ARTICLE

JOURNAL`S UNIVERSITY OF BABYLON FOR
ENGINEERING SCIENCES (JUBES)
مـــجلـة جـــامعة بـــابـل للعلـــــوم الهندسية

Vol. 3٣, No.6. \ 202٥ ISSN: 2616 - 9916

Using Combination of Independent Component Analysis and Decision Tree algorithm for Botnet detection in IoT Devices

Zaid Raheem mal.[1a] **Salam Mohammed Salam** [1b] **Ali Asghar Safaei** [1c]

[1] *Ministry of electricity, Iraq*

[2] *Ministry of Health, Iraq*

[3] *Azid University, Tehran, Iran*

**Corresponding Author: Zaid Raheem mal., Salam Mohammed Salam,*

Dr. Ali Asghar Safaei

[1] Zaidraheem215@gmail.com ,[2] salammohammed427@gmail.com

Received:	**13/10/2025**	**Accepted:**	**29/12//2025**	**Published:**	**31/12/2025**

Abstract

Bot malware and botnet are two widely understood concepts in the cyber security literature. Specifically, a botnet is a geographically dispersed network of infected bots (such as any computing device, including an Internet of Things (IoT) device such as a smart TV, that is compromised by bot malware), which is remotely controlled by a bot. The master is controlled. Such botnets are commonly used to carry out a wide range of malicious cyber activities, from sending spam to launching distributed denial of service (DDoS) attacks to spreading malicious programs (malware) to distributing illegal material (such as child abuse material). In this research, botnet detection systems were investigated and methods to optimize and increase their efficiency in order to identify botnets developed. The chosen method for this work is the combination of independent component analysis and the decision tree algorithm and the single decision tree 15 method, which we discussed with the classification of the data used, i.e., UNSW-NB features, and the 40 dataset contains 15 The chosen method for this work is the combination of independent component analysis and decision tree algorithm and single decision tree 15method, which we discussed with the classification of the data used i.e. UNSW-NB features and the 40 dataset contains 15 Since the UNSW. NB2018and CiCIDS features, we investigated the use of the two approaches in 78 datasets has 2018CiCIDS terms of accuracy, correctness and error criteria. The accuracy obtained for botnet detection according to the approach of combining independent component analysis percent 99.994data set is equal to 15and decision tree algorithm on the UNSW-NB. Also, the amount of RMSE error 99.177 data set is equal to 2018 and on the CiCIDS and on the 0.0076 data set is equal to a small value of 15obtained on the UNSW-NB 0.0076 data set equal to 2018CiCIDS This research focuses on studying and improving robot network detection systems to enhance their identification efficiency. The proposed methodology uses a hybrid approach combining Independent Component Analysis (ICA) and Decision Tree (DT) algorithms, and compares its performance to an independent decision tree model.

Keywords: Botnet, Cyber, Intrusion detection, Machine learn.

1. Introduction

On the serious growth of problems caused by social botnets. Bushmoff et al.[1] observed that defending against malicious bots raises a set of unique challenges related to the web, including online and offline authentication automation and security. They concluded that a successful penetration campaign has three critical security issues:

First, the target structure of online social networks can be dangerous and contaminated with a large number of unreal connections.

Second, an intruder can destroy users' privacy by collecting large amounts of private data.

Third, Finally, the enemy can use the infiltrated OSNs of Orbit Showtime Network to spread false information. Researchers have proposed several constructive approaches to analyze and identify social botnets, as well as to deal with these challenges, which can be divided into two categories by the location of detection: Host- based detection method: This method tries to identify malware on the host computer. In this method, the internal state of the resources and the behavior of the host are checked and monitored. The pattern of commands in the binary code is a set of system.

Although they may exist in bot binary code, they can also be used as malware signatures. Botnets are one of the biggest Internet threats of our time and their importance is increasing day by day. New design features, attack methods and targets will likely make the next generation of botnets equally dangerous and more difficult to counter. By carrying out this project, we are trying to identify botnets, penetration methods and provide a suitable solution to deal with this malware.

Botnet has a strong negative effect on many computers and through commands and controls, it creates a wide network of infected devices, the purpose of which is to spread malicious codes, launch DDoS attacks, send spam, phishing, and add spyware. The big problem with botnet happens when it attacks the target. Botnet is even capable of threatening national security and causes panic disorder and excessive use of network resources. Therefore, the government should spend a lot of money to prevent and treat botnet attack. Botnet attack is carried out using thousands and even millions of computers, so the impact of the attack is very high. Several authors have cited botnet spamming as a major concern due to the widespread distribution of spam that consumes many resources on the network.

2. Related Work

The Internet of Things is an information technology based on the Internet that facilitates the exchange of goods and services. The purpose of IOT is to provide an information technology infrastructure and facilitate the exchange of things in a safe and reliable way, its function to overcome the gap between objects in the physical world and their representation in information systems. IOT in common words means connecting physical objects and humans by sensors and using the collected information to monitor, detect or predict physical situations and events. In the IOT environment, physical and virtual environments "have identity and characteristics and can use intelligent interfaces as an information network [2]. Basically, IOT can be considered as a wonderful set of connecting devices that are connected by technology. Identify the existing unique communications. The word "Internet" and "things" mean a global network connected around the world based on sensitivity communication, network and information processing technology, which may be the new version of information and communication technology. The term connected devices is named when the provision of technology (RFID) to track items in a supply chain is known as the communication system where the Internet is and through the widespread use of devices to the physical world is connected [3]. The main goal of the Internet of Things is to integrate heterogeneous devices.

3.Necessity of Research

The purpose of this research is to use the combination of independent component analysis and decision tree algorithm to identify botnets in Internet of Things devices. The presented method is new[4].

4.The Research Questions

1. Is it possible to identify botnets in Internet of Things devices by using the combination of independent component analysis and decision tree algorithm?
2. Is this method more effective than similar methods?

3. Proposed Methodology

Based on the above flowchart, the steps to achieve the goals by implementing the proposed method are as follows:

Part I: Preprocessing

If the values of the features of the data set are in a different range, the possibility of errors in the findings increases. Placing the data of a statistical population in the same domain is called normalization. In the proposed model, the method of normalization is to use the following relationship. The standard form is to put all the data between d1 and d2 using the following formula:(3-1) According to the data, d1=0 and d2=+1. In other words, using this relationship, all the data are placed in the [0,1] range. In a data set, there is a possibility of missing values for records. The data in a dataset must be complete without missing values or incomplete data when it enters the algorithm. Also, cases where possibly wrong values have been assigned to the attributes of a record should be corrected and if not corrected, removed from the data set. Unfortunately, there are missing values in the DDoS attacks dataset. In this study, the maximum possible value was used for the missing values [5]. In the maximum possible value method, among the acceptable values for a particular feature, its maximum value is selected for replacement.

part 2: feature selection

After the stage of reading the data set of botnet attacks in the Internet of Things and the pre-processing operation on the data, the feature selection operation is performed. In this algorithm, N is the number of parameters and D is the number of decision variables or dimensions of the optimization problem. Therefore, feature selection is simulated by an N*D matrix. Each row corresponds to a possible solution of the optimization problem. In the proposed model, N is the number of data set records and D is the number of features and is defined according to equation (3-2). In the proposed model, the working method for the botnet dataset is such that the proposed algorithm consists of 40 features. Each iteration is defined by the 22 available in the database.

□ $Population\ of\ GWO = x11 \quad x12 \quad x22 \quad x22 \vdots \quad \vdots xn1 \quad xn2 \cdots x1d \ldots x2d \quad \vdots \ldots \quad xnd$ (3-2)

□ In the set $\boldsymbol{xi} = \boldsymbol{xi1}, \boldsymbol{xi2}, \ldots, \boldsymbol{xid}, \boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \ldots., \boldsymbol{n}$ Each x_i represents a possible solution in the solution space Is. The evaluation of each answer is calculated based on the objective function according to equation (3-3).

$fiti = 1 - Obji{-}worst(Obj)$

$best\ Obj - worst(Obj)$ (3-3)

□ In equation (3-3), fit is the fitness of the solution. The Obj _i parameter is the value of the objective function for the answer. To convert numbers to binary.

The Proposed Method

part 3: classification

For classification, first, it is necessary to divide the data set into two parts: training (80% of samples) and testing (20% of samples). The data of the training section produces the evaluation model, and the data of the test section tests the produced model with the help of a number of records and determines the corresponding label of the said records and determines their class. In the decision tree algorithm, according to the data, they start creating a tree structure that works like the IF and ELSE rules and finally reach the labels learned from the training data. In fact, the learning operation in the decision tree is the construction of elements and leaves of a tree. In this research, the ID3 decision tree method is used to classify the trained features Simulation and Results.

The botnet detection system is implemented and the results of independent component analysis algorithm and decision tree algorithm in two modes of using deep neural networks and shallow neural networks approaches on two UNSW-NB15 botnet detection datasets are discussed. and CICIDS2018 based on evaluation criteria. In order to evaluate each of these approaches, evaluation criteria are used.

It should be noted that all the results of program coding have been done in Python software and in a system with Core i5 processor power and 8 gigs of RAM. In this project, in order to achieve detection and improve the detection of botnet attacks in the intrusion detection system using the proposed method, first the data set related to different attacks should be identified and collected. Therefore, in order to implement the method, we have used two data sets UNSW-NB15 and CICIDS2018. The UNSW-NB15 dataset was developed in a hybrid environment by the Cyber Security Research Group at the Australian Cyber Security Center (ACCS) in 2015. The UNSW- NB15 network traffic

Contains real modern natural hybrid non-conventional network traffic. The network traffic records of the UNSW-NB15 dataset contain 40 adaptive features and one tagged feature. There are ten types of network traffic records. Attack logs are grouped into nine families according to the nature of the attacks, Recon Nais sane, Backdoors, DOS, Exploits.

On the other hand, the CICIDS2018 dataset was created in 2018 by the Cyber Security Institute (CIC) of Canada, which contains updated attack records [6]. Attacks implemented in this database include Brute Force, DoS, Botnet, DDoS, Port Scan, Web Attack, Infiltration. The dataset extracts 78 features from the simulated network traffic, the last column being the traffic label. Before presenting the results, it is necessary to mention that due to the high volume of each of the aforementioned data sets, which in the original version contained more than 1 million data series, and considering that the processing of this volume of data requires a supercomputer. In this research, in both shallow learning and deep learning processing modes, we used only a part of this data that was randomly taken from the whole data. Based on this, we have used 257,673 records for the UNSW-NB15 dataset and 57,513 records for the CiCIDS2018 dataset in order to analyses and evaluate the results.

Based on this, in the following sections, the results obtained from the application of the neural network algorithm with the shallow learning approach and with the deep network approach are presented on both databases. Independent component analysis (ICA) is a sophisticated computational technique used in signal processing and machine learning to separate a multivariable signal into statistically independent, associative subcomponents, focusing on maximizing variance, while ICA aims to achieve statistical independence [7]. This paper explores the underlying theory, mathematical framework, and diverse applications.

Simulation and Results.

Tables 1 and 2 present the results of the decision tree approach on UNSW-NB15 and CiCIDS2018 datasets respectively.

Table 1: Evaluation result of decision tree method on UNSW-NB15 dataset

Value	Evaluation criteria
50796	TP
734	TN
6	FP
0	FN
99.988	Precision
1	Sensitivity
99.988	Accuracy
68.6419	Specification
0.99994	F1_Score

Table 2: Evaluation result of decision tree method on CiCIDS2018 dataset.

value	Evaluation criteria
4665	TP
502	TN
80	FP
8	FN

98.314	Precision
0.99829	Sensitivity
98.325	Accuracy
8.0155	Specification
0.99066	F1_Score
□. □□□□ × □□−□	MSE
0.0147	R M S E

4. Simulation and Results

Based on the results obtained in the above two tables, it is clear that the accuracy obtained for detecting Ethernet attacks according to the decision tree approach in the UNSW-NB15 dataset is 99.988%. Also, the obtained RMSE error is equal to a small value of 0.0108.

Also, according to the application of the shallow neural network method on the CiCIDS2018 data set, it was found that the performance of this method in botnet detection has an accuracy of 98.325% and the RMSE error is equal to 0.0147. The reason that the CiCIDS2018 dataset has less accuracy than the UNSW-NB15 dataset can be seen in two reasons: the number of data is much lower and the number of features is much higher.

Now, in order to more accurately evaluate the integrated methods of independent component analysis and decision tree algorithm and decision tree on CiCIDS2018 and UNSW-NB15 datasets, we use the bar chart of Figure (3) to compare the accuracy

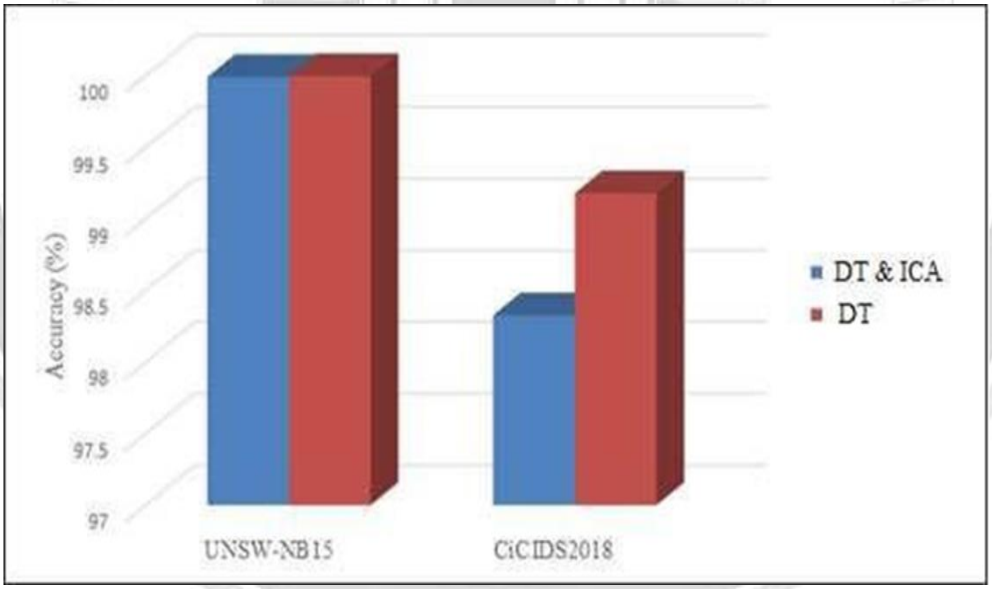


Figure 3: Comparing the RMSE of the methods used on both databases

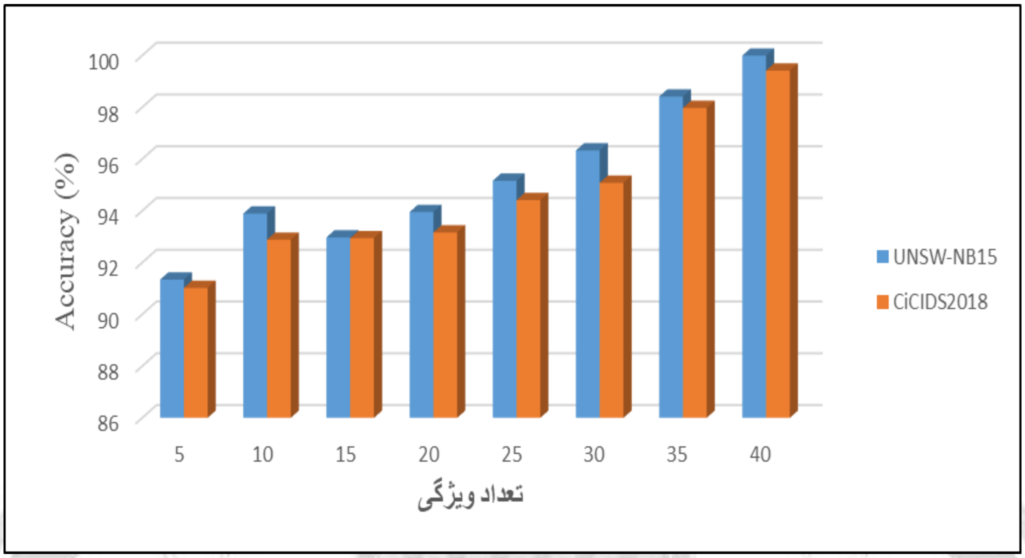
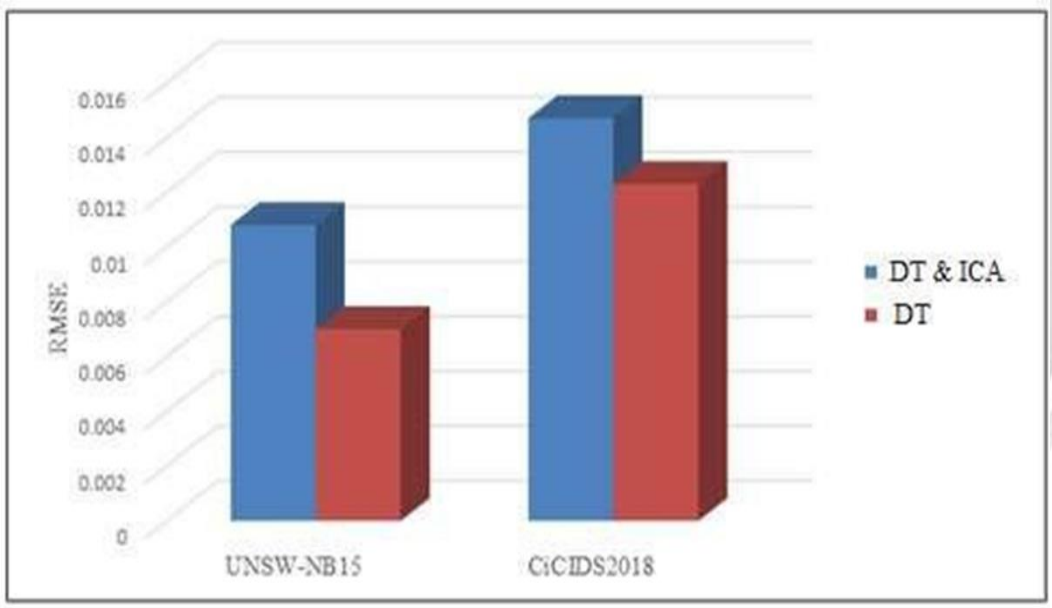


Figure 4: Comparing the impact of the number of features on the accuracy of detecting cyber attacks



As can be seen in Figures (3 & 4) using the integrated approach of independent component

Analysis and decision tree algorithm has performed better than the decision tree algorithm method. Because he can learn all the tasks at the same time and determine how to change the

Characteristics on them in a more acceptable way [8]. Since the difference between both databases was in the number of their features, the accuracy of detection was completely different for these two. In other words, since the UNSW-NB15 dataset contains 40 features and the

CiCIDS2018 dataset has 78 features, the accuracy of the proposed method on the UNSW-NB15 database is higher [9]. Based on this, Table 5-4 examines the change in the number of features on the accuracy of each of the two databases.

5. Conclusion

In this research, botnet detection systems were investigated and methods of optimizing and increasing their efficiency in order to detect botnets were discussed. The chosen method for this work is the combination of independent component analysis and decision tree algorithm and single decision tree method, which we discussed with the classification of the data used i.e. UNSW-NB15 and CiCIDS2018. Since the UNSW-NB15 dataset contains 40 features and the CiCIDS2018 dataset has 78 features, we investigated the use of the two approaches in terms of accuracy, correctness and error criteria. Based on this, the results obtained in this research can be stated as follows:

The accuracy obtained for botnet detection according to the decision tree approach on the UNSW- NB15 data set is equal to 99.988 percent and on the CiCIDS2018 data set is equal to 98.325. Also, the RMSE error obtained on the UNSW-NB15 data set is equal to a small valu21e of .0108 and on the CiCIDS201 data set is equal to 0.0147.

The accuracy obtained for botnet detection according to the approach of combining independent component analysis and decision tree algorithm on the UNSW-NB15 data set is equal to 99.994 percent and on the CiCIDS2018 data set is equal to 99.177. Also, the amount of RMSE error obtained on the UNSW-NB15 data set is equal to a small value of 0.0076 and on the CiCIDS2018 data set it is equal to 0.0076.

Finally, the comparison results with the CFA method show that the method of combining independent component analysis and the decision tree algorithm has improved by about 4% compared to CFA. In general, the selection of the combination algorithm of independent component analysis and the decision tree algorithm is in the selection of features and their better classification in line with botnet detection.

6. Future Work

In order to continue this research, the following can be followed:

1. Using the integrated algorithm of decision tree and gray wolf in order to determine the desired features and compare with the obtained results.
2. Botnet detection based on data mining on NIST data

7. References

[1] Davis, C., Fernandez, J., Neville, S., Optimizing sybil attacks against p2p-based botnets. In: Malicious and Unwanted Software (MALWARE), 2009 4th International Conference on, pp. 78–87, 2009. https://doi.org/10.1109/MALWARE.2009.5403016.

[2] Dittrich, D., Dietrich, S., P2p as botnet command and control: a deeper insight. In: 3rd International Conference on Malicious and Unwanted Software, pp. 41–48, 2008. https://doi.org/10.1109/MALWARE.2008.4690856.

[3] Dittrich, D., Dietrich, S., P2p as botnet command and control: a deeper insight. In: 3rd International Conference on Malicious and Unwanted Software, pp. 41–48, 2008. https://doi.org/10.1109/MALWARE.2008.4690856.

[4] Dries, A., Rckert, U., Adaptive concept drift detection. Stat. Anal. Data Min. 2(5-6), 311–327, 2009. https://doi.org/10.1002/sam.10054.

[5] Dries, A., Rickert, U., 2009. Adaptive concept drift detection. Stat. Anal. Data Min. 2(5-6), 311–327, https://doi.org/10.1002/sam.10054

[6] Dshieldorg, Most attacked port reports. 2013. http://www.dshield.org/portreport.html.

[7] Espinosa, J., Vandewalle, J., Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. IEEE Transactions on Fuzzy Systems 8 (5), 1998. https://doi.org/10.1109/91.873582.

[8] Espinosa, J., Vandewalle, J., Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. IEEE Transactions on Fuzzy Systems 8 (5), 1998. https://doi.org/10.1109/91.873582.

[9] Fawcett, T., An introduction to roc analysis. Pattern Recognit. Lett. 27 (8), 861–874, 2006. https://doi.org/10.1016/j.patrec.2005.10.010.

ARTICLE

JOURNAL`S UNIVERSITY OF BABYLON FOR
ENGINEERING SCIENCES (JUBES)
مـــــجلــة جـــــامعة بـــــابــل للعلـــــوم الهندسية

Vol. 3٣, No.6. \ 202٥ ISSN: 2616 - 9916

استخدام مزيج من تحليل المكونات المستقلة وخوارزمية شجرة القرار للكشف عن شبكات الروبوتات في أجهزة إنترنت الأشياء

زيد رحيم مال الله[١] سلام محمد سلام[٢] علي أصغر صفائي[٣]

1 - وزارة الكهرباء، العراق

2 - وزارة الصحة، العراق

3 - جامعة عزيد، طهران، ايران

salammohammed427@gmail.com Zaidraheem215@gmail.com

الخلاصة

يُعدّ مفهوما بوت نت وبوت نت مفهومين شائعين في أدبيات الأمن السيبراني. بوت نت هي شبكة موزعة جغرافيًا من بوت نات مصابة (مثل أي جهاز حاسوبي، بما في ذلك أجهزة إنترنت الأشياء (IoT) مثل التلفزيون الذكي، والتي تُخترق ببوت نت)، والتي يتحكم بها بوت عن بُعد. يتم التحكم في الجهاز الرئيسي. تُستخدم هذه بوت نت عادةً لتنفيذ مجموعة واسعة من الأنشطة السيبرانية الخبيثة، بدءًا من إرسال البريد العشوائي وشن هجمات حجب الخدمة الموزعة (DDoS) ونشر البرامج الضارة (البرمجيات الخبيثة) وتوزيع المواد غير القانونية (مثل مواد إساءة معاملة الأطفال). في هذا البحث، تم التحقيق في أنظمة الكشف عن بوت نت وطرق تحسين كفاءتها لتحديدها. الطريقة المختارة لهذا العمل هي الجمع بين تحليل المكونات المستقلة وخوارزمية شجرة القرار، بالإضافة إلى طريقة شجرة القرار المفردة ١٥، والتي ناقشناها مع تصنيف البيانات المستخدمة، أي ميزاتUNSW-NB ، وتحتوي مجموعة البيانات ٤٠ على ١٥ من . UNSW

ميزات NB2018 وCiCIDS، بحثنا في استخدام النهجين في ٧٨ مجموعة بيانات تحتوي على ٢٠١٨ CiCIDS من حيث الدقة والصحة ومعايير الخطأ. كانت الدقة التي تم الحصول عليها للكشف عن شبكات الروبوتات، وفقًا لنهج الجمع بين تحليل المكونات المستقلة، ٩٩.٩٩٤% من مجموعة البيانات، تساوي ١٥% وخوارزمية شجرة القرار على .UNSW-NB كذلك، فإن مقدار خطأ RMSE 99.177في مجموعة بيانات CiCIDS يساوي ٢٠١٨، وفي مجموعة بيانات ٠.٠٠٧٦ يساوي قيمة صغيرة قدرها ١٥ تم الحصول عليها في مجموعة بيانات0.0076 UNSW-NB وهي تساوي CiCIDS.٢٠١٨ تستخدم المنهجية المقترحة نهجًا هجينًا يجمع بين خوارزميات تحليل المكونات المستقلة (ICA) وخوارزميات شجرة القرار (DT) وتقارن أدائها بنموذج شجرة القرار المستقل.

الكلمات الدالة: شبكة بوت نت، سيبرانية، كشف التسلل، تعلم الآلة.