

Intrusion detection in WSN using Supervised Machine Learning Techniques

Deepa Jeevaraj¹, Dr. B. Karthik^{2*}, Dr. M. Sriram³, Dr. S. P. Vijayaragavan⁴ and Dr. D. Gokulakrishnan⁵

Submitted: 11/10/2023

Revised: 14/12/2023

Accepted: 23/12/2023

Abstract: Wireless sensor network (WSN) is becoming increasingly one of the trendiest research areas in Computer Science applications. It finds wide applications department of Defence, banking, hospital, marketing, education, and all prioritized government sectors. Applications that have created many problems especially in security levels and hindrance caused due to the intrusion in WSN based communication. In proposed system depending upon the security and dependability of this article builds model on IoT is established using machine learning algorithms. This intrusion detection system is very compatible and characteristics of determining the interactions in any dataset have given an exemplary classification, performance level and receiver of operator characteristics. This paper uses specialized data set of WSN to detect and classify different class attributes like black hole flooding and scheduling attacks. This paper considers the use of novel Framework that is trained using a dataset to detect and classify different attacks. Output results of the model show that WSN has improved ability for the intrusion detection system using higher classification and accuracy rate of 99.45% for IBk classifier using Weka tool. The precision rate for the built model is 97% and the area under the curve is also gave an optimum result ranging from 0.77 to 0.985 for Naïve Bayes multinominal and IBk classifier. An optimum model is built using the Weka tool which is trained using the dataset different types of attacks using some selected classifiers. The attacks like black hole, flooding, scheduling, and grey hole were predicted in WSN.

Keywords: Wireless Sensor Network, Intrusion Detection, Machine Learning, Naïve Bayes, WEKA, IBK, One R

1. Introduction

The IDS and intrusion preventing systems is established to find the anomalies, threats that take place in the network. Wireless sensor networks have become the heart of information exchange in wide range of applications in real time, in the department of defence surveillance security, environmental monitoring, forecasting and health care. A WSN consists of number of nodes that is distributed all over the geographical area connected with multiple systems [1]. IDS system is important security monitoring component that is very much useful in the trending world of information interchange and things to happen. A wireless sensor network is a powerful area where information exchange should happen only between the authenticated persons. Some

Malware signatures and threats in the network should be identified by these IDS and safeguarded using some special firewalls and prevent the network from any intrusions and hacks [2,3].

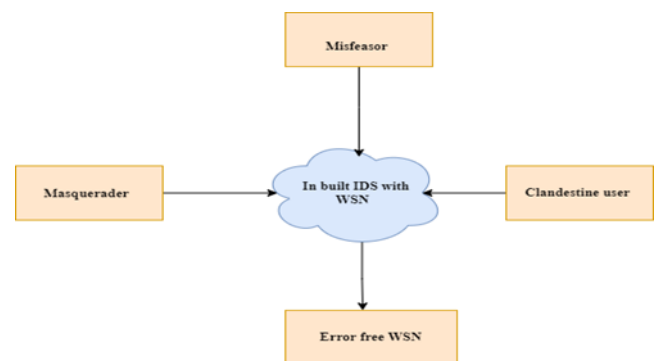


Fig 1 WSN with IDS

Fig 1 shows a block diagram of a wireless sensor network which is built in intrusion detection system using machine learning Framework. The three main classes of intruders are masquerade, misfeasor and clandestine user [5-8]. The first one type of intrusion is an unauthorized person who enter into the system and misuse the whole network as an authenticated Network and spoil the authenticated user's account from outside. Misfeasor intrusion is a person from outside and he is an authenticated person but he is not authorized to access or use all his privileges. The third one clandestine is a person whose supervise control over the whole and other from inside or outside who has a super

¹Research Scholar, Department of ECE, Bharath Institute of Higher Education and Research, India. Email: jdeepainbox@gmail.com
ORCID ID : 0000-0001-8495-495X

²Associate Professor, Department of ECE, Bharath Institute of Higher Education and Research, India. Email: karthik.ece@bharathuniv.ac.in, karthikguru33@gmail.com
ORCID ID : 0000-0003-1890-0640

³Associate Professor, Department of CSE, Bharath Institute of Higher Education and Research, India. Email: msr1sriram@gmail.com
ORCID ID : 0000-0002-9866-6374

⁴Professor, Department of EEE, Bharath Institute of Higher Education and Research, India. Email: vijayaragavan.eee@bharathuniv.ac.in
ORCID ID : 0000-0001-8638-7645

⁵Assistant professor, Department of Computing Technology, SRM Institute of Science and Technology, Chennai, India.

Email: gokulakd@srmist.edu.in ORCID ID : 0000-0002-6777-7610

*Correspondence: Email: karthik.ece@bharathuniv.ac.in

power network and take the supervisory control of the network [9].

Wireless sensor network are the essential media in all application. An attacker can overcome the sensor nodes to corrupt the network resources. Daniel of service attacks is one of the most general intrusions of WSN security. Traditional methods like Cryptography is the security technique adopted for protecting the WSN from external intrusions. It also ensures many type of identification of intrusions in the network buy used metric in Compression or public key cryptography and many other methods [10].

These techniques are not going to help in detecting the internal attacks by just security of cryptography methods. This is the first line of Defence Services in networks whereas the actual task is to detect the internal intrusions like black hole flooding gray hole etc. Some specialized framework like machine learning concepts is applied in identifying the IDS. This contributes the efficiency that can be performed in the second line of Defence Services known as attacks which are suspicious and Malware activities that are detected using the proposed methods.

IDS is mainly considered in this article is network based intrusion detection system that is used to monitor analyze and capture the malicious data. A predefined list of all possible attacks is there which is to notify the system administrator about the type of attack happened in the network [11]. The type of anomalies that threaten the reliability, the secrecy, and the accessibility of the networks. The nodes of the systems that is interconnected to produce very large manipulations of digital data in leading WSN [12]. The progress in any network system is not exploited. In order to secure information that is made by building a model that can detect the type of attacks happened in the network system.

A machine learning [13-18] based Model is built using a supervised learning technique with selected classifiers to identify the type of attacks that occur in a wireless sensor network in a very less time.

S. S. Shivaji et al. (2015)[19] The monitoring wireless sensor network has a wide range of applications in environment, health, military, and 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 maximum lifetime of network. In most of the Daniel of service attacks that destroys the network and loss of its energy rapidly is identified using a novel approach. Efficient intrusion detection system or scheme is designed in such a way that malicious node is identified by 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 machine learning algorithm.

S. Jiang et al. (2020)[20] WSN is key object in any cyber physical systems. It is composed of many stationary as well as mobile parts like sensors that transmits and receives information through WSN. The intrusion that affects the WSN has to be prevented using a special mechanism in smart environment. The novel approach is sequence backward selection algorithm that detects the attacks in a faster rate. The experiment results based on this approach has given an efficient F-measure of 0.96 0.99 for all kinds of network attacks.

S. Pundir et al. (2020) [21] A new protocol developed in WSN integrated to IoT deployment. In today's activities every common man has advancement in information communication and Technology ICT. Advancement is also suffering from various attacks that occur in WSN and IoT. As we all progress to this fast moving environment and more vulnerable security threats. In future everyone is connected to internet with numerous smart objects and for a smooth progression we are in need of IDS and IPS. This article gives in emerging intrusion detection system with the new approach. A privacy preservation protocol is integrated in WSN and IoT to address the intrusion detection Protocol in wireless sensor networks that is integrated to internet-of-things.

A. Halbouni et al. (2022)[22] Network security is an unavoidable event in our daily interactions and networks. Intrusions are also developing more and more critical as Technology also grow. Techniques employed using machine learning algorithm to detect intrusions. However there is an advantage of deep learning algorithms and AI to generate special features that automatically detect the attacks without any human intervention. Long short term memory network with spatial features is employed to detect 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 of very high and effective.

D. Thomas et al. (2021) [23] A secured energy efficient barrier coverage schedule has been developed using 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 subset of sensor nodes are overlapped to meet all the quality of service requirements. Expected node failures due to barrier security attacks such as Daniel of service are a challenging in maintaining the quality of service levels. Smart proposal using machine learning algorithm is proposed to detect The Attacks in an efficient way. WSN based IoT applications that utilizes kNN machine learning algorithm to detect the malicious attacks.

However, data standardization should be done before implementing PCA in order to find optimal principal components. Koli et al. [8] have suggested a novel technique in this research T-IDS developed an RDPLM (Randomized Data Partitioned Learning Model), which is based on feature set, and the method for feature selection, simplified sub spacing and several randomized meta learning methods. It is observed that the other machine learning models such as deep neural network, decreased error pruning the tree detection task sequential minimal optimization, and random tree.

This paper organized as follows. By giving introduction about the intrusion detecting systems for WSN and the difficulties faced due to various degrees of intrusions like black hole, grey hole, flooding etc. In Section 2 we examine Data set description by preprocessing the data set, by changing the format in attribute relation file format. In Section 3, some of selected Machine learning algorithms like Bayes category ML algorithms, One R and IBk (Instance base classifier). The investigational examinations are carried and tabulated.

In Section 4 we highlight proposed methodology of research, Section 5, analysis of Result and discussion and finally in Section 6 we describe the conclusion and future work.

2. Data description

The data set is collected from kaggle.com [14] in order to monitor, service with minimum cost. Data Set consists of nearly 18 different attributes and 5 classes of output showing the attack on the network. Attributes are namely ID, time whether the CH (node) is not 0 or 1. The received signal strength indicator and the distance to the CH. The average distance to the CH and the energy consumption of the node and the number of advertisers CH messages received from the nodes. The number of joined request messages or the request interrupt given by the CH. Acknowledgement for the number of advertisers using TDMA broadcast messages to the nodes.

TDMA messages were received from the CH. Order for the ranking of nodes within the time division multiplexing access. The data is sent and received from the nodes. The number of data packets sent to the base station and the distance between the CH and the base station. Code sent by the cluster and finally, we have nearly the output class of 5 different classes namely black hole, flooding, grey hole, normal, and TDMA the total number of instances is nearly 374662 instances.

3. Selected ML Algorithms

The selected software is Weka tool box the cross validation with 10 fold cross validation with training and testing of data

set are applied. Weka is a machine learning software with Java implementation. It performs the entire data mining task like regression, clustering, Association rule mining and visualization. It is basically from the University of Waikato, Hamilton New Zealand named with the flightless bird within inquisitive nature.

Weka is open source software under general public license. It also supports deep learning it performs all the Preprocessing, classifying, clustering, association, and visualization of the data set. Classifiers for performing the intrusion detection system model comprises of Bayes category algorithms [21], rules-based category classifiers and Instance base classifiers.

Naïve Bayes: Naive Bayes classifier basically works on unsupervised machine learning algorithm based on base year. It works on the principle of probability for classify and it is often used for NLP (Natural linguistic processing) task that comprises of text sentiment tone or opinion. The bayes theorem is basically a hypothesis with their prior knowledge available this probability is the given formula (1).

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right)P(A)}{P(B)} \quad (1)$$

NB algorithm is basically works on supervised learning algorithm it is also a probabilistic classifier. It works on the basis of probability of any object, on the principle like filtration, analysis, and classifying articles and so on. Naive means occurrence of any feature in independent with other features. Bayes means dependence on the principle, the frequency of occurrence is tabulated.

The probabilities of likelihood are framed for the conditions applying Bayes theorem to design the output of the model. Advantage of this classifier is it can handle high volume data set. Used for binary as well as multi-class classification. It is most popular for text based classification problems. Drawback of NB classifier is features independent so it cannot learn the relationship between features. Is used in real time prediction proposed methodology.

Bayesnet: Bayes net algorithm works on the principle of predictions of any detections and Diagnostics. It works on the principle of directed cyclic graph with some conditional probabilities. It is composed of a network that comprises of nodes and arcs, which are called directed links. It works on a joint probability distribution and conditional probability (2).

$$P(A_i|A_{i-1}, \dots, A_1) = P(A_i | \text{Parents}(A_i)) \quad (2)$$

where A = Events

In Weka Bayesian network learning using various search

algorithms and quality measures base. Network that uses the data structure with some conditional probability distributions. Bayes network learning algorithms has the capability of handling data which are of binary class, missing value class and nominal class. The attribute selection is of binary attributes, nominal attributes, and can also find missing values and attributes. Interfaces the drawable and weighted instances handler

One R : One Rule is a best, accurate classifying algorithm that generates one rule for each predicted data. One R algorithm selects the rule with smallest error percentage and hence called one rule algorithm. It makes the rule for each value it predicts and find the most frequent class that appears and assign a rule for each prediction and calculate the total error of the rules of each predictor. Predictor with the smallest total error.

In Weka 0 - R classifier is a rules based classifier that which builds a class by predicting the mean or the mode for a numeric and a nominal class. Zero R operates or has a capability to handle binary class, date class missing value class, nominal class, and numeric class. It can handle binary, empty nominal value, relational, and string attributes. It can interface source able and weighted instance handler.

IBk: Instant base learning algorithm which predicts with the instances close to the data point. KNN Means K nearest neighbor which is instance based learning methodology or a case based learning technique [12]. It explores using graphical user interface and capable of predicting both numerical and nominal values. It is one of the best classifier called meta classifier for classification purposes. Models are build with less time, the training as per the lazy rules is carried out, and the performance of classification is verified.

In Weka the likelihood of the nearest neighbor classifier that is able to select appropriate value of k based on cross validation. It also manipulates the distance and has the capability to handle binary, date missing value, nominal, and numeric classes. Attribute selection is also similar to the above mentioned classifiers. IBk can interface undatable classifier and Weighted instances handler.

Naive Bayes Multinomial: Multinomial Naive Bayes algorithm is also used for data with problems and multiple classes. This algorithm works by understanding the bayes theorem concept that is probability of an event with prior knowledge under related conditions for an event. It is working on the principle by satisfying the given equation (3):

$$P(A/B) = P(A) * P(B/A) \div P(B) \quad (3)$$

In Weka the NB multi nominal class is used for building classifying a model by comparison of event for Naive Bayes text classification. The capabilities of this classifier is to

handle binary, missing values, and nominal classes. It handles only the numeric attributes and interfaces the weighted instances handler.

4. Experimental Flow Procedures

The experimental flow procedures shown in Fig 2 as follows: Firstly, the data has been collected from common public platform like kaggle.com to find solutions for the WSN in predicting the intrusions using machine learning Framework. Secondly the collected data file is formatted access in Weka tool [15] in building a model using the above mentioned Framework.

The file formats are changed in attribute relation format or comma delimited value format to access in Weka tool. The next step is to build a model by the selected classifiers and train the model using 66% of the dataset and the classification task performed on the remaining 44% of the dataset. The next step is to check the following characteristics of the selected classifier for its performances percentage.

If the required characteristics rate is high enough then the model stated as optimum model. Further, we can also change the kernel size, batch size and the number of neighbors in IBk classifier to attain maximum performance.

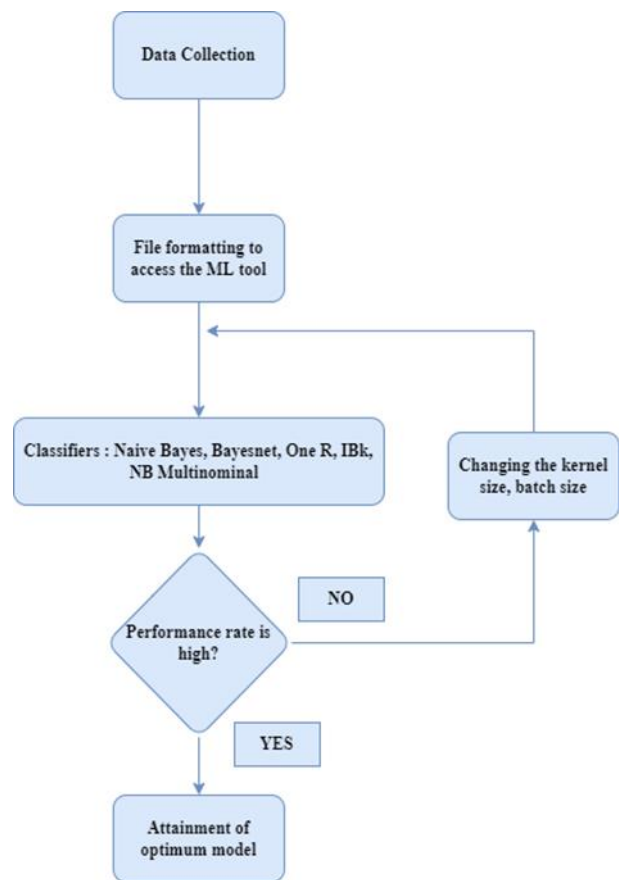


Fig 2 Experiment flow procedures

5. Result and Discussion

Table 1 gives the performance characteristics of the selected base classifier. The accuracy of the classifiers ranges from 77.7% to 99.45% for Naive Bayes multi nominal classifier and the instance based classifier. Accuracy is predictions done by the built model in a true positive way.

Table 1: Performance characteristics of the selected base classifiers

S.No	Classifier	Accuracy	Precision	Recall	F-Measure	ROC Area
1	Naive Bayes	95.3734 %	0.967	0.954	0.958	0.980
2	Bayesnet	96.6159 %	0.978	0.966	0.970	0.991
3	OneR	94.1996 %	0.933	0.942	0.934	0.771
4	IBk	99.448 %	0.994	0.994	0.994	0.985
5	NaiveBayesMultinomial	77.7006 %	0.944	0.777	0.839	0.926

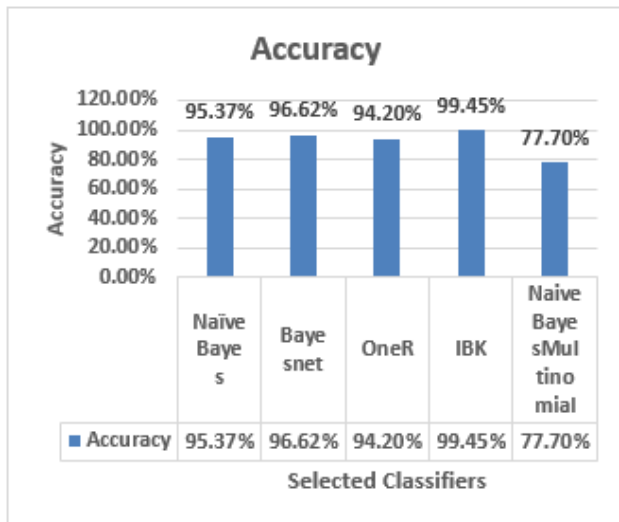


Fig 3 Classifier versus % Accuracy

Accuracy: Fig 3 indicates the accuracy percentage characteristics of the selected classifiers. The accuracy investigational report clearly support that the established model has given an optimum percentage of performance. The classifier that gives at higher rate of accuracy is instance based classifier.

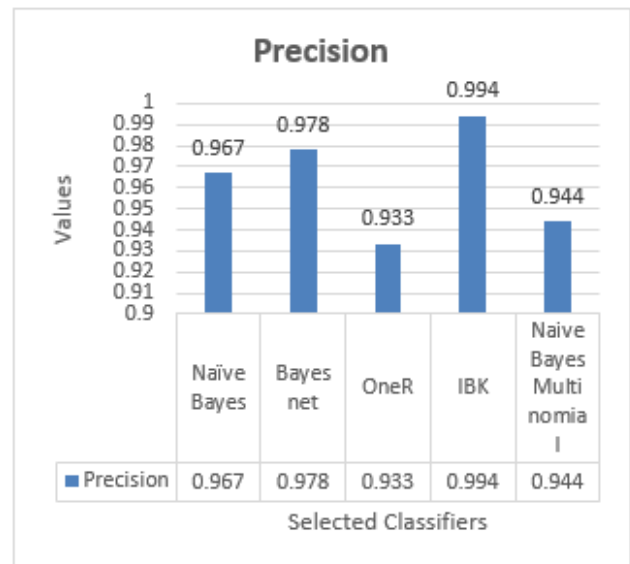


Fig 4 Classifier versus precision values

Precision: The quality of performance of the built model which has predicted in true positive direction. 100% position value is mentioned by 1 or 100% percentage depreciation value attained is less means it has incorrectly classified. The position level for the selected classifiers ranges from 0.93 to 0.99 for One R and IBk instance base classifiers. In fig 4 is the characteristic of the selected classifiers precision value. The output shows the novelty of the model with higher value for all the selected classifiers.

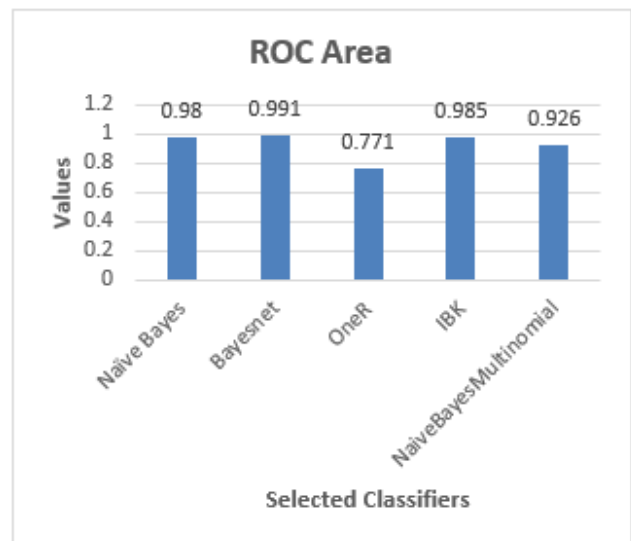


Fig 5 AUC versus Base classifier

ROC Value: The area under the curve and Receiver operator characteristics gives a cross measure of the performances done by the built model in classification. ROC is also the probability that gives the degree of measure of separately. Higher the area the better the model is predicting the classes. As shown in fig 5 the AUC value is in the range of 0.771 to 0.991 for one R and Bayes net classifiers.

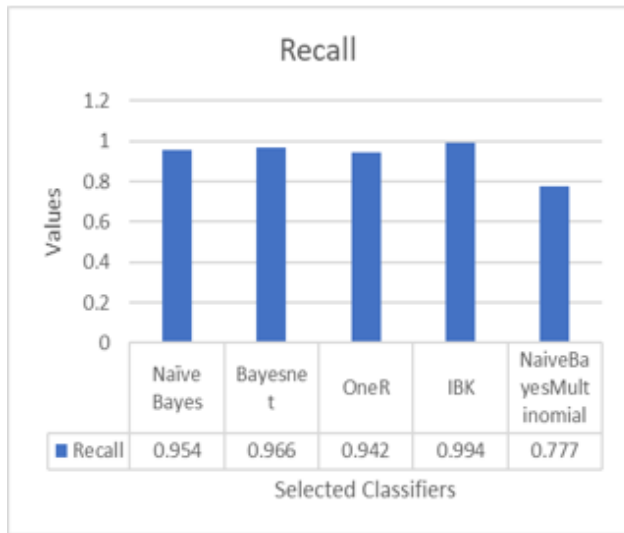


Fig 6 Classifier versus Recall values

Recall: As shown in fig 6 the recall value is also similar to the ROC values. The false negative classification of 10% is allowed for an optimum model. It follows the below mentioned equation. The span for recall in the investigational report is 0.777 to 0.994 for NB multinomial and IBk classifiers.

F-Measure: As shown in fig 7 the F-Measure value is similar to the Recall values. The false negative classification of 10% is allowed for an optimum model. It follows the below mentioned equation. The span for F-Measure in the investigational report is 0.839 to 0.994 for IBk classifiers.

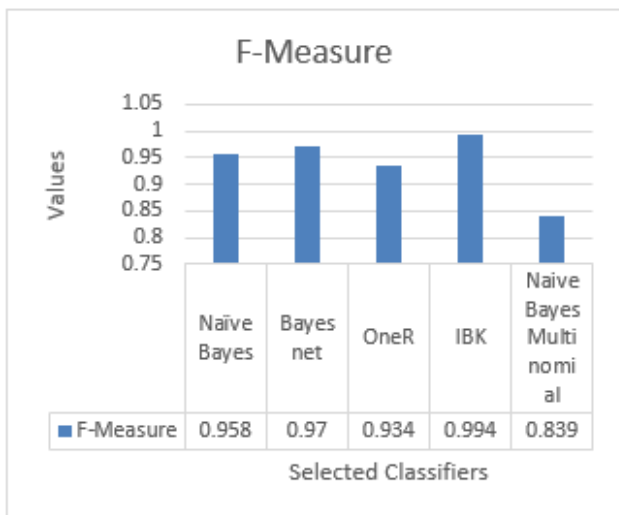


Fig 7 Classifier Vs F measure values

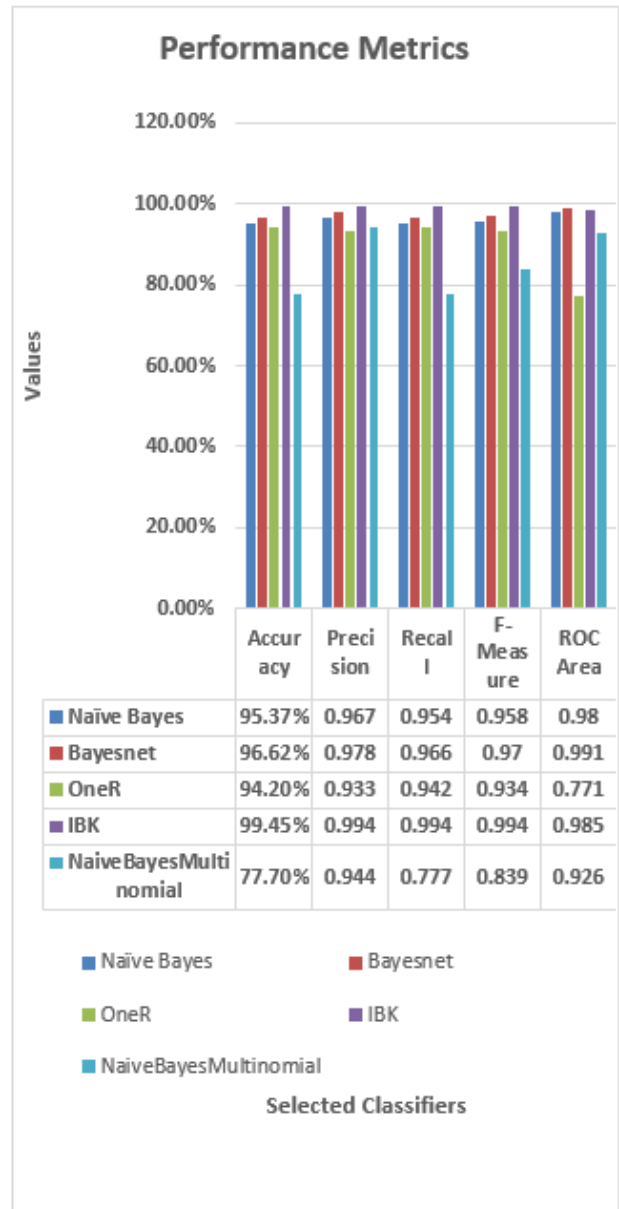


Fig 8 Overall performances of the 5 different classifiers

The overall performance of the built model with five different classifiers in which IBk classifier to attain maximum performance is as shown in fig 8. The result exhibited by the IBk classifier has shown exemplary performance characteristics after building the model which has given at highest rate of accuracy. Highest Precision and recall the f measure is also obtained. The receiver operating characteristic also exhibit I am unique or normal value of greater than 0.9.

6. Conclusions

Currently this research article in intelligent module in recommending. The Attacks that occur in wireless sensor networks. This also serve as a guide for any unknown attacks that may happen in wireless sensor network using machine learning algorithm we can sort out the intrusion in the system. Trending machine learning algorithm has given a start up to build a model in this sector. This can further

extended using deep learning methodologies to explore any types of attacks that make the system to fail in essential situations.

Author contributions

Deepa Jeevaraj and B.Karthik: Conceptualization, Methodology, Software, Field study **Deepa Jeevaraj B.Karthik and M.Sriram:** Data curation, Writing-Original draft preparation, Software, Validation., Field study **Deepa Jeevaraj, S.P.Vijayaragavan and D.Gokulakrishnan** Visualization, Investigation, Writing-Reviewing and Editing.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflicts of interest.

References

- [1] R. Ruskone, S. Airault, and O. Jamet, "Vehicle Detection on Aerial Images", *International Journal of Intelligent Engineering and Systems*, Vol. 1, No. 1, pp. 123-456, 2009.
- [2] Liu Zhiqiang, Ghulam Mohiuddin, Zheng Jiangbin, Muhammad Asim, Wang Sifei, "Intrusion detection in wireless sensor network using enhanced empirical based component analysis," *Future Generation Computer Systems*, Volume 135, 2022, Pages 181-193.
- [3] Creech, G.; Hu, J. A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns. *IEEE Trans. Comput.* 2014, 63, 807–819.
- [4] Vokorokos, L.; Baláž, A. Host-Based Intrusion Detection System. In *Proceedings of the 2010 IEEE 14th International Conference on Intelligent Engineering Systems*, Las Palmas, Spain, 5–7 May 2010; pp. 43–47
- [5] Mohammadi, S., Mirvaziri, H., Ghazizadeh-Ahsae, M. and Karimipour, H., Cyber intrusion detection by combined feature selection algorithm. *Journal of information security and applications*, 44, (2017), 80-88.
- [6] Koli, MS & Chavan, MK 2017, 'An Advanced Method for Detection of Botnet Traffic using Intrusion Detection System', *IEEE International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 481-485.
- [7] Sun, Y & Liu, F, 2016, 'SMOTE-NCL: A Re-Sampling Method with Filter for Network Intrusion Detection', *IEEE International Conference on Computer and Communications (ICCC)*, pp. 1157-1161.
- [8] Vijayan T, Sangeetha M, A. Kumaravel, Karthik B, "Feature selection for Simple Color Histogram Filter based on Retinal Fundus Images for Diabetic Retinopathy recognition," *IETE Journal of Research*, 2020. <https://doi.org/10.1080/03772063.2020.1844082>
- [9] H. Elbahadır and E. Erdem, "Modeling Intrusion Detection System Using Machine Learning Algorithms in Wireless Sensor Networks," 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021, pp. 401-406, DOI: 10.21817/indjce/2021/v12i6/211206061.
- [10] Geetha Hari Priya ,K. Amuthavalli, T.Vijayan et al., "Classifiers with synthetic oversampling preprocess for In Vitro Fertilization predictions," *Indian Journal of Computer Science and Engineering (IJCSSE)*, Vol. 12, No. 6, pp. 1532-1541 Nov-Dec 2021. DOI: 10.21817/indjce/2020/v11i5/201105266.
- [11] Sajjad, S.M.; Bouk, S.H.; Yousaf, M. Neighbor Node Trust Based Intrusion Detection System for WSN. *Procedia Comput. Sci.* 2015, 63, 183–188.
- [12] <https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds>
- [13] <https://www.cs.waikato.ac.nz/ml/weka/>
- [14] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 266–282, 2014.
- [15] Tan, Xiaopeng et al. "Wireless Sensor Networks Intrusion Detection Based on SMOTE and the Random Forest Algorithm." *Sensors (Basel, Switzerland)* vol. 19,1 203. 8 Jan. 2019, doi:10.3390/s19010203
- [16] Jeevaraj, D., Karthik, B., Sriram, M., Vijayaragavan, S.P., A Second Order Inductive Deduction Approach for Identification of Room Occupancy by Wireless Sensors, *Indian Journal of Computer Science and Engineering*, 13(5), pp. 1702–1715, 2022.
- [17] Mohammed S. Alsahli, Marwah M. Almasri, Mousa Al-Akhras, Abdulaziz I. Al-Issa and Mohammed Alawairdhi, "Evaluation of Machine Learning Algorithms for Intrusion Detection System in WSN" *International Journal of Advanced Computer Science and Applications(IJACSA)*, 12(5), 2021. <http://dx.doi.org/10.14569/IJACSA.2021.0120574>.
- [18] Jeevaraj, D., Karthik, B., Sriram, M., Vijayaragavan, S.P., A Novel Model for Network Intrusion Detection by Using Statistical Learning Approach, *Indian Journal of Computer Science and Engineering*, 13(5),

pp. 1688–1701, 2022.

- [19] S. S. Shivaji and A. B. Patil, "Energy Efficient Intrusion Detection Scheme Based on Bayesian Energy Prediction in WSN," 2015 Fifth International Conference on Advances in Computing and Communications (ICACC), 2015, pp. 114-117, doi: 10.1109/ICACC.2015.107.
- [20] S. Jiang, J. Zhao and X. Xu, "SLGBM: An Intrusion Detection Mechanism for Wireless Sensor Networks in Smart Environments," in IEEE Access, vol. 8, pp. 169548-169558, 2020, doi: 10.1109/ACCESS.2020.3024219.
- [21] S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. J. P. C. Rodrigues and Y. Park, "Intrusion Detection Protocols in Wireless Sensor Networks Integrated to Internet of Things Deployment: Survey and Future Challenges," in IEEE Access, vol. 8, pp. 3343-3363, 2020, doi: 10.1109/ACCESS.2019.2962829.
- [22] Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi and R. Ahmad, "CNN-LSTM: Hybrid Deep Neural Network for Network Intrusion Detection System," in IEEE Access, vol. 10, pp. 99837-99849, 2022, doi: 10.1109/ACCESS.2022.3206425
- [23] D. Thomas, R. Shankaran, M. A. Orgun and S. C. Mukhopadhyay, "SEC2: A Secure and Energy Efficient Barrier Coverage Scheduling for WSN-Based IoT Applications," in IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 622-634, June 2021, doi: 10.1109/TGCN.2021.3067606.