed for the implementation of intelligent governance applications. In their study, Lee et al. [24] examine the experiences of Seoul and San Francisco to derive significant insights that might inform the creation of efficient frameworks for the development of smart cities.

In addition to the advancements in smart buildings and the Internet of Things (IoT), Bai et al. [25] provide an extensive overview of acoustic-based sensing applications. In their study, Li et al. [26] examine the difficulties and potential opportunities associated with multi-user activity recognition. Reduan and Jamil [27] provide a comprehensive evaluation of application characteristics and traffic requirements within the domain of smart grid communications. The study conducted by Mumtaz et al. [28] focuses on the optimization of energy consumption in smart direct-LTE networks. Rahhal et al. [29] shed light on a common viewpoint about the health of both humans and machines.

In their study, Woźniak et al. [30] propose a type-2 fuzzy logic model to enhance driving support, thereby expanding the existing range of approaches in this field. In their study, Mohanty and Pani [31] propose a novel approach to livestock health monitoring that utilizes neural networks enabled by the Internet of Things (IoT). In their study, Raja and Chakraborty [32] introduce a wearable healthcare system that utilizes Internet of Things (IoT) technology, specifically designed to cater to healthcare needs in distant regions. In their study, Raut et al. [33] provide a novel adaptive event detection system designed specifically for streaming Internet of Things (IoT) data. In their study, Sharma and Mehra [34] conducted a comprehensive examination of the topic of secure communication within unmanned aerial vehicle (UAV) networks.

Zhao et al. [35] investigate the integration of the Internet of Things (IoT) and digital twin technologies as a means to increase safety management. Hou and Bergmann [36] adeptly combine inertial navigation with neural networks. In their study, Adarsh and Kumar [37] propose the use of wireless medical sensor networks as a means to significantly transform the field of e-healthcare. In his work, Nethercote [38] critically examines the concept of platform landlords and their role in exerting control over urban spaces within the context of the digital era. The authors of the study conducted by Lee et al. [39] examine the integration of artificial intelligence (AI) in the healthcare sector for the older population. In conclusion, Sampaio et al. [40] emphasize the significance of autonomous energy management within the context of Fog Computing. This comprehensive compilation enhances our comprehension across several technological disciplines, augmenting our understanding and ramifications.

Rani et al. [41] presented a voice-activated home automation system built on natural language processing (NLP) and artificial intelligence methods. Voice instructions are sent through a cell phone to operate household appliances, and the preset natural language processing medium interprets the commands. The system's application was limited to controlling household appliances; it was not utilized for additional home automation functions like environmental monitoring, motion detection, intruder detection, or control. Jaihar et al. [42] introduced an intelligent smart home automation system that controls lighting, music systems, and other household appliances by performing tasks based on the user's emotional state. To anticipate actions and reduce user engagement, several machine learning algorithms were integrated and utilized to assess the user's wants as well as the environment. Depending on the mood that the machine learning model detects, the house appliances are turned ON or OFF. Their methodology improved domestic energy efficiency.

Khan et al. [43] suggested a real-time algorithm for house monitoring and control, including ambient factors, motion sensors, electrical appliances, and home appliances. The motion sensors' algorithmically derived inferences determined whether to turn on or off the lights. Using the WiFi module, the suggested algorithm was also utilized to track the power usage of different household appliances and to set

off an alert depending on the gas pressure in the house. In an experiment conducted by Popa et al. [44] with two deep neural network models for a smart home, anomalous patterns of energy usage could be identified. A modular framework for gathering, combining, and preserving data in the context of smart homes was developed through the use of cloud computing services. By using less energy, the authors demonstrated how the recommended machine learning techniques improved smart home automation.

The energy-saving strategy for equipment in a smart home setting presented by Machorro-Cano et al. [45] makes use of big data and machine learning approaches. Using methods to identify the home energy usage level by learning user behavior and consumption patterns, the J48 machine learning algorithm and Weka API were used to assure the energy efficiency of the system. Using a smartphone app called HEMS-IoT that the authors created, the system recorded and presented real-time data as well as suggestions for energy-saving measures throughout the house. Their strategy addressed home comfort and the safety of people and gadgets in addition to conserving energy by enabling system users to communicate with their houses and request the necessary IoT service. Singh et al. [46] presented a smart home automation system for managing electrical appliances, and doors, and detecting activity in a house, in addition to monitoring energy use in the home and delivering frequent notifications about it. The device might also notify the user if sensors detect low quantities of gas in a cylinder or the presence of a human. An Arduino Uno board, a Node MCU ESP8266, and IR and LDR sensor modules were used to prototype the system.

3 Smart home identification method based on WiFi data frame features

This paper focuses on studying smart home devices such as home cameras, smart doorbells, smart TVs, smart locks, smart speakers, robotic vacuum cleaners, and smart gateways. It proposes a smart home identification method based on WiFi data frame features, with an overall process depicted in Fig. 1. It consists of three steps: traffic collection, traffic processing, and device identification. The traffic collection module gathers traffic data in WiFi environments, the traffic processing module filters, extracts features, and performs feature subset analysis and representation for smart home devices' network traffic, and the device identification module trains the recognition model, implementing device identification using an improved CART algorithm for decision trees.

Connecting commonplace equipment like sensors and actuators for automation that are included in household appliances is made possible by smart homes. The core of a smart home is the convergence of several technological platforms,

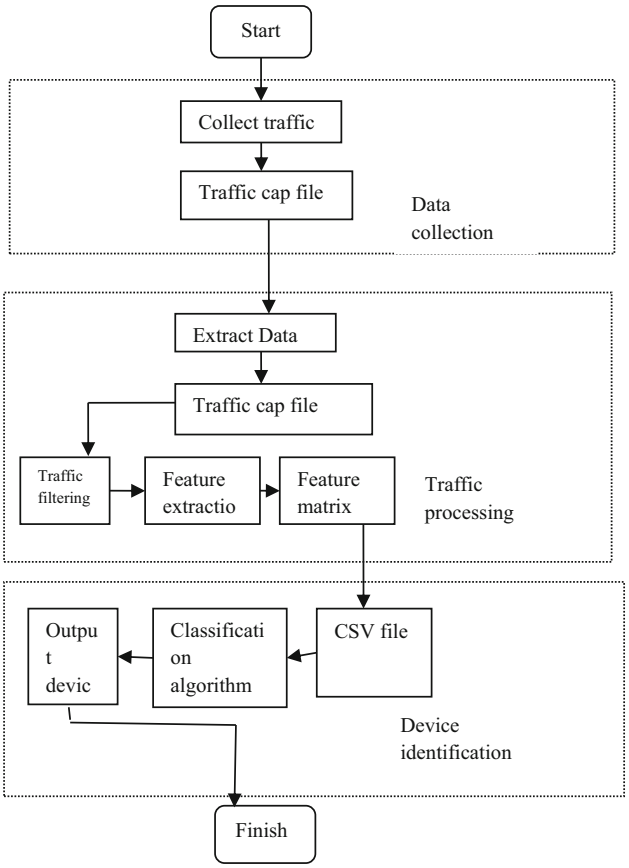


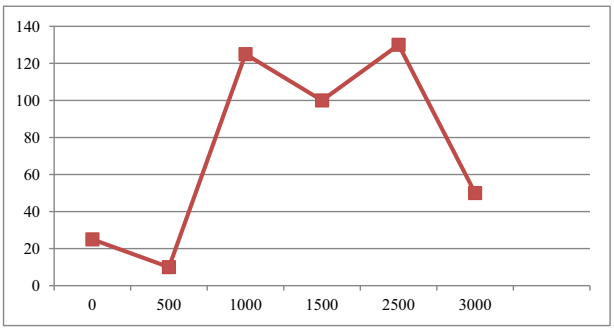
Fig. 1 WiFi data frame-based smart house identification technique

often involving three tiers: the application, network, and perception layers. In addition to acting as an interface between people and the linked items, the perception layer collects information from the environment. Novel and pervasive sensing approaches have been developed in response to the need for an interface that is more user-friendly and pleasant. WiFi sensing, whether active or passive, does away with the need for physical touch by using invisible radio waves to feel the environment. It causes no discomfort since it can do the sensing functions without the user realizing it.

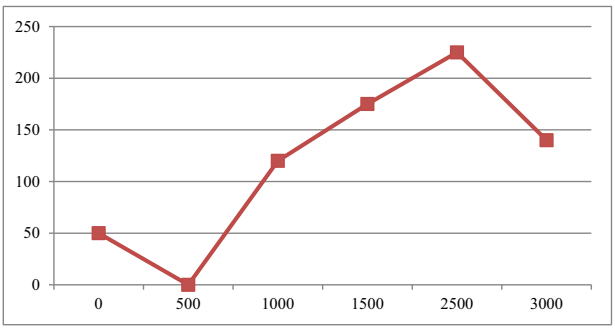
3.1 Data frame feature selection

Under WiFi, 802.11 data frames are divided into three types: management frames, data frames, and control frames. Since control frames have a simple structure and few extractable valid features, and not all devices generate management frames, which also lack universality in extracted features, this paper selects data frame types with rich protocol field information as experimental data.

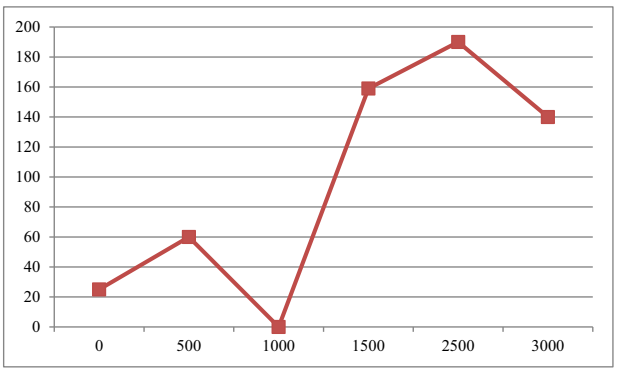
To acquire traffic features of data frames, this paper analyzed traffic from various models of smart home devices. "YiSight C6P" represents camera devices, which maintain



(a) "YiSight C6P" Traffic I/O Graph



(b) "360 Smart Doorbell" Traffic I/O Graph



(c) "Huawei Smart Speaker" Traffic I/O Graph

Fig. 2 Traffic I/O graphs

surveillance and provide real-time alerts for specific captured information. "360 Smart Doorbell" is a smart doorbell device that remains silent after activation until the doorbell is pressed. "Huawei Smart Speaker" represents smart speaker devices, primarily used for playing music, adjusting volume, and switching songs. Using these three types as examples, the paper further analyzes traffic differences among devices of different types and models.

Figure 2 depicts the I/O traffic graphs of the three smart home devices. It can be observed that the "YiSight C6P" device, while in monitoring mode, exhibits activity below 25 frames per second (fps), showing regular fluctuations. Burst traffic, indicating the capture of specific information, occurs at around 100 fps. The "360 Smart Doorbell" device, while in silent mode, transmits at a lower rate of approximately 5 frames per second. Upon pressing the doorbell, burst traffic surges to over 200 frames per second. The "Huawei Smart Speaker" device sends around 50 frames per second cyclically during music playback. Transitioning volume and switching songs result in burst traffic peaking at up to 1500 frames per second. The three devices display distinct trends in the number of frames arriving per second in the I/O graphs, underscoring the significant differences in traffic among different types and models.

Table 1 Distribution statistics of frame lengths

Device frame length	Distributed/%		
	Fluorite C6P	360 Smart Doorbell	Huawei Speaker
0 ~ 39	34.25	22.7	0.2
40 ~ 79	1.35	2.36	1.88
80 ~ 159	57.39	34.67	19.8
160 ~ 319	2.9	0.27	1.14
320 ~ 639	0.75	0.21	41.08
640 ~ 1279	3.36	0.63	0.18
1280 ~ 2559	0	39.17	35.71
Average frame length/byte	100.89	632.9	790.89

Table 1 lists statistical information about frame lengths for the three smart home devices. The "YiSight C6P" device has shorter frame lengths, with most falling in the range of 80 to 159 bytes. The "360 Smart Doorbell" device exhibits higher occurrences in the ranges of 80 to 159 bytes and above 1280 bytes, with an average frame length of 632.90 bytes. The "Huawei Smart Speaker" device's frame lengths are mainly distributed between 320 to 639 bytes and above 1280 bytes, with an average frame length reaching 790.89 bytes. Consequently, there are distinct differences in frame lengths among different devices.

Based on the aforementioned analysis and the comparison of 802.11 data frame traffic from various smart home devices in WiFi environments, this paper extracts 11 features from device traffic, including frame length, frame interval time, frame arrival time, duration, frame sequence number, frame type, frame subtype, transmission direction, retransmission flag, QoS traffic identifier, and data length. The specific descriptions of these features are presented in Table 2.

3.2 Improved decision tree CART algorithm

The CART (Classification and Regression Trees) algorithm for decision trees employs a binary tree structure and a recursive binary partitioning technique. It uses the Gini index

Table 2 Feature descriptions

Frame feature	Feature description
Frame length	The length of the data frame
Interframe time	time interval between two consecutive frames
Frame arrival time	The arrival time of the data frame
Duration	The time the data frame and its acknowledgment frame occupy the channel
Frame sequence number	Data frame sequence control bits, used to reassemble frame fragments and discard duplicate frames
Frame type	Types of 802.11 Data Frames
Frame subtype	Subtypes of 802.11 Data Frames
Transmission direction	DS flag, indicating the transmission direction of the frame BSS and DS
Retransmission flag	Retransmission flag, indicating that the frame is a retransmission frame of the transmission segment
QoS traffic identifier	Types of QoS
Data length	The number of bytes occupied by the data portion of the data frame

to represent the purity of a model, where a smaller Gini index signifies higher purity and better feature representation. The Gini index reduces the logarithmic calculations involved in entropy models, thus lowering computational costs. For classification problems, assuming a given sample E with K classes, where the number of samples in the k-th class is D_k, and the probability of the k-th class is q_k, the expression for the Gini index is shown in Formula (1); while, the Gini index expression for sample E is shown in Formula (2).

$$\text{Gini}(q) = \sum_{k=1}^{K} p_k(1 - p_k) = 1 - \sum_{k=1}^{K} p_k^2 \tag{1}$$

$$\text{Gini}(E) = 1 - \sum_{k=1}^{K} \left(\frac{|D_k|}{|E|} \right)^2 \tag{2}$$

This study improves the Decision Tree CART algorithm through parameter optimization. Parameter optimization aims to minimize the objective function, enhancing the fit between model output and actual data results to achieve higher accuracy and reliability in the final recognition outcome. To reduce interference from experimental sample noise and further enhance recognition accuracy, this study optimizes parameters related to the decision tree's maximum depth, internal nodes, leaf nodes, and minimum impurity decrease for node splitting. These parameter settings are designed to prevent model overfitting and enhance the robustness of the recognition model.

Grid Search CV (Grid Search Cross-Validation) is a commonly used parameter tuning method. It sequentially adjusts parameters within specified ranges, traversing all possible parameter combinations to train models and select the set of parameters that yield the highest accuracy. This study employs Grid Search CV for parameter optimization. However, since Grid Search CV requires traversing all parameter combinations within the given range, it can consume a considerable amount of time and computational resources, especially when dealing with large datasets and numerous parameters. Therefore, this study first conducts score curve analysis for each parameter to identify the approximate optimal parameter range. Subsequently, the Grid Search CV method is used to determine the optimal parameter combination, achieving the recognition model. This approach reduces the computational burden and time cost of grid search while improving the accuracy of device recognition to a certain extent.

4 Experimental analysis

This research focuses on the recognition of smart homes in WiFi environments. Due to the lack of publicly available datasets for mainstream Chinese brands such as Huawei and Xiaomi, this study sets up a practical IoT environment for experimental analysis. The Decision Tree algorithm is optimized to ensure data stability during training, and the effectiveness of extracting data frame features is verified. Through parameter optimization, this work enhances the Decision Tree CART method. To increase the accuracy and dependability of the final recognition result, parameter optimization seeks to minimize the objective function and improve the fit between the model output and real data results. This work improves the decision tree's maximum depth, internal nodes, leaf nodes, and minimal impurity reduction for node splitting to lessen interference from experimental sample noise and improve identification accuracy even more. These parameter configurations aim to keep the recognition model more robust and avoid overfitting. One popular technique for fine-tuning parameters is Grid Search CV (Grid Search Cross-Validation). To train models and choose the combination of parameters that produce the best accuracy, it iteratively modifies parameters within predetermined limits.

For parameter optimization, Grid Search CV is used in this study. To determine the estimated ideal parameter range, this study first analyzes the score curve for each parameter. The best parameter combination is then found using the Grid Search CV technique, which results in the recognition model. With a little increase in device detection accuracy, this method lessens the grid search's computational load and

Fig. 3 Data collection topology

Table 3 Data collection environment configuration

Configuration	Parameter
virtual machine	VMware Workstation 16
virtual machine operating system	Kali Linux 2021.3a
raspberry pie	Raspberry Pi 4
Raspberry Pi OS	Kali Linux 2021.3 rpi4
Wireless network card	ALFA RTL 8812AU
Python	python3.9.0

time cost. The sniffer's wireless port is configured in monitoring mode. To capture all network traffic broadcast over the air in the same channel of the WiFi network, the Linux system's aerodump terminal command is used to define the channel, BSSID, duration, and other parameters. The two sorts of captured data packets are those that occur during the idle time after a smart home device's WiFi connection and those that occur during the live interactions between users and smart home devices. As experimental data samples, data are gathered for one hour during both the idle and interaction phases, and recorded separately. cap files, and then analyzed.

4.1 Traffic data collection

The topology of the data collection setup is shown in Fig. 3. A Raspberry Pi and a virtual machine on a laptop with a wireless USB adapter are used to build a traffic sniffer for data capture. The hardware and environmental configurations are detailed in Table 3.

Specific information about the smart home devices used in the experiment is presented in Table 4. The smart home devices are connected to the test WiFi, and the network environment is configured to keep the devices operational.

The wireless port of the sniffer is set to monitoring mode. The aerodump terminal command in the Linux system is used to specify the channel, BSSID, duration, and other parameters, capturing all network traffic transmitted over the air in the same channel of the WiFi network. Captured data packets are categorized into two types: packets during the idle period after smart home devices connect to WiFi, and packets during real-time interactions between users and smart home devices. Data are collected in both idle and interaction periods for an hour each time, saved separately as.cap files, serving as experimental data samples. Information regarding these data samples is presented in Table 5.

4.2 Traffic feature processing

Experimental samples were collected interactively over one hour, resulting in over 190,000 traffic data entries. The data frame type was extracted using the frame control field, and the traffic was classified and filtered by the Mac address of the smart home devices. For each device's traffic, features such as frame length, frame interval time, frame arrival time, duration, frame sequence number, frame type, frame subtype, transmission direction, retransmission flag, QoS traffic identifier, and data length were extracted frame by frame. Some features were subjected to normalization; the two directions in the transmission direction field were normalized to 1 and 2, missing values in the data length field were set to 0, and missing values in the QoS traffic identifier field were set to 16. The final feature representation consisted of feature vector matrices for different devices.

After extracting data frame features, feature importance was measured, and the feature ranking results are presented in Fig. 4. It can be observed that features like retransmission, type, and subtype have relatively low importance indicators. Consequently, this study selects a subset of 8 features comprising frame length, frame arrival time, duration, frame sequence number, transmission direction, interval time, data length, and QoS traffic identifier as the final feature set for data frame recognition.

4.3 Algorithm parameter configuration

Based on the Decision Tree CART recognition algorithm, parameters were set with a maximum depth single increment of 10, and internal nodes and leaf nodes with a single increment of 1. The scores were computed for different parameter values, and the score curves are plotted as shown in Fig. 5. In the figure, the x-axis represents different parameter values, and the y-axis represents the coefficient of determination S^2 under that value, expressed as in Eq. (3):

$$S^2 = \frac{\sum_i \left(\hat{x}_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}{\sum_i \left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \tag{3}$$

where x represents the actual result and $\hat{x}$ represents the model's predicted result. S^2 is a commonly used metric for evaluating the goodness of fit of a model; the closer its value is to 1, the better the model's fit.

Table 4 List of smart home devices

Serial number	Device name	Physical address	Category
1	Fluorite C6P	C0:E4:34:29:89:09	Camera
2	Fluorite C3W	EC:9C:32:C5:7C:EA	Camera
3	Fluorite C6CN	D4:E8:53:05:89:BB	Camera
4	Fluorite C6C	EC:9C:32:A0:D4:E8	Camera
5	TP-Link IPC55a	7C:B5:9B:E2:D9:7F	Camera
6	Hikvision Dome Camera	00:95:69:D0:49:3E	Camera
7	Xiongmai Robot Camera	7C:A7:B0:4E:F4:2D	Camera
8	Fluorite Doorbell	DC:F5:05:F1:0E:8B	Doorbell
9	360 Smart Doorbell	B2:59:47:00:4E:1B	Doorbell
10	Ding Zero Doorbell	90:E8:68:28:8B:79	Doorbell
11	Millet Doorbell	EC:2E:98:22:94:7D	Doorbell
12	Kim Jong Tv	DC:29:19:64:8A:F8	Television
13	Haier TV Yunos	1C:30:08:67:DD:F5	Television
14	Konka TV	08:38:69:00:2E:48	Television
15	Kaidis Smart Door Lock	F4:CF:A2:F0:67:6D	Door Lock
16	Deschmann Smart Door Lock	70:3A:2D:2B:C1:D6	Door Lock
17	Huawei Speaker	78:85:F4:EC:D0:3C	Speakers
18	Xiaoai Speaker	9C:9D:7E:A6:06:61	Speakers
19	Roborock Sweeping Robot S51	04:CF:8C:F8:D0:DC	Sweeping Robot
20	Lumi Multimode Gateway	54:EF:44:20:0A:17	Gateway

Table 5 Information about data samples

State	The amount of data	File size/106
Stand still	4,935,057	1146.88
Interactive	1,978,838	275

From Fig. 5, it can be observed that the peak of the score curve for the parameter "maximum depth" is around 370, with an $R2$ value of 0.958. This indicates that the optimal parameter range for maximum depth is approximately 360–380. The score curves for the parameters "internal nodes" and "leaf nodes" show a general decreasing trend. The preliminary optimal range for internal nodes is around 2–5, and for leaf nodes, it's around 1–5. Based on these parameter ranges, a grid search was conducted using GridSearchCV, resulting in the optimal parameter set: {'max_depth': 376, 'internal_nodes': 2, 'leaf_nodes': 1, 'min_impurity_decrease': 0.0}.

4.4 Device recognition analysis

4.4.1 Smart home model recognition

The feature matrices of smart home devices were divided into training and testing sets in a 7:3 ratio. A supervised training was performed to generate the recognition model, and the experimental results are presented in Fig. 6 with Table 6. The average accuracy of recognizing 20 different device models from the dataset reached 91.3%, with recognition rates exceeding 85% for 17 device models. Additionally, Table 7 shows a comparison between the test results when frame types are not extracted and when data frame types are extracted. The results indicate that extracting data frame types significantly improves the accuracy of device model recognition.

With a recall of 0.93, 93% of real-world abnormalities are properly identified by the model. The F1 score of 0.91, which represents the harmonic mean of the model's recall and accuracy, serves as a summary of its effectiveness. The model has a strong discriminative ability to distinguish outliers from the overall population, as evidenced by the AUC-ROC value of 0.94. With a precision level of 0.82, 82% of the time a face prediction is accurate. The algorithm properly recognizes 86% of real-world faces, according to a recall value of 0.86. The model's overall effectiveness at recognizing human faces is indicated by its F1 score of 0.84. Furthermore, the AUC-ROC score of 0.90 indicates how well the model classified face occurrences. The model correctly identifies 89% of all occurrences with an accuracy of 0.89. 86 percent of the time, with an accuracy of 0.86, the outliers are, in fact, outliers. With a recall value of 0.91, the model accounts for 91% of true outliers. The F1 score, which stands at 0.88, is an

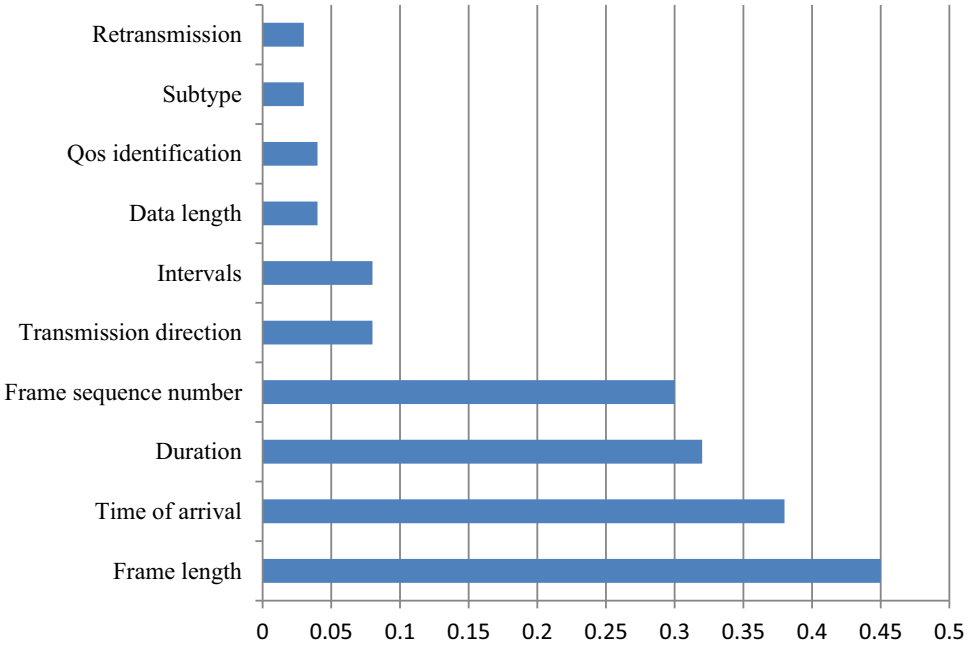


Fig. 4 Feature importance measure

attempt to balance recall and accuracy. A further indication of the model's capacity to differentiate abnormal from ordinary data is its AUC-ROC of 0.92. The model successfully recognizes 83% of the tested faces with an accuracy of 0.83. Eighty percent of the examples with an accuracy rating of 0.80 are faces. Recalling 85% of real-world faces correctly, the model has a 0.85 recall score.

4.4.2 Comparative analysis

The traffic data used in this study represents passive traffic in a WiFi environment, which differs from the plaintext traffic often examined in mainstream research. In reference [5], the recognition of WiFi encrypted traffic using the duration field was tested, yielding an accuracy of 62.2%. Reference [10] utilized encrypted WiFi traffic, directly transforming it into traffic images without extracting frame features. Multiple algorithms were combined to recognize specific IoT device models, and the accuracy of the decision tree algorithm was 78.1%. In contrast, this study achieved a device model recognition accuracy of 91.3% using the proposed method. Additionally, this study tested device type recognition, and the comparative results are detailed in Table 8. The proposed method enhances the recognition rate of IoT devices, validating the effectiveness of extracting data frame features and addressing the issue of smart home device model recognition in situations where router-specific information is inaccessible due to environmental constraints.

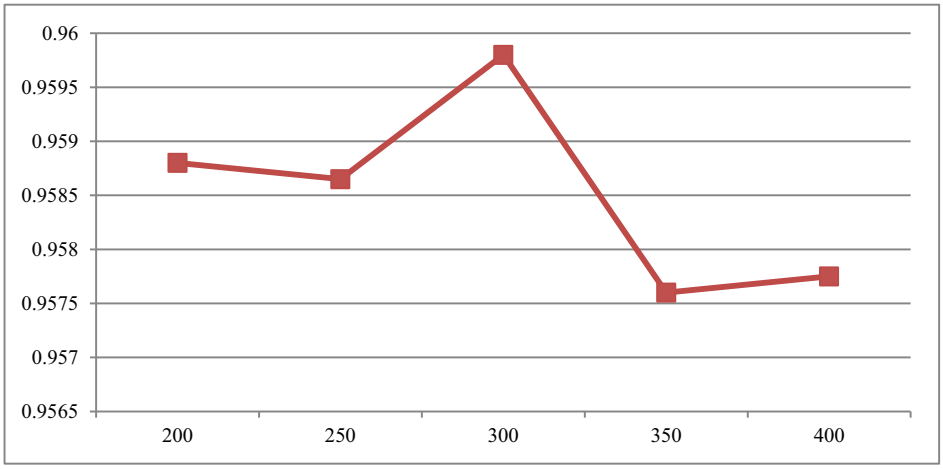
In this experiment, due to limitations in the experimental environment, only 20 different types of smart home devices were tested. In future research, it is intended to apply the method proposed in this study to a wider range of scenarios. This would involve expanding the number of devices in the training model, encompassing a greater variety of types, brands, and models of IoT devices. This expansion would help address the management challenges posed by IoT devices and contribute to enhancing convenience in the network space environment.

4.5 Discussion

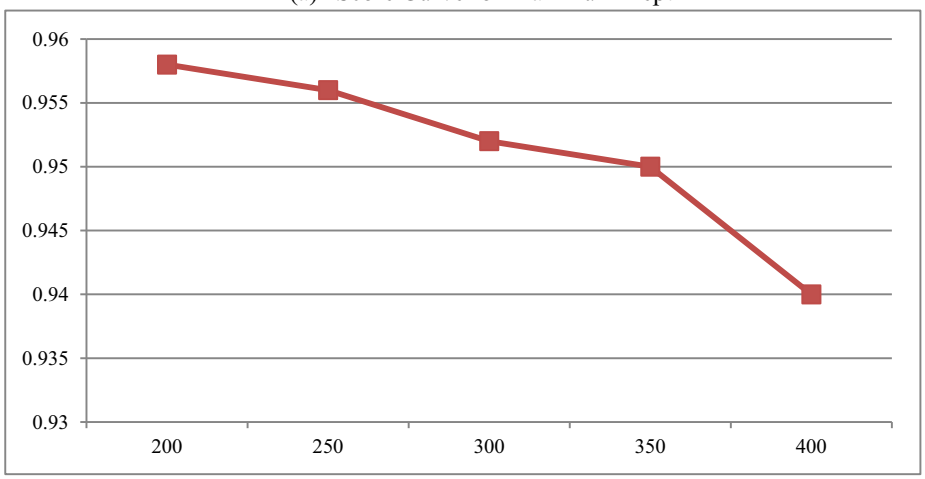
The dataset's 20 distinct device models were recognized with an average accuracy of 91.3%; 17 of the device models had identification rates higher than 85%. The maximum depth single increment of 10 and the single increment of 1 for internal and leaf nodes were the parameters selected based on the Decision Tree CART recognition method. Score curves were presented after the scores were calculated for various parameter values. For each device, feature vector matrices made up the final feature representation. Following the extraction of data frame features, the results of the feature ranking are shown together with a measurement of feature relevance. It is noted that characteristics with relatively low relevance indications include retransmission, type, and subtype.

Because of this, the final feature set for data frame recognition in this study consists of a subset of 8 features: frame length, frame arrival time, duration, frame sequence number, transmission direction, interval time, data length, and QoS traffic identifier. Over 190,000 traffic data entries were produced via interactively gathering experimental samples over one hour. The traffic was categorized and filtered based on the Mac address of the smart home devices, and the data frame

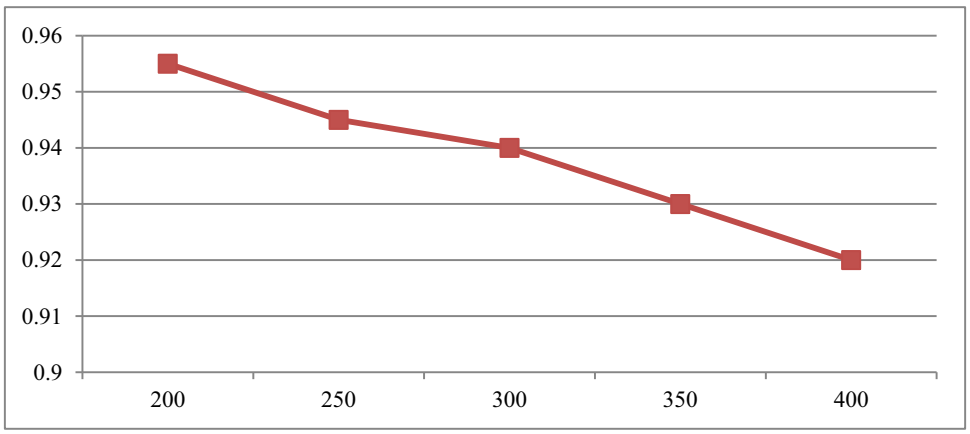
Fig. 5 Scoring curve



(a) Score Curve for Maximum Depth



(b) Score Curve for Internal Nodes



(C) Score Curve for Leaf Nodes

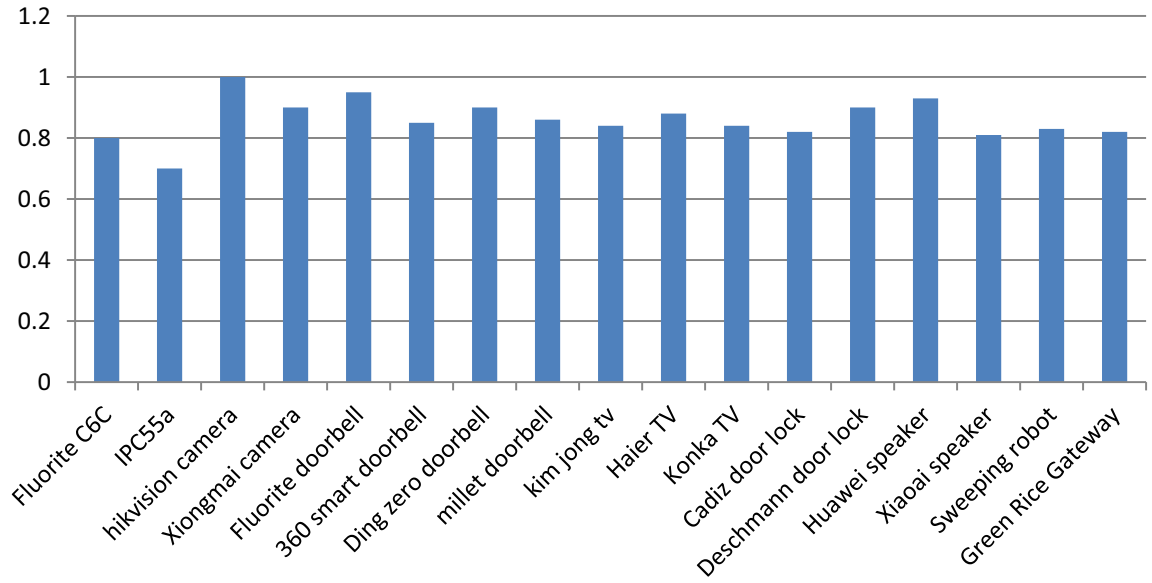


Fig. 6 Different models of device recognition accuracy

Table 6 Different models of device recognition accuracy

Device	Accuracy
Fluorite C6C	0.8
IPC55a	0.7
Hikvision camera	1
Xiongmai camera	0.9
Fluorite doorbell	0.95
360 smart doorbell	0.85
Ding zero doorbell	0.9
millet doorbell	0.86
kim jong TV	0.84
Haier TV	0.88
Konka TV	0.84
Cadiz door lock	0.82
Deschmann door lock	0.9
Huawei speaker	0.93
Xiaoai speaker	0.81
Sweeping robot	0.83
Green Rice Gateway	0.82

type was retrieved using the frame control field. Frame by frame, information about each device's traffic was retrieved, including the QoS traffic identity, transmission direction, frame type, frame subtype, frame length, frame interval time, frame arrival time, duration, frame sequence number, and transmission type.

Table 7 Comparison of recognition with and without data frame type extraction

Frame type	Accuracy	Recall rate	F1-Score
All frames	79	79.4	79.2
Data frame	91.3	91.3	91.3

Normalization was applied to a few characteristics; the two directions in the transmission direction field were set to 1 and 2, and the missing values in the data length and QoS traffic identification fields were set to 0 and 16, respectively. With an R2 value of 0.958, it can be seen that the parameter "maximum depth" has a peak on the score curve of about 370. This suggests that about 360–380 is the ideal parameter range for the greatest depth. A general declining tendency can be seen in the score curves for the parameters "internal nodes" and "leaf nodes." For internal nodes, the preliminary ideal range is around 2–5, while for leaf nodes, it is approximately 1–5. To identify certain IoT device types, many algorithms were merged; the decision tree algorithm's accuracy was 78.1%. On the other hand, this study used the suggested strategy to obtain 91.3% accuracy in device model recognition. This study also evaluated the recognition of device type, and the comparison outcomes are presented in detail. By improving the recognition rate of IoT devices, the suggested approach validates the efficacy of obtaining data frame attributes and solves the problem of smart home device model recognition when environmental restrictions prevent access to router-specific information.

Table 8 Comparison of evaluation indicators

Recognition result	traffic data set	Adapting methods	Accuracy	Recall rate	F1-score
Device specific model	Literature [10]	Literature [10] (DT)	78.1	78.3	78.4
	Proposed Work	Literature [5] (DT)	62.2	37.3	40.1
		Literature [10] (DT)	86.8	90.4	88.6
		The method in this paper	91.3	91.3	91.3
Device specific type	Proposed Work	The method in this paper	88.2	87.1	87.7

5 Conclusion

This paper introduced a smart home device recognition method based on 802.11 data frame features in a WiFi environment, enabling the identification of smart home device models. The primary contributions of this research are as follows: it introduced a data frame feature set suitable for smart home device recognition in WiFi environments and improved the recognition algorithm using the Decision Tree CART approach. Furthermore, practical experiments were conducted within a real smart home environment to validate the proposed method. The study demonstrated the applicability of this method to commonly used domestic smart home devices, achieving a device model recognition accuracy of 91.3%. In the future, we expect to widen our horizons by incorporating machine learning into a mobile application to identify photographs obtained by the camera and tell the user of the specific identification of the photographed object. The technique given in this paper may also be used for security systems in big communities such as smart cities, office buildings, hotels, shopping malls, and university settings to improve the security system of the unique environment. It is also claimed that machine learning makes prediction easier. Machine learning may also be used to anticipate weather and house conditions in the environmental module of smart home automation. By carrying out research on usability, attending to computing efficiency, and taking deployment issues into account, models may be optimized for real-world situations and useful insights can be gained into practical matters. Sustained investigation and advancement within this domain may augment the safety, confidentiality, and general user experience inside smart home settings.

Author contributions All authors have equally contributed to this research.

Funding This is self-funded research.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors do not have any conflict of interest.

Ethical approval All ethical issues including human or animal participation have been done.

Consent participation There is no such participation.

References

1. Xia, Z., Chong, S.: WiFi-based indoor passive fall detection for the medical Internet of Things. Comput. Electr. Eng. **109**, 108763 (2023). https://doi.org/10.1016/j.compeleceng.2023.108763
2. Omran, M.A., Hamza, B.J., Saad, W.K.: The design and fulfillment of a Smart Home (SH) material powered by the IoT using the Blynk app. Mater. Today Proc. **60**, 1199–1212 (2022). https://doi.org/10.1016/j.matpr.2021.08.038
3. Castelo Gómez, J.M., Carrillo-Mondéjar, J., MartínezMartínez, J.L., Navarro García, J.: Forensic analysis of the Xiaomi Mi Smart Sensor Set. Forensic Sci. Int. Digit. Investig. **42–43**, 301451 (2022). https://doi.org/10.1016/j.fsidi.2022.301451
4. Roy Chowdhury, R., Aneja, S., Aneja, N., Abas, P.E.: Packet-level and IEEE 802.11 MAC frame-level network traffic traces data of the D-Link IoT devices. Data Brief **37**, 107208 (2021). https://doi.org/10.1016/j.dib.2021.107208
5. Han, S.: Congestion-aware WiFi offload algorithm for 5G heterogeneous wireless networks. Comput. Commun. **164**, 69–76 (2020). https://doi.org/10.1016/j.comcom.2020.10.006
6. Javed, A.R., Shahzad, F., Urrehman, S., Zikria, Y.B., Razzak, I., Jalil, Z., Xu, G.: Future smart cities: requirements, emerging technologies, applications, challenges, and future aspects. Cities **129**, 103794 (2022). https://doi.org/10.1016/j.cities.2022.103794
7. S, M., M, R.: MUD enabled deep learning framework for anomaly detection in IoT-integrated smart building. e-Prime Adv. Electr. Eng. Electron. Energy **5**, 100186 (2023). https://doi.org/10.1016/j.prime.2023.100186

8. Yao, Y., Zhang, H., Xia, P., Liu, C., Geng, F., Bai, Z., Du, L., Chen, X., Wang, P., Han, B., Yang, T., Fang, Z.: Signature: semi-supervised human identification system based on millimeter wave radar. Eng. Appl. Artif. Intell.Artif. Intell. **126**, 106939 (2023). https://doi.org/10.1016/j.engappai.2023.106939
9. Alhamed, K.M., Iwendi, C., Dutta, A.K., Almutairi, B., Alsaghier, H., Almotairi, S.: Building construction based on video surveillance and deep reinforcement learning using a smart grid power system. Comput. Electr. Eng. **103**, 108273 (2022). https://doi.org/10.1016/j.compeleceng.2022.108273
10. Gaber, T., El-Ghamry, A., Hassanien, A.E.: Injection attack detection using machine learning for smart IoT applications. Phys. Commun. **52**, 101685 (2022). https://doi.org/10.1016/j.phycom.2022.101685
11. Sharma, A., Gupta, A.K., Shabaz, M.: Categorizing threat types and cyber-assaults over Internet of Things-equipped gadgets. Paladyn J. Behav. Robotics **13**(1), 84–98 (2022). https://doi.org/10.1515/pjbr-2022-0100
12. Prentow, T.S., Ruiz-Ruiz, A.J., Blunck, H., Stisen, A., Kjærgaard, M.B.: Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring. Pervasive Mob. Comput.Comput. **16**, 305–316 (2015). https://doi.org/10.1016/j.pmcj.2014.12.006
13. Abdulsalam, K.A., Adebisi, J., Emezirinwune, M., Babatunde, O.: An overview and multicriteria analysis of communication technologies for smart grid applications. e-Prime Adv. Electr. Eng. Electron. Energy **3**, 100121 (2023). https://doi.org/10.1016/j.prime.2023.100121
14. Chowdhury, R.R., Abas, P.E.: A survey on device fingerprinting approach for resource-constraint IoT devices: comparative study and research challenges. Internet of Things **20**, 100632 (2022). https://doi.org/10.1016/j.iot.2022.100632
15. Sun, X., Yuan, L., Wang, X.: Intelligent monitoring of home movement based on fuzzy control theory. Microprocess. Microsyst. **82**, 103943 (2021). https://doi.org/10.1016/j.micpro.2021.103943
16. Kaur, B., Dadkhah, S., Shoeleh, F., Neto, E.C.P., Xiong, P., Iqbal, S., Lamontagne, P., Ray, S., Ghorbani, A.A.: Internet of Things (IoT) security dataset evolution: challenges and future directions. Internet of Things **22**, 100780 (2023). https://doi.org/10.1016/j.iot.2023.100780
17. Ma, C., Man Lee, C.K., Du, J., Li, Q., Gravina, R.: Work engagement recognition in smart office. Proc. Comput. Sci. **200**, 451–460 (2022). https://doi.org/10.1016/j.procs.2022.01.243
18. Huseien, G.F., Shah, K.W.: A review of 5G technology for smart energy management and smart buildings in Singapore. Energy AI **7**, 100116 (2022). https://doi.org/10.1016/j.egyai.2021.100116
19. Khalil, N., Benhaddou, D., Gnawali, O., Subhlok, J.: Nonintrusive ultrasonic-based occupant identification for energy-efficient smart building applications. Appl. Energy **220**, 814–828 (2018). https://doi.org/10.1016/j.apenergy.2018.03.018
20. Malkawi, A., Ervin, S., Han, X., Chen, E.X., Lim, S., Ampanavos, S., Howard, P.: Design and applications of an IoT architecture for data-driven smart building operations and experimentation. Energy Build. **295**, 113291 (2023). https://doi.org/10.1016/j.enbuild.2023.113291
21. Gowda, V.D., Sharma, A., Rao, B.K., Shankar, R., Sarma, P., Chaturvedi, A., Hussain, N.: Industrial quality healthcare services using the Internet of Things and fog computing approach. Meas. Sens. **24**, 100517 (2022). https://doi.org/10.1016/j.measen.2022.100517
22. Nauman, A., Jamshed, M.A., Ahmad, Y., Saad, M., Bilal, M., Shanmuganathan, V., Kim, S.W.: Injecting cognitive intelligence into beyond-5G networks: a MAC layer perspective. Comput. Electr. Eng. **108**, 108717 (2023). https://doi.org/10.1016/j.compeleceng.2023.108717
23. Wirtz, B.W., Weyerer, J.C., Schichtel, F.T.: An integrative public IoT framework for smart government. Gov. Inf. Q. **36**(2), 333–345 (2019). https://doi.org/10.1016/j.giq.2018.07.001
24. Lee, J.H., Hancock, M.G., Hu, M.-C.: Towards an effective framework for building smart cities: lessons from Seoul and San Francisco. Technol. Forecast. Soc. Chang. **89**, 80–99 (2014). https://doi.org/10.1016/j.techfore.2013.08.033
25. Bai, Y., Lu, L., Cheng, J., Liu, J., Chen, Y., Yu, J.: Acoustic-based sensing and applications: a survey. Comput. Netw. **181**, 107447 (2020). https://doi.org/10.1016/j.comnet.2020.107447
26. Li, Q., Gravina, R., Li, Y., Alsamhi, S.H., Sun, F., Fortino, G.: Multi-user activity recognition: challenges and opportunities. Inf. Fusion **63**, 121–135 (2020). https://doi.org/10.1016/j.inffus.2020.06.004
27. Khan, R.H., Khan, J.Y.: A comprehensive review of the application characteristics and traffic requirements of a smart grid communications network. Comput. Netw. **57**(3), 825–845 (2013). https://doi.org/10.1016/j.comnet.2012.11.002
28. Mumtaz, S., Lundqvist, H., Huq, K.M.S., Rodriguez, J., Radwan, A.: Smart Direct-LTE communication: an energy saving perspective. Ad Hoc Netw. **13**, 296–311 (2014). https://doi.org/10.1016/j.adhoc.2013.08.008
29. Rahhal, M., Adda, M., Atieh, M., Ibrahim, H.: Health of humans and machines in a common perspective. Proc. Comput. Sci. **177**, 415–422 (2020). https://doi.org/10.1016/j.procs.2020.10.055
30. Woźniak, M., Zielonka, A., Sikora, A.: Driving support by type-2 fuzzy logic control model. Expert Syst. Appl. **207**, 117798 (2022). https://doi.org/10.1016/j.eswa.2022.117798
31. Mohanty, R., Pani, S.K.: Livestock health monitoring using a smart IoT-enabled neural network recognition system. In: Cognitive Big Data Intelligence with a Metaheuristic Approach, pp. 305–321. Elsevier (2022). https://doi.org/10.1016/b978-0-323-85117-6.00007-8
32. Raja, G.B., Chakraborty, C.: Internet of things based effective wearable healthcare monitoring system for remote areas. In: Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain, pp. 193–218. Elsevier (2023). https://doi.org/10.1016/b978-0-323-91916-6.00004-7
33. Raut, A., Shivhare, A., Chaurasiya, V.K., Kumar, M.: AEDS-IoT: adaptive clustering-based event detection scheme for IoT data streams. Internet of Things **22**, 100704 (2023). https://doi.org/10.1016/j.iot.2023.100704
34. Sharma, J., Mehra, P.S.: Secure communication in IOT-based UAV networks: a systematic survey. Internet of Things **23**, 100883 (2023). https://doi.org/10.1016/j.iot.2023.100883
35. Zhao, Z., Shen, L., Yang, C., Wu, W., Zhang, M., Huang, G.Q.: IoT and digital twin-enabled smart tracking for safety management. Comput. Oper. Res. **128**, 105183 (2021). https://doi.org/10.1016/j.cor.2020.105183
36. Hou, X., Bergmann, J.H.M.: HINNet: Inertial navigation with head-mounted sensors using a neural network. Eng. Appl. Artif. Intell.Artif. Intell. **123**, 106066 (2023). https://doi.org/10.1016/j.engappai.2023.106066
37. Adarsh, A., Kumar, B.: Wireless medical sensor networks for smart e-healthcare. In: Intelligent Data Security Solutions for e-Health Applications, pp. 275–292. Elsevier (2020). https://doi.org/10.1016/b978-0-12-819511-6.00015-7
38. Nethercote, M.: Platform landlords: renters, personal data, and new digital footholds of urban control. Digit. Geogr. Soc. **5**, 100060 (2023). https://doi.org/10.1016/j.diggeo.2023.100060
39. Lee, C.-H., Wang, C., Fan, X., Li, F., Chen, C.-H.: Artificial intelligence-enabled digital transformation in the elderly healthcare field: a scoping review. Adv. Eng. Inform. **55**, 101874 (2023). https://doi.org/10.1016/j.aei.2023.101874

40. Sampaio, H.V., Westphall, C.B., Koch, F., Do Nascimento Boing, R., Santa Cruz, R.N.: Autonomic energy management with Fog Computing. Comput. Electr. Eng. **93**, 107246 (2021). https://doi.org/10.1016/j.compeleceng.2021.107246
41. Rani, P.J., Jason, B., Praveen, K.U., Praveen, K.U., Santhosh, K.: Voice controlled home automation system using natural language processing (NLP) and Internet of things (IoT). In: Proceedings of the Third International Conference on Science Technology Engineering and Management. IEEE, Chennai, India (2017)
42. Jaihar, J., Lingayat, N., Vijaybhai, P.S., Venkatesh, G., Upla, K.P.: Smart home automation using machine learning algorithms. In: Proceedings of the International Conference for Emerging Technology, IEEE, Belgaum, India (2020)
43. Khan, S.A., Farhad, A., Ibrar, M., Arif, M.: Real time algorithm for the smart home automation based on the Internet of things. Int. J. Comput. Sci. Inf. Secur. **14**(7), 94–99 (2016)
44. Popa, D., Pop, F., Serbanescu, C., Castiglione, A.: Deep learning model for home automation and energy reduction in a smart home environment platform. Neural Comput. Appl. 1–21 (2018)
45. Machorro-Cano, I., Alor-Hernandez, G., Paredes-Valverde, M.A., Rodriguez-Mazahua, L., Sanchez-Cervantes, J.L., Olmedo-Aguirre, J.O.: HEMS-IoT: a big data and machine learning-based smart home system for energy saving. Energies **13**(1097), 1–24 (2020)
46. Singh, H., Pallagani, V., Khandelwal, V., Venkanna, U.: IoT-based smart home automation system using sensor node. In: Proceedings of the Fourth International Conference on Recent Advances in Information Technology. IEEE, Dhanbad, India (2018)

International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-023-00489-3

REGULAR PAPER

Enhancing smart home device identification in WiFi environments for futuristic smart networks-based IoT

Hassan Falah Fakhruldeen[1] · Mohamed J. Saadh[2] · Samiullah Khan[3] · Nur Agus Salim[4] · Naveed Jhamat[5] · Ghulam Mustafa[5]

Received: 25 September 2023 / Accepted: 28 November 2023

Abstract
The recognition of smart home devices within WiFi environments stands as a pivotal focus within contemporary Internet of Things (IoT) security, especially in the context of Futuristic Smart Networks-based IoT. The inherent encryption feature of the 802.11 protocol in WiFi settings renders conventional identification methods, reliant on plaintext traffic patterns, ineffective for IoT devices. Through an examination of the 802.11 protocol, distinctive traits within data frames of various smart home devices are revealed. Building on these insights, this research selects attributes like frame length, frame arrival time, duration, and frame sequence number as salient traffic characteristics. Leveraging an enhanced decision tree CART algorithm, the study achieves robust device identification for smart home devices operating within WiFi environments. Experimental outcomes affirm the method's efficacy by accurately discerning device models, achieving an impressive identification accuracy of 91.3%.

Keywords Device identification · WiFi environments · Futuristic smart networks · IoT Security · Smart home devices · Traffic characteristics

✉ Samiullah Khan
samikhan@aup.edu.pk

Hassan Falah Fakhruldeen
hassan.fakhruldeen@gmail.com

Mohamed J. Saadh
msaadeh@meu.edu.jo

Nur Agus Salim
nuragussalim@uwgm.ac.id

Naveed Jhamat
naveed.jhamat@pugc.edu.pk

Ghulam Mustafa
gmustafa@pugc.edu.pk

1 Computer Techniques Engineering Department, Faculty of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq10011

2 Faculty of Pharmacy, Middle East University, Amman 11831, Jordan

3 Department of Mathematics, Statistics & Computer Science, The University of Agriculture Peshawar, Peshawar, KP, Pakistan

4 Universitas Widya Gama Mahakam Samarinda, Samarinda, Indonesia

5 Department of Information Technology, University of Punjab, Gujranwala Campus, Gujranwala, Pakistan

1 Introduction

The Internet of Things (IoT) has become an important component of the new generation of information technology, with widespread applications in various fields such as smart homes, intelligent healthcare, and smart cities. The quantity and variety of IoT devices have grown exponentially due to increasing market demand. It is projected that by the year 2025, there will be over 75 billion IoT devices connected to the Internet [1], and this number is expected to reach 125 billion by 2030 [2]. Presently, due to the advantages of wide wireless network coverage, strong mobility, and low construction costs, a large number of IoT devices are connected via WiFi networks. Especially in the case of smart home devices, connecting to WiFi allows for quick information exchange between devices, enhancing convenience in daily life.

Most of the research on IoT device identification is based on extracting traffic features at the TCP/IP layers. Author [3] extracted 23 features including protocols, packet sizes, IP addresses, and port numbers from different network layers. Author [4] identified devices by extracting 67 characteristics such as TTL and TCP window size from packet sequences. These methods utilize privileged access to router and network

Published online: 07 January 2024

protocol headers, enabling the extraction of a wide range of protocol feature fields. However, in unfamiliar WiFi environments, the captured traffic consists of encrypted link layer data based on the 802.11 protocol. Among the fastest growing segments of the communications sector today is 802.11g WLAN technology. Naturally, it does this without a network cable and offers constant network access. Workers from home or those who work remotely can build up networks without worrying about how to put cables through homes that were not intended to accommodate network equipment. A collection of specifications known as 802.11 covers computer communications over wireless local area networks (WLANs) operating in the 2.4, 3.6, and 5 GHz bands. For big enterprise wireless systems and household wireless access points, the most popular standards are 802.11a, b, and g. With data transmission rates of up to 54 Mbps, 802.11a is faster than 802.11b. A privacy mechanism called WEP is defined in 802.11 to safeguard connection data that is sent via WLAN. This speaks to the intention of giving wireless LAN users a privacy service like what a traditional LAN's built-in physical security offers. The RC4 symmetric stream cipher with 40-bit and 104-bit encryption keys is used for the WEP encryption. The 802.11 standard does not specify 104-bit encryption keys. However, several wireless AP manufacturers do support them.

Extracting specific fields of IP layers and above is challenging in such cases. Moreover, stability and interference resistance of data transmission in WiFi environments are weak, leading to significant difficulties in device identification. Currently, research on device identification in WiFi environments is relatively limited. Author [5] analyzed the duration field in 802.11 traffic for device identification author [6] used encrypted WiFi traffic's destination address, network name, frame size, and MAC protocol fields as implicit identifiers for wireless devices. Author [7] differentiated devices through temporal analysis of 802.11 probe request frames. Author [8] evaluated features like transmission rate, frame size, media access time, transmission time, and inter-frame arrival time in 802.11 traffic, with transmission time and inter-frame arrival time proving most effective. The author [9] utilized similar hashing algorithms to generate device fingerprints for 802.11 management frames. Author [10] transformed traffic from IoT devices in wireless networks into traffic grayscale images for identification experiments. Given the limited information obtainable from WiFi-based traffic, effectively identifying IoT devices in this environment remains a challenging aspect that requires further research.

Most IoT device identification algorithms primarily employ machine learning techniques, constructing classification models based on feature attributes extracted from network traffic. Author [3] utilized a random forest algorithm to identify devices using a $23 \times N$ feature matrix. The author [11] generated feature vectors with features like TCP window size and payload length and employed a gradient boosting tree algorithm for identification. The author [12] used the J48 decision tree algorithm to identify 23 IoT devices. Author [13] compared identification algorithms including random forest, k-nearest neighbors (KNN), decision tree (DT), and support vector machine (SVM), concluding that random forest and decision tree algorithms excel in recognition rate and speed, respectively. The author [14] used a naive Bayesian optimization algorithm and clustering algorithm for IoT device and non-IoT device identification.

Author [15] employed KNN for identifying 22 IoT devices based on features like average packet length, average arrival time, and packet length. The author [16] utilized an SVM algorithm for IoT device identification using various protocol features and device-specific attributes. Based on existing research, this paper further investigates link layer traffic features of the 802.11 standard under IoT environments by monitoring smart home device traffic under WiFi and using machine learning algorithms for device identification.

However, as the heterogeneous smart home devices connected to WiFi networks rapidly increase, the resulting wireless encrypted traffic becomes increasingly complex and chaotic. This leads to the growth of network complexity and scale, making the management of smart home devices more intricate. Additionally, since smart home devices often transmit sensory information and user privacy data, they are susceptible to attacks, thereby posing a serious threat to the security and privacy of IoT systems. Therefore, achieving smart home device identification in WiFi environments is currently a prominent area of research. Device identification methods can associate information about IoT devices, users, functionalities, and data within the network space. This aids in situational analysis, event tracing, and targeted control of these devices, thereby holding significant importance in addressing the management and security issues of smart home devices.

The Internet of Things (IoT) has become increasingly significant in the dynamic field of information technology, as it has made its presence felt in several sectors such as smart homes, healthcare, and urban environments. Given the anticipated increase in the quantity of Internet of Things (IoT) devices, it is imperative to explore effective methods for device identification. The principal objective of this study is to tackle the issue of discerning Internet of Things (IoT) devices within wireless fidelity (WiFi) networks, with a specific emphasis on the complex smart home setting.

The work makes a valuable contribution by:

1. The objective of this study is to explore and provide reliable methods for accurately identifying Internet of Things (IoT) devices within WiFi networks.
2. This paper aims to examine the intricacies and security concerns associated with smart home setups.

3. Facilitating situational analysis and monitoring of events.
4. Improving the administration and security of smart home devices.

2 Related work

This extensive literature review explores a wide range of advanced subjects encompassing smart buildings, the Internet of Things (IoT), and related disciplines. The study conducted by Ma et al. [17] focuses on the improvement of job engagement in smart offices as a means to enhance staff productivity. In their study, Huseien and Shah [18] conducted a thorough examination of the incorporation of 5G technology to enhance energy management and smart infrastructures, with a specific focus on the context of Singapore. In this study, Khalil et al. [19] present an innovative approach to nonintrusive occupant identification through the use of ultrasonic sensors. The primary objective of this method is to enhance energy efficiency in smart buildings. In their study, Malkawi et al. [20] provide a novel Internet of Things (IoT) architecture that aims to improve data-driven operations and experimentation in the context of smart buildings.

In the realm of healthcare, Gowda et al. [21] proposes a novel integration of the Internet of Things (IoT) and fog computing as a means to transform and enhance the delivery of high-quality industrial healthcare services. In their study, Nauman et al. [22] explore the integration of cognitive intelligence into post-5G networks, with a specific focus on enhancing network efficiency at the MAC layer. The authors Wirtz et al. [23] provide a valuable contribution by developing a comprehensive public Internet of Things (IoT) infrastructure specifically designed for the implementation of intelligent governance applications. In their study, Lee et al. [24] examine the experiences of Seoul and San Francisco to derive significant insights that might inform the creation of efficient frameworks for the development of smart cities.

In addition to the advancements in smart buildings and the Internet of Things (IoT), Bai et al. [25] provide an extensive overview of acoustic-based sensing applications. In their study, Li et al. [26] examine the difficulties and potential opportunities associated with multi-user activity recognition. Reduan and Jamil [27] provide a comprehensive evaluation of application characteristics and traffic requirements within the domain of smart grid communications. The study conducted by Mumtaz et al. [28] focuses on the optimization of energy consumption in smart direct-LTE networks. Rahhal et al. [29] shed light on a common viewpoint about the health of both humans and machines.

In their study, Woźniak et al. [30] propose a type-2 fuzzy logic model to enhance driving support, thereby expanding the existing range of approaches in this field. In their study, Mohanty and Pani [31] propose a novel approach to livestock health monitoring that utilizes neural networks enabled by the Internet of Things (IoT). In their study, Raja and Chakraborty [32] introduce a wearable healthcare system that utilizes Internet of Things (IoT) technology, specifically designed to cater to healthcare needs in distant regions. In their study, Raut et al. [33] provide a novel adaptive event detection system designed specifically for streaming Internet of Things (IoT) data. In their study, Sharma and Mehra [34] conducted a comprehensive examination of the topic of secure communication within unmanned aerial vehicle (UAV) networks.

Zhao et al. [35] investigate the integration of the Internet of Things (IoT) and digital twin technologies as a means to increase safety management. Hou and Bergmann [36] adeptly combine inertial navigation with neural networks. In their study, Adarsh and Kumar [37] propose the use of wireless medical sensor networks as a means to significantly transform the field of e-healthcare. In his work, Nethercote [38] critically examines the concept of platform landlords and their role in exerting control over urban spaces within the context of the digital era. The authors of the study conducted by Lee et al. [39] examine the integration of artificial intelligence (AI) in the healthcare sector for the older population. In conclusion, Sampaio et al. [40] emphasize the significance of autonomous energy management within the context of Fog Computing. This comprehensive compilation enhances our comprehension across several technological disciplines, augmenting our understanding and ramifications.

Rani et al. [41] presented a voice-activated home automation system built on natural language processing (NLP) and artificial intelligence methods. Voice instructions are sent through a cell phone to operate household appliances, and the preset natural language processing medium interprets the commands. The system's application was limited to controlling household appliances; it was not utilized for additional home automation functions like environmental monitoring, motion detection, intruder detection, or control. Jaihar et al. [42] introduced an intelligent smart home automation system that controls lighting, music systems, and other household appliances by performing tasks based on the user's emotional state. To anticipate actions and reduce user engagement, several machine learning algorithms were integrated and utilized to assess the user's wants as well as the environment. Depending on the mood that the machine learning model detects, the house appliances are turned ON or OFF. Their methodology improved domestic energy efficiency.

Khan et al. [43] suggested a real-time algorithm for house monitoring and control, including ambient factors, motion sensors, electrical appliances, and home appliances. The motion sensors' algorithmically derived inferences determined whether to turn on or off the lights. Using the WiFi module, the suggested algorithm was also utilized to track the power usage of different household appliances and to set

off an alert depending on the gas pressure in the house. In an experiment conducted by Popa et al. [44] with two deep neural network models for a smart home, anomalous patterns of energy usage could be identified. A modular framework for gathering, combining, and preserving data in the context of smart homes was developed through the use of cloud computing services. By using less energy, the authors demonstrated how the recommended machine learning techniques improved smart home automation.

The energy-saving strategy for equipment in a smart home setting presented by Machorro-Cano et al. [45] makes use of big data and machine learning approaches. Using methods to identify the home energy usage level by learning user behavior and consumption patterns, the J48 machine learning algorithm and Weka API were used to assure the energy efficiency of the system. Using a smartphone app called HEMS-IoT that the authors created, the system recorded and presented real-time data as well as suggestions for energy-saving measures throughout the house. Their strategy addressed home comfort and the safety of people and gadgets in addition to conserving energy by enabling system users to communicate with their houses and request the necessary IoT service. Singh et al. [46] presented a smart home automation system for managing electrical appliances, and doors, and detecting activity in a house, in addition to monitoring energy use in the home and delivering frequent notifications about it. The device might also notify the user if sensors detect low quantities of gas in a cylinder or the presence of a human. An Arduino Uno board, a Node MCU ESP8266, and IR and LDR sensor modules were used to prototype the system.

3 Smart home identification method based on WiFi data frame features

This paper focuses on studying smart home devices such as home cameras, smart doorbells, smart TVs, smart locks, smart speakers, robotic vacuum cleaners, and smart gateways. It proposes a smart home identification method based on WiFi data frame features, with an overall process depicted in Fig. 1. It consists of three steps: traffic collection, traffic processing, and device identification. The traffic collection module gathers traffic data in WiFi environments, the traffic processing module filters, extracts features, and performs feature subset analysis and representation for smart home devices' network traffic, and the device identification module trains the recognition model, implementing device identification using an improved CART algorithm for decision trees.

Connecting commonplace equipment like sensors and actuators for automation that are included in household appliances is made possible by smart homes. The core of a smart home is the convergence of several technological platforms,

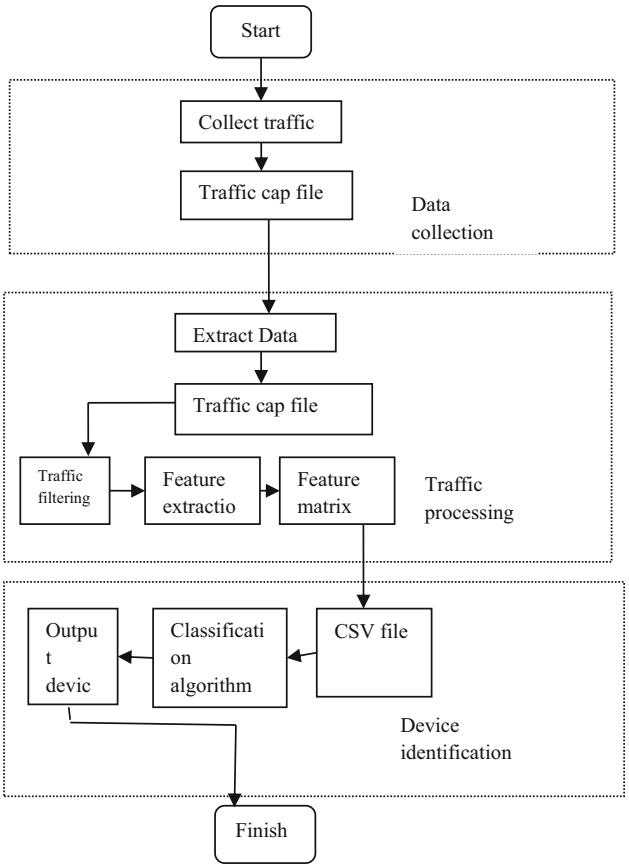


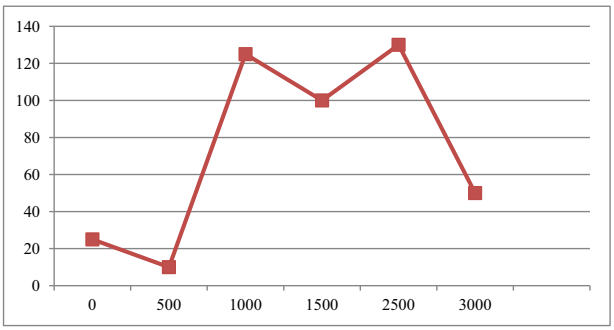
Fig. 1 WiFi data frame-based smart house identification technique

often involving three tiers: the application, network, and perception layers. In addition to acting as an interface between people and the linked items, the perception layer collects information from the environment. Novel and pervasive sensing approaches have been developed in response to the need for an interface that is more user-friendly and pleasant. WiFi sensing, whether active or passive, does away with the need for physical touch by using invisible radio waves to feel the environment. It causes no discomfort since it can do the sensing functions without the user realizing it.

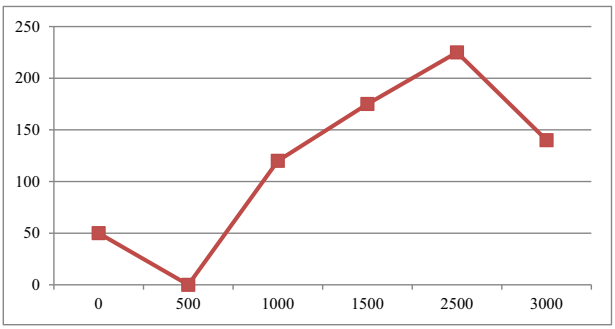
3.1 Data frame feature selection

Under WiFi, 802.11 data frames are divided into three types: management frames, data frames, and control frames. Since control frames have a simple structure and few extractable valid features, and not all devices generate management frames, which also lack universality in extracted features, this paper selects data frame types with rich protocol field information as experimental data.

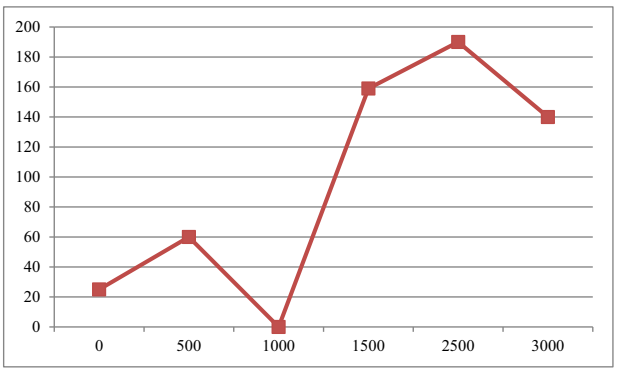
To acquire traffic features of data frames, this paper analyzed traffic from various models of smart home devices. "YiSight C6P" represents camera devices, which maintain



(a) "YiSight C6P" Traffic I/O Graph



(b) "360 Smart Doorbell" Traffic I/O Graph



(c) "Huawei Smart Speaker" Traffic I/O Graph

Fig. 2 Traffic I/O graphs

surveillance and provide real-time alerts for specific captured information. "360 Smart Doorbell" is a smart doorbell device that remains silent after activation until the doorbell is pressed. "Huawei Smart Speaker" represents smart speaker devices, primarily used for playing music, adjusting volume, and switching songs. Using these three types as examples, the paper further analyzes traffic differences among devices of different types and models.

Figure 2 depicts the I/O traffic graphs of the three smart home devices. It can be observed that the "YiSight C6P" device, while in monitoring mode, exhibits activity below 25 frames per second (fps), showing regular fluctuations. Burst traffic, indicating the capture of specific information, occurs at around 100 fps. The "360 Smart Doorbell" device, while in silent mode, transmits at a lower rate of approximately 5 frames per second. Upon pressing the doorbell, burst traffic surges to over 200 frames per second. The "Huawei Smart Speaker" device sends around 50 frames per second cyclically during music playback. Transitioning volume and switching songs result in burst traffic peaking at up to 1500 frames per second. The three devices display distinct trends in the number of frames arriving per second in the I/O graphs, underscoring the significant differences in traffic among different types and models.

Table 1 Distribution statistics of frame lengths

Device frame length	Distributed/%		
	Fluorite C6P	360 Smart Doorbell	Huawei Speaker
0 ~ 39	34.25	22.7	0.2
40 ~ 79	1.35	2.36	1.88
80 ~ 159	57.39	34.67	19.8
160 ~ 319	2.9	0.27	1.14
320 ~ 639	0.75	0.21	41.08
640 ~ 1279	3.36	0.63	0.18
1280 ~ 2559	0	39.17	35.71
Average frame length/byte	100.89	632.9	790.89

Table 1 lists statistical information about frame lengths for the three smart home devices. The "YiSight C6P" device has shorter frame lengths, with most falling in the range of 80 to 159 bytes. The "360 Smart Doorbell" device exhibits higher occurrences in the ranges of 80 to 159 bytes and above 1280 bytes, with an average frame length of 632.90 bytes. The "Huawei Smart Speaker" device's frame lengths are mainly distributed between 320 to 639 bytes and above 1280 bytes, with an average frame length reaching 790.89 bytes. Consequently, there are distinct differences in frame lengths among different devices.

Based on the aforementioned analysis and the comparison of 802.11 data frame traffic from various smart home devices in WiFi environments, this paper extracts 11 features from device traffic, including frame length, frame interval time, frame arrival time, duration, frame sequence number, frame type, frame subtype, transmission direction, retransmission flag, QoS traffic identifier, and data length. The specific descriptions of these features are presented in Table 2.

3.2 Improved decision tree CART algorithm

The CART (Classification and Regression Trees) algorithm for decision trees employs a binary tree structure and a recursive binary partitioning technique. It uses the Gini index

Table 2 Feature descriptions

Frame feature	Feature description
Frame length	The length of the data frame
Interframe time	time interval between two consecutive frames
Frame arrival time	The arrival time of the data frame
Duration	The time the data frame and its acknowledgment frame occupy the channel
Frame sequence number	Data frame sequence control bits, used to reassemble frame fragments and discard duplicate frames
Frame type	Types of 802.11 Data Frames
Frame subtype	Subtypes of 802.11 Data Frames
Transmission direction	DS flag, indicating the transmission direction of the frame BSS and DS
Retransmission flag	Retransmission flag, indicating that the frame is a retransmission frame of the transmission segment
QoS traffic identifier	Types of QoS
Data length	The number of bytes occupied by the data portion of the data frame

to represent the purity of a model, where a smaller Gini index signifies higher purity and better feature representation. The Gini index reduces the logarithmic calculations involved in entropy models, thus lowering computational costs. For classification problems, assuming a given sample E with K classes, where the number of samples in the k-th class is D_k, and the probability of the k-th class is q_k, the expression for the Gini index is shown in Formula (1); while, the Gini index expression for sample E is shown in Formula (2).

$$\text{Gini}(q) = \sum_{k=1}^{K} p_k(1 - p_k) = 1 - \sum_{k=1}^{K} p_k^2 \tag{1}$$

$$\text{Gini}(E) = 1 - \sum_{k=1}^{K} \left(\frac{|D_k|}{|E|} \right)^2 \tag{2}$$

This study improves the Decision Tree CART algorithm through parameter optimization. Parameter optimization aims to minimize the objective function, enhancing the fit between model output and actual data results to achieve higher accuracy and reliability in the final recognition outcome. To reduce interference from experimental sample noise and further enhance recognition accuracy, this study optimizes parameters related to the decision tree's maximum depth, internal nodes, leaf nodes, and minimum impurity decrease for node splitting. These parameter settings are designed to prevent model overfitting and enhance the robustness of the recognition model.

Grid Search CV (Grid Search Cross-Validation) is a commonly used parameter tuning method. It sequentially adjusts parameters within specified ranges, traversing all possible parameter combinations to train models and select the set of parameters that yield the highest accuracy. This study employs Grid Search CV for parameter optimization. However, since Grid Search CV requires traversing all parameter combinations within the given range, it can consume a considerable amount of time and computational resources, especially when dealing with large datasets and numerous parameters. Therefore, this study first conducts score curve analysis for each parameter to identify the approximate optimal parameter range. Subsequently, the Grid Search CV method is used to determine the optimal parameter combination, achieving the recognition model. This approach reduces the computational burden and time cost of grid search while improving the accuracy of device recognition to a certain extent.

4 Experimental analysis

This research focuses on the recognition of smart homes in WiFi environments. Due to the lack of publicly available datasets for mainstream Chinese brands such as Huawei and Xiaomi, this study sets up a practical IoT environment for experimental analysis. The Decision Tree algorithm is optimized to ensure data stability during training, and the effectiveness of extracting data frame features is verified. Through parameter optimization, this work enhances the Decision Tree CART method. To increase the accuracy and dependability of the final recognition result, parameter optimization seeks to minimize the objective function and improve the fit between the model output and real data results. This work improves the decision tree's maximum depth, internal nodes, leaf nodes, and minimal impurity reduction for node splitting to lessen interference from experimental sample noise and improve identification accuracy even more. These parameter configurations aim to keep the recognition model more robust and avoid overfitting. One popular technique for fine-tuning parameters is Grid Search CV (Grid Search Cross-Validation). To train models and choose the combination of parameters that produce the best accuracy, it iteratively modifies parameters within predetermined limits.

For parameter optimization, Grid Search CV is used in this study. To determine the estimated ideal parameter range, this study first analyzes the score curve for each parameter. The best parameter combination is then found using the Grid Search CV technique, which results in the recognition model. With a little increase in device detection accuracy, this method lessens the grid search's computational load and

Fig. 3 Data collection topology

Table 3 Data collection environment configuration

Configuration	Parameter
virtual machine	VMware Workstation 16
virtual machine operating system	Kali Linux 2021.3a
raspberry pie	Raspberry Pi 4
Raspberry Pi OS	Kali Linux 2021.3 rpi4
Wireless network card	ALFA RTL 8812AU
Python	python3.9.0

time cost. The sniffer's wireless port is configured in monitoring mode. To capture all network traffic broadcast over the air in the same channel of the WiFi network, the Linux system's aerodump terminal command is used to define the channel, BSSID, duration, and other parameters. The two sorts of captured data packets are those that occur during the idle time after a smart home device's WiFi connection and those that occur during the live interactions between users and smart home devices. As experimental data samples, data are gathered for one hour during both the idle and interaction phases, and recorded separately. cap files, and then analyzed.

4.1 Traffic data collection

The topology of the data collection setup is shown in Fig. 3. A Raspberry Pi and a virtual machine on a laptop with a wireless USB adapter are used to build a traffic sniffer for data capture. The hardware and environmental configurations are detailed in Table 3.

Specific information about the smart home devices used in the experiment is presented in Table 4. The smart home devices are connected to the test WiFi, and the network environment is configured to keep the devices operational.

The wireless port of the sniffer is set to monitoring mode. The aerodump terminal command in the Linux system is used to specify the channel, BSSID, duration, and other parameters, capturing all network traffic transmitted over the air in the same channel of the WiFi network. Captured data packets are categorized into two types: packets during the idle period after smart home devices connect to WiFi, and packets during real-time interactions between users and smart home devices. Data are collected in both idle and interaction periods for an hour each time, saved separately as.cap files, serving as experimental data samples. Information regarding these data samples is presented in Table 5.

4.2 Traffic feature processing

Experimental samples were collected interactively over one hour, resulting in over 190,000 traffic data entries. The data frame type was extracted using the frame control field, and the traffic was classified and filtered by the Mac address of the smart home devices. For each device's traffic, features such as frame length, frame interval time, frame arrival time, duration, frame sequence number, frame type, frame subtype, transmission direction, retransmission flag, QoS traffic identifier, and data length were extracted frame by frame. Some features were subjected to normalization; the two directions in the transmission direction field were normalized to 1 and 2, missing values in the data length field were set to 0, and missing values in the QoS traffic identifier field were set to 16. The final feature representation consisted of feature vector matrices for different devices.

After extracting data frame features, feature importance was measured, and the feature ranking results are presented in Fig. 4. It can be observed that features like retransmission, type, and subtype have relatively low importance indicators. Consequently, this study selects a subset of 8 features comprising frame length, frame arrival time, duration, frame sequence number, transmission direction, interval time, data length, and QoS traffic identifier as the final feature set for data frame recognition.

4.3 Algorithm parameter configuration

Based on the Decision Tree CART recognition algorithm, parameters were set with a maximum depth single increment of 10, and internal nodes and leaf nodes with a single increment of 1. The scores were computed for different parameter values, and the score curves are plotted as shown in Fig. 5. In the figure, the x-axis represents different parameter values, and the y-axis represents the coefficient of determination S^2 under that value, expressed as in Eq. (3):

$$S^2 = \frac{\sum_i \left(\hat{x}_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}{\sum_i \left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \tag{3}$$

where x represents the actual result and $\hat{x}$ represents the model's predicted result. S^2 is a commonly used metric for evaluating the goodness of fit of a model; the closer its value is to 1, the better the model's fit.

Table 4 List of smart home devices

Serial number	Device name	Physical address	Category
1	Fluorite C6P	C0:E4:34:29:89:09	Camera
2	Fluorite C3W	EC:9C:32:C5:7C:EA	Camera
3	Fluorite C6CN	D4:E8:53:05:89:BB	Camera
4	Fluorite C6C	EC:9C:32:A0:D4:E8	Camera
5	TP-Link IPC55a	7C:B5:9B:E2:D9:7F	Camera
6	Hikvision Dome Camera	00:95:69:D0:49:3E	Camera
7	Xiongmai Robot Camera	7C:A7:B0:4E:F4:2D	Camera
8	Fluorite Doorbell	DC:F5:05:F1:0E:8B	Doorbell
9	360 Smart Doorbell	B2:59:47:00:4E:1B	Doorbell
10	Ding Zero Doorbell	90:E8:68:28:8B:79	Doorbell
11	Millet Doorbell	EC:2E:98:22:94:7D	Doorbell
12	Kim Jong Tv	DC:29:19:64:8A:F8	Television
13	Haier TV Yunos	1C:30:08:67:DD:F5	Television
14	Konka TV	08:38:69:00:2E:48	Television
15	Kaidis Smart Door Lock	F4:CF:A2:F0:67:6D	Door Lock
16	Deschmann Smart Door Lock	70:3A:2D:2B:C1:D6	Door Lock
17	Huawei Speaker	78:85:F4:EC:D0:3C	Speakers
18	Xiaoai Speaker	9C:9D:7E:A6:06:61	Speakers
19	Roborock Sweeping Robot S51	04:CF:8C:F8:D0:DC	Sweeping Robot
20	Lumi Multimode Gateway	54:EF:44:20:0A:17	Gateway

Table 5 Information about data samples

State	The amount of data	File size/106
Stand still	4,935,057	1146.88
Interactive	1,978,838	275

From Fig. 5, it can be observed that the peak of the score curve for the parameter "maximum depth" is around 370, with an $R2$ value of 0.958. This indicates that the optimal parameter range for maximum depth is approximately 360–380. The score curves for the parameters "internal nodes" and "leaf nodes" show a general decreasing trend. The preliminary optimal range for internal nodes is around 2–5, and for leaf nodes, it's around 1–5. Based on these parameter ranges, a grid search was conducted using GridSearchCV, resulting in the optimal parameter set: {'max_depth': 376, 'internal_nodes': 2, 'leaf_nodes': 1, 'min_impurity_decrease': 0.0}.

4.4 Device recognition analysis

4.4.1 Smart home model recognition

The feature matrices of smart home devices were divided into training and testing sets in a 7:3 ratio. A supervised training was performed to generate the recognition model, and the experimental results are presented in Fig. 6 with Table 6. The average accuracy of recognizing 20 different device models from the dataset reached 91.3%, with recognition rates exceeding 85% for 17 device models. Additionally, Table 7 shows a comparison between the test results when frame types are not extracted and when data frame types are extracted. The results indicate that extracting data frame types significantly improves the accuracy of device model recognition.

With a recall of 0.93, 93% of real-world abnormalities are properly identified by the model. The F1 score of 0.91, which represents the harmonic mean of the model's recall and accuracy, serves as a summary of its effectiveness. The model has a strong discriminative ability to distinguish outliers from the overall population, as evidenced by the AUC-ROC value of 0.94. With a precision level of 0.82, 82% of the time a face prediction is accurate. The algorithm properly recognizes 86% of real-world faces, according to a recall value of 0.86. The model's overall effectiveness at recognizing human faces is indicated by its F1 score of 0.84. Furthermore, the AUC-ROC score of 0.90 indicates how well the model classified face occurrences. The model correctly identifies 89% of all occurrences with an accuracy of 0.89. 86 percent of the time, with an accuracy of 0.86, the outliers are, in fact, outliers. With a recall value of 0.91, the model accounts for 91% of true outliers. The F1 score, which stands at 0.88, is an

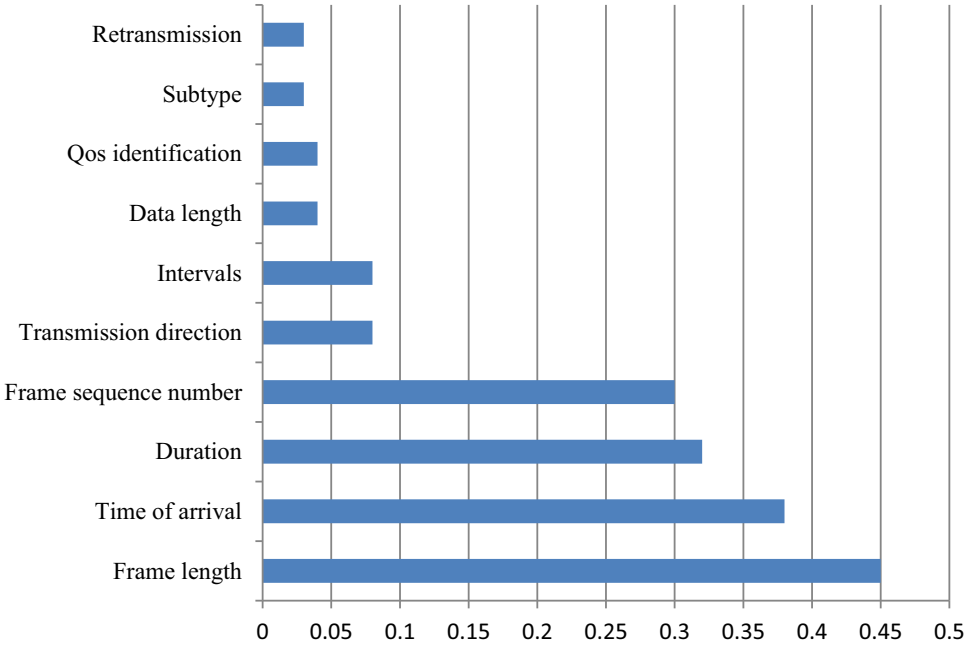


Fig. 4 Feature importance measure

attempt to balance recall and accuracy. A further indication of the model's capacity to differentiate abnormal from ordinary data is its AUC-ROC of 0.92. The model successfully recognizes 83% of the tested faces with an accuracy of 0.83. Eighty percent of the examples with an accuracy rating of 0.80 are faces. Recalling 85% of real-world faces correctly, the model has a 0.85 recall score.

4.4.2 Comparative analysis

The traffic data used in this study represents passive traffic in a WiFi environment, which differs from the plaintext traffic often examined in mainstream research. In reference [5], the recognition of WiFi encrypted traffic using the duration field was tested, yielding an accuracy of 62.2%. Reference [10] utilized encrypted WiFi traffic, directly transforming it into traffic images without extracting frame features. Multiple algorithms were combined to recognize specific IoT device models, and the accuracy of the decision tree algorithm was 78.1%. In contrast, this study achieved a device model recognition accuracy of 91.3% using the proposed method. Additionally, this study tested device type recognition, and the comparative results are detailed in Table 8. The proposed method enhances the recognition rate of IoT devices, validating the effectiveness of extracting data frame features and addressing the issue of smart home device model recognition in situations where router-specific information is inaccessible due to environmental constraints.

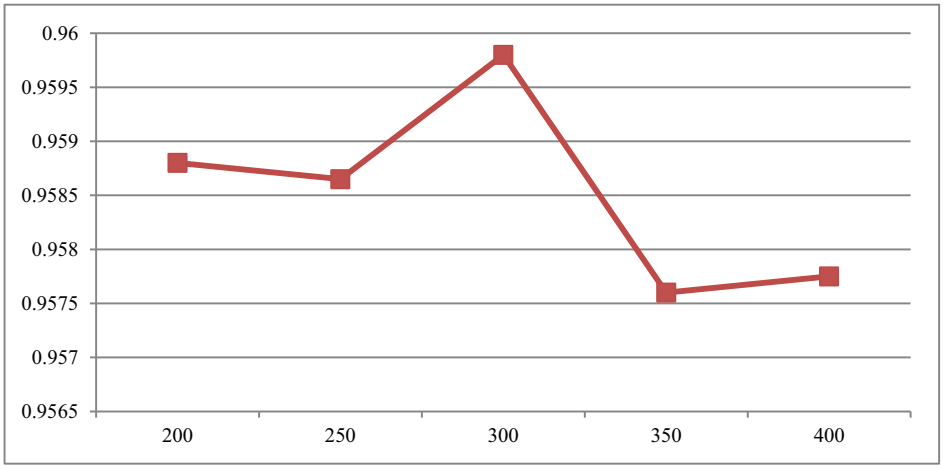
In this experiment, due to limitations in the experimental environment, only 20 different types of smart home devices were tested. In future research, it is intended to apply the method proposed in this study to a wider range of scenarios. This would involve expanding the number of devices in the training model, encompassing a greater variety of types, brands, and models of IoT devices. This expansion would help address the management challenges posed by IoT devices and contribute to enhancing convenience in the network space environment.

4.5 Discussion

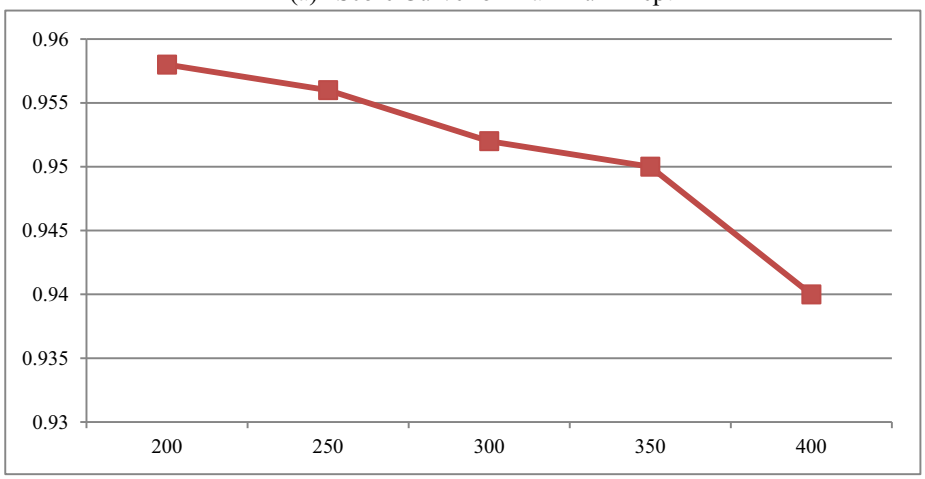
The dataset's 20 distinct device models were recognized with an average accuracy of 91.3%; 17 of the device models had identification rates higher than 85%. The maximum depth single increment of 10 and the single increment of 1 for internal and leaf nodes were the parameters selected based on the Decision Tree CART recognition method. Score curves were presented after the scores were calculated for various parameter values. For each device, feature vector matrices made up the final feature representation. Following the extraction of data frame features, the results of the feature ranking are shown together with a measurement of feature relevance. It is noted that characteristics with relatively low relevance indications include retransmission, type, and subtype.

Because of this, the final feature set for data frame recognition in this study consists of a subset of 8 features: frame length, frame arrival time, duration, frame sequence number, transmission direction, interval time, data length, and QoS traffic identifier. Over 190,000 traffic data entries were produced via interactively gathering experimental samples over one hour. The traffic was categorized and filtered based on the Mac address of the smart home devices, and the data frame

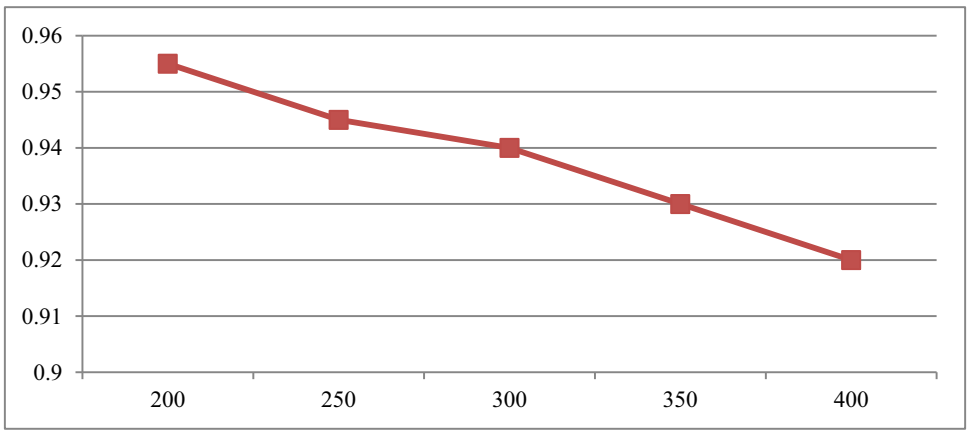
Fig. 5 Scoring curve



(a) Score Curve for Maximum Depth



(b) Score Curve for Internal Nodes



(C) Score Curve for Leaf Nodes

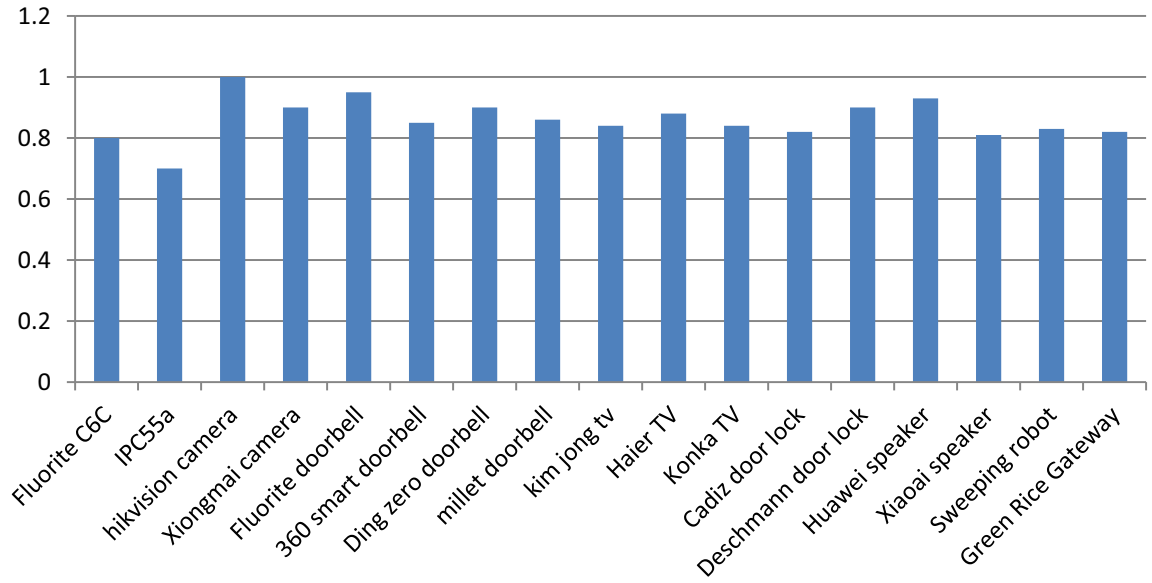


Fig. 6 Different models of device recognition accuracy

Table 6 Different models of device recognition accuracy

Device	Accuracy
Fluorite C6C	0.8
IPC55a	0.7
Hikvision camera	1
Xiongmai camera	0.9
Fluorite doorbell	0.95
360 smart doorbell	0.85
Ding zero doorbell	0.9
millet doorbell	0.86
kim jong TV	0.84
Haier TV	0.88
Konka TV	0.84
Cadiz door lock	0.82
Deschmann door lock	0.9
Huawei speaker	0.93
Xiaoai speaker	0.81
Sweeping robot	0.83
Green Rice Gateway	0.82

type was retrieved using the frame control field. Frame by frame, information about each device's traffic was retrieved, including the QoS traffic identity, transmission direction, frame type, frame subtype, frame length, frame interval time, frame arrival time, duration, frame sequence number, and transmission type.

Table 7 Comparison of recognition with and without data frame type extraction

Frame type	Accuracy	Recall rate	F1-Score
All frames	79	79.4	79.2
Data frame	91.3	91.3	91.3

Normalization was applied to a few characteristics; the two directions in the transmission direction field were set to 1 and 2, and the missing values in the data length and QoS traffic identification fields were set to 0 and 16, respectively. With an R2 value of 0.958, it can be seen that the parameter "maximum depth" has a peak on the score curve of about 370. This suggests that about 360–380 is the ideal parameter range for the greatest depth. A general declining tendency can be seen in the score curves for the parameters "internal nodes" and "leaf nodes." For internal nodes, the preliminary ideal range is around 2–5, while for leaf nodes, it is approximately 1–5. To identify certain IoT device types, many algorithms were merged; the decision tree algorithm's accuracy was 78.1%. On the other hand, this study used the suggested strategy to obtain 91.3% accuracy in device model recognition. This study also evaluated the recognition of device type, and the comparison outcomes are presented in detail. By improving the recognition rate of IoT devices, the suggested approach validates the efficacy of obtaining data frame attributes and solves the problem of smart home device model recognition when environmental restrictions prevent access to router-specific information.

Table 8 Comparison of evaluation indicators

Recognition result	traffic data set	Adapting methods	Accuracy	Recall rate	F1-score
Device specific model	Literature [10]	Literature [10] (DT)	78.1	78.3	78.4
	Proposed Work	Literature [5] (DT)	62.2	37.3	40.1
		Literature [10] (DT)	86.8	90.4	88.6
		The method in this paper	91.3	91.3	91.3
Device specific type	Proposed Work	The method in this paper	88.2	87.1	87.7

5 Conclusion

This paper introduced a smart home device recognition method based on 802.11 data frame features in a WiFi environment, enabling the identification of smart home device models. The primary contributions of this research are as follows: it introduced a data frame feature set suitable for smart home device recognition in WiFi environments and improved the recognition algorithm using the Decision Tree CART approach. Furthermore, practical experiments were conducted within a real smart home environment to validate the proposed method. The study demonstrated the applicability of this method to commonly used domestic smart home devices, achieving a device model recognition accuracy of 91.3%. In the future, we expect to widen our horizons by incorporating machine learning into a mobile application to identify photographs obtained by the camera and tell the user of the specific identification of the photographed object. The technique given in this paper may also be used for security systems in big communities such as smart cities, office buildings, hotels, shopping malls, and university settings to improve the security system of the unique environment. It is also claimed that machine learning makes prediction easier. Machine learning may also be used to anticipate weather and house conditions in the environmental module of smart home automation. By carrying out research on usability, attending to computing efficiency, and taking deployment issues into account, models may be optimized for real-world situations and useful insights can be gained into practical matters. Sustained investigation and advancement within this domain may augment the safety, confidentiality, and general user experience inside smart home settings.

Author contributions All authors have equally contributed to this research.

Funding This is self-funded research.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors do not have any conflict of interest.

Ethical approval All ethical issues including human or animal participation have been done.

Consent participation There is no such participation.

References

1. Xia, Z., Chong, S.: WiFi-based indoor passive fall detection for the medical Internet of Things. Comput. Electr. Eng. **109**, 108763 (2023). https://doi.org/10.1016/j.compeleceng.2023.108763
2. Omran, M.A., Hamza, B.J., Saad, W.K.: The design and fulfillment of a Smart Home (SH) material powered by the IoT using the Blynk app. Mater. Today Proc. **60**, 1199–1212 (2022). https://doi.org/10.1016/j.matpr.2021.08.038
3. Castelo Gómez, J.M., Carrillo-Mondéjar, J., MartínezMartínez, J.L., Navarro García, J.: Forensic analysis of the Xiaomi Mi Smart Sensor Set. Forensic Sci. Int. Digit. Investig. **42–43**, 301451 (2022). https://doi.org/10.1016/j.fsidi.2022.301451
4. Roy Chowdhury, R., Aneja, S., Aneja, N., Abas, P.E.: Packet-level and IEEE 802.11 MAC frame-level network traffic traces data of the D-Link IoT devices. Data Brief **37**, 107208 (2021). https://doi.org/10.1016/j.dib.2021.107208
5. Han, S.: Congestion-aware WiFi offload algorithm for 5G heterogeneous wireless networks. Comput. Commun. **164**, 69–76 (2020). https://doi.org/10.1016/j.comcom.2020.10.006
6. Javed, A.R., Shahzad, F., Urrehman, S., Zikria, Y.B., Razzak, I., Jalil, Z., Xu, G.: Future smart cities: requirements, emerging technologies, applications, challenges, and future aspects. Cities **129**, 103794 (2022). https://doi.org/10.1016/j.cities.2022.103794
7. S, M., M, R.: MUD enabled deep learning framework for anomaly detection in IoT-integrated smart building. e-Prime Adv. Electr. Eng. Electron. Energy **5**, 100186 (2023). https://doi.org/10.1016/j.prime.2023.100186

8. Yao, Y., Zhang, H., Xia, P., Liu, C., Geng, F., Bai, Z., Du, L., Chen, X., Wang, P., Han, B., Yang, T., Fang, Z.: Signature: semi-supervised human identification system based on millimeter wave radar. Eng. Appl. Artif. Intell.Artif. Intell. **126**, 106939 (2023). https://doi.org/10.1016/j.engappai.2023.106939
9. Alhamed, K.M., Iwendi, C., Dutta, A.K., Almutairi, B., Alsaghier, H., Almotairi, S.: Building construction based on video surveillance and deep reinforcement learning using a smart grid power system. Comput. Electr. Eng. **103**, 108273 (2022). https://doi.org/10.1016/j.compeleceng.2022.108273
10. Gaber, T., El-Ghamry, A., Hassanien, A.E.: Injection attack detection using machine learning for smart IoT applications. Phys. Commun. **52**, 101685 (2022). https://doi.org/10.1016/j.phycom.2022.101685
11. Sharma, A., Gupta, A.K., Shabaz, M.: Categorizing threat types and cyber-assaults over Internet of Things-equipped gadgets. Paladyn J. Behav. Robotics **13**(1), 84–98 (2022). https://doi.org/10.1515/pjbr-2022-0100
12. Prentow, T.S., Ruiz-Ruiz, A.J., Blunck, H., Stisen, A., Kjærgaard, M.B.: Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring. Pervasive Mob. Comput.Comput. **16**, 305–316 (2015). https://doi.org/10.1016/j.pmcj.2014.12.006
13. Abdulsalam, K.A., Adebisi, J., Emezirinwune, M., Babatunde, O.: An overview and multicriteria analysis of communication technologies for smart grid applications. e-Prime Adv. Electr. Eng. Electron. Energy **3**, 100121 (2023). https://doi.org/10.1016/j.prime.2023.100121
14. Chowdhury, R.R., Abas, P.E.: A survey on device fingerprinting approach for resource-constraint IoT devices: comparative study and research challenges. Internet of Things **20**, 100632 (2022). https://doi.org/10.1016/j.iot.2022.100632
15. Sun, X., Yuan, L., Wang, X.: Intelligent monitoring of home movement based on fuzzy control theory. Microprocess. Microsyst. **82**, 103943 (2021). https://doi.org/10.1016/j.micpro.2021.103943
16. Kaur, B., Dadkhah, S., Shoeleh, F., Neto, E.C.P., Xiong, P., Iqbal, S., Lamontagne, P., Ray, S., Ghorbani, A.A.: Internet of Things (IoT) security dataset evolution: challenges and future directions. Internet of Things **22**, 100780 (2023). https://doi.org/10.1016/j.iot.2023.100780
17. Ma, C., Man Lee, C.K., Du, J., Li, Q., Gravina, R.: Work engagement recognition in smart office. Proc. Comput. Sci. **200**, 451–460 (2022). https://doi.org/10.1016/j.procs.2022.01.243
18. Huseien, G.F., Shah, K.W.: A review of 5G technology for smart energy management and smart buildings in Singapore. Energy AI **7**, 100116 (2022). https://doi.org/10.1016/j.egyai.2021.100116
19. Khalil, N., Benhaddou, D., Gnawali, O., Subhlok, J.: Nonintrusive ultrasonic-based occupant identification for energy-efficient smart building applications. Appl. Energy **220**, 814–828 (2018). https://doi.org/10.1016/j.apenergy.2018.03.018
20. Malkawi, A., Ervin, S., Han, X., Chen, E.X., Lim, S., Ampanavos, S., Howard, P.: Design and applications of an IoT architecture for data-driven smart building operations and experimentation. Energy Build. **295**, 113291 (2023). https://doi.org/10.1016/j.enbuild.2023.113291
21. Gowda, V.D., Sharma, A., Rao, B.K., Shankar, R., Sarma, P., Chaturvedi, A., Hussain, N.: Industrial quality healthcare services using the Internet of Things and fog computing approach. Meas. Sens. **24**, 100517 (2022). https://doi.org/10.1016/j.measen.2022.100517
22. Nauman, A., Jamshed, M.A., Ahmad, Y., Saad, M., Bilal, M., Shanmuganathan, V., Kim, S.W.: Injecting cognitive intelligence into beyond-5G networks: a MAC layer perspective. Comput. Electr. Eng. **108**, 108717 (2023). https://doi.org/10.1016/j.compeleceng.2023.108717
23. Wirtz, B.W., Weyerer, J.C., Schichtel, F.T.: An integrative public IoT framework for smart government. Gov. Inf. Q. **36**(2), 333–345 (2019). https://doi.org/10.1016/j.giq.2018.07.001
24. Lee, J.H., Hancock, M.G., Hu, M.-C.: Towards an effective framework for building smart cities: lessons from Seoul and San Francisco. Technol. Forecast. Soc. Chang. **89**, 80–99 (2014). https://doi.org/10.1016/j.techfore.2013.08.033
25. Bai, Y., Lu, L., Cheng, J., Liu, J., Chen, Y., Yu, J.: Acoustic-based sensing and applications: a survey. Comput. Netw. **181**, 107447 (2020). https://doi.org/10.1016/j.comnet.2020.107447
26. Li, Q., Gravina, R., Li, Y., Alsamhi, S.H., Sun, F., Fortino, G.: Multi-user activity recognition: challenges and opportunities. Inf. Fusion **63**, 121–135 (2020). https://doi.org/10.1016/j.inffus.2020.06.004
27. Khan, R.H., Khan, J.Y.: A comprehensive review of the application characteristics and traffic requirements of a smart grid communications network. Comput. Netw. **57**(3), 825–845 (2013). https://doi.org/10.1016/j.comnet.2012.11.002
28. Mumtaz, S., Lundqvist, H., Huq, K.M.S., Rodriguez, J., Radwan, A.: Smart Direct-LTE communication: an energy saving perspective. Ad Hoc Netw. **13**, 296–311 (2014). https://doi.org/10.1016/j.adhoc.2013.08.008
29. Rahhal, M., Adda, M., Atieh, M., Ibrahim, H.: Health of humans and machines in a common perspective. Proc. Comput. Sci. **177**, 415–422 (2020). https://doi.org/10.1016/j.procs.2020.10.055
30. Woźniak, M., Zielonka, A., Sikora, A.: Driving support by type-2 fuzzy logic control model. Expert Syst. Appl. **207**, 117798 (2022). https://doi.org/10.1016/j.eswa.2022.117798
31. Mohanty, R., Pani, S.K.: Livestock health monitoring using a smart IoT-enabled neural network recognition system. In: Cognitive Big Data Intelligence with a Metaheuristic Approach, pp. 305–321. Elsevier (2022). https://doi.org/10.1016/b978-0-323-85117-6.00007-8
32. Raja, G.B., Chakraborty, C.: Internet of things based effective wearable healthcare monitoring system for remote areas. In: Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain, pp. 193–218. Elsevier (2023). https://doi.org/10.1016/b978-0-323-91916-6.00004-7
33. Raut, A., Shivhare, A., Chaurasiya, V.K., Kumar, M.: AEDS-IoT: adaptive clustering-based event detection scheme for IoT data streams. Internet of Things **22**, 100704 (2023). https://doi.org/10.1016/j.iot.2023.100704
34. Sharma, J., Mehra, P.S.: Secure communication in IOT-based UAV networks: a systematic survey. Internet of Things **23**, 100883 (2023). https://doi.org/10.1016/j.iot.2023.100883
35. Zhao, Z., Shen, L., Yang, C., Wu, W., Zhang, M., Huang, G.Q.: IoT and digital twin-enabled smart tracking for safety management. Comput. Oper. Res. **128**, 105183 (2021). https://doi.org/10.1016/j.cor.2020.105183
36. Hou, X., Bergmann, J.H.M.: HINNet: Inertial navigation with head-mounted sensors using a neural network. Eng. Appl. Artif. Intell.Artif. Intell. **123**, 106066 (2023). https://doi.org/10.1016/j.engappai.2023.106066
37. Adarsh, A., Kumar, B.: Wireless medical sensor networks for smart e-healthcare. In: Intelligent Data Security Solutions for e-Health Applications, pp. 275–292. Elsevier (2020). https://doi.org/10.1016/b978-0-12-819511-6.00015-7
38. Nethercote, M.: Platform landlords: renters, personal data, and new digital footholds of urban control. Digit. Geogr. Soc. **5**, 100060 (2023). https://doi.org/10.1016/j.diggeo.2023.100060
39. Lee, C.-H., Wang, C., Fan, X., Li, F., Chen, C.-H.: Artificial intelligence-enabled digital transformation in the elderly healthcare field: a scoping review. Adv. Eng. Inform. **55**, 101874 (2023). https://doi.org/10.1016/j.aei.2023.101874

40. Sampaio, H.V., Westphall, C.B., Koch, F., Do Nascimento Boing, R., Santa Cruz, R.N.: Autonomic energy management with Fog Computing. Comput. Electr. Eng. **93**, 107246 (2021). https://doi.org/10.1016/j.compeleceng.2021.107246
41. Rani, P.J., Jason, B., Praveen, K.U., Praveen, K.U., Santhosh, K.: Voice controlled home automation system using natural language processing (NLP) and Internet of things (IoT). In: Proceedings of the Third International Conference on Science Technology Engineering and Management. IEEE, Chennai, India (2017)
42. Jaihar, J., Lingayat, N., Vijaybhai, P.S., Venkatesh, G., Upla, K.P.: Smart home automation using machine learning algorithms. In: Proceedings of the International Conference for Emerging Technology, IEEE, Belgaum, India (2020)
43. Khan, S.A., Farhad, A., Ibrar, M., Arif, M.: Real time algorithm for the smart home automation based on the Internet of things. Int. J. Comput. Sci. Inf. Secur. **14**(7), 94–99 (2016)
44. Popa, D., Pop, F., Serbanescu, C., Castiglione, A.: Deep learning model for home automation and energy reduction in a smart home environment platform. Neural Comput. Appl. 1–21 (2018)
45. Machorro-Cano, I., Alor-Hernandez, G., Paredes-Valverde, M.A., Rodriguez-Mazahua, L., Sanchez-Cervantes, J.L., Olmedo-Aguirre, J.O.: HEMS-IoT: a big data and machine learning-based smart home system for energy saving. Energies **13**(1097), 1–24 (2020)
46. Singh, H., Pallagani, V., Khandelwal, V., Venkanna, U.: IoT-based smart home automation system using sensor node. In: Proceedings of the Fourth International Conference on Recent Advances in Information Technology. IEEE, Dhanbad, India (2018)