

Random Forest for IoT Smart the home Attack and Anomaly Detection

Bushra N. Abdul Razzaq

Department of computer science, University of Kufa,

Article Info

Article history:

Received July, 5, 2025
Revised July, 20, 2025
Accepted Aug., 15, 2025

Keywords:

Security
Intrusion
Anomaly detection
Malware extenuation
Organization and Random
Forest

ABSTRACT

The Internet of Things' growth (IoT) has also led to an increase in damaging attacks that seriously jeopardize unprotected IoT equipment. IoT technologies therefore raise a number of security and privacy concerns. Anomalies in IoT systems may enable an attacker to infiltrate a system, often resulting in unforeseen interruptions or sensor malfunctions. Consequently, it is indispensable to monitor the unforeseen occurrences resulting from sensor inputs. This study evaluates the efficacy of an ensemble mechanism knowledge model against a conventional learning method for detecting attacks and anomalies inside an IoT smart home environment, using a categorical dataset of mainSimulationAccessTraces traffic data. To achieve optimal anomaly detection outcomes, we use the Random Forest methodology to consolidate the most effective models subsequent to training each model using the training dataset. This paper elucidates many metrics and evaluation models while also providing a complete explanation of the training strategy. We also compare our findings to those of comparable model versions. Finally, we discuss difficulties and future projects.

Corresponding Author:

Bushra N. Abdul Razzaq
Department of Computer Science, University of Kufa
Email: bushran.almafrachi@uokufa.edu.iq

1. INTRODUCTION

Significant progress has been made in the creation and use of Internet of Things plans. In large part, this is because of recently created, sophisticated IoT device features that enable applications across various businesses and disciplines.

IoT enables real-time Internet-based machine-to-machine communication. Because of enabling technologies such as remote servers, practically all IoT devices in use today can send data from their sensors to authorized cloud infrastructures. Sensors in an Internet of Things device may detect changes in their surroundings.

A temperature sensor, for instance, monitors its surroundings and is useful for a variety of applications, including water temperature management, building temperature control, and refrigeration. Similar to this, there are a number of sensors, such as GPS, accelerometer, pressure, humidity, and more. Depending on the needs of the device, each Internet of Things application has a certain function.

IoT has a wide range of uses and domains, including the following: the field of health [1], for tracking and identifying human action; battles[2], to counter emerging dangers; the manufacturing sector [3], to interpret and intelligence the status and functionality of machinery; smart families [4], to manage smart illumination systems, PCs, Use applications include studying climate change, weather, and disaster predictions in environmental monitoring [6], home applications, safety cameras, and lights in the transportation sector [5], and many more. The increased availability of these smart devices may lead to a rise in possible security flaws. Among the primary objectives of IoT technology are unavoidable monitoring, anomaly identification, and environmental change or drift detection.

A popular area of research, anomaly detection in IoT devices is a relatively recent discipline. Anything that is "other than" what is considered "normal" behavior can be considered an anomaly. Our goal is to detect the development of novel events that cannot be explained by the existing models in order to maximize IoT adoption.

The goal of anomaly detection is to find anomalous activity in high-dimensional data for applications like smart IoT cameras in IoT smart homes. Additionally, data entry problems can be automatically identified and corrected by anomaly detection. In this article, we focus on seven types of anomalies: malicious activity, data probing, DDoS assaults, network scanning, illegal control, misconfiguration, and monitoring occur inside a smart home setting. These irregularities provide an attacker the opportunity to undermine the security and privacy of a smart home user. In light of this, we address the following two research issues:

- 1) In an Internet of Things setting, how can we spot anomalies and dangers?
- 2) How can we develop a machine learning-based model that outperforms the current state-of-the-art methods and is tailored for a categorical IoT dataset?

This study employs ensemble learning to mitigate volatility and bias while enhancing classification outcomes for anomaly detection across several sensor types often seen in an IoT environment. In our analysis, we employ a variety of classification and regression methodologies, including Naive Bayes, Decision Trees (DT), Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Linear Discriminant Analysis (LDA), Random Forests, Logistic Regression, Multi-Layer Perceptrons (MLP), and Artificial Neural Networks (ANN). The same training data is then amalgamated with the outcomes to construct a conclusive prediction model.

Our approach then determines an ensemble model for the categorical dataset that produces the best anomaly detection and the fewest false positives. Our method produces satisfactory results by using ensemble learning in categorical data. Therefore, the following are our contributions:

- 1) We present and put into practice an attack and anomaly detection mechanism based on ensemble learning, employing multi-class classification on categorical IoT traffic trace datasets instead of binary classification.
- 2) We demonstrate how, for attack and anomaly detection systems in IoT contexts, ensemble learning models perform better than conventional machine learning.
- 3) We compare Our approach's performance in comparison to other existing approaches.
- 4) Our proposed method identifies an ensemble model on the categorical dataset that achieves optimal anomaly detection with minimal false positives.

This paper's remaining sections are organized as follows: The relevant literature on methods for anomaly identification in diverse applications is reviewed in section II. The many evaluation models that are employed in our system are described in depth in Section III. We outline the assessment metrics that were employed in the paper to assess our model's performance in section IV. The suggested method is presented in part V, and the system details, dataset evaluation, and results are discussed in section VI. Section VII presents our findings and recommendations for further research.

2. RELATED WORK

Despite technological advancements, implementation of IoT still faces security, privacy, and authentication risks [7]-[8]-[9]-[10]. An Internet of Things device can be heterogeneous by combining numerous sensors, such as smart wearables with GPS, gyroscope, and accelerometer, clever drones with GPS and camera, and so on.

In addition, an Internet of Things device is connected to other devices and applications, which makes it vulnerable to cybercrimes. The quantity and frequency of such attacks have increased, leading to a notable increase in study in this field [11]-[12]-[13]-[14]. In spite of this, hackers try to exploit the network and private information of individuals and/or companies by breaking into IoT systems.

In order to lessen the detrimental belongings on these schemes, it is now more suitable to detect these intrusions early on. A multi-platform for monitoring and identifying irregularities in IoT systems that takes heterogeneity into account was proposed by Stiawan et al. in [15]. Their method addresses the problem of broken devices by keeping an eye on the Internet of Things network and creating a comprehensive model for early anomaly identification.

Aebebe et al. assert that deep knowledge replicas excel in accuracy, untrue alarm taxes, and scalability, as shown in [16]. Sai et al. in [17] demonstrate that using carbon-based sensors with RFID is an effective method to bolster the security of the Internet of Things ecosystem.

To detect anomalies in resource-constrained IoT devices, Al-kadi in [18] aimed to create an exceptionally lightweight deep packet method capable of distinguishing between normal and abnormal payloads. Hoang et al. [19] introduced a detection approach that employs the Hidden Markov Model (HMM) to identify all denial-of-service (DoS) attacks in test traces and reduces the training duration by 60%. In [20], Sophia et al. introduce an intrusion detection system that employs sliding windows and support vector machines (SVMs). The technology recognized selective forwarding attacks and black holes. Heba et al. [21] developed a multi-layer model that employed a C5 decision tree to improve the accuracy of intrusion detection. Lyu et al. proposed a distributed system for Fog-Empowered anomaly detection that capitalizes on the hyper-ellipsoidal clustering method and the characteristics of the Fog computing platform.

They use fog-to-things computing for testing purposes. Their study aims to minimize latency and energy usage by using fog topologies for anomaly detection. Ting et al. [23] used entity embedding and pairwise interaction to identify anomalies in complicated categorical data and developed a probabilistic model for anomaly detection to reveal outliers that deviate from the overall data distribution model. Anomaly detection is often used in time series data to find outliers or irregularities related to prior data points [24]–[25]–[26].

There are a number of techniques that can be applied to quantitative or real-world datasets. Additionally, it is easy to define anomalies in quantitative data. When minority samples in a dataset deviate from the suggested pattern of the bulk of samples, anomalies are identified.

As such, the process of creating irregularities in measurable data, both with and deprived of abnormalities, is quite straightforward. Despite being widely used in a variety of disciplines, including politics, sociology, biology, education, and so on, categorical data is given less consideration in IoT ecosystems. In other technological fields, such as network breaches, anomalies have been found using categorical datasets. [27], Social Networks [29], [30], Health and Medicine [28], Credit Fraud, Legislation [31], etc. In this study, we analyze continuous categorical datasets of IoT traffic traces within a smart house, emphasizing numerous anomalies and hazards in a monitored environment.

3. EVALUATION MODELS

We classify the attack and abnormalities found in the gathered category data using the following machine learning models.

3.1. k-Nearest Neighbour

In the k-Nearest Neighbour (KNN) classification new examples are placed together with the most related cases. In this supervised machine learning approach, a set of true classifications serves as the basis for a prediction. The algorithm determines how far apart the two locations are. The nearest neighbors are found based on pairwise distances.

The class label prediction using the closest neighbor list is then decided by a majority vote. The distance metric for kNN that we have used in this study is Euclidean distance, which is provided by:

$$\text{Euclidean Distance} = F(y) = \frac{1}{1 + e^{-(my+d)}} \quad (1)$$

Where,

(a1,b1) = first point and (a2,b2) =second point

3.2. Linear Discriminant Analysis

A linear decision boundary is established in Linear Discriminant Analysis (LDA) by fitting class-conditional densities to the data. It implements the Gaussian distribution and Bayes' theorem for each category. LDA is a dependable classifier that can also be employed to reduce dimensionality. LDA effectively ascertains whether the logarithm of the likelihood ratio is below or above a specified threshold in the presence of only two classes. When only two classes are present, the dot product is defined as follows:

$$\begin{aligned} &\vec{w} \cdot \vec{x} > c \\ &\text{where,} \\ &\vec{w} = \sum^{-1} (\vec{\mu}_1 - \vec{\mu}_2) \\ &\vec{x} \\ &\text{and} \\ &x = \text{set of observations (features, attributes, variables, measurements)} \end{aligned} \quad (2)$$

It is usual practice to use the "one-against-the-rest" method in multi-class categorization. In this method, there are κ _ binary classifiers for each κ _ class. Another prevalent method is the "pairwise" technique, which

employs $\kappa(\kappa - 1)/2$ classifiers for each pair of classes. In both scenarios, the final classification is generated as an output by combining all of the classifiers in some capacity.

3.3. Decision Tree

One type of supervised learning classifier is a decision tree. Using Attribute Selection Measures (ASM), the algorithm divides the data and selects the best attribute. The dataset is separated into smaller subsets, and the chosen property serves as a decision node. This process is repeated for every kid until specific conditions are met, resulting in a recursive decision tree.

3.4. Random Forest

Many decision trees are incorporated into random forests to minimize overfitting. k features are chosen at random from a training set that has m features in total, where $k \ll m$. The procedure is carried out n times to create n decision trees. The test object's final class is determined by adding up the votes from n decision trees.

3.5. Logistic Regression

Logistic Regression is the most widely used predictive linear machine learning model. The algorithm illustrates the correlation between a dependent binary variable and one or more independent variables. By optimizing the parameters m and d , the logistic sigmoid function in (3) produces a value between 0 and 1:

$$F(y) = \frac{1}{1 + e^{-(my+d)}} \quad (3)$$

where,

$F(y)$ = output between 0 and 1

y = input m , d = slope and intercept

3.6. Support Vector Machine

Support Vector Machines (SVMs) are supervised machine learning methods that are both reliable and versatile, and they are used for regression and classification. Frequently formed hyperplanes are the most effective method for distinguishing between the classes. The hyperplane that is selected based on its ability to suit the classes is known as the maximum marginal hyperplane (MMH). The SVM algorithm is executed by a kernel that converts a low-dimensional input space into a higher-dimensional space. The Radial Basis Function (RBF) serves primarily as our kernel for Support Vector Machine (SVM) classifications. Equation (4) provides a mathematical representation, whereby gamma is established at 0.1 throughout the learning process and varies between 0 and 1.

$$K(g,gi) = \exp(-\gamma * \sum(g - gi)^2) \quad (4)$$

where $,g \in \mathbb{R}^2$

3.7. ANN - MLP classifier

An Artificial Neural Network (ANN) that uses feed-forward technology is the Multi-Layer Perceptron (MLP) classifier. It is an algorithm for supervised learning. The network's output comprises one or more output nodes, many hidden units including nonlinear activations, and numerous input nodes that denote the characteristics. The information from the input nodes is analyzed by the units in each hidden layer, which are governed by a set of associated weights. In light of the following phrases:

x_1, \dots, x_n = input variables, converging to the unit k w_{k1}, \dots, w_{kn} = weights connecting unit k

v_k = net input

y_k = output of the unit where v_{k0} is a bias term and $\phi(\cdot)$ = the activation function

$y(i)$ = predicted value b

The output node from each hidden layer is calculated as [32]:

$$v_k = \sum_{j=1}^n w_{kj} x_j \quad (5)$$

and

$$y_k = \phi(v_k + (v_{k0})) \quad (6)$$

where ,

$\phi(v) = \frac{1}{1+e^{-v}}$ for sigmoid activation function

In feed-forward neural networks, the sigmoid node serves as the fundamental unit. Ultimately, the ANN's output receives the updated information. (7) computes the loss function based on the true and anticipated values. The multi-layer perceptron is frequently trained using the simple Backpropagation approach [33]. Gradient descent is employed to determine the weights of the neural network, with the objective of minimizing the error resulting from the loss function. The loss function is defined as follows:

$$L(y(i),y(i)) = -(y(i)\log(y(i))+(1-y(i)\log(1-y(i))) \quad (7)$$

4. IV. EVALUATION METRICS

4.1. Precision

Accurate positive identifications or values are represented by precision. The ratio of accurate positive forecasts to all positive predictions determines the quantity of information that a value conveys. In other words, it is the proportion of positive examples that are genuinely positive. While FN is less of a concern during evaluation, precision is considered more important. The precision calculation is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

4.2. Recall

The percentage of accurately labeled positive examples is called recall. The model's insight is provided by calculating the number of right predictions. Mors tatter et al. in [34] present an intriguing method to improve the recall in bot detection. Remembering is based on:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

4.3. F1-score

The F1-score computes the vocal mean by taking into account both Precision and Recall values. It is extensively employed in the information retrieval sector for credit scoring, document classification, and search measurement [35] [36]. It is between 0 and 1, where 1 is the classifier that exploits recall and exactness. The F1-score is computed as follows:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

4.4. Accuracy

The accuracy of a model is determined by computing the percentage of the time that the model's output matches the unique output in the test data, or the right findings that the classifier has produced. Accuracy is determined by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

5. PROPOSED APPROACH

5.1. Data Analysis

The suggested anomaly detection architecture for an IoT traffic trace database is shown in Figure 1. Before being input into any mechanism knowledge classifier, raw sensor data is examined and modified. Alongside arithmetical and insignificant data, the timestamp attribute is included in the mainSimulationAccessTraces data collection, which constitutes a categorical data set. The "Value" and "Accessed Node Type" attributes, which include incessant and categorical numerical values, respectively, are deficient in data. In the "Node Type Accessed," it was determined that 148 rows with "NaN" values were anomalous, along with their corresponding

labels. Thus, the "Malicious" value was used to substitute the absent "NaN" entries in the "Accessed Node Type" to protect the essential data. Correspondingly, non-continuous values such as "False," "Twenty," "True," and "none" were substituted with "0," "20," "1," and "0," correspondingly, under the "Value" attribute.

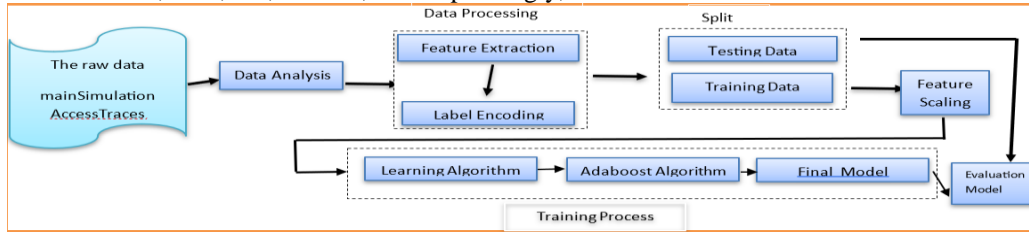


Figure 1 depicts our suggested anomaly detection framework

5.2. Data Processing

Multiple Correspondence Analysis (MCA) is used in this model to extract categorical characteristics from the categorical dataset. MCA is often referred to as Main Constituent Analysis (PCA) for definite data. PCA is a dimensionality discount method that maps data after a high-dimensional space to a lower-dimensional subspace to discern significant characteristics within a heterogeneous dataset. MCA yields the variance and factor scores that most accurately reflect categorical data.

I have performed Automated Correspondence Analysis (MCA) on the taxonomic data, displayed the results in a graph, selected the taxonomic columns from the cleaner data frame, created an MCA model with two entities, built a model on the taxonomic data, transformed it to obtain the MCA data components, and plotted the first MCA component in a graph Figure 2.

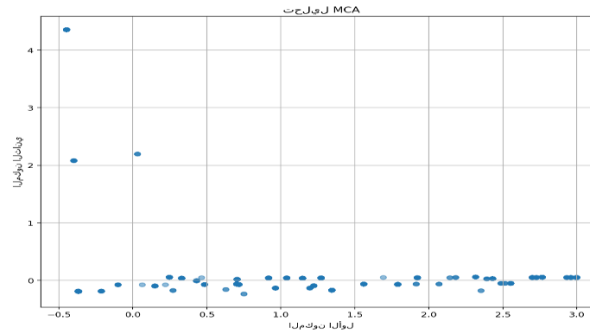


Figure 2 Multiple Correspondence Analysis (MCA).

We originate that the "timestamp" variable in the mainSimulationAccessTraces dataset has a nominal correlation with the dataset's forecast normalcy. Thus, 12 features are selected from the dataset's 85 characteristics. Once features have been identified, the categorical data must be transformed into vector form. The feature segmentation approach is currently used [37]. Because the label encoding approach uses less disk space, we choose to utilize it. One-hot encoding faces the "curse of dimensionality" as the amount of features upsurges dramatically.

5.3. Feature Scaling

After encoding the taxonomic column labels, we separated the dataset into exercise and test data. To scale the characteristics, we utilize the standard scaling approach, which is the most used scaling algorithm. When the characteristics are changed via this procedure, the distribution has a mean of 0 and a variance of 1. To avoid the exercise and test data being scaled about a mean value, it is essential to partition the data before the feature scaling phase. The performance of the models is significantly improved by standardization. All characteristics are assumed to have a variance of the same order and to be centered around 0 in models such as the RBF kernel in SVMs. This might take over the role and contribute to the complexity of learning. The traditional method centers the data by means of the following formulation:

$$V_{scaled} = \frac{(V - M)}{S} \quad (12)$$

where,
S = standard deviation
M = mean

5.4. Training Process

Section III delineates the machine learning techniques used to train the training data. The labels of an output indicate its affiliation with the anticipated class. A training dataset comprises this. Subsequently, a learning model is adeptly trained to align with the example data and discern the accurate place inside the fresh dataset. It is not always the case that a certain learning model produces the greatest results or the fewest mistakes. Because of this, we explore an ensemble learning approach in our technique, where the example's correct location is determined by applying a number of assumptions that are built on the training data. Several models' judgments are integrated in this stage, which boosts the model's overall effectiveness and results in more precise output. It also creates a model that is more stable and strong than individual models.

We employed AdaBoost as an ensemble learning boosting technique in our methodology. by constructing a pipeline from section III of successive models. To create a resilient ensemble model, weight is given to the data for faulty predictions, which are then followed by correct predictions by the next model. The classification of the dataset's anomalies is then predicted by the finished model.

6. EVALUATION

6.1. Experiment Setup

We evaluated the mainSimulationAccessTraces [38] dataset from the Distributed Smart Space Orchestration System, which is publically accessible. Twenty-one sites are included in the dataset, each of which is equipped with twenty-one movement sensors, twenty temperature control sensors, six battery sensors, five door lock sensors, four heating control sensors, three washing machine sensors, three visitor service sensors, and twenty illumination control sensors. In figure 3a, denial of service (DoS) assaults account for 58% of all anomalous data, rendering them the most prevalent type of aberrant data.

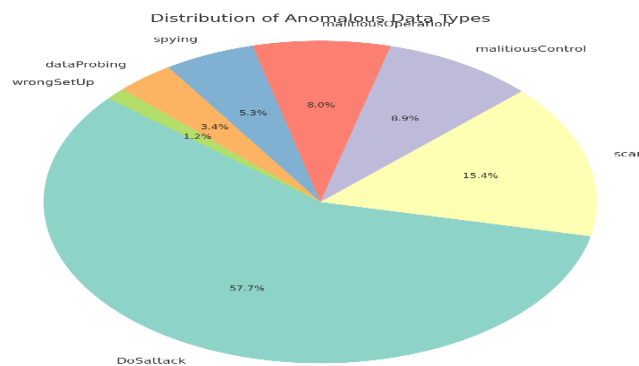


Figure 3a: distribution of anomalous data

0.43%, 0.25%, 0.22%, 0.15%, 0.10%, and 0.03%, respectively, are the other irregular data, which include scan, hateful control, malicious operation, espionage, data searching, and incorrect setup. Figure 3b shows the frequency of each of the eight groups, which include the normal data and the data that is abnormal for more than seven.

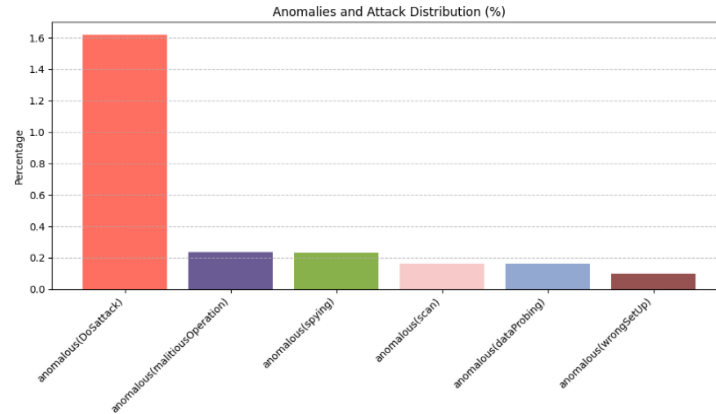


Figure 3b : The frequency of all the 8 classes which includes above 7 anomalous data and normal data is illustrated

To enhance our understanding of the data, we analyzed a histogram delivery. It contains anomalous data for each frequency of the sensor type. Figures 4a and 4b illustrate the categories of irregular data in the Light Switch (LC) and Crusade Switch (MC) Sensors.

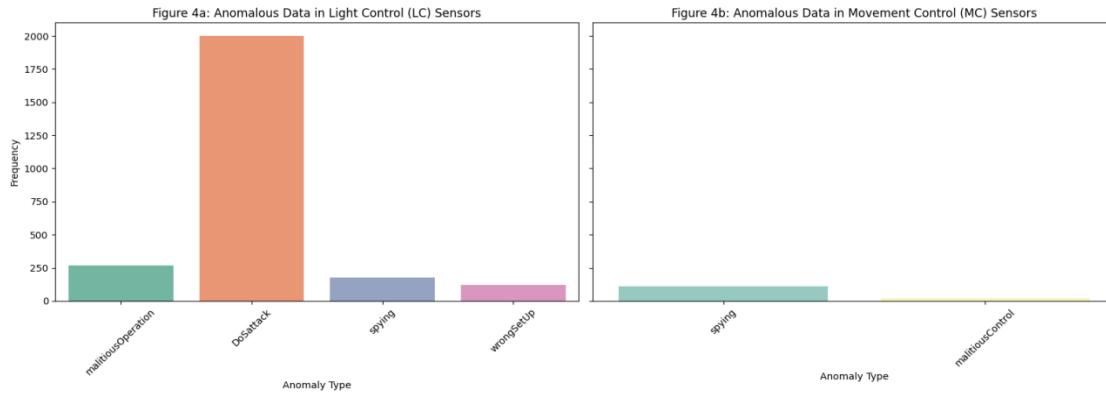


Figure 4a,4b: anomalous data in light control sensors and anomalous data in movement control sensors

We can see that from total of Number of sensors that showed abnormal data.

6.2. Results and Discussion

The assessment metrics of several machine learning techniques used to the mainSimulationAccessTraces categorical dataset are shown in Table I. These techniques are compared to the Random Forest model that is recommended in this table. This RF is the finest, in my opinion, even if machine learning techniques like KNN, ANN, and DT outperform other evaluation models like LDA, LR, SVM, ANN, and Adaboost Learning, according to the table. TABLE I : Evaluation findings from a number of machine learning algorithms on mainSimulationAccessTraces datasets

Table 1: Evaluation findings from a number of machine learning algorithms on mainSimulationAccessTraces datasets

Model	Accuracy	Precision	Recall	F1	Average
KNN	0.99	0.99	1.00	0.99	0.99
LDA	0.97	0.729190	0.43	0.54	0.67
DT	0.99	0.99	1.00	0.99	0.99
RF	1.00	1.00	1.00	1.00	1.00
LR	0.98	0.91	0.40	0.56	0.71
SVM	0.97	0.95	0.20	0.33	0.61
ANN	0.99	0.98	1.00	0.98	0.99
AdaBoost	0.99	0.96	0.74	0.83	0.88

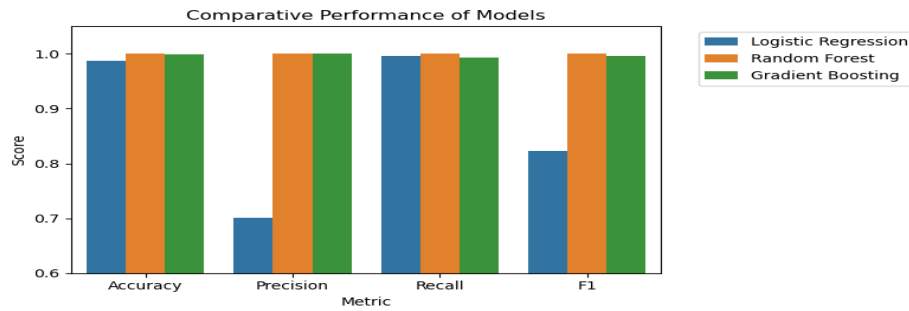


Figure 5 : Comparative Results

When learning on categorical data, the accuracy, precision, recall, and F1 are all 0.99.

Figure 2 illustrates our strategy in comparison to others. Pahl et al. employed their own synthetic dataset in [39]. In multi-class classification, the accuracy was attained at 96.3% using K-means and BIRCH clustering. By employing a neural network approach for multi-class classification on the NSL-KDD dataset, Diro et al. achieved an accuracy of 98.27% in [40]. D'Angelo et al. employed the U-BRAIN model for binary classification on NSL-KDD and actual traffic data, achieving accuracies of 97.4% and 94.1%, respectively. Unlike other studies, ours offers a comprehensive and intelligible explanation. a thorough description of the procedures involved in our recommended approach, as well as the assessment models and criteria utilized to test the suggested technique.

6.3. Challenges

Finding anomalies in categorical data presents a number of difficulties. There is contention in the literature over the definition of an outlier. Only a limited number of datasets pertaining to the IoT-sensor category are now available. Making synthetic definite data and including anomalies is more challenging than producing quantitative data. [42]. Identifying anomalies in the nominal dataset is more difficult due to its absence of a natural order. The quantity of features or observations and categorical variables inside categorical data, together with the number of categories each contains, complicates the identification of patterns in a categorical dataset. Therefore, when anomaly detection algorithms are used, they are increasing the computing complexity of the category dataset.

7. CONCLUSION

In this paper, To ensure that answer our first research query from Section I, we use conventional mechanism learning techniques like KNN, LDA, DT, SVM, and so on. In order to solve the second research problem, we use traditional machine learning models to detect anomalies and attacks by using an ensemble learning technique to a mainSimulationAccessTraces categorical IoT dataset. Our approach processes data using feature extraction and label encoding methods. The training procedure then employs the Random Forest model. Lastly, the results are evaluated using several evaluation criteria by comparing the test and training data. According to the data shown in Figure 2, our method is superior to the current one. A variety of analysis-relevant assessment models and metrics are also described in this work. In an IoT system, identifying anomalies is essential for preventing sudden disruptions like sensor failures, identifying attacker incursions, identifying unidentified security concerns, and monitoring odd user behavior. This study compares the efficacy of an collaborative machine learning model with traditional learning methods for anomaly identification in a keen home environment inside the Internet of Things framework.

We want to use a similar methodology on a much bigger IoT category dataset in our next effort. We also want to test a number of other ensemble learning strategies. Lastly, we want to test the approach in an actual IoT system to identify abnormalities and assaults

ACKNOWLEDGEMENTS (10 PT)


Many thanks to the entire respected editorial team.

REFERENCES

- [1] J. K. Hart and K. Martinez, 'Sensor Networks and Geohazards', in *Treatise on Geomorphology*, Elsevier, 2022, pp. 100–120. doi: 10.1016/B978-0-12-818234-5.00037-7.
- [2] A. Kott, A. Swami, and B. J. West, 'The Internet of Battle Things', *Computer* (Long Beach, Calif.), vol. 49, no. 12, pp. 70–75, 2016, doi: 10.1109/MC.2016.355.
- [3] Y. Liao, E. D. F. R. Loures, and F. Deschamps, 'Industrial Internet of Things: A Systematic Literature Review and Insights', *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4515–4525, 2018, doi: 10.1109/IIOT.2018.2834151.
- [4] T. Malche and P. Maheshwary, 'Internet of Things (IoT) for building smart home system', *Proc. Int. Conf. IoT Soc. Mobile, Anal. Cloud, I-SMAC 2017*, no. February, pp. 65–70, 2017, doi: 10.1109/I-SMAC.2017.8058258.
- [5] M. Saifuzzaman, N. N. Moon, and F. N. Nur, 'IoT based street lighting and traffic management system', *5th IEEE Reg. 10 Humanit. Technol. Conf. 2017, R10-HTC 2017*, vol. 2018-Janua, no. March, pp. 121–124, 2018, doi: 10.1109/R10-HTC.2017.8288921.
- [6] J. Shah and B. Mishra, 'IoT enabled environmental monitoring system for smart cities', *2016 Int. Conf. Internet Things Appl. IOTA 2016*, pp. 383–388, 2016, doi: 10.1109/IOTA.2016.7562757.
- [7] A. Riahi Sfar, E. Natalizio, Y. Challal, and Z. Chtourou, 'A roadmap for security challenges in the Internet of Things', *Digit. Commun. Networks*, vol. 4, no. 2, pp. 118–137, 2018, doi: 10.1016/j.dcan.2017.04.003.
- [8] T. Heer, O. Garcia-Morchon, R. Hummen, S. L. Keoh, S. S. Kumar, and K. Wehrle, 'Security challenges in the IP-based Internet of Things', *Wirel. Pers. Commun.*, vol. 61, no. 3, pp. 527–542, 2011, doi: 10.1007/s11277-011-0385-5.
- [9] Xu, Teng, James B. Wendt, and Miodrag Potkonjak, 'Xu, Teng, James B. Wendt, and Miodrag Potkonjak. "Security of IoT systems: Design challenges and opportunities." 2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD). IEEE, 2014.', *IEEE/ACM Int. Conf. Comput. Des. (ICCAD)*. IEEE, pp. 417–423, 2014.
- [10] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, 'Future internet: The internet of things architecture, possible applications and key challenges', *Proc. - 10th Int. Conf. Front. Inf. Technol. FIT 2012*, pp. 257–260, 2012, doi: 10.1109/FIT.2012.53.
- [11] S. Jabbar, F. Ullah, S. Khalid, M. Khan, and K. Han, 'Semantic interoperability in heterogeneous IoT infrastructure for healthcare', *Wirel. Commun. Mob. Comput.*, vol. 2017, 2017, doi: 10.1155/2017/9731806.
- [12] A. Darwish, A. E. Hassanien, M. Elhoseny, A. K. Sangaiah, and K. Muhammad, 'The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems', *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 10, pp. 4151–4166, 2019, doi: 10.1007/s12652-017-0659-1.
- [13] S. Sasirekha, S. Swamynathan, and S. Suganya, 'An ECC-based algorithm to handle secure communication between heterogeneous IoT devices', *Lect. Notes Electr. Eng.*, vol. 443, pp. 351–362, 2018, doi: 10.1007/978-981-10-4765-7_37.
- [14] M. Noura, M. Atiquzzaman, and M. Gaedke, 'Interoperability in Internet of Things: Taxonomies and Open Challenges', *Mob. Networks Appl.*, vol. 24, no. 3, pp. 796–809, 2019, doi: 10.1007/s11036-018-1089-9.
- [15] D. Stiawan, M. Y. Idris, R. F. Malik, S. Nurmaini, and R. Budiarto, 'Anomaly detection and monitoring in Internet of Things communication', *Proc. 2016 8th Int. Conf. Inf. Technol. Electr. Eng. Empower. Technol. Better Futur. ICITEE 2016, 2017*, doi: 10.1109/ICITEED.2016.7863271.
- [16] A. Abeshu and N. Chilamkurti, 'Deep Learning: The Frontier for Distributed Attack Detection in Fog-To-Things Computing', *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 169–175, 2018, doi: 10.1109/MCOM.2018.1700332.
- [17] S. Jiao and R. P. Liu, 'A survey on physical authentication methods for smart objects in IoT ecosystem', *Internet of Things (Netherlands)*, vol. 6, p. 100043, 2019, doi: 10.1016/j.iot.2019.02.003.
- [18] A. Alkadi, 'Anomaly Detection in RFID Networks', 2017, [Online]. Available: <https://digitalcommons.unf.edu/etd/768/%0Ahttps://digitalcommons.unf.edu/cgi/viewcontent.cgi?article=1813&context=etd>
- [19] X. D. Hoang and J. Hu, 'An efficient hidden markov model training scheme for anomaly intrusion detection of server applications based on system calls', *Proc. - IEEE Int. Conf. Networks, ICON*, vol. 2, no. May, pp. 470–474, 2004, doi: 10.1109/ICON.2004.1409210.
- [20] S. Kaplantzis, A. Shilton, N. Mani, and Y. A. Şekerciğlu, 'Detecting selective forwarding attacks in wireless sensor networks using support vector machines', *Proc. 2007 Int. Conf. Intell. Sensors, Sens. Networks Inf. Process. ISSNIP*, no. January, pp. 335–340, 2007, doi: 10.1109/ISSNIP.2007.4496866.
- [21] H. EzzatIbrahim, S. M. Badr, and M. A. Shaheen, 'Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems', *Int. J. Comput. Appl.*, vol. 56, no. 7, pp. 10–16, 2012, doi: 10.5120/8901-2928.

- [22] L. Lyu, J. Jin, S. Rajasegarar, X. He, and M. Palaniswami, 'Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering', *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1174–1184, 2017, doi: 10.1109/JIOT.2017.2709942.
- [23] T. Chen, L. A. Tang, Y. Sun, Z. Chen, and K. Zhang, 'Entity embedding-based anomaly detection for heterogeneous categorical events', *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-Janua, pp. 1396–1403, 2016.
- [24] N. Laptev, S. Amizadeh, and I. Flint, 'Generic and scalable framework for automated time-series anomaly detection', *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 2015-Augus, pp. 1939–1947, 2015, doi: 10.1145/2783258.2788611.
- [25] P. K. Chan and M. V. Mahoney, 'Modeling multiple time series for anomaly detection', *Proc. - IEEE Int. Conf. Data Mining, ICDM*, no. May, pp. 90–97, 2005, doi: 10.1109/ICDM.2005.101.
- [26] F. Giannoni, M. Mancini, and F. Marinelli, 'Anomaly Detection Models for IoT Time Series Data', 2018, [Online]. Available: <http://arxiv.org/abs/1812.00890>
- [27] A. Sari, 'A Review of Anomaly Detection Systems in Cloud Networks and Survey of Cloud Security Measures in Cloud Storage Applications', *J. Inf. Secur.*, vol. 06, no. 02, pp. 142–154, 2015, doi: 10.4236/jis.2015.62015.
- [28] F. Ieva and A. M. ari. Paganoni, 'Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models', *Health Care Manag. Sci.*, vol. 18, no. 2, pp. 166–172, 2015, doi: 10.1007/s10729-013-9264-9.
- [29] G. O. Campos, W. Meira, and A. Zimek, 'Outlier Detection in Graphs', pp. 1–12, 2018, doi: 10.1145/3227609.3227646.
- [30] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, 'Kdd10_Coda.Pdf'.
- [31] V. Aggarwal et al., 'Outcomes of mechanically ventilated critically III geriatric patients in intensive care unit', *J. Clin. Diagnostic Res.*, vol. 11, no. 7, pp. OC01–OC03, 2017, doi: 10.7860/JCDR/2017/23931.10126.
- [32] L. J. Lancashire, C. Lemetre, and G. R. Ball, 'An introduction to artificial neural networks in bioinformatics - Application to complex microarray and mass spectrometry datasets in cancer studies', *Brief. Bioinform.*, vol. 10, no. 3, pp. 315–329, 2009, doi: 10.1093/bib/bbp012.
- [33] M. W. Gardner and S. R. Dorling, 'Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences', *Atmos. Environ.*, vol. 32, no. 14–15, pp. 2627–2636, 1998, doi: 10.1016/S1352-2310(97)00447-0.
- [34] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, 'A new approach to bot detection: Striking the balance between precision and recall', *Proc. 2016 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2016*, pp. 533–540, 2016, doi: 10.1109/ASONAM.2016.7752287.
- [35] G. Hripesak and A. S. Rothschild, 'Agreement, the F-measure, and reliability in information retrieval', *J. Am. Med. Informatics Assoc.*, vol. 12, no. 3, pp. 296–298, 2005, doi: 10.1197/jamia.M1733.
- [36] W. Chen and L. Shi, 'Credit scoring with F-score based on support vector machine', *Proc. - 2013 Int. Conf. Mechatron. Sci. Electr. Eng. Comput. MEC 2013*, pp. 1512–1516, 2013, doi: 10.1109/MEC.2013.6885307.
- [37] D. K. Barupal and O. Fiehn, 'Generating the blood exposome database using a comprehensive text mining and database fusion approach', *Environ. Health Perspect.*, vol. 127, no. 9, pp. 2825–2830, 2019, doi: 10.1289/EHP4713.
- [38] D. Rani, N. S. Gill, P. Gulia, F. Arena, and G. Pau, 'Design of an Intrusion Detection Model for IoT-Enabled Smart Home', *IEEE Access*, vol. 11, pp. 52509–52526, 2023, doi: 10.1109/ACCESS.2023.3276863.
- [39] M. O. Pahl and F. X. Aubet, 'All Eyes on You: Distributed Multi-Dimensional IoT Microservice Anomaly Detection', *14th Int. Conf. Netw. Serv. Manag. CNSM 2018 Work. 1st Int. Work. High-Precision Networks Oper. Control. HiPNet 2018 1st Work. Segm. Routing Serv. Funct. Chain. SR+SFC 2*, pp. 72–80, 2018.
- [40] A. A. Diro and N. Chilamkurti, 'Distributed attack detection scheme using deep learning approach for Internet of Things', *Futur. Gener. Comput. Syst.*, vol. 82, pp. 761–768, 2018, doi: 10.1016/j.future.2017.08.043.
- [41] G. D'Angelo, F. Palmieri, M. Ficco, and S. Rampone, 'An uncertainty-managing batch relevance-based approach to network anomaly detection', *Appl. Soft Comput. J.*, vol. 36, pp. 408–418, 2015, doi: 10.1016/j.asoc.2015.07.029.
- [42] G. O. Campos et al., 'On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study', vol. 30, no. 4, *Springer US*, 2016, doi: 10.1007/s10618-015-0444-8.

BIOGRAPHIES OF AUTHORS

 A portrait of a woman wearing a blue hijab with a gold and white patterned border. She is looking directly at the camera with a neutral expression.	<p>Assistant Lecturer Bushra Naeem Abdul Razzaq, obtained her bachelor's degree from the University of Babylon / College of Science - Department of Computers in 1996 and her master's degree from Iran / Azad University - Department of Computer Network Engineering in 2023. I am currently working as an assistant lecturer at the University of Kufa / College of Education for Girls - Department of Computers. He can be contacted at Email: bushran.almafrachi@uokufa.edu.iq</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------