第 49 卷 第 4 期
2022 年 4 月

湖南大学学报（自然科学版）
Journal of Hunan University (Natural Sciences)

Vol. 49 No. 4
April 2022

Open Access Article https://doi.org/10.55463/issn.1674-2974.49.4.39

An Intelligent Approach for Preserving the Privacy and Security of a Smart Home Based on IoT Using LogitBoost Techniques

Asif Rahim[1], Yanru Zhong[2*], Tariq Ahmad[3], Umar Islam[4]

[1] School of Computer and Information Security, Guilin University of Electronic Technology, Guilin, China

[2] Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphic, Guilin University of Electronic Technology, Guilin, China

[3] School of Information and Communication Engineering, Guilin University of Electronic Technology, Guilin, China

[4] Institute of Computer Science and IT, The University of Agriculture, Peshawar, KPK, Pakistan

Abstract: Development and use of IoT devices have grown significantly in recent years. Many departments such as smart homes, smart healthcare, smart sports analysis, and different smart industries use IoT-based devices. In IoT devices, traffic is a very important part. IoT device traffic is distinct from traditional device traffic in various respects. In this study, 41 Internet-of-Things (IoT) devices were used. IoT devices provided 13 network traffic attributes to construct a multiclass classification model. Pre-processing techniques such as Normalization and Scaling of Dataset were used to pre-process the raw data acquired. Features can be extracted from text data using feature engineering algorithms. After stratification, the dataset contains 117,423 feature vectors utilized to develop the classification model further. Multiple performance metrics were used to demonstrate how well LogitBoost algorithms perform in this research. Using ensemble-based hybrid machine learning models to detect network anomalies in this research is an early step in developing an intrusion detection system (IDS). The main objective of this study is to detect attacks and anomalies in an IoT environment in a smart home. We have proposed a novel approach to developing LogitBoost algorithms, i.e., Logi-XGB, Logi-GBC, Logi-ABC, Logi-CBC, Logi-LGBM, and Logi-HGBC. After applying LogitBoost algorithms to the dataset for the classification, Logi-XGB scored 80.20% accuracy, and Logi-GBC scored 77.80% accuracy. Logi-ABC scored 80.33% accuracy. Logi-CBC scored the highest accuracy of 85.66%. Logi-LGBM and Logi-HGBC scored the same accuracy of 81.37%. Compared with previous LogitBoost algorithms implemented in previous studies, our proposed Logi-CBC has scored the highest accuracy on the given dataset.

Keywords: logistic regression, boosting models, machine learning.

基于物联网的罗吉特促进技术保护智能家居隐私和安全的智能方法

摘要：近年来，物联网设备的开发和使用显着增长。智能家居、智能医疗、智能运动分析等许多部门以及不同的智能行业都使用基于物联网的设备。在物联网设备中，流量是一个非常重要的部分。物联网设备流量在各个方面都不同于传统设备流量。在这项研究中，使用了 41 台物联网 (物联网) 设备。物联网设备提供了 13 个网络流量属性来构建多类分类模型。数据集的标准化和缩放等预处理技术用于预处理采集的原始数据。可以使用特征工程算法从文本数据中提取特征。分层后，数据集包含 117,423 个特征向量，用于进一步开发分类模型。多个性能指标用于展示罗吉特促进 算法在本研究中的表现。在这项研究中使用基于集成的

Received: February 19, 2022 / Revised: March 17, 2022 / Accepted: March 26, 2022 / Published: April 30, 2022

Fund Project: The National Natural Science Foundation of China (No. 62166011, No. 62033001), Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphic

About the authors: Asif Rahim, School of Computer and Information Security, Guilin University of Electronic Technology, Guilin, China; Yanru Zhong, Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphic, Guilin University of Electronic Technology, Guilin, China; Tariq Ahmad, School of Information and Communication Engineering, Guilin University of Electronic Technology, Guilin, China; Umar Islam, Institute of Computer Science and IT, The University of Agriculture, Peshawar, Pakistan

Corresponding author Yanru Zhong, rosezhong@guet.edu.cn

混合机器学习模型来检测网络异常是开发入侵检测系统 (身份识别系统) 的早期步骤。本研究的主要目的是检测智能家居物联网环境中的攻击和异常情况。我们提出了一种开发罗吉特促进 算法的新方法，即罗技-XGB、罗技-GBC、逻辑美国广播公司、逻辑加拿大广播公司、罗技-LGBM 和罗技-HGBC。将罗吉特促进 算法应用于数据集进行分类后，罗技-XGB 的准确度为 80.20%，罗技-GBC 的准确度为 77.80%。逻辑美国广播公司的准确率为 80.33%。逻辑加拿大广播公司得分最高，为 85.66%。罗技-LGBM 和罗技-HGBC 的准确率相同，均为 81.37%。与之前研究中实现的罗吉特促进算法相比，我们提出的逻辑加拿大广播公司在给定数据集上的准确度最高。

关键词: 逻辑回归，增强模型，机器学习。

1. Introduction

The scope of the study is described in this section, which serves as the introduction. It provides context for the study's findings. After that, the research problem statement is briefly presented to illustrate with what we work. After defining the scope of this study, the research statement outlines what this research is about and how it will benefit the field of study. For example, recognizing genuine and timely gadgets and security flaws that could be exploited for bad purposes are all issues that arise with the Internet of Things (IoT), a new idea in technological growth.

IoT device traffic is distinct from traditional device traffic in various respects. Regardless of their purpose or performance, devices having certain characteristics can be grouped into categories. A smart house or other dynamic and heterogeneous setting is required for this system to work. 41 Internet-of-Things (IoT) devices are used in this study. Unsupervised machine learning's LogitBoost enhanced the logistic regression method for constructing classification models. IoT devices provided 13 network traffic attributes to construct a multiclass classification model [1],

By the end of 2021, the Internet of Things is expected to impact the success of 92 percent of enterprises significantly. The Internet of Things concept is viewed as the most difficult to execute by businesses. Security, privacy, cost, and regulatory difficulties are among the issues that need to be addressed. According to research [2] conducted in 1,430 companies, the vast majority of Internet of Things users (95 percent) recognize numerous benefits from the concept (small, medium, and big).

Based on the number of devices placed in smart building environments previous to 2017, Gartner estimates that the Internet of Things concept was the most often used in smart building environments before 2017. After 2017, there are more Internet of Things devices in a smart home concept than in any other area. In a study conducted by [3], one can find detailed information regarding IoT device representation by certain application sectors. Compared to other application areas, the smart home idea has the biggest number of Internet of Things devices placed on its premises (822.6 million). According to forecasts, the Industrial Internet of Things concept (which is predicted to develop at a rate of 23.4 percent) will be the fastest-growing sector of IoT application by 2021, with a growth rate of 19.6 percent per annum. It is necessary to categorize Internet of Things devices for various reasons. As explained in this article, an Internet of Things device that fails to operate as expected or unexpectedly may indicate that a security event has occurred within the system, making it critical to correctly identify IoT devices in a specific situation and environment as described in this article. [4].

The classification and identification of new and previously unknown devices in IoT contexts and those of previously unknown devices can improve traffic management and network capacity.

The Internet of Things (IoT) is a system where items may be connected and monitored over the internet remotely [5]. The IoT concept has changed significantly in recent years and is currently being applied in different fields, such as intelligent homes, telemedicine, and industries. IoT-integrated wireless sensor network technologies offer a global connection of intelligent devices with enhanced features [6, 31]. The key to creating intelligent homes is a wireless home automation network consisting of sensors and actuators that share resources and are interconnected [7]. An "intelligent house" is part of the IoT paradigm and seeks to incorporate home automation [8], [9]. Fig. 1 depicts an example of an intelligent home with several IoT-connected utilities.

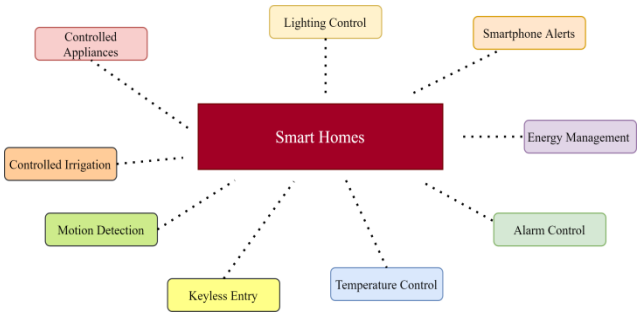


Fig. 1 Smart homes with different IoT devices connected

One of the biggest advantages of home automation systems is that they are easy to handle with various devices such as smartphones, laptops, desktops, tablets, smartwatches, or voice helpers. In addition, they enhance safety through equipment and lighting management, secure the home through automatic door locks, increase awareness by security cameras, increase convenience by adjusting temperatures, save valuable time, regulate and save money. As a result, in recent years, IoT-based home automation component shipments have been increased to a wide number of ranges (Fig. 2).

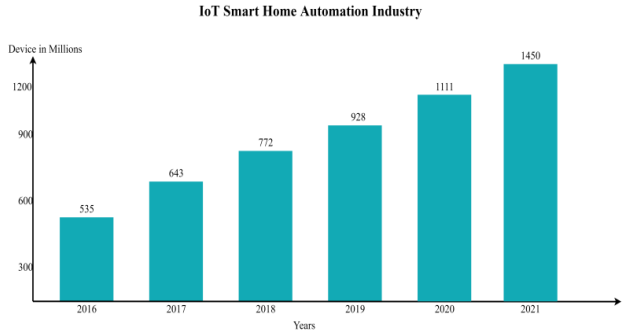


Fig. 2 IoT-based smart home automation industry sale

Using the Internet of Things (IoT) in various business areas is becoming increasingly important. However, security, privacy, pricing, and regulatory difficulties are the biggest hurdles to enterprises' implementation and use of the IoT [10], [11].

The vast majority of IoT users (95 percent) identify many advantages to the concept, according to a study done in 1,430 firms (small, medium, and big). So, 53 percent claim significant benefits from using the IoT in business. At the same time, 79% of those polled believe they accomplish positive results in different work areas that they would not otherwise be able to achieve by using the IoT concept. According to Gartner, the IoT concept was most widely deployed in smart building environments until 2017, based on the number of IoT devices used. There are more IoT devices in a smart home idea than in any other location after 2017. IHS Markit's research better understands how IoT devices are represented by individual application areas [12].

The smart home idea has the most installed IoT devices (822.6 million) compared to other applications [1]. The industrial IoT concept (CAGR of 23.4 percent) is predicted to be the fastest-expanding area of IoT application by 2021, with a CAGR of 19.6 percent for smart homes [13]. For a variety of reasons, the classification of IoT devices is critical. In a specific situation or environment, illegitimate or unauthorized IoT devices, unwanted devices, devices that do not perform as expected, and potential security threats can be identified through successful IoT device identification. In addition, the classification and identification of new, previously unknown devices can help improve traffic management and network capacity in IoT contexts.

In light of the preceding, it is hypothesized that an efficient classification model for IoT devices can be developed based on the features of the generated traffic flows. An ensemble-supervised machine learning method will be used to construct a classification model that can assign IoT devices to predefined classes based on their traffic flow values. Individual devices are the focus of current research in this field. However, such an approach is ineffective in a fast-changing, dynamic, and heterogeneous environment like the Internet of Things (IoT).

Because of this, the research in this study is original and provides the ability to identify a new and previously unidentified class of IoT devices based on their network traffic patterns. Our strategy is to generalize the identified problem and design a solution tailored to the IoT environment, which we do. For solving a problem such as managing IoT devices, detecting network anomalies created by IoT devices, or determining whether an unapproved IoT device is in the network, IoT devices need not be inspected individually. Instead, a classification model that can assign previously unknown devices to generic behavior patterns is required. According to this study, more devices were watched, more data were collected, an innovative way to classify IoT devices was used, and better results were achieved using the proposed classification model. To address the above-stated problems, the main objectives of this study are:

To detect attacks and anomalies of an IoT environment in a Smart Home

To design ensemble-based novel hybrid machine learning classification models, i.e., LogiXGBoost classifier, LogiGradientBoost classifier, LogiAdaBoost classifier, LogiCatBoost classifier, LogiLGBoost classifier, and LogiHistGBoost classifier for the prediction of attacks and anomalies in IoT-based smart home for preserving the privacy and security.

2. Literature Review

This section briefly describes the previous state of art techniques used to preserve the privacy and security of smart homes using IoT devices based on machine learning techniques.

To safeguard IoT security, Li et al. [14] devise a detecting method in this work. The system's primary roles are accomplished through supervised learning, which categorizes the malicious traffic created and

identifies the various attack types. In addition, the authors present a lightweight feature selection method that evaluates the two functions using a limited amount of characteristics. This approach automatically extracts 29 and 9 from 88 features, and then the built system has a high accuracy rate of 98.7% and 98.99% in the classification experiments. As a result of using a modest number of characteristics, our method is nevertheless very accurate.

Anthi et al. [15] proposed a three-layered IDS for smart homes. A three-layer IDS detects a variety of prevalent network-based cyber-attacks on IoT networks. The system has three major functions: 1) categorize and profile each IoT device connected to the network, 2) identify malicious packets on the network when an attack occurs, and 3) classify the attack type. An 8-device smart home testbed is used to evaluate the system. An IDS architecture is evaluated by deploying 12 attacks from 4 major network attack categories: Denial of Service (DoS), MITM/Spoofing, Reconnaissance, and Replay. The system is also tested against four multi-stage assaults with complicated event chains. The system's three basic operations perform at 96.2 percent, 90.0 percent, and 98.0 percent, respectively. In this example, an IoT device linked to the network successfully detects an attack and the suggested architecture automatically distinguishes between harmful and benign network traffic.

An alternative technique in [16] for IoT networks uses a hybrid feature selection engine that only selects the most important features and a random forest algorithm to classify traffic as normal or abnormal. The performance was examined using IoTID20, a new IoT anomaly detection dataset. The suggested approach detects DoS (99.95%), MITM (99.97%), and scanning (99.96%) attacks with excellent accuracy. Furthermore, this study can identify an IoT device using flow-based features [8].

Full-feature, reduced-feature, and flow-based-feature datasets yield 100% precision, recall, and F-score. Based on our dataset, we show that the suggested model can accurately categorize IoT devices.

Hedge et al. [17] show that the classifiers can distinguish between harmful and non-malicious behavior within a modest IoT network dataset. Additionally, the classifiers can distinguish between harmful and non-malicious activities in progressively huge datasets. Experiments have shown gradual improvements in accuracy, detection probability, and false alarm rate. 99.9 percent accuracy, 99.8 percent detection probability, and 0 percent false alarm rate are the top outcomes in terms of performance. Additionally, this paper demonstrates how these classifiers improve performance when training datasets grow larger.

In this work [17], two machine learning classification models are utilized, and the results are compared. Classification techniques based on logistic regression and artificial neural networks are used. Two distinct ways are being tested because there are about 3.5 lakh data sets. When using the algorithm described above, it is applied to 3.5 million datasets. However, the feature "value" is omitted when applying other algorithms. 75% of the available data is used in the training set, while 25% is used in the testing set. In the first scenario, an ANN achieves a precision of 99.4 percent, whereas the approach described above achieves a precision of 99.99 percent.

Aivodji et al. [18] present IOTFLA, a new smart home architecture that integrates federated learning with secure data aggregation, focusing on security and privacy.

Using a set of rules table, [19] first classifies data collected by traditional smart home Internet of Things manually, including opening and closing doors and windows, starting and stopping motors, connecting and interrupting the system, time of sending each data to label, and uses support-vector machine (SVM) algorithm to classify, build, and train models.

Early prediction on an intrusion detection system (IDS) is presented in this research [20], which uses Extreme Learning Machine and Artificial Immune System to detect anomalies in the smart home network (AIS-ELM). AIS uses the clonal algorithm to optimize input parameters for improved detection of aberrant activity, whereas ELM examines the input parameter. With this strategy in conjunction with a push notification system, the home network gateway may alert the owner of any anomalies in their network and take immediate action.

[21] aims to construct a smart home system that can take offline and online attendance and identify offenders using an image recognition algorithm. Because the epidemic has caused enormous changes in human lives, jobs and services are moving from offline to online, and digitization is rapidly expanding in education, home automation, and security. For many new emerging areas of interest, this might be the realization.

The Alibaba ECS simulated a smart home system [22]. The hardware architecture used edge computing technologies. The method develops a clear classifier to distinguish between regular and mutation codes. It can be used to detect network mutation codes. The project employed the dataset vector to categorize them positively and negatively, and the RBF-function SVM algorithm performed the best. This research improved network security detection in IoT systems and expanded machine learning applications.

Majumder et al. [24] designed an IoT security system for the smart house. Using a NoIR Pi Camera Module, a Raspberry Pi works as a security system recording movies and photos [30]. A PIR Motion Sensor detects motion. Using motion sensor data and

photos from the NoIR Pi Camera Module, the authors propose forecasting a security danger using our method's facial recognition classification. The technology can alert the user in an emergency. The proposed system can detect any security threat with 95.5 percent accuracy and 91% precision.

On a categorical DS2oS traffic traces dataset, the performance of an ensemble machine learning model is compared to a classical learning strategy for attack and anomaly detection. To optimize outcomes for anomaly detection, Khare et al. [1] merged the best models using AdaBoost ensemble learning. This paper describes the training procedure and numerous evaluation models and metrics. The dynamic surface control (DSC) and minimal learning parameter (MLP) approaches are combined [23] using neural network approximates to create an adaptive neural controller for smart home security.

The blockchain can run a machine learning model in a decentralized manner using several nodes to perform some computation. User modifications for IoT devices can be generated using our technology in a smart home IoT setup. Using a distributed association rule mining algorithm, Singla et al. [24] extract the user activity rules from the logs of the devices. The system architecture is described, and the system simulation is performed utilizing tools based on the Ethereum blockchain.

The purpose of [25] is to create a new smart OPH system that is less intrusive and less expensive while assessing individual users' life cycles. The new method uses inexpensive IoT devices to capture non-intrusive user and environmental data within the home. The technology then recognizes the user's daily actions. Eventually, the technology will quantify the user's daily cycles and provide practical advice for living a healthy life. Table 1 shows the comparative analysis of the previous state-of-art studies regarding the accuracy scores of machine learning classification models.

Table 1 Comparative analysis of previous studies

References	Dataset	Techniques	Outcomes	Accuracy
[3]	IoT Based SH Dataset	Supervised Learning Methods	Anomaly Detection in Smart Homes (SH)	81%
[26]	IoT Based SH Dataset	OCTAVE Methodology	Cyberattack detection	80%
[27]	IoTID20	Random Forests	DoS and MITM	81.04%
[28]	IoT Network	Machine Learning Models	False Alarm Detection	84%
[29]	NoIR Based IoT Security System Dataset	Machine Learning Models	Security Danger Detection	80.67%

3. Methodology

This section explains the dataset, technique, and evaluation measures in detail. Pre-processing techniques such as Normalization and Scaling of Dataset were used to further pre-process the raw data acquired. Features can be extracted from text data using feature engineering algorithms. With the pre-processed data, machine learning approaches are used to assess the model performance.

3.1. Framework

The study's three phases are shown in the figure below, along with the events represented there. After determining the study question and setting up the laboratory, the first part is complete. The dataset was built using both primary and secondary sources of data. Classifying devices into IoT classes was done in the second phase of this study. Based on their Cu index as part of the pre-processing of the datasets, features were designed and data adjusted (dealing with null and categorical values). The dataset was balanced and built a classification model in the third and final phase. Models were built using the ensemble supervised machine learning approach. Confidence matrix, accuracy, true positive ratio, false positive ratio, F-measure, and other measures were utilized to evaluate the performance of the developed classification model.

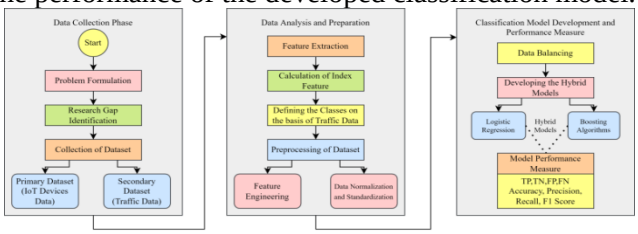


Fig. 3 Proposed framework

3.2. Phase 1 (Data Collection Phase)

Data from an open-source website, kaggle.com, has been gathered. There are several independent variables in the dataset and a single dependent one (Outcome). One hundred three pcap files are included in this study's primary dataset, including all network activity. The secondary dataset contains 144 network traffic files generated by various SHIoT devices, including 41 files in the same format as the first dataset. Table 1 shows a three-line representation of traffic on the network. The data in each of the 144 files represents the volume of web traffic generated for a particular day. The following table lists each device's categories and identifies primary or secondary in its power supply.

Table 2 Categorical distribution of devices

Device name	**Source**	**Category**
Phillips hue starter kit 2xe26	Primary	Comfort and lighting
Phillip hue starter kit 4xe26	Secondary	Comfort and lighting
Wiz colors esp_0531f3	Primary	Comfort and lighting
Wiz colors esp_0506b0	Primary	Comfort and lighting
Light bulbs LIFX smart bulb	Secondary	Comfort and lighting
Withings aura sleep tracking mat	Secondary	Comfort and lighting
Google Chromecast	Primary	Home entertainment
Invoxia Triby speaker	Secondary	Home entertainment
Pix-star photo-frame	Secondary	Home entertainment
Amazon Alexa Dot	Primary	Home entertainment
Amazon Alexa Echo	Secondary	Home entertainment
Google home mini	Primary	Home entertainment
Tplink smart plug hs110	Primary	Control and connectivity
Tplink smart plug hs105	Secondary	Control and connectivity
Mystrom switch	Primary	Control and connectivity
D-link DSP-W245 plug	Secondary	Control and connectivity
D-link DSP-W115 plug	Secondary	Control and connectivity
iHome power plug	Primary	Control and connectivity
Belkin WeMo Switch	Primary	Control and connectivity
Samsung smart things	Secondary	Control and connectivity
Blipcare blood pressure meter	Secondary	Control and connectivity
Awair air quality monitor	Secondary	Control and connectivity
iRobot Roomba 896	Secondary	Smart appliances
iRobot Roomba 895	Secondary	Smart appliances
Withings body	Secondary	Smart appliances
Smartwares c923ip camera	Primary	Security
Blink xt2 camera	Primary	Security
Canary view camera	Secondary	Security
Netatmo welcome camera	Secondary	Security
TP-Link Day/Night Cloud Camera NC220	Secondary	Security
Samsung smartcam	Secondary	Security
Nest dropcam	Secondary	Security
Belkin netcam camera	Secondary	Security
Insteon hd wifi camera	Secondary	Security
Withings Smart Baby Monitor	Secondary	Security
Belkin WeMo motion sensor	Secondary	Security
Nest protect smoke alarm	Secondary	Security

Continuation of Table 2

Device name	Source	Category
Phillips hue starter kit 2xe26	Primary	Comfort and lighting
Phillip hue starter kit 4xe26	Secondary	Comfort and lighting
Wiz colors esp_0531f3	Primary	Comfort and lighting

Standard deviation, minimum, and maximum values were determined for several variables, including the number of collected packets, file size, total data collected, average data transfer rate, and average packet size for each logical component. We also calculated the mean. These metrics represent the characteristics of the collected data. According to these examples, the secondary dataset may have a smaller average packet size or a greater overall size than the original dataset. All datasets and attributes of IoT devices have a high degree of device heterogeneity and diversity. Table 3 shows the properties of the original data. The total number of files, traffic packets, and data bytes collected over 24 hours represent all of this information. A network traffic collection technique impacts the file's size and the amount of traffic (collected data) it contains (Wireshark).

3.3. Phase 2 (Data Analysis)

The raw data was acquired. As a result, the data has been cleaned using various strategies, such as removing duplicates and null values.

3.3.1. Feature Engineering

Feature engineering is a process that leverages data from a specific domain to build functions used by learning machines. It analyses raw data and transforms it into machine-learning formats by extracting the most important attributes. The correlation matrix is utilized in this study to determine the relationship between variables. The SHIoT device categorization model was built using individual traffic filtering and a pcap file created using the device's MAC address. The IP address supplied to a device by DHCP servers (Dynamic Host Configuration Protocol) can change over time and is therefore not a reliable feature for accurately filtering traffic to a single device over time. The traffic characteristics of 41 SHIoT devices included in the study are monitored at the traffic flow level. For classifying traffic flows, packets with the same source and destination addresses as well as communication ports and protocols (TCP (Transmission Control Protocol) or UDP (User Datagram Protocol)) are grouped. According to the packet header's aggregated (statistical) data, traffic flow is chosen as the observation and analysis level best portrays communication between source and destination. Packet-level traffic analysis demands more processing power and storage capacity to store and analyze extra data. There is a correlation between the number of traffic flows and the number of packets that Google Chromecast (the device analyzed in this study) sends over 24 hours.

3.3.2. Calculation of Index Feature

Predictability in IoT device behavior is a phenomenon that has emerged as a result of research into IoT devices' communication activities. Given that SHIoT devices have limited capabilities, their behavior will be very predictable over time. A limited number of applications can be run on devices not connected to the Internet of Things (IoT).

On the other hand, IoT devices rely only on their end-users for communication activities. SHIoT devices, as a result, can be predicted using the index of the amount of predictability of IoT devices (Cu index) over time. The closer the index (Cu) gets to 0, the more predictable it is, and the less it differs from the quantity of data received and sent. It is possible to calculate an index feature:

$$C_u = Cvar_u \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x_{i*})^2}}{\frac{1}{N}\sum_{i=1}^{N} x_i} \qquad (1)$$

3.3.3. Data Pre-Processing

An important step in data mining is transforming raw data into something usable. However, our data is often partial, mismatched, or lacking altogether when it comes to some behaviors and patterns. The Cu index value was used to classify the device classes using the coefficients of variation classification approach. It assumes a normal distribution of the data. Because the derived values (Cu index) distribution is biased to the left, the data are transformed. Using the Ladder of Powers method, researchers were able to find the best data transformation function for the study to construct a normal distribution.

3.3.4. Data Normalization

Normalization is a common practice in data preparation for machine learning. To be consistent, one must normalize the data to a standard scale without distorting the range of numbers or surrendering any information.

3.4. Phase 3 (Classification Model Development)

3.4.1. Data Balancing

An obstacle to accurate predictive modeling is an

unbalanced set of classifications. It is common for machine learning algorithms to use the same number of examples in each class. It results in inaccurate models, especially for minorities. It is a problem since the minority group is more important and more susceptible to mistakes in classification than the majority group. As a result, we could eliminate the outliers from the sample and bring the data set back into line. Many more sophisticated resampling algorithms have been proposed due to this research. For example, we can aggregate most class records under sampling to conserve information by extracting records from each cluster. Instead of making exact replicas of minority class data, we can introduce small changes to these versions during the sampling process, resulting in more diversified synthetic samples. Data mining research requires a well-balanced and uniform dataset. In a dataset, "outliers" can be found. Outliers are the values in a data set that are different from the rest. The outlier has been normalized using SMOTE technique to handle the imbalanced dataset.

The outliers can be produced by misreading, faulty devices, or human mistakes. It must be omitted from the data before conducting any research or statistical tests. Any information outliner can generate partial or incorrect results, affecting the analysis and subsequent processing. The IQR approach eliminates outliers when the data boxplot exceeds the specified range. The discrepancy is the difference between the upper and lower quartiles' IQRs. Statistical approaches such as IQR, Z-Score, and Data Smoothing are used in this study to find outliers in the data. The first quartile (Q1) and third quartile (Q3) of a data set, i.e., the 25th and 75th percentiles, are used to calculate the IQR.

$$\text{IQR} = \text{Q3} - \text{Q1} \quad (2)$$

3.5. Phase 3 (Hybrid Classification Algorithms)

It relies on a limited number of complementary ways of categorization. The classification conclusion is based on a single method that solves various tasks. IoT devices can be categorized by the amount of data they send and receive. Each model's explanation is provided below.

3.5.1. Logi-XGB

This model has been developed by ensembling the logistic regression model into XGBoost Classifier to improve both models' accuracy. The mathematical model of the Logi-XGB classification model is as follows:

$$y = \alpha \sum_{t_i \epsilon\, Tree}^{T} \eta^i * leaf(t_i) \quad (a)$$

$$\ln \frac{P}{1-P} = a + by \quad (b)$$

$$\frac{P}{1-P} = e^{a+by} \quad (c)$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \quad (d)$$

Here, P is the probability function of logistic regression, and Y is the output of the XGBoost classification model. $\sum_{k=1}^{n} f(x)$ shows the boosting function of the XGB classifier. When XGB takes the output of y, it will be sent to the probability function of logistic regression for classification. Fig. 4 shows the hybrid model of the Logi-XGB classification model.

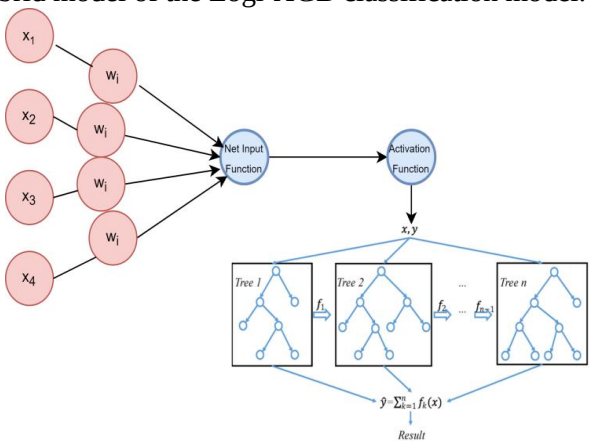


Fig. 3 Hybrid classifier (Logi-XGB classification) model

3.5.2. Logi-GBC

This model has been developed by ensembling the logistic regression model into Gradient Boosting Classifier to improve both models' accuracy. The mathematical model of the Logi-GBC classification model is as follows:

$$y = y^i = y^i + \alpha * \frac{\partial \sum (y_i - {y_i}^p)^2}{\partial y_p^i} \quad (a)$$

$$\ln \frac{P}{1-P} = a + by \quad (b)$$

$$\frac{P}{1-P} = e^{a+by} \quad (c)$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \quad (d)$$

Here, P is the probability function of logistic regression, and y^i is the output of the GBC classification model. $\frac{\partial \sum (y_i - {y_i}^p)^2}{\partial y_p^i}$ shows the sum of residuals in trees, and α is the learning rate of GBC. When GBC takes the output of y, it will be sent to the probability function of logistic regression for classification. Fig. 5 shows the hybrid model of the Logi-GBC classification model.

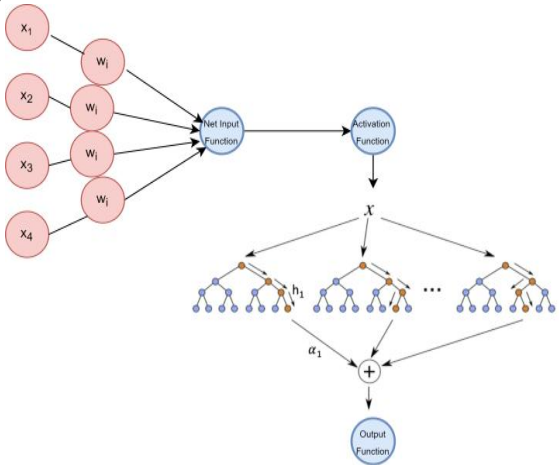


Fig. 4 Hybrid classifier (Logi-GBC classification) model

3.5.3. Logi-ABC

This model has been developed by ensembling the logistic regression model into AdaBoost Classifier to improve both models' accuracy. The mathematical model of the Logi-ABC classification model is as follows:

$$y = significance \sum_{t=1}^{T} \alpha_t h_t(x) \quad (a)$$

$$\ln \frac{P}{1-P} = a + by \quad (b)$$

$$\frac{P}{1-P} = e^{a+by} \quad (c)$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \quad (d)$$

Here, P is the probability function of logistic regression, and y is the output of the ABC classification model. $\sum_{t=1}^{T} \alpha_t h_t(x)$ shows the sum of residual in trees with significance α. When ABC takes the output of y, it will be sent to the probability function of logistic regression for classification. Fig. 6 shows the hybrid model of the Logi-ABC classification model.

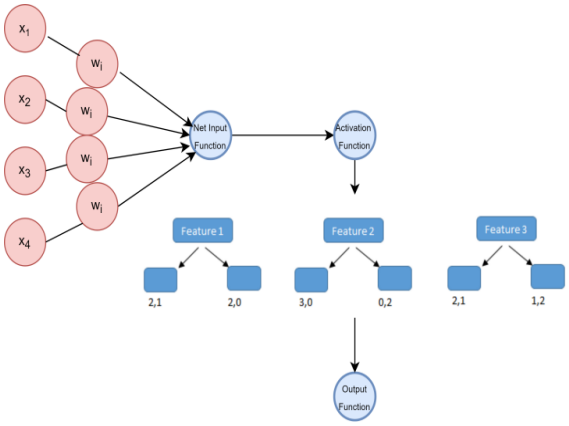


Fig. 5 Hybrid classifier (Logi-ABC classification) model

3.5.4. Logi-CBC

This model has been developed by ensembling the logistic regression model into CatBoost Classifier to improve both models' accuracy. The mathematical model of the Logi-CBC classification model is as follows:

In the first step, we will initialize the model,

$$F_o(x) = argmin_\gamma \sum_{i=1}^{n} L(y, \gamma) \quad (a)$$

For m = 1 to M, we will compute the residuals.

$$\gamma_{im} = -\left[\frac{\partial L[y, F(x_i)]}{\partial F x_i}\right]_{F(x)=F_{M-1}(x)} \quad (b)$$

Then we will fit the base leaner to compute it with pseudo residuals:

$$\gamma_{im} = argmin_\gamma \sum_{xi}^{n} L\big(y, F_{M-1}(x)\big) \quad (c)$$

Updated Model will be:

$$F_m(x) = F_{M-1}(x) + \alpha \sum_{i=1}^{n} \gamma_{im} \quad (d)$$

$$ln \frac{P}{1-P} = a + bF_m(x) \quad (e)$$

$$\frac{P}{1-P} = e^{a+bF_m(x)} \quad (f)$$

$$P = \frac{e^{a+bF_m(x)}}{1+e^{a+bF_m(x)}} \quad (g)$$

Here, P is the probability function of logistic regression, and y is the output of the CBC classification model. $\left[\frac{\partial L[y, F(x_i)]}{\partial F x_i}\right]_{F(x)=F_{M-1}(x)}$ shows the sum of residual in trees with significance α. When CBC takes the output of y as $argmin_\gamma \sum_{xi}^{n} L\big(y, F_{M-1}(x)\big)$, it will be sent to the probability function of logistic regression for classification. Fig. 7 shows the hybrid model of the Logi-CBC classification model.

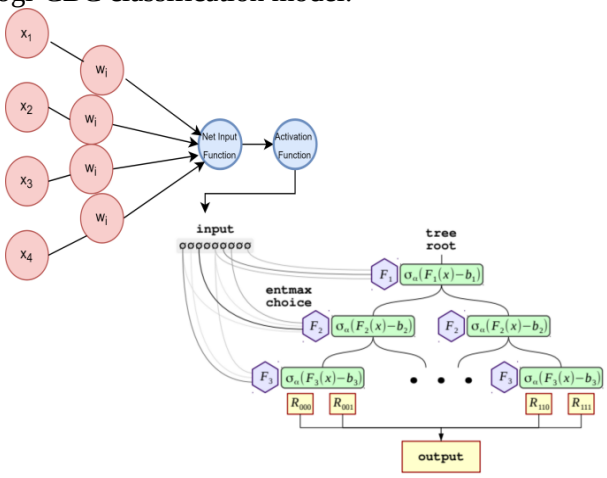


Fig. 6 Hybrid classifier (Logi-CBC classification) model

3.5.5. Logi-LGBM

This model has been developed by ensembling the logistic regression model into light-gradient boosting model classifier to improve both models' accuracy. The mathematical model of the Logi-LGBM classification model is as follows:

$$y = \alpha \sum_{t_i \epsilon\, Tree}^{T} \eta^i * leaf(t_i) \quad (a)$$

$$\ln \frac{P}{1-P} = a + by \quad (b)$$

$$\frac{P}{1-P} = e^{a+by} \quad (c)$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \quad (d)$$

Here, P is the probability function of logistic regression, and y is the output of the LGBM classification model. $\sum_{t_i \epsilon\, Tree}^{T} \eta^i * leaf(t_i)$ shows the sum of residual in leaves with learning rate α. When LGBM takes the output of y, it will be sent to the probability function of logistic regression for classification. Fig. 8 shows the hybrid model of the Logi-LGBM classification model.

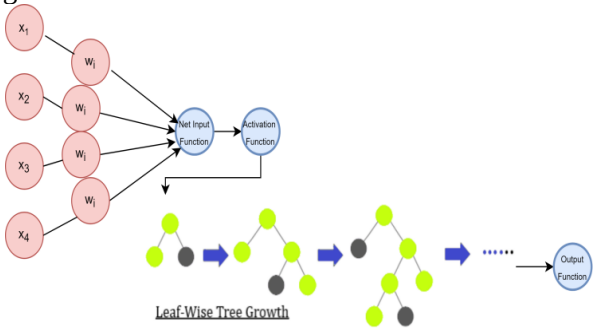


Fig. 7 Hybrid classifier (Logi-LGBM classification) model

3.5.6. Logi-HGBC

This model has been developed by ensembling the logistic regression model into histogram-based gradient

boosting classifier to improve both models' accuracy. The mathematical model of the Logi-HGBC classification model is as follows:

$$y = \frac{sum\ of\ residuals}{sum\ of\ each\ (1-p)\ for\ each\ sample\ \ in\ the\ leaf} \quad (a)$$

$$\ln\frac{P}{1-P} = a + by \quad (b)$$

$$\frac{P}{1-P} = e^{a+by} \quad (c)$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \quad (d)$$

Here, P is the probability function of logistic regression, and y is the output of the HGBC classification model. $\frac{sum\ of\ residuals}{sum\ of\ each\ (1-p)\ for\ each\ sample\ \ in\ the\ leaf}$ shows the sum of residual in trees. When HGBC takes the output of y, it will be sent to the probability function of logistic regression for classification. Fig. 9 shows the hybrid model of the Logi-HGBC classification model.

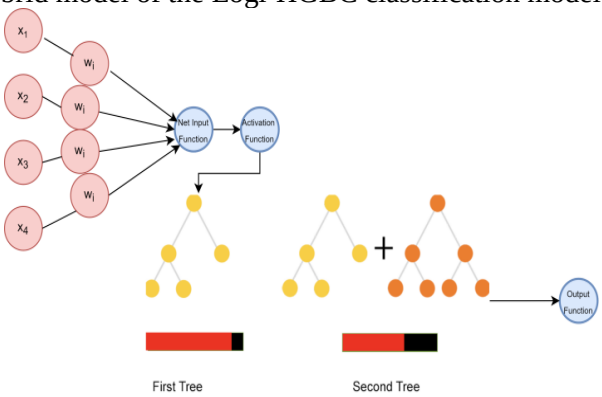
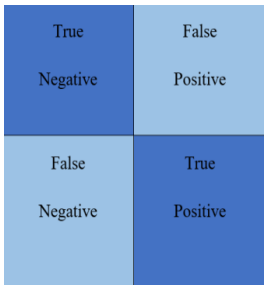


Fig. 8 Hybrid classifier (Logi-HGBC classification) model

3.6. Performance Parameters

F1 score and accuracy measures were used to evaluate the system's accuracy. The confusion matrix indicated the classified and misclassified clauses. The metrics utilized in this investigation are shown in Table 3.

Table 3 Description of the metrics

Metric	Description
Accuracy	$\text{Accuracy} = \frac{\text{TP}}{(\text{TP}+\text{TN})*100}$ True-Positive (TP): the feature result is 1, and the sample is present in this data file. True-Negative (TN): the feature result is 0, and the sample is absent in the data file.
Confusion Matrix	

4. Results

A device's class is determined by analyzing network flow data for ten days. Traffic flow feature vectors for SHIoT devices are categorized by class. The number of SHIoT traffic flows that a device creates in a given period depends on its SHIoT characteristics. Six hundred eighty-one thousand six hundred eighty-four feature vectors are separated into four classes for the initial dataset, as indicated above. Class imbalance and class balance are shown in the figure below for traffic flow classes:

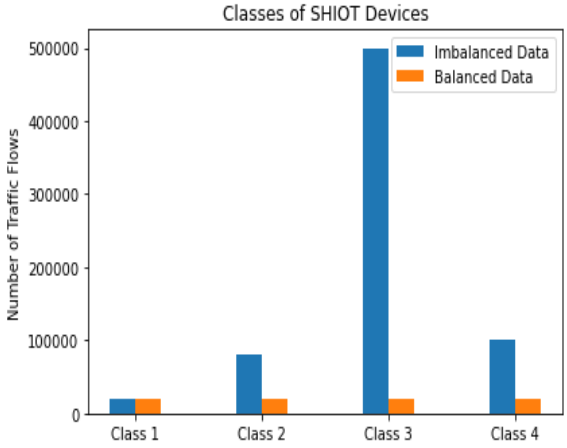


Fig. 9 Classes of SHIoT devices

As a result, the majority class was under-sampled in the dataset used to develop a classification model. The original dataset took into account the traffic flow of each unique device. Before creating a model, stratifying classes is essential to avoid model bias in classes with the most feature vectors. After stratification, the dataset contains 117,423 feature vectors that will further develop the classification model. The performance of LogitBoost algorithms has been demonstrated in this section by displaying various performance metrics. Logi-XGB scored 80.20% accuracy, and Logi-GBC scored 77.80% accuracy. Logi-ABC scored 80.33% accuracy. Logi-CBC scored the highest accuracy of 85.66%. Logi-LGBM and Logi-HGBC scored the same accuracy of 81.37%. Compared with previous LogitBoost algorithms implemented in previous studies, our proposed Logi-CBC has scored the highest accuracy on the given dataset.

4.1. Hybrid Model Logi-XGB

These two models have been combined to simultaneously increase their accuracy by using the XGBoost classifier. Logistic regression's probability function will be fed the data received from y by XGB. An independent LOGISTIC REGRESSION study found that 69.2 percent of the time, the hybrid classifier raised this accuracy to 80.20 percent. Below is a graphic depicting a hybrid model of Logi-XGB classification model performance.

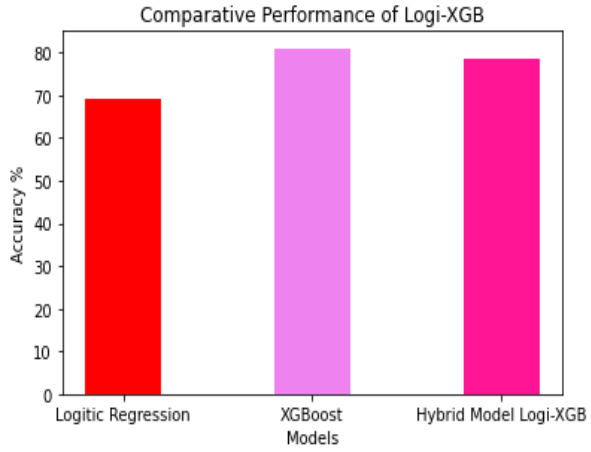


Fig. 10 Logi-XGB classification model performance

Fig. 12 shows the confusion matrix of the Logi-XGB classification model with 19 true negative, two false positive, nine false negative, and three true positive values.

Fig. 11 The confusion matrix of the Logi-XGB classification model

4.2. Hybrid Model Logi-GBC

Combining the logistic regression and gradient boosting classifier models was necessary to increase the accuracy of both models, which is how this model was constructed immediately upon receipt of y's output by the GBC. Fig. 13 illustrates the performance of the hybrid model of the Logi-GBC classification model, which achieved an accuracy of 77.8 percent.

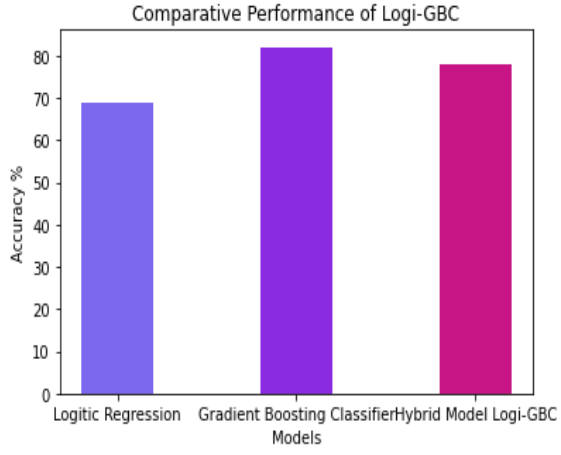


Fig. 12 Logi-GBC classification model performance

Fig. 14 shows the confusion matrix of the Logi-GBC classification model with 18 true negative, two false positive, nine false negative, and seven true positive values.

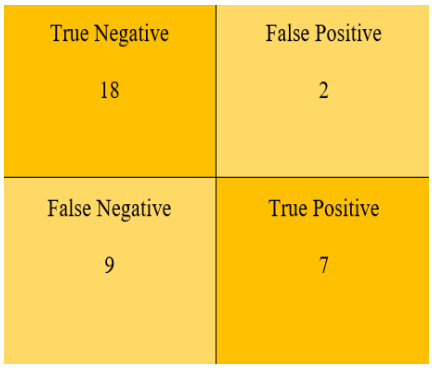


Fig. 13 The confusion matrix of the Logi-GBC classification model

4.3. Hybrid Model Logi-ABC

AdaBoost Classifier was used with the logistic regression model to increase both models' accuracy. It is submitted to the probability of logistic regression. With an accuracy rate of 80.33 percent, the hybrid Logi-ABC classification model performance model is shown in the diagram below:

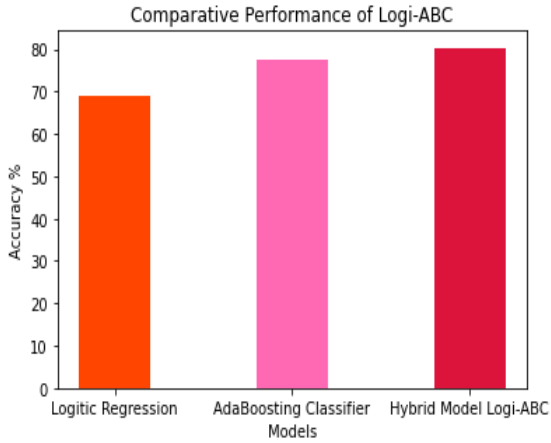


Fig. 14 Logi-ABC classification model performance

Fig. 16 shows the confusion matrix of the Logi-ABC classification model with 19 true negative, eight false positive, two false negative, and two true positive values.

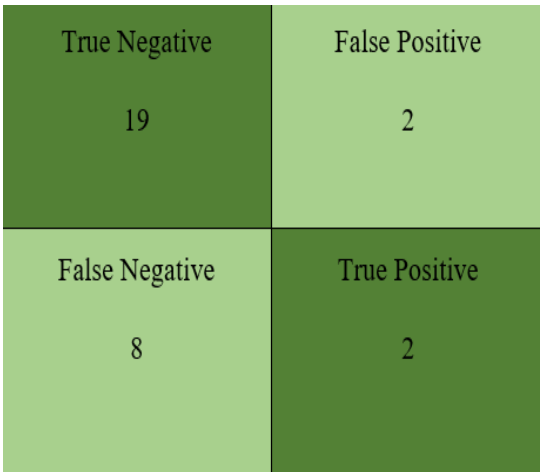


Fig. 15 Confusion matrix of the Logi-ABC classification model

4.4. Hybrid Model Logi-CBC

In order to improve the accuracy of both models, the CatBoost Classifier was utilized to combine the logistic regression model with the CatBoost Classifier. When CBC takes the output of y as L(y, F (M-1) (x)), it will be sent to the logistic regression's probability function for classification. As of this writing, Logi-CBC is the most accurate at 85.66%. Logi-CBC Classification Model is depicted as a hybrid model in the figure:

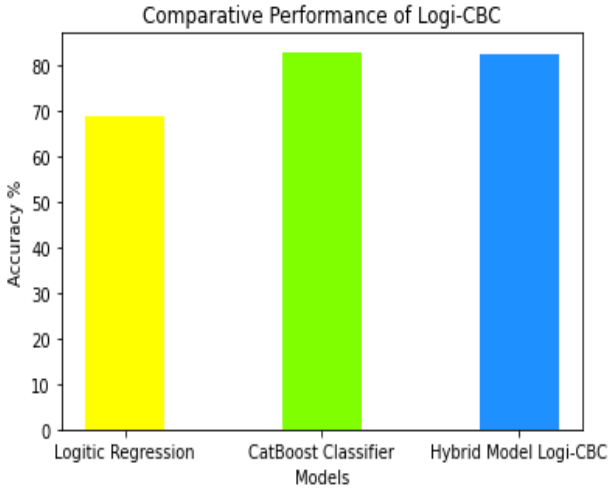


Fig. 16 Logi-CBC classification model performance

Fig. 18 shows the confusion matrix of the Logi-CBC classification model with 19 true negative, two false positive, seven false negative, and seven true positive values.

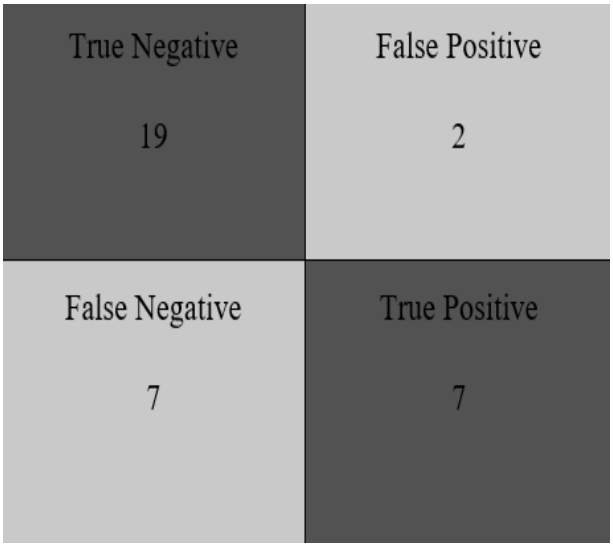


Fig. 17 The confusion matrix of the Logi-CBC classification model

4.5. Hybrid Model Logi-LGB

The accuracy of both models can be improved by combining them. The LGBM will next use the likelihood function of logistic regression to classify the attacks. The graph below shows the hybrid Logi-LGBM classification model performance with an accuracy of 81.37%.

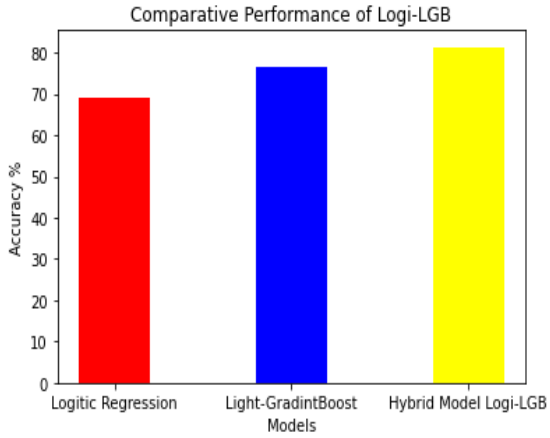


Fig. 18 Logi-LGB classification model performance

Fig. 20 shows the confusion matrix of the Logi-LGB classification model with 18 true negative, two false positive, two false negative, and seven true positive values.

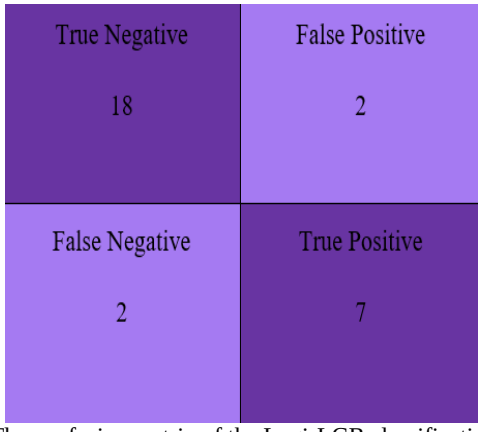


Fig. 19 The confusion matrix of the Logi-LGB classification model

4.6. Hybrid Model Logi-HBC

It was developed by merging the logistic regression model with the histogram-based gradient boosting classifier to improve both models' accuracy. As soon as the probability function of logistic regression is received by HGBC, it will be evaluated to identify whether or not a class has changed. Below is a schematic of the Logi-HGBC Classification Model's hybrid model performance with an accuracy of 81.37%.

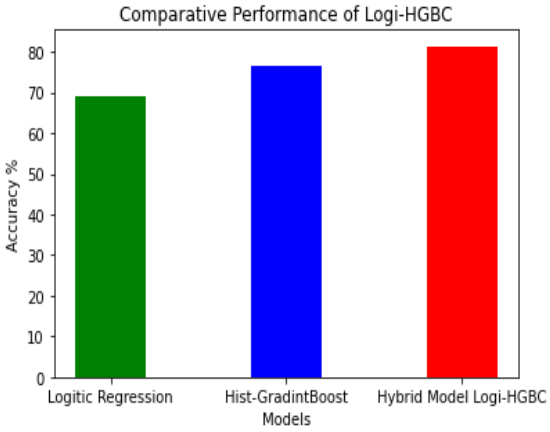


Fig. 20 Logi-HGBC classification model performance

Fig. 22 shows the confusion matrix of the Logi-HGBC classification model with 19 true negative, two false positive, nine false negative, and three true positive values.

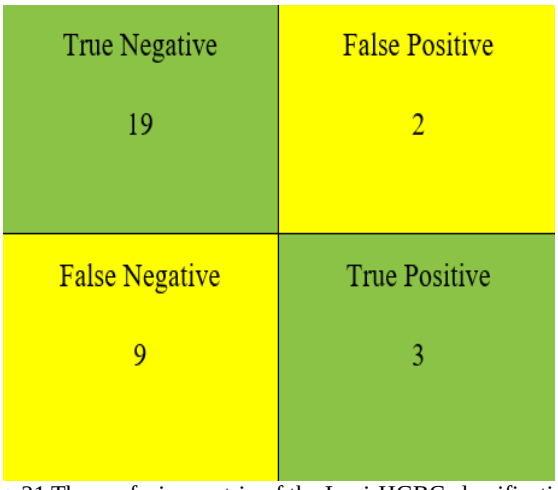


Fig. 21 The confusion matrix of the Logi-HGBC classification model

4.7. Comparative Analysis

Table 4 shows the accuracy percentage of each model. Comparatively, Logi-CBC has scored the highest accuracy in hybrid models, i.e., 85.66%. After this model, Logi-LGBM and Logi-HGBC scored 81.37% accuracy. The accuracy of Logi-GBC was poor, i.e., 77.80. Logi-XGB scored an average accuracy of 80.20%, while Logi-ABC achieved 80.33%. Comparative analysis of proposed models can be seen in the table below.

Table 4 Comparative analysis

Model	Accuracy
Logi-XGB	80.20%
Logi-GBC	77.80%
Logi-ABC	80.33%
Logi-CBC	85.66%
Logi-LGBM	81.37%
Logi HGBC	81.37%

5. Conclusion

5.1. Findings

According to the findings of a study, machine learning is capable of efficiently classifying new and unfamiliar devices and their traffic flow. It also complies with the emerging Internet of Things specifications (IoT). Because the number of devices is expanding exponentially, it is impossible to know the traffic profile of any individual device. However, it is sufficient to understand the device's class. If this novel strategy proves successful, it could pave the way for further Internet of Things-related activities and research projects. Furthermore, these discoveries will be used in a future study to detect anomalies in communication networks related to the Internet of Things (IoT devices).

Additionally, the established model may have real-world use as a software solution to improve existing IoT devices and network monitoring and control systems. This approach is capable of monitoring and managing device groupings that communicate similarly. It is also capable of predicting future device capacity and doing comparable functions. However, this method is only applicable in a few specific situations. The findings of this study will serve as a foundation for future research, real-world applications, and service provision.

5.2. Comparison

The performance of LogitBoost algorithms has been demonstrated in this study by utilizing a variety of metrics. Overall, Logi-CBC was the most accurate model in hybrid models, with an accuracy rate of 85.66 percent. Logi-LGBM and Logi-HGBC both achieved 81.37 percent accuracy while operating in this mode. However, Logi-GBC received 77.80 percent, Logi-XGB received 80.20 percent, and Logi-ABC received 80.33 percent. Comparative analysis shows that the current study has potential and is strong enough, compared to previous studies. Furthermore, we have compared our model with previous LogitBoost algorithms used in the previous state-of-art models (Table 5).

Table 5 Comparative analysis of previous studies

Reference	Dataset	Techniques	Accuracy
[3]	IoT Based SH Dataset	LogitBoost Algorithms	81%
[28]	IoT Network	LogitBoost Algorithms	84%.
[29]	NoIR Based IoT Security System Dataset	LogitBoost Algorithms	80.67%
Our Proposed Work	IoT Dataset for Smart Home	LogitBoost Algorithms	85.66%

5.3. Implications

The first stage is completed when the research question has been determined, and the laboratory has been set up. The dataset was created by combining information from both primary and secondary sources. The second phase of this study involved categorizing

devices into IoT classes. As part of the pre-processing of the datasets, features were constructed, and data was changed based on their Cu index, which was calculated as part of the Cu index calculation (dealing with null and categorical values). In the third and final phase, the dataset was balanced, and a classification model was constructed using it. This study used the supervised ensemble machine learning approach to build models. The accuracy of the classification model, the confidence matrix, the True Positive Ratio, the False Positive Ratio, the F-measure, and other metrics were used to evaluate the performance of the newly built classification model.

5.4. Strengths and Limitations

Many sectors, such as smart homes, smart healthcare, smart sports analysis, and various smart industries, rely on Internet of Things (IoT)-based gadgets to function properly. Traffic is a critical component of Internet-of-Things devices. Device traffic from the Internet of Things differs from regular device traffic. The Internet-of-Things devices supplied 13 network traffic parameters to develop a multiclass classification model. Pre-processing techniques such as normalization and scaling of the dataset were utilized to pre-process the raw data collected previously. With feature engineering algorithms, it is possible to extract features from textual data. The novel technology uses low-cost Internet-of-Things devices to collect non-intrusive user and environmental data while the user is at home. After then, the technology learns about the user's daily activities. After some time, technology will be developed that will quantify the user's daily cycles and provide practical advice on living a healthy life. However, this study must be limited in internal network intrusion detection because it trains on only traffic and external data.

5.5. Recommendations

This article suggests a novel technique for designing LogitBoost algorithms, such as Logi-XGB, Logi-GBC, Logi-ABC, Logi-CBC, Logi-LGBM, and Logi-HGBC. We also discuss the advantages and disadvantages of the proposed approach. The unique solution uses low-cost Internet-of-Things devices to collect non-intrusive user and environmental data while the user is at home, allowing the user to remain completely anonymous. The technology then begins to gather information on the user's everyday activities.

Acknowledgment

This research was funded by the National Natural Science Foundation of China (No. 62166011, No. 62033001), Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphic.

References

[1] KHARE S., and TOTARO M. Ensemble Learning for Detecting Attacks and Anomalies in IoT Smart Home. Proceedings of the 3rd International Conference on Data Intelligence and Security, South Padre Island, Texas, 2020, pp. 56–63. http://dx.doi.org/10.1109/ICDIS50059.2020.00014

[2] MACHORRO-CANO I., ALOR-HERNÁNDEZ G., PAREDES-VALVERDE M. A., RODRÍGUEZ-MAZAHUA L., SÁNCHEZ-CERVANTES J. L., and OLMEDO-AGUIRRE J. O. HEMS-IoT: A big data and machine learning-based smart home system for energy saving. *Energies*, 2020, 13(5): 1097. https://doi.org/10.3390/en13051097

[3] SPANOS G., GIANNOUTAKIS K. M., VOTIS K., and TZOVARAS D. Combining statistical and machine learning techniques in IoT anomaly detection for smart homes. Proceedings of the IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Limassol, 2019, pp. 1–6. https://doi.org/10.1109/CAMAD.2019.8858490

[4] SIVANATHAN A., GHARAKHEILI H. H., LOI F., RADFORD A., WIJENAYAKE C., VISHWANATH A., and SIVARAMAN V. Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. *IEEE Transactions on Mobile Computing*, 2019, 18(8): 1745–1759. https://doi.org/10.1109/TMC.2018.2866249

[5] STOLOJESCU-CRISAN C., CRISAN C., and BUTUNOI B. P. An iot-based smart home automation system. *Sensors*, 2021, 21(11): 3784. https://doi.org/10.3390/s21113784

[6] YAR H., IMRAN A. S., KHAN Z. A., SAJJAD M., and KASTRATI Z. Towards Smart Home Automation Using IoT-Enabled Edge-Computing Paradigm. *Sensors*, 2021, 21(14): 4932. https://doi.org/10.3390/s21144932

[7] MAHMUD S., AHMED S., and SHIKDER K. A smart home automation and metering system using internet of things (IoT). Proceedings of the International Conference on Robotics, Electrical and Signal Processing Techniques, Dhaka, 2019, pp. 451–454. https://doi.org/10.1109/ICREST.2019.8644232

[8] ULLAH I., and MAHMOUD Q. H. Network Traffic Flow Based Machine Learning Technique for IoT Device Identification. Proceedings of the IEEE International Systems Conference, Vancouver, 2021, pp. 1-8. https://doi.org/10.1109/SysCon48628.2021.9447099

[9] RATHORE A. S., XU C., ZHU W., DAIYAN A., WANG K., LIN F., REN K., and XU W. Scanning the Voice of Your Fingerprint with Everyday Surfaces. *IEEE Transactions on Mobile Computing*, 2021. https://doi.org/10.1109/TMC.2021.3049217

[10] CVITIĆ I., PERAKOVIĆ D., PERIŠA M., and GUPTA B. Ensemble machine learning approach for classification of IoT devices in smart home. *International Journal of Machine Learning and Cybernetics*, 2021, 12(11): 3179–3202. https://doi.org/10.1007/s13042-020-01241-0

[11] ZAINAB A., REFAAT S. S., and BOUHALI O. Ensemble-based spam detection in smart home IOT devices time series data using machine learning techniques. *Information*, 2020, 11(7): 344. https://doi.org/10.3390/info11070344

[12] MAKKAR A., and KUMAR N. An efficient deep

learning-based scheme for web spam detection in IoT environment. *Future Generation Computer Systems*, 2020, 108: 467–487. https://doi.org/10.1016/j.future.2020.03.004
[13] IBRAHIM A. Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-based Prediction Model. Proceedings of the IEEE International IOT, Electronics and Mechatronics Conference, Toronto, 2021, pp. 1-5. https://doi.org/10.1109/IEMTRONICS52119.2021.9422647
[14] LI T., HONG Z., and YU L. Machine Learning-based Intrusion Detection for IoT Devices in Smart Home. Proceedings of the IEEE 16th International Conference on Control & Automation, Singapore, 2020, pp. 277–282. https://doi.org/10.1109/ICCA51439.2020.9264406
[15] ANTHI E., WILLIAMS L., SLOWINSKA M., THEODORAKOPOULOS G., and BURNAP P. A Supervised Intrusion Detection System for Smart Home IoT Devices. *IEEE Internet of Things Journal*, 2019, 6(5): 9042–9053. https://doi.org/10.1109/JIOT.2019.2926365
[16] MANIRIHO P., NIYIGABA E., BIZIMANA Z., TWIRINGIYIMANA V., MAHORO L. J., and AHMAD T. Anomaly-Based Intrusion Detection Approach for IoT Networks Using Machine Learning. Proceedings of the International Conference on Computer Engineering, Network, and Intelligent Multimedia, Surabaya, 2020, pp. 303–308.
https://doi.org/10.1109/CENIM51130.2020.9297958
[17] SAHU N. K., and MUKHERJEE I. Machine learning based anomaly detection for IoT network: (Anomaly detection in IoT network). Proceedings of the 4th International Conference on Trends in Electronics and Informatics, Tirunelveli, 2020, pp. 787-794. https://doi.org/10.1109/ICOEI48184.2020.9142921
[18] AÏVODJI U. M., GAMBS S., and MARTIN A. IOTFLA: A secured and privacy-preserving smart home architecture implementing federated learning. Proceedings of the IEEE Security and Privacy Workshops, San Francisco, California, 2019, pp. 175–180. https://doi.org/10.1109/SPW.2019.00041
[19] HSU H. T., JONG G. J., CHEN J. H., and JHE C. G. Improve IoT security system of smart-home by using support vector machine. Proceedings of the IEEE 4th International Conference on Computer and Communication Systems, Singapore, 2019, pp. 674–677. https://doi.org/10.1109/CCOMS.2019.8821678
[20] ALALADE E. D. Intrusion Detection System in Smart Home Network Using Artificial Immune System and Extreme Learning Machine Hybrid Approach. Proceedings of the IEEE 6th World Forum on Internet of Things, New Orleans, Louisiana, 2020, pp. 20–21. https://doi.org/10.1109/WF-IoT48130.2020.9221151
[21] PANDIMURUGAN V., JAIN A., and SINHA Y. IoT based face recognition for smart applications using machine learning. Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, Thoothukudi, 2020, pp. 1263–1266.
https://doi.org/10.1109/ICISS49785.2020.9316089
[22] HOU S., and HUANG X. Use of Machine Learning in Detecting Network Security of Edge Computing System. Proceedings of the IEEE 4th International Conference on Big Data Analytics, Suzhou, 2019, pp. 252–256. https://doi.org/10.1109/ICBDA.2019.8713237
[23] LIU Z., DONG X., XUE J., LI H., and CHEN Y. Adaptive Neural Control for a Class of Pure-Feedback Nonlinear Systems via Dynamic Surface Technique. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(9): 1969–1975. https://doi.org/10.1109/TNNLS.2015.2462127
[24] SINGLA K., BOSE J., and KATARIYA S. Machine Learning for Secure Device Personalization Using Blockchain. Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Bangalore, 2018, pp. 67–73. https://doi.org/10.1109/ICACCI.2018.8554476
[25] NAKAMURA M. Improving Health and Quality of Life in One-Person Households Using IoT and Machine Learning. Proceedings of the IEEE 17th International Conference on Software Engineering Research, Management and Applications, Honolulu, Hawaii, 2019, pp. 1–1. https://doi.org/10.1109/sera.2019.8886791
[26] ALI B., and AWAD A. I. Cyber and physical security vulnerability assessment for IoT-based smart homes. *Sensors*, 2018, 18(3): 817. https://doi.org/10.3390/s18030817
[27] LEI X., TU G. H., LI C. Y., XIE T., and ZHANG M. SecWIR: Securing smart home IoT communications via wi-fi routers with embedded intelligence. Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services, Toronto, 2020, pp. 260–272. https://doi.org/10.1145/3386901.3388941
[28] AWAN N., KHAN S., RAHMANI M. K. I., TAHIR M., ALAM N., ALTURKI R., and ULLAH I. Machine Learning-Enabled Power Scheduling in IoT-Based Smart Cities. *Computers, Materials & Continua*, 2021, 67(2): 2449–2462. https://doi.org/10.32604/cmc.2021.014386
[29] HOQUE M. A., and DAVIDSON C. Design and implementation of an IoT-based smart home security system. *International Journal of Networked and Distributed Computing*, 2019, 7(2): 85–92. https://doi.org/10.2991/ijndc.k.190326.004
[30] PERDANA D., PAMUNGKAS P. A., and IRAWAN A. I. Smart Door System Prototype with a Control Based on Biometric Palmprint and the Internet of Things. *Journal of Southwest Jiaotong University*, 2020, 55(6). https://doi.org/10.35741/issn.0258-2724.55.6.33
[31] AL-MASHHADI H. M., and HASSAN K. R. Design and Implementation of a Smart Integrated Framework to Monitor and Control the Smart City using the Internet of Things. *Journal of Southwest Jiaotong University*, 2019, 54(6). https://doi.org/10.35741/issn.0258-2724.54.6.61

参考文:

[1] KHARE S. 和 TOTARO M. 用于检测物联网智能家居中的攻击和异常的集成学习。第三届数据智能与安全国际会议论文集，德克萨斯州南帕德里岛，2020 年，第 56-63 页。
http://dx.doi.org/10.1109/ICDIS50059.2020.00014
[2] MACHORRO-CANO I.、ALOR-HERNÁNDEZ G.、PAREDES-VALVERDE M.A.、RODRÍGUEZ-MAZAHUA L.、SÁNCHEZ-CERVANTES J.L. 和 OLMEDO-AGUIRRE J. O. HEMS-物联网：基于大数据和机器学习的智能家居节能系统。能源，2020，13(5)：1097.

https://doi.org/10.3390/en13051097

[3] SPANOS G.、GIANNOUTAKIS K. M.、VOTIS K. 和 TZOVARAS D. 在智能家居的物联网异常检测中结合统计和机器学习技术。IEEE 第 24 届通信链路和网络计算机辅助建模与设计国际研讨会(卡马德)会议记录，利马索尔，2019 年，第 1-6 页。https://doi.org/10.1109/CAMAD.2019.8858490

[4] SIVANATHAN A.、GHARAKHEILI H. H.、LOI F.、RADFORD A.、WIJENAYAKE C.、VISHWANATH A. 和 SIVARAMAN V. 使用网络流量特征对智能环境中的物联网设备进行分类。IEEE 移动计算汇刊, 2019, 18(8): 1745–1759。https://doi.org/10.1109/TMC.2018.2866249

[5] STOLOJESCU-CRISAN C.、CRISAN C. 和 BUTUNOI B. P. 基于物联网的智能家居自动化系统。传感器, 2021, 21(11): 3784. https://doi.org/10.3390/s21113784

[6] YAR H.、IMRAN A.S.、KHAN Z.A.、SAJJAD M. 和 KASTRATI Z. 使用支持物联网的边缘计算范式实现智能家居自动化。传感器，2021, 21(14): 4932. https://doi.org/10.3390/s21144932

[7] MAHMUD S.、AHMED S. 和 SHIKDER K. 使用物联网(物联网)的智能家居自动化和计量系统。机器人、电气和信号处理技术国际会议论文集，达卡，2019 年，第 451-454 页。https://doi.org/10.1109/ICREST.2019.8644232

[8] ULLAH I. 和 MAHMOUD Q. H. 基于网络流量流的物联网设备识别机器学习技术。IEEE 国际系统会议论文集，温哥华，2021 年，第 1-8 页。https://doi.org/10.1109/SysCon48628.2021.9447099

[9] RATHORE A.S.、XU C.、ZHU W.、DAIYAN A.、WANG K.、LIN F.、REN K. 和 XU W. 使用日常表面扫描指纹的声音。IEEE 移动计算汇刊，2021 年。https://doi.org/10.1109/TMC.2021.3049217

[10] CVITIĆ I.、PERAKOVIĆ D.、PERIŠA M. 和 GUPTA B. 用于智能家居中物联网设备分类的集成机器学习方法。国际机器学习与控制论杂志, 2021, 12(11): 3179–3202。https://doi.org/10.1007/s13042-020-01241-0

[11] ZAINAB A.、REFAAT S. S. 和 BOUHALI O. 使用机器学习技术在智能家居物联网设备时间序列数据中基于集成的垃圾邮件检测。信息，2020, 11(7): 344. https://doi.org/10.3390/info11070344

[12] MAKKAR A. 和 KUMAR N. 一种基于深度学习的高效物联网环境垃圾邮件检测方案。下一代计算机系统，2020 年，108：467–487。https://doi.org/10.1016/j.future.2020.03.004

[13] IBRAHIM A. 使用推特的情绪分析预测比特币的早期市场走势：基于集合的预测模型。IEEE 国际物联网、电子和机电一体化会议论文集，多伦多，2021 年，第 1-5 页。https://doi.org/10.1109/IEMTRONICS52119.2021.9422647

[14] LI T.、HONG Z. 和 YU L. 基于机器学习的智能家居物联网设备入侵检测。IEEE 第 16 届国际控制与自动化会议论文集，新加坡，2020 年，第 277-282 页。https://doi.org/10.1109/ICCA51439.2020.9264406

[15] ANTHI E.、WILLIAMS L.、SLOWINSKA M.、THEODORAKOPOULOS G. 和 BURNAP P. 智能家居物联网设备的监督入侵检测系统。IEEE 物联网杂志，2019，6(5)：9042–9053。https://doi.org/10.1109/JIOT.2019.2926365

[16] MANIRIHO P.、NIYIGABA E.、BIZIMANA Z.、TWIRINGIYIMANA V.、MAHORO L. J. 和 AHMAD T. 使用机器学习的物联网网络基于异常的入侵检测方法。计算机工程、网络和智能多媒体国际会议论文集，泗水，2020 年，第 303-308 页。https://doi.org/10.1109/CENIM51130.2020.9297958

[17] SAHU N. K. 和 MUKHERJEE I. 基于机器学习的物联网网络异常检测：（物联网网络中的异常检测）。第四届电子与信息学趋势国际会议论文集，蒂鲁内尔维利，2020 年，第 787-794 页。https://doi.org/10.1109/ICOEI48184.2020.9142921

[18] AĬVODJI U. M.、GAMBS S. 和 MARTIN A. IOTFLA：实现联合学习的安全且保护隐私的智能家居架构。IEEE 安全和隐私研讨会论文集，加利福尼亚州旧金山，2019 年，第 175-180 页。https://doi.org/10.1109/SPW.2019.00041

[19] HSU H. T., JONG G. J., CHEN J. H., 和 JHE C. G. 利用支持向量机改进智能家居物联网安全系统。IEEE 第四届计算机与通信系统国际会议论文集，新加坡，2019 年，第 674-677 页。https://doi.org/10.1109/CCOMS.2019.8821678

[20] ALALADE E. D. 使用人工免疫系统和极限学习机混合方法的智能家居网络入侵检测系统。IEEE 第 6 届世界物联网论坛论文集，路易斯安那州新奥尔良，2020 年，第 20-21 页。https://doi.org/10.1109/WF-IoT48130.2020.9221151

[21] PANDIMURUGAN V.、JAIN A. 和 SINHA Y. 基于物联网的人脸识别，用于使用机器学习的智能应用。第三届智能可持续系统国际会议论文集，托图库迪，2020 年，第 1263-1266 页。https://doi.org/10.1109/ICISS49785.2020.9316089

[22] HOU S., 和 HUANG X. 机器学习在边缘计算系统网络安全检测中的应用. IEEE 第四届大数据分析国际会议论文集，苏州，2019 年，第 252-256 页。https://doi.org/10.1109/ICBDA.2019.8713237

[23] LIU Z., DONG X., XUE J., LI H., 和 CHEN Y. 基于动态表面技术的一类纯反馈非线性系统的自适应神经控制。IEEE 神经网络和学习系统汇刊, 2016, 27(9): 1969–1975。https://doi.org/10.1109/TNNLS.2015.2462127

[24] SINGLA K.、BOSE J. 和 KATARIYA S. 使用区块链实现安全设备个性化的机器学习。计算、通信和信息学进展国际会议论文集，班加罗尔，2018 年，第 67-73 页。https://doi.org/10.1109/ICACCI.2018.8554476

[25] NAKAMURA M. 使用物联网和机器学习改善单人家庭的健康和生活质量。IEEE 第 17 届软件工程研究、管理和应用国际会议论文集，夏威夷檀香山，2019 年，第 1-1 页。https://doi.org/10.1109/sera.2019.8886791

[26] ALI B.和 AWAD A.I. 基于物联网的智能家居的网络和物理安全漏洞评估。传感器, 2018, 18(3): 817. https://doi.org/10.3390/s18030817

[27] LEI X.、TU G. H.、LI C. Y.、XIE T. 和 ZHANG M. 秒 WIR：通过具有嵌入式智能的无线上网路由器保护智能家居物联网通信。第 18 届移动系统、应用和服务国际会议论文集，多伦多，2020 年，第 260-272 页。https://doi.org/10.1145/3386901.3388941

[28] AWAN N.、KHAN S.、RAHMANI M.K.I.、TAHIR M.、ALAM N.、ALTURKI R. 和 ULLAH I. 基于物联网的智能城市中基于机器学习的电力调度。计算机、材料与连续剧，2021 年，67（2）：2449–2462。https://doi.org/10.32604/cmc.2021.014386

[29] HOQUE M. A. 和 DAVIDSON C. 基于物联网的智能家居安全系统的设计和实现。国际网络和分布式计算杂志，2019，7(2)：85-92。https://doi.org/10.2991/ijndc.k.190326.004

[30] PERDANA D.、PAMUNGKAS P. A. 和 IRAWAN A. I. 基于生物识别掌纹和物联网控制的智能门系统原型。西南交通大学学报，2020, 55(6). https://doi.org/10.35741/issn.0258-2724.55.6.33

[31] AL-MASHHADI H. M. 和 HASSAN K. R. 使用物联网监测和控制智能城市的智能集成框架的设计和实施。西南交通大学学报，2019, 54(6). https://doi.org/10.35741/issn.0258-2724.54.6.61