Feature Selection Model using Naive Bayes ML Algorithm for WSN Intrusion Detection System

Original Scientific Paper

Deepa Jeevaraj
Department of ECE, Bharath Institute of Higher Education and Research, India
jdeepainbox@gmail.com

T. Vijayan
Department of ECE, Bharath Institute of Higher Education and Research, India.
tvij16@gmail.com

B. Karthik
Department of ECE, Bharath Institute of Higher Education and Research, India.
karthikguru33@gmail.com*, karthik.ece@bharathuniv.ac.in

M. Sriram
Department of CSE, Bharath Institute of Higher Education and Research, India.
msr1sriram@gmail.com

Abstract *– Intrusion detection models using machine-learning algorithms are used for Intrusion prediction and prevention purposes. Wireless sensor network has a possibility of being attacked by various kinds of threats that will de-promote the performance of any network. These WSN are also affected by the sensor networks that send wrong information because of some environmental causes in-built disturbances misaligned management of the sensors in creating intrusion to the wireless sensor networks. Even though signified routing protocols cannot assure the required security in wireless sensor networks. The idea system provides a key solution for this kind of problem that arises in the network and predicts the abnormal behavior of the sensor nodes as well. But built model by the proposed system various approaches in detecting these kinds of intrusions in any wireless sensor networks in the past few years. The proposed system methodology gives a phenomenon control over the wireless sensor network in detecting the inclusions in its early stages itself. The Data set pre-processing is done by a method of applying the minimum number of features for intrusion detection systems using a machine learning algorithm. The main scope of this article is to improve the prediction of intrusion in a wireless sensor network using AI-based algorithms. This also includes the finest feature selection methodologies to increase the performance of the built model using the selected classifier, which is the Bayes category algorithm. Performance accuracy in the prediction of different attacks in wireless sensor networks is attained at nearly 95.8% for six selected attributes, a Precision level of 0.958, and the receiver operating characteristics or the area under the curve is equal to 0.989.*

Keywords*: IDS, WSN, Machine learning, ROC, Precision, Naïve Bayes*

1. INTRODUCTION

Trending widespread application of wireless sensor networks in getting the solution for intrusions cum detection for these networks has become a challenging task nowadays. There is an urgent need to develop a system that detects intrusions, malicious node that breaks down the wireless sensor networks [1-3]. Anomaly-based intrusion detection systems are the trending demand in this widespread application. According to the behavior of some effective detection methods using a machine-learning algorithm in this article. The main concern of this article is to build a model using the Bayes category ML algorithm to predict and prevent various types of intrusion attacks that create a breakdown of wireless sensor network applications [4-8]. This is deployed by monitoring various parameters of the wireless sensor Network and the output based on their weights and concentrations and energy consumed at various nodes. The built model is used to identify the intrusions that create attacks on the WSN as well as increase energy consumption or loss of energy consumption. The model built provides a higher rate of intrusion detection rate and reduces the loss of energy.

WSN is used by Defence Services, biotelemetry health care, automation Industries so on. The physical attributes [9] where human activity finds it difficult to supervise these wireless sensor networks the sensor nodes that are deployed in specific areas. Nodes transmit and receive data continuously through the base stations. There are many issues that come in contact with these wireless sensor networks including the attacks and energy consumption in not identifying the malicious node. The specific protocol used for routing, their efficiency in energy consumption the cluster head

selection, and the Novelty of the wireless sensor network, etc [9-11]. Trending constraints in this article is to build a model to optimize and detect all the intrusion cum attacks created to these wireless sensor networks using a machine learning Framework. Figure 1 shows the sensor nodes that monitor all the physical parameters like temperature, pollution, and Connected devices. The data collected all are synced through the internet and get the information from the nodes is shown in Fig.1.

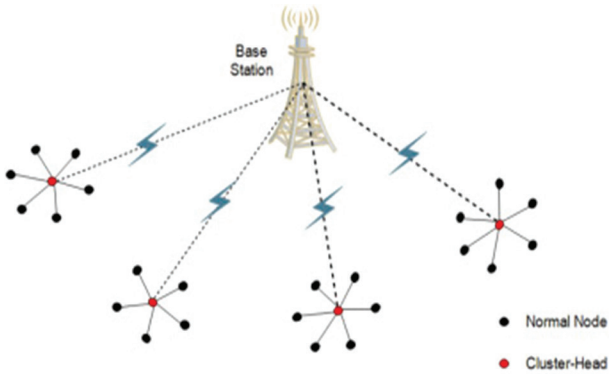

Fig. 1. WSN architecture with a base station

Organization of the paper: The article is composed of five sections such as the related works that support greatly in including the machine learning concepts in the research article. Secondly the data type and its description of all the dataset parameters. The next section is about the materials and the experiment methodology handled in the research proposal. The next section is the experimental investigations and inferences discussed. The next section is the conclusion part, the future scope, and enhancements.

2. RELATED WORKS

In [12] authors presented the present day Scenario integrated internet using a wireless sensor network is presently having a great impact on today's life. The privacy and security of a network in preventing and detecting intrusion in WSN is a challenging issue. The different types of threats which are prominent and very hard to detect the device attacks are taken into the act. Protection in communication connected with wireless sensor networks used encryption-based techniques traditionally. Which has proved to be inefficient in recent days. This proposal has given an intrusion detection cum protection of WSN using a new direction towards internet integrated wireless sensor network.

In [13] authors prescribe Wireless sensor network comes in contact with compact size and inexpensive sensor nodes. The place of usage of WSN with a sensor undergoes arbitrary Placement in open areas. In this kind of situation, there is a higher rate of attacks. The innovative idea behind the intrusion detection system in this proposal is building a model using machine learning algorithms like support vector machines to detect intrusion in the wireless sensor network. The result of this portrayed high results of accuracy nearly 94.09% and a detection rate of 95%.

In [14] authors monitor wireless sensor network has a wide range of applications in the environment, health, military, industries, etc. WSN has Limited source and energy concerns. A challenging task that is designed in such a way that it utilizes minimum energy consumption and gives a maximum lifetime of the network. In most of the Daniel of service attacks that destroys the network and loss its energy rapidly is identified using a novel approach. An efficient intrusion detection system or scheme is designed in such a way that malicious node is identified with very little energy conservation. All the nodes are continuously monitored whose energy consumption is monitored and by comparing the actual and the predicted energy the malicious node is identified. This malicious node is identified by using a bayesian approach of a machine learning algorithm.

In [15] authors propose WSN is a key object in any cyber-physical system. It is composed of many stationary as well as mobile parts like sensors that transmit and receives information through WSN. The intrusion that affects the WSN has to prevent using a special mechanism in a smart environment. The novel approach is a sequence backward selection algorithm that detects the attacks at a faster rate. The experiment results based on this approach have given an efficient F-measure of 0.96 0.99 for all kinds of network attacks.

In [16] authors present a new protocol developed in WSN integrated into IoT deployment. In today's activities, every common man has an advancement in information communication and Technology ICT. Advancement is also suffering from various attacks that occur in WSN and IoT. Because of trending progress in this fast-moving environment and more vulnerable security threats. In the future, everyone is connected to the internet with numerous smart objects and for a smooth progression there is a need for IDS and IPS. This article gives an emerging intrusion detection system with a new approach. A privacy preservation protocol is integrated with WSN and IoT to address the intrusion detection Protocol in wireless sensor networks that is integrated into the internet-of-things.

In [17] authors prescribe Network security as an unavoidable event in our daily interactions and networks. Intrusions are also developing more and more critical as Technology also grows. Techniques employed using machine learning algorithms to detect intrusions. However, there is an advantage of deep learning algorithms and AI to generate special features that automatically detect attacks without any human intervention. Long short term memory network with spatial features is employed to detect a hybrid intrusion detection system with a model that is built using this deep learning methodology. The investigational report specifies high accuracy, Precision, and detection rate as very high and effective.

In [18] authors propose a secured energy-efficient barrier coverage schedule that has been developed using a machine learning algorithm to maintain the quality of service. A barrier coverage schedule is also energy conserving scheme. In spite of a wide range of areas called the barriers and a subset of sensor nodes overlapped to meet all the quality of service requirements. Expected node failures due to barrier security attacks such as Daniel of service is a challenging in maintaining the quality of service levels. A smart proposal using a machine learning algorithm is proposed to detect The Attacks in an efficient way. WSN-based IoT applications that utilize kNN machine learning algorithms to detect malicious attacks.

In [19,20] authors present the Wireless sensor network as one of the third Millennium technology that had a wide range of applications in the surrounding medium or environment. The main reason for the application of WSN application is the low production cost, the installation, unattended operations, anonymous and longtime operations that occur. WSN integrated with IoT in sensor nodes and sensing ability using internet-connected devices is a recent advancement taking place in WSN. The absence of physical in-line security defense gateways that comes in contact with network security with IoT is a big concern to the scientific community. A novel technique for the prevention, detection, and mitigation of all the attacks is proposed in this article. Recent integration and collaboration of WSN and IoT are facing open challenges in terms of security. A system should be developed with security administrators and network managers to predict all the threats and attacks to detect the malicious nodes Machine learning tool is a powerful tool in predicting the intrusion caused in a WSN in less time. However, the prediction is accurate with the only parameter being the perfect dataset with the required attributes. So that the trained model will very well perform in predicting the intrusions timely.

3. DATA DESCRIPTION

The data set WSN is collected from public platforms like kaggle.com [21]. Table 1 shows the description of the data set which consists of nearly three lakh seventy-four thousand and six hundred and sixty-two instances with 18 attributes. The attributes are namely the ID the channel is present or absent, the Signal strength indicator, the average distance of the channel and the energy consumption of the nodes, and the number of messages and advertisers that are received from the nodes.

That acknowledgment of the number of advertisers using time division multiplexing broadcast messages to the nodes. Data is transmitted and received from the nodes. The packet is sent to the base stations and the distance between the channel and the base station. The code is finally in the cluster where there are nearly five output classes that designate the type of attack in the WSN undergone.

Table 1. Experimental dataset for WSN attack prediction

S.No	Attributes	Instances count	Class description of Level
1	18	374662	5, TDMA, Black hole, Flooding, Grey hole, and Normal

The selected machine learning algorithm for this article is a Naive Bayes classifier. The basic principle of this algorithm is it works on the probability of the events X and Y. It also comprises some text sentiments and opinions for processing. The Bayes Theorem basically gives the hypothesis with the prior acquired knowledge available from the previous experiences as given in the formula (1).

$$P\left(\frac{X}{Y}\right) = \frac{P\left(\frac{Y}{X}\right)P(X)}{P(Y)} \quad (1)$$

Where P(X) and P(Y) are the probability of events X and Y.

4. METHODOLOGY

The experiment proceeds in the following ways as shown in Fig. 2 in attaining the optimum model.

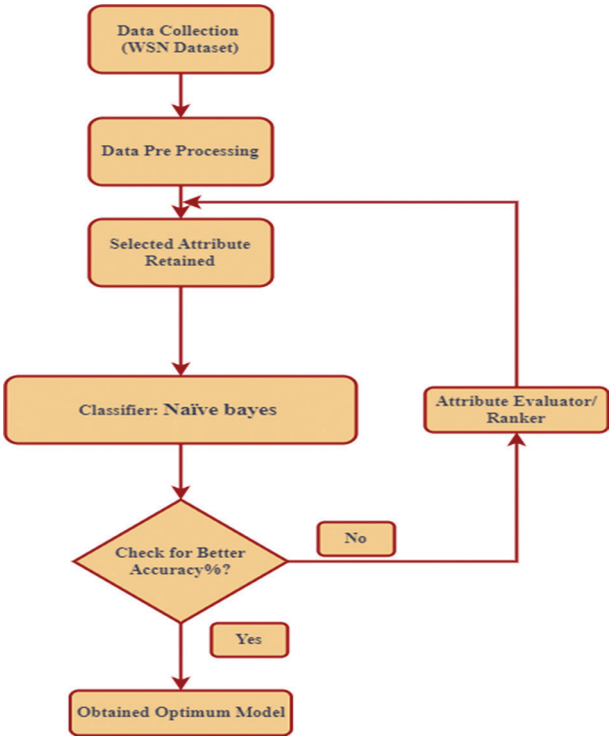

Fig. 2. workflow diagram of WSN attack prediction

The work show illustrates the data collection in the initial state from a public platform like kaggle.com. The data format is changed according to the usable format in the attribute relation file format. Once it is accessible in a tool used like Weka 3.8.5 [22] The selected attributes are retained for classification purposes nearly 18 attributes are retained for the first classification procedures. After the classification by done using the select-

ed machine learning algorithm called Naive Bayes. The next step is to check for the accuracy percentage of the f-measure, the receiver operating characteristics, and the Precision tabulated. The procedure is the full data set process tree process some of the attributes according to the information present in the attributes using and ranker attribute in the Weka tool. Once the ranks are obtained according to the information gain some of the attributes with high again and go with the classification process. Again the procedure is repeated for selected attributes and removing the attributes with less information.

The final optimum model is then concluded with maximum accuracy, precision, and receiver operator characteristics. The investigation was supported by giving maximum efficiency for the build model and output accuracy with nearly 95.854% with 6 attributes itself and yielding maximum area under the curve of 0.979.

5. EXPERIMENTS RESULTS AND DISCUSSION

The experimental results are as follows: The attribute accessor search methodology is applied. The retained attributes are classified whose accuracy, precision, Recall, F-Measure, and receiver operator characteristics are analyzed and tabulated. Nearly five different types of attribute are selected which is classified from retaining all the attributes to removing the attributes [9] according to their rank. Table 1 Investigation results after classifying all attributes to retained attributes.

Table 1. Experimental Results of Attribute Selection method

S.NO	Attribute Selection	Attribute selection session (Attribute assessor/Search method)	Classifier	Accuracy	Precision	Recall	F-Measure	ROC
1.	Present all attributes(18), ADV_S, Is_CH, Expaned Energy, DATA_S, Rank, send code, JOIN_S, Dist_To_CH, ADV_R, SCH_R, SCH_S, who CH, Data_Sent_To_BS, id, JOIN_R, dist_CH_To_BS, Time, DATA_R	InfoGain AttributeEval / Ranker	Naïve Bayes	95.3734 %	0.967	0.954	0.958	0.980
2.	selected attributes (15), ADV_S, Is_CH, Expaned Energy, DATA_S, Rank, send code, JOIN_S, Dist_To_CH, ADV_R, SCH_R, SCH_S, who CH, Data_Sent_To_BS, id, JOIN_R, **(Removed last 3 attributes)**	InfoGain AttributeEval / Ranker	Naïve Bayes	95.3216 %	0.965	0.953	0.957	0.980
3.	selected attributes (12), ADV_S, Is_CH, Expaned Energy, DATA_S, Rank, send code, JOIN_S, Dist_To_CH, ADV_R, SCH_R, SCH_S, who CH, **(Removed last 6 attributes)**	InfoGain AttributeEval / Ranker	Naïve Bayes	94.2433 %	0.958	0.942	0.949	0.983
4.	selected attributes (9), ADV_S, Is_CH, Expaned Energy, DATA_S, Rank, send code, JOIN_S, Dist_To_CH, ADV_R, **(Removed last 9 attributes)**	InfoGain AttributeEval / Ranker	Naïve Bayes	92.328 %	0.933	0.923	0.925	0.979
5.	selected attributes (6), ADV_S, Is_CH, Expaned Energy, DATA_S, Rank, send code, **(Removed last 12attributes)**	InfoGain AttributeEval / Ranker	Naïve Bayes	95.8154 %	0.968	0.958	0.961	0.989

Table 1 shows the experimental results of the built model by applying the information gain cum attribute evaluator which ranks for attributes in the dataset. The ranked attributes with maximum information are retained attributes. 6th attribute ADV_S is ranked first according to the information gain. Attribute 3 is_CH to channel is present or not and the energy conservation is the third attribute and so on. The Second attribute (time) and the fourteenth attribute data_R are attributes with very less information as per the attribute information evaluator.

Figure 3 shows the graph for the number of attributes retains to the percentage of accuracy performance. The percentage of accuracy is 95.82%stage for 6 attributes

itself. The performance accuracy was 95.37% stage for retaining all the attributes. Therefore, feature selection using the selected attributes gives A Remarkable performance using this information gain evaluator.

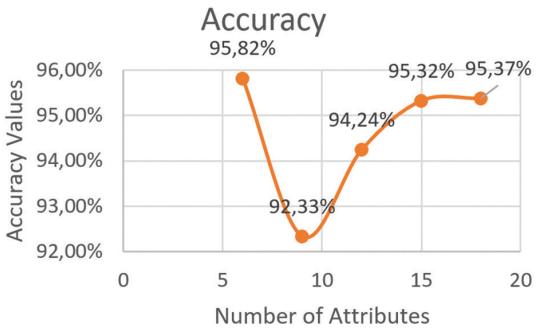

Fig. 3. No of the retained attribute to % of Accuracy

Attribute versus Precision characteristics clearly gives the built model has given good precise values for retaining 6 attributes out of 18 as shown in Fig. 4. The precision value for retaining all 18 attributes has given 0.967 and for retaining six different attributes as given 0.968. Good precision value characteristics for using the feature selection option in the Weka tool.

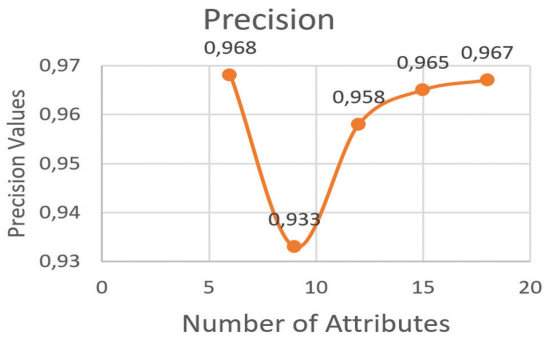

Fig. 4. Attributes versus Precision

The retained attribute versus recall characteristics also gives a good performance by attaining a value of 0.954 for 18 attributes as shown in Fig. 5. Just for six attributes the recall value of 0.958 using the information gain evaluator.

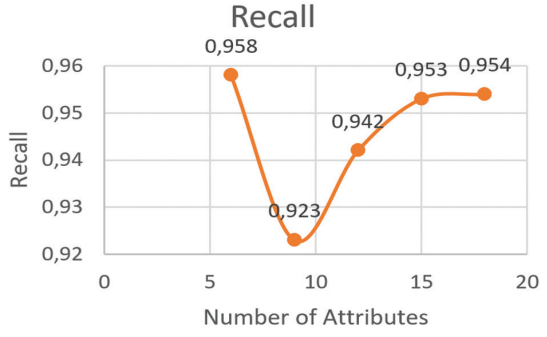

Fig. 5. Attributes versus Recall

The consolidated measure of both Precision and recall is done using F-measure. The attribute versus F -measure characteristics are shown in Fig. 6. The consolidated effect of both Precision and recall values of the built model. This specifies the model classifying the output has given a remarkable output for retaining 6 attributes.

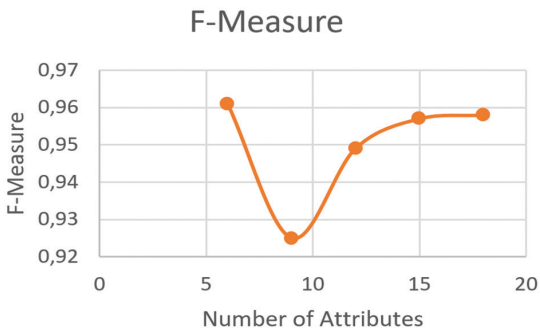

Fig. 6. Attributes versus F -Measure

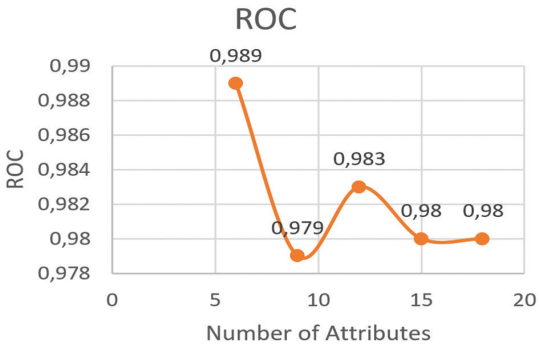

Fig. 7. Attributes versus ROC

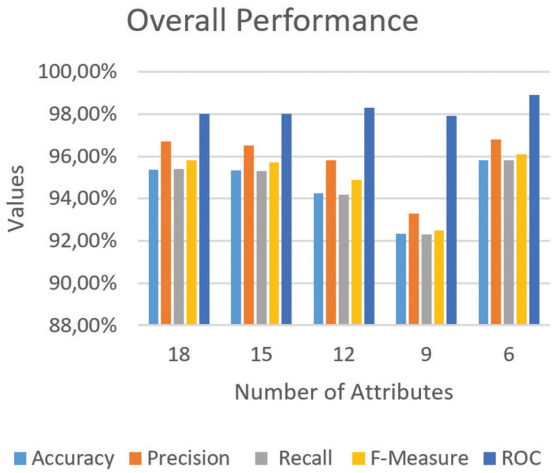

Fig. 8. Overall performance characteristics of a built model

The overall performance characteristics of the built model using the Naive Bayes algorithm are shown in Fig. 8. The characteristics clearly mark that the built model gives a remarkable output in classifying the output intrusion in WSN. The difference between retaining all the attributes and retaining six attributes. The performance characteristics support predicting the attacks in a WSN in less time and preventing the WSN from intrusions. The feature selection method using an information gain evaluator really works in an efficient manner.

6. CONCLUSION AND FUTURE SCOPE

In the first iteration, the accuracy percentage with maximum performance was attained to be 95.37 percent with eighteen attributes. The result obtained with six attributes was 95.82 percent using feature selection methods. These findings are the first of their type in this structure for intrusion prediction utilizing the WSN dataset that is based on real-time data acquisition. In the future, this can be implemented in finding the intrusion cum preventing system in WSN with deep learning methodologies. The future scope of the proposed methodology is a fast-growing field. The intrusion detection in WSN faces greater demand in the future and the proposed method can be established in deep learning technique in meeting the above-mentioned demand in the future.

7. REFERENCES

[1] L. Zhiqiang, G. Mohiuddin, Z. Jiangbin, M. Asim, W. Sifei," Intrusion detection in wireless sensor network using enhanced empirical based component analysis", Future Generation Computer Systems, Vol. 135, 2022, pp. 181-193.

[2] G. Creech, J. Hu, "A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguousand Discontiguous System Call Patterns", IEEE Transactions on Computers, Vol. 63, 2014, pp. 807-819.

[3] L. Vokorokos, A. Baláz, "Host-Based Intrusion Detection System", Proceedings of the IEEE 14th International Conference on Intelligent Engineering Systems, Las Palmas, Spain, 5-7 May 2010, pp. 43-47

[4] A. H. Farooqi, F. A. Khan, "A Survey of Intrusion Detection Systems for Wireless Sensor Networks", International Journal of Ad Hoc and Ubiquitous Computing, Vol. 9. 2012, pp. 69-83.

[5] G. Liu, H. Zhao, F. Fan, G. Liu, Q. Xu, S. Nazir, "An Enhanced Intrusion Detection Model Based on Improved kNN in WSNs", Sensors, Vol. 22, 2022, p. 1407.

[6] Y. Canbay, S. Sagiroglu, "A hybrid method for intrusion detection", Proceedings of the IEEE 14th International Conference on Machine Learning and Applications, Miami, FL, USA, 9-11 December 2015, pp. 156-161.

[7] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, H. Karimipour, "Cyber intrusion detection by combined feature selection algorithm", Journal of Information Security and Applications, Vol. 44, 2017, pp. 80-88.

[8] M. S. Koli, M. K. Chavan, "An Advanced Method for Detection of Botnet Traffic using Intrusion Detection System", Proceedings of the IEEE International Conference on Inventive Communication and Computational Technologies, Coimbatore, India, 10-11 March 2017, pp. 481-485.

[9] T. Vijayan, M. Sangeetha, A. Kumaravel, B. Karthik, "Feature selection for Simple Color Histogram Filter based on Retinal Fundus Images for Diabetic Retinopathy recognition", IETE Journal of Research, 2020.

[10] Y. Sun, F. Liu, "SMOTE-NCL: A Re-Sampling Method with Filter for Network Intrusion Detection", Proceedings of the IEEE International Conference on Computer and Communications, Chengdu, 14-17 October 2016, pp. 1157-1161.

[11] H. Elbahadır, E. Erdem, "Modeling Intrusion Detection System Using Machine Learning Algorithms in Wireless Sensor Networks", Proceedings of the 6th International Conference on Computer Science and Engineering, Ankara, Turkey, 15-17 September 2021, pp. 401-406.

[12] B. J. S. Kumar, S. Sinha, "An Intrusion Detection and Prevention System against DOS Attacks for Internet-Integrated WSN", Proceedings of the 7th International Conference on Communication and Electronics Systems, Coimbatore, India, 22-24 June 2022, pp. 793-797.

[13] S. Amaran, R. M. Mohan, "Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks", Proceedings of the International Conference on Artificial Intelligence and Smart Systems, 2021, pp. 1100-1104.

[14] S. S. Shivaji, A. B. Patil, "Energy Efficient Intrusion Detection Scheme Based on Bayesian Energy Prediction in WSN", Proceedings of the Fifth International Conference on Advances in Computing and Communications, Kochi, India, 2-4 September 2015, pp. 114-117.

[15] S. Jiang, J. Zhao, X. Xu, "SLGBM: An Intrusion Detection Mechanism for Wireless Sensor Networks in Smart Environments", IEEE Access, Vol. 8, 2020, pp. 169548-169558.

[16] S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. J. P. C. Rodrigues, Y. Park, "Intrusion Detection Protocols in Wireless Sensor Networks Integrated to Internet of Things Deployment: Survey and Future Challenges", IEEE Access, Vol. 8, 2020, pp. 3343-3363.

[17] Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, R. Ahmad, "CNN-LSTM: Hybrid Deep Neural Network for Network Intrusion Detection System", IEEE Access, Vol. 10, 2022, pp. 99837-99849.

[18] D. Thomas, R. Shankaran, M. A. Orgun, S. C. Mukhopadhyay, "SEC2: A Secure and Energy Efficient Barrier Coverage Scheduling for WSN-Based IoT Applications", IEEE Transactions on Green Communications and Networking, Vol. 5, No. 2, 2021, pp. 622-634.

[19] Butun, P. Österberg and H. Song, "Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures", IEEE Communications Surveys & Tutorials, Vol. 22, No. 1, 2020, pp. 616-644.

[20] D. Hemanand, G. V. Reddy, S. S. Babu, K. R. Balmuri, T. Chitra, S. Gopalakrishnan, "An intelligent intrusion detection and classification system using CSGO-LSVM model for wireless sensor networks (WSNS)", International Journal of Intelligent Systems and Applications in Engineering, Vol.10, No. 3, 2022, pp. 285-293.

[21] Kaggle, https://www.kaggle.com/datasets/bassamkasasbeh 5 1/wsnds (accessed: 2022)

[22] Weka, https://www.cs.waikato.ac.nz/ml/weka/ (accessed: 2022)