SURVEY ON SECURITY OF WIRELESS SENSOR NETWORKS USING MACHINE LEARNING TECHNIQUES

[1]**K.NIRMALA**
[1]Research Scholar,
Department of Computer Science and Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India, 517501.
[2]**Dr.CH.D.V.SUBBA RAO**
[2]Professor,
Department of Computer Science and Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India, 517501.

ABSTRACT: Security of communication systems has become a crucial issue. Wireless Sensor Networks are basically a collection of sensor nodes scattered in a large area such that the desired information can be collected. Widespread use of Wireless Sensor Networks (WSNs) introduced many security threats due to the nature of such networks, particularly limited hardware resources and infrastructure less nature. As the sensor nodes undergo arbitrary placement in the open areas, there is a higher possibility of affected by distinct kinds of attacks. An effective Intrusion Detection System (IDS) is essential for ensuring network security. Intrusion detection systems include pattern analysis techniques to discover useful patterns of system features. These patterns describe user behavior. This work is focusing on survey of different machine learning based attack detection which is an essential task to secure the network and the data. The presenting survey paper uses the genetic k-means algorithm (GKA), optimal support vector machine (OSVM), K-nearest neighbor (KNN) and Intrusion detection model based on information gain ratio and online passive aggressive algorithm (ID- GOPA). We provide a deep insight survey on existing schemes of different machine learning methods which are providing security to the wireless sensor networks (WSN). In the results section performance measures are described to evaluate their consequent advantages and disadvantages.

KEYWORDS: Wireless Sensor Networks (WSNs), K-nearest neighbor (KNN), Machine learning techniques.

I. INTRODUCTION

There is a widespread use of networking systems over the globe by most companies and even individual developers to create innovative solutions and products that can help organizations and citizens utilize different new technologies to satisfy their various needs. Sensors are one of the relatively new technologies that started early in the market and now is being used in the Internet of Things (IoT) [1]. Wireless sensor network (WSN) composed of sensor nodes is deployed in the target area where a phenomenon of interest occurs [2]. The typical goal of a WSN is to monitor and collect information from the target area and transfer the detected information to a base station or a sink node. While fulfilling a task to achieve the goal is critical for WSNs in general, cost-efficiency and energy-efficiency are increasingly important for larger scale WSNs [3]. A large number of sensors deployed in a large coverage area cause many technical issues that should be addressed while developing algorithms to achieve the goal.

Wireless sensor networks represent a special class of ad hoc networks. They are made up of many smart sensor nodes of small sizes, limited power, at low-cost, and multi-functional (also called nano computers). In principle, these network nodes have a spontaneous mode of organization because they are intended to be deployed quickly and arbitrarily in a space of interest. They are powered by a power unit (battery) of limited capacity. They can capture (or collect) physical quantities from the environment such as temperature, wind speed, relative humidity, etc. They are also able to detect real-world events, process data, and communicate with each other to bring the information collected to a collection point called sink node or Base Station (BS). This information is then transmitted via a transport network to a processing center where possible analyzes, interpretations, and decision-making are carried out by an end-user.

Intrusion detection systems (IDS) are one of the most flexible and useful tools to guard WSNs from known and unknown attacks [4]. IDS observe and analyze the events generated in the network to detect anything unusual and alert sensor nodes about the intruder. This concept was originally proposed by Anderson. The strategies broadly utilized to develop IDS used for attack detection nowadays are vastly related to machine learning techniques [5]. Most approaches, however, are based on offline learning which requires all, or at least a sample, of historical data to be kept in memory.

ML technique facilitates the development of complex models solely based on data, without the need of specialized human intervention. ML techniques are usually categorized into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (RL) [6]. Over the past decades, ML techniques have been applied to diverse applications including social media, medical systems, transportation, computer vision, and wireless communication. The ML techniques have been applied successfully to many WSN applications such as classification, clustering, dimensionality reduction, feature extraction, and forecasting [7]. The advantages of applying ML techniques to WSNs can be enumerated as finding an optimal solution, e.g., optimal location for the placement of sensors, reducing computational complexity, e.g., reducing the required bandwidth to transmit collected sensor data, and flexibility, e.g., capability to react to dynamic inputs. The following are the challenges or problems related with designing of WSN.

Adverse Environment: Being deployed randomly into a space, the environmental related parameters can lead to non-functionality of the SN and attacker can take advantage of this situation and can manipulate it.

No Surveillance: SNs are deployed in a territory where its continuous surveillance is not possible which easily gives attackers a way to physical tampering and attack on higher levels afterwards.

Limited Resources: For the SN to work, it requires energy, memory etc. for data collection and transmission. Since these resources are limited for a SN and hence exhaustion of these resources can lead to non-functionality of the sensor which can lead to either packet drop or corrupted message, systems lag etc.

Reliability in Wireless Communication: In WSN the sensor data may be distorted due to channel errors which may lead to conflicts, and at highly busy node the data may also be and thus Denial-of Service (DoS) attack can be easily launched. Due to the greater congestion at a single node the overload results in increasing the latency in the sensor network thus causing synchronization errors and lag in the system, including sensor nodes.

The paper is organized as follows: In the next section, we present in literature survey on machine learning based security analysis of WSN. The simulation results used to evaluate the performance of the individual machine learning models in Section 3. Finally, the paper ends with a conclusion and perspectives.

II. LITERATURE SURVEY

Mourabit, et al. [8] presented a comparative evaluation of a series of attack detection techniques to be applied in wireless sensor networks. Besides the authors proposed enhancements of random forest and other techniques for effective detection of specific anomalies on KDD'99 datasets. It also provides recommendations on improvement of the performance of intrusion detection mechanisms in wireless sensor networks.

As indicated by Feurer et al. [9], machine learning and artificial intelligence are considered very effective tools. Most researchers highlight the fact that they continue to improve, which is another strength. The more data they experience or receive, the more they will continue to improve in terms of knowledge. Algorithms continue to improve based upon what they learn and the accuracy they obtain. They have become faster at predicting and identifying system threats. Other strengths include handling huge amounts of data and having a wide range of applications in various security fields. Finding the right dataset for a project can be difficult, and several factors must be considered, such as the reliability of the data, their validity, and the legally binding aspects behind behind use of the data.

Rajs, et al. [10] compare non-functional limitations, including performance indicators, which are especially important when organizing security equipment in wireless sensor networks. Both techniques are compared by using an accuracy indicator and attack identification time. In particular, the attacks associated with a violation of synchronization in the operation of wireless sensor network nodes are simulated. In fact this simulation is performed by using RTS/CTS packets and leads to a collision state.

Yu, et al. [11] proposed a generalized architecture for building a framework of machine learning based intrusion detection system for wireless sensor networks. A set of features that can be used to analyze and detect a number of attackers in such networks, including RTS packets rate, Neighbor count, Power consumption rate, etc. were defined.

2.1 Intrusion Detection in Wireless Sensor Network Using Genetic K-Means Algorithm:

Sandhya G et. al. [12] proposes the Intrusion Detection in Wireless Sensor Network Using Genetic K-Means Algorithm (GKA). Figure 1 shows the flowchart of Genetic K-Means algorithm according to which GKA is implemented in our system.

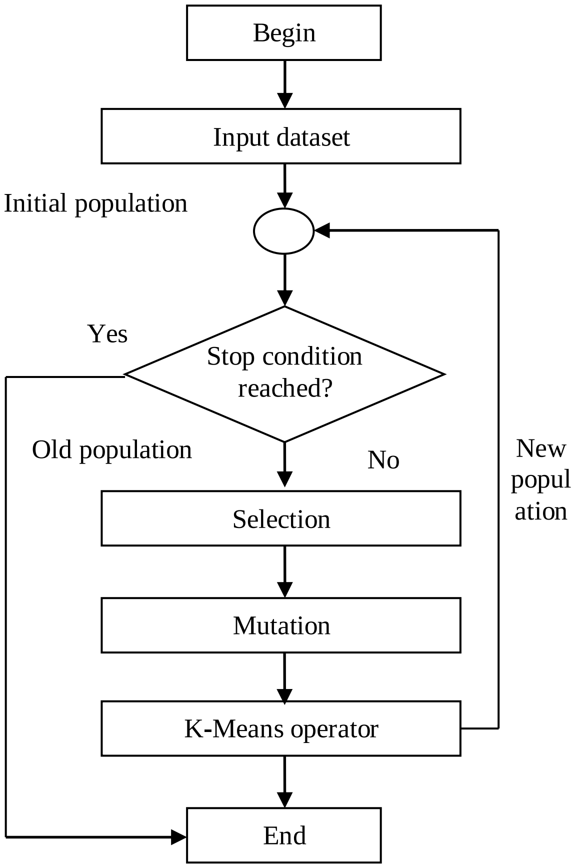

Fig. 1: FLOWCHART OF IMPLEMENTING GENETIC K-MEANS ALGORITHM

This paper, propose an approach for intrusion detection that employs genetic k-means algorithm. Genetic KMeans algorithm is applied to differentiate normal and abnormal intrusion behavior and the rule base of intrusion detection is updated. Finally, a real-time intrusion detection rule base is set. Clustering helps in finding patterns in unlabeled data of many dimensions. The major advantage of clustering algorithm is the ability to learn from and detect intrusions in the audit data without explicit signatures. One of the simplest and efficient clustering algorithms in machine learning is k-means clustering algorithm. It can automatically detect groups of similar objects in data training. The algorithm classifies instances to a pre-defined number of clusters.

Identification of attacks:
Genetic k-means algorithm IS implemented to identify attacks based on the following:
Parameters used- Destination Sequence Number (DSN), Source Sequence Number (SSN), Node ID (NID), Malicious Node ID (MN_ID) and RR (Request Reply).

1) Initialization: Start discovery phase with the source node S. Assign current time and time required to execute the PriorReceive Reply.
2) Storing: To store all the Route Replies DSN and NID in the RR Table. Repeat the above process until the time exceeds. While ((current time <= (current time + wait time)) {Store the route replies DSN and NID in the RR Table.}
3) Identification and Removal of Attack Node: Retrieve the first entry from RR Table. Check the DSN with SSN. If DSN is greater than SSN, then discard the first selected entry from the RR Table.
If (DSN > SSN) {MN_ID = NID; Discard entry from table}
4) Node Selection: Sort the contents of RR Table entries according to the DSN. Select the NID having highest value of DSN among the RR Table entries
5) Continue Default Process: Continue step 3 and step 4 until we have to find the destination node. Call Receive Reply method of default AODV Protocol.

2.2 A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks:

Eliel Marlon de Lima Pinto et. al. [13] proposes the A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks. In order to improve the detection rate when the legitimate node and the attacking node are at the same distance or at a very close distance from each other in relation to the landmark, we propose a new strategy. The proposed spoofing detection system has four phases which are presented in Figure 2.

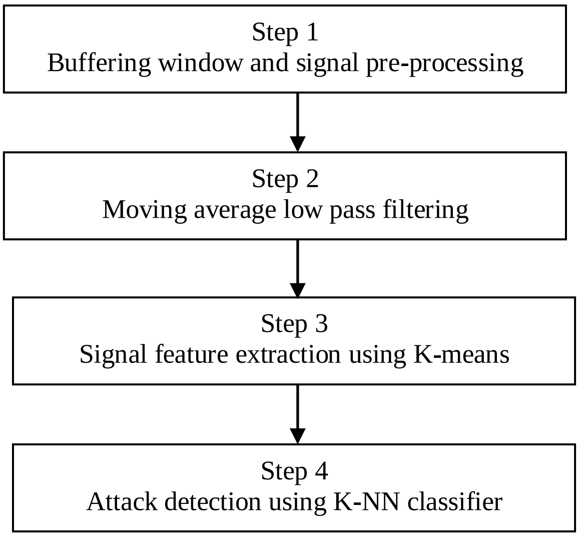
Step 1
Buffering window and signal pre-processing
Step 2
Moving average low pass filtering
Step 3
Signal feature extraction using K-means
Step 4
Attack detection using K-NN classifier

Fig. 2: DETECTION SYSTEM ARCHITECTURE

During Step 1, the RSS samples read at the landmark are buffered and grouped into l non overlapping sub-windows.

$$\omega = \{\omega_0, \omega_1, \dots\dots.\omega_{t-1}\}$$

With n samples each,

$$\omega_j = \{r_0, r_1, \dots\dots.r_{n-1}\}$$

These samples are filtered in Step 2 using a low pass moving average filter in order to reduce the variance of the RSS samples, which is caused by the channel fading. In Step 3 we employ the k-means clustering algorithm over the set ω_j and use the Euclidean distance $D_j = ||c_1 - c_2||$ as the dissimilarity measure to determine the distance between two centroids (c1 and c2). These statistics are the features employed in Step 4 of the proposed method, where we apply the supervised k-NN classifier to attack detection. These four steps are executed continuously at the landmark using the current window of RSS samples. Step 4 includes a supervised k-NN classifier. Based on a preliminary analysis to determine the more significant features to be used by the k-NN algorithm, we selected the minimum and maximum values of µc. Some preliminary simulation results on feature selection showed us that the minimum and maximum values of µc are good choices for the k-NN classifier because this lower and upper metrics can better indicate the presence or not of an attacker. They are denoted as min µc and max µc, respectively. RSS samples required to train (training set) and to test (testing set) the k-NN classifier in order to identify the scenarios with or without attacker.

2.3 Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks:

Samir Ifzarne et. al. [14], proposes the Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks. WSN Intrusion detection model based on information gain ratio and online passiveaggressive algorithm, the shortened form is ID-GOPA. The main purpose of the proposed model, Figure 3, is to apply the study of the online classifier for the streaming data of the network. ID-GOPA inspects all events circulating in the network by observing abnormal activities and it consists of two phases: the offline and the online phase.

- In the offline phase (training dataset), the model is trained by the online classifier PA to be more familiar and more learnable for existing activities in the network flow, where the processed and labeled learning records are introduced for build a learnable model capable of being tested.
- In the online phase, using the trained model from the offline phase with the same prepossessing engine selecting only the relevant attribute based on information gain ratio algorithm and classifying every packet as either normal or attack in real-time detection.

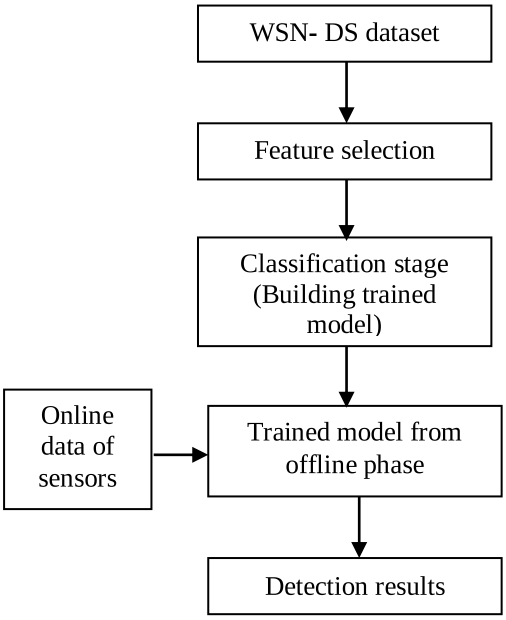
WSN- DS dataset
Feature selection
Classification stage (Building trained model)
Online data of sensors
Trained model from offline phase
Detection results

Fig. 3: THE STRUCTURE OF THE PROPOSED ID-GOPA MODEL

2.4 Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks:

Sibi Amaran et. al. [15], proposed an Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks. The overall working principle is demonstrated in Fig. 4.

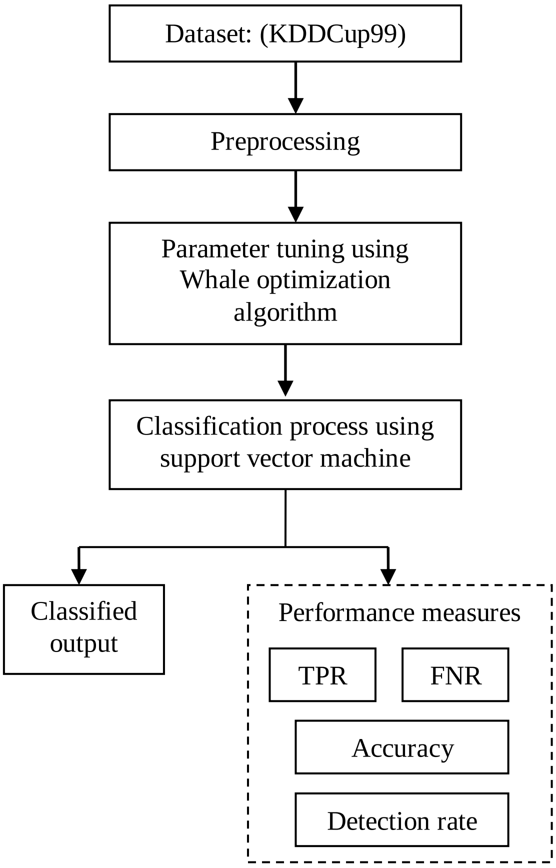
Dataset: (KDDCup99)
Preprocessing
Parameter tuning using Whale optimization algorithm
Classification process using support vector machine
Classified output
Performance measures
TPR
FNR
Accuracy
Detection rate

Fig. 4: WORKING PROCESS OF OSVM MODEL

The presented OSVM model incorporates intrusion detection using three sub-processes such as pre-processing, classification, and kernel selection. Primarily, the input network data is preprocessed to transform it into a useful format. Followed by, the OSVM model is applied for classification of the intrusions. New OSVM based IDS in WSN. Thirdly, the presented OSVM model involves the proficient selection of optimal kernels in the SVM. Here, intrusion prediction is carried out using SVM. The Modified Whale Optimization Algorithm (MWOA) has been applied for selecting the best kernel in SVM classification model. A kernel function employed in SVM, $K(X_n, X_j)$, converted the actual data space a novel space with maximum dimension. Hence, the accuracy of WOA models is defined can be improvised by allocating objective function values to random measures.

III. RESULT ANALYSIS

3.1 Intrusion Detection in Wireless Sensor Network Using Genetic K-Means Algorithm:

The Network Simulator ns-2.34 is used for implementing genetic k-means algorithm to detect intrusions in wireless sensor network using AODV protocol. The effectiveness of our proposed intrusion detection system using genetic k-means algorithm is evaluated using Detection Rate and False Positive Rate. The genetic kmeans algorithm has the highest detection rate when compared to Enhanced Intrusion Detection Algorithm (EIDA) and kmeans algorithm. System also showed low false positive rate as the false alarms detected remain minimum.

3.2 A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks:

First, we investigated the performance of the proposed strategy in the three scenarios as

- Scenario 1 (Without Attack) - Only the legitimate node is present and transmitting at a distance d1 from the landmark.
- Scenario 2 (With Attack) - The legitimate node and the attacker node are transmitting and they are located at the same distance from the landmark (d1 = d2).
- Scenario 3 (With Attack) - The legitimate and attacker nodes are transmitting and they are located at different, but close, distance from the landmark (d1 ≈ d2).

We evaluated the probability of true positive detection, which in this case is the probability of recognizing the presence of only the legitimate node, which is 87.21%. The algorithm can detect an attack regardless of whether the attacker is in Scenario 2 or Scenario 3, with probability 87.06%. The method proposed in this paper still performs well in all scenarios with an attacker. The accuracy improves in at least 8% regarding the strategy.

3.3 Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks:

The experiment uses a simulated wireless sensor network-detection system (WSN-DS) dataset and the network simulator NS-2 was used to simulate wireless sensor network environment based on the LEACH routing protocol to collect data from network. The results of this study are evaluated according to four criteria, namely accuracy (ACC), precision (PR), f1-score (F), and recall (RE).

The whole accuracy amount is 96%, as we analyze each class label to observe each individual performance, we see that the detection performance of normal cases is very high compared to abnormal cases with detection rate of 99%. The RE of the proposed method reaches 96%, which is higher than that of existing models. the accuracy of Online PA classifier is high because it works better for large stream dataset and we have used large dataset for our experimentation in addition it gives better results, as it's learning rate does not decrease with respect to time since most of the online algorithms their concepts might change through time.

3.4 Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks:

The experiment is carried out on an Intel®-core ™ i7-7500 2.70-2.90 GHz CPU processor, 8 GB memory, and running Windows 10 OS (64-bit). The software environment is MATLAB R2014b version. The performance of the proposed OSVM model is validated employing KDDCup99 dataset. The detailed comparative results analysis of the OSVM model with existing methods is processed. The OSVM model has obtained a maximum accuracy of 94.09%, TPR 95.53% and FNR 4.47. However, the OSVM model has obtained a maximum detection rate of 95.02% than the existing models as SVM and ELM models have resulted to lower detection rate values of 74.74% and 75.53% respectively. Besides, the MK-ELM methodology has depicted considerable outcomes with a detection rate of 83.81%.

Table 1: COMPARATIVE ANALYSIS

S. No.	Using classification	Platform	Performance measures
1	Genetic K-means	NS2.34	Detection rate, False positive rate
2	KNN	Used three scenarios	Detection rate, Accuracy
3	ID-GOPA	NS-2	Accuracy, Precision, Recall, F1-Score
4	Optimal SVM (OSVM)	MATLAB2014b	Accuracy, true positive rate, false negative rate

IV. CONCLUSION

Providing security services in WSN based on intrusion detection systems to identify attacks with high accuracy is a challenging task. An effective intrusion detection system should provide reliable and continuous detection service. However, many of the current systems generate high false positive and false negative rates. In this paper we have presented the survey on wireless sensor networks security by using different machine learning classifications. These machine learning models are Genetic K-means algorithm (GKA), K-Nearest Neighbor (KNN), Intrusion detection model based on information gain ratio and online passiveaggressive algorithm (ID-GOPA) and Optimal Support Vector Machine (OSVM). The model determines the presence of an intrusion, and classifies the type of attack in real-time environment. According to performance measures are these machine learning models based security of wireless sensor network is analyzed.

V. REFERENCES

[1] Dong Seong Kim, Kok Onn Chee, Mengmeng Ge, "A Novel Graphical Security Model for Evolving Cyber Attacks in Internet of Things", 2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S), Year: 2020

[2] Guoping You, Yingli Zhu, "Structure and Key Technologies of Wireless Sensor Network", 2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC), Year: 2020

[3] A. Karthikeyann, V.P. Arunachalam, S. Karthik, P. Dhivya, "Energy Efficient Structure Free and Location Based Routing Protocol in WSN", 2018 International Conference on Soft-computing and Network Security (ICSNS), Year: 2018

[4] Umashankar Ghugar, Jayaram Pradhan, "NL-IDS: Trust Based Intrusion Detection System for Network layer in Wireless Sensor Networks", 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Year: 2018

[5] Skhumbuzo Zwane, Paul Tarwireyi, Matthew Adigun, "Performance Analysis of Machine Learning Classifiers for Intrusion Detection", 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC), Year: 2018

[6] Scott Ruoti, Scott Heidbrink, Mark O'Neill, Eric Gustafson, Yung Ryn Choe, "Intrusion Detection with Unsupervised Heterogeneous Ensembles Using Cluster-Based Normalization", 2017 IEEE International Conference on Web Services (ICWS), Year: 2017

[7] Ehsan Kargaran, Danilo Manstretta, Rinaldo Castello, "A 30μW, 3.3dB NF CMOS LNA for wearable WSN applications", 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Year: 2017

[8] Y. E. Mourabit, A. Toumanari, A. Bouirden, and N. E. Moussaid, "Intrusion Detection Techniques in Wireless Sensor Network using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection," Int. Journal of Advanced Computer Science and Applications, vol. 6, no 9, pp. 164–172, 2015. DOI: DOI:10.14569/IJACSA.2015.060922.

[9] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and robust automated machine learning," in Advances in neural information processing systems, 2015.

[10] A. B. Raj, M. V. Ramesh, and R. V. Kulkarni, "Security Enhancement in Wireless Sensor Networks Using Machine Learning," IEEE 14th Int. Conf. on High Performance Computing and Communication and IEEE 9th Int. Conf. on Embedded Software and Systems, pp. 1264–1269, 2012. DOI: 10.1109/HPCC.2012.186.

[11] Z. Yu and J. J. P. Tsai, "A Framework of Machine Learning Based Intrusion Detection for Wireless Sensor Networks," IEEE Int. Conf. on Sensor Networks, Ubiquitous, and Trustworthy Computing, pp. 272– 279, 2008. DOI: 10.1109/SUTC.2008.39.

[12] Sandhya G, Anitha Julian, "Intrusion Detection in Wireless Sensor Network Using Genetic K-Means Algorithm", IEEE International Conference on Advanced Communication Control and Computing Teclmologies (ICACCCT), 2014.

[13] E. M. d. L. Pinto, R. Lachowski, M. E. Pellenz, and M. C. Penna, “A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks,” IEEE 32nd Int. Conf. on Advanced Information Networking and Applications, pp. 752–758, 2018. DOI: 10.1109/AINA.2018.00113.

[14] Samir Ifzarne, Hiba Tabbaa, Imad Hafidi, Nidal Lamghari, “Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks”, The International Conference on Mathematics & Data Science (ICMDS), Year: 2020.

[15] Sibi Amaran, Dr. R. Madhan Mohan, “Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks”, Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Year: 2021.