Effects of Feature Selection Techniques In Detecting Gray Hole Attacks In Ad-Hoc Wireless Sensor Networks Using Supervised Machine Learning Algorithms

Navneet[1], Akash Deep[2], Anshul kalia[3]
[1]Dept of CSE
[2, 3]Professor, Dept of CSE
[1, 2, 3] HPU

Abstract- *Wireless Sensor networks (WSNs) have become immensely popular due to their simplicity, low cost, ease of deployment, and wide application area.WSNs are a group of tiny autonomous sensor-equipped devices that are deployed in physical or environmental conditions for information gathering. Some of the applications of WSN are forest fire detection, the Establishment of smart roads, tracking parking zones, etc. WSN introduces numerous security threads as a result of its widespread use. The most frequent attack that can harm WSNs is a DOS attack. Grayhole attack is one of the popular attacks against WSNs. The Gray Hole attack is extremely harmful to sensor node networks and causes widespread network malfunctions as well as communication issues across all sensor networks. In this paper, a security mechanism is been proposed to detect grayhole attacks in WSN using the Machine Learning model. Moreover, multi-class classification has been performed on the WSN-DS dataset aiming for gray hole attack detection, 3 different feature selection techniques have been performed and 7 different supervised machine learning classifiers have been implemented on the 3 different feature sets. Parameters such as Precision, Accuracy, Recall, F1-support, and validation are used for evaluation and comparison purposes.Out of all the feature selection techniques, the Univariate Statistical method performed the best with the highest accuracy of 99.8% in the RFC and DTC model.*

Keywords- Machine learning, multi-classification, Intrusion Detection, WSN-DS Dataset..

I. INTRODUCTION

Wireless Sensor networks can be regarded as a group of autonomous sensor-equipped devices that are deployed in physical or environmental conditions for information-gathering and are well organized in an Ad-hoc manner. This type of network works with the collaboration of finite sensor nodes and sink nodes. The sensor nodes gather useful information from the surroundings and forward it to the sink node. The sink node is used to capture the data from the various sensor nodes and acts as a gateway to the external systems. WSNs can work in a centralized as well as decentralized manner [1] . WSNs are bi-directional allowing control of sensor activity from the base station to the sensor as well as the transfer of information that can be traced from nodes to a central node or base station [2]. A GPS component is also equipped inside the device for tracking their location. The WSN plays an important role in modeling the smart world. Our day-to-day activities involve many WSN devices; we often use them to make our work easier and faster. Some of the major applications of WSN are forest fire detection mechanism that helps in monitoring combustion gases and creating alert zones in the suspected area, the establishment of smart roads by deploying sensor network on roads that help in generating warning messages to the drivers, and help them to avoid unexpected events like accidents or traffic jams, tracking parking zones by deploying wireless sensor nodes, water purity measurement using sensor nodes in the river system, measuring noise level of surroundings, landslide detection. Eventually, the wireless sensor node has been appointed to play important roles in these types of areas. WSN introduces numerous security threads as a result of its widespread use. These sensor nodes are compact in size which makes it difficult to fit resources inside them. The compact nature of the devices causes insufficient storage space therefore, it becomes very difficult to run big codes. Indeed, to establish effective security techniques, it is necessary to limit the size of the security algorithm code [3]. The most frequent attack that can harm WSNs is a DOS attack [4]. DOS cyberattacks have gained popularity recently, but still, it is very difficult to mitigate their impact on the networks. A DOS attack can harm a network service by flooding its resources with numerous fake requests and preventing legitimate traffic from entering the network [5] A severe security concern that not only partially loses a packet but also compromises communication is the grayhole attack. Gray hole attack is the modified version of the blackhole attack. In blackhole attack. When the RREQ request message is put on by the destination node, the blackhole node sends a fake RREP message, and this is how the fake node drops all the packets of the traffic. Whereas, gray hole attack drops selective packets and this makes it difficult to detect them [6].

Presently, there are certain drawbacks to the use of standard wireless network intrusion detection techniques, such as low detection accuracy, low precision rate, and a high false positive rate [7]**.** Machine learning technique has gained increasing attention as a result of advancements in machine learning prediction models. In particular, when working with huge datasets, machine learning techniques enhance automating and reforming attack detection. Moreover, you can choose the algorithm based on your problem and even combine various techniques for the best results. ML has a significant potential to improve the security of WSNs. In this paper, multi-class classification has been performed on the WSN-DS dataset aiming for grayhole attack detection, 3 different feature selection techniques have been performed and 7 different supervised machine learning classifiers have been implemented on the 3 different feature sets. Parameters such as Precision, Accuracy, Recall, F1-support, and validation are used for evaluation and comparison purposes.

The Structure of the paper is as follows, Section 2 contains Literature Review, Section 3 contains Research Methodology, Section 4 contains the Results, Section 5 contains an Analysis of the Results, and Section 6 contains the Conclusion and Future Scope.

II. LITERATURE REVIEW

In current times, with the advancement in machine learning technology, the automation of WSNs operations has been experienced. Various machine learning strategies have been used in WSNs to handle the information and increase the performance of the network. The integration of multiple autonomous, tiny, low-cost, and low-power sensor nodes adds up to the formation of a wireless sensor network [3]. Sensor networks usually contain small sensor nodes which are deployed in various applications like animal tracking, traffic monitoring and control**,** forest fire detection, and habitat monitoring [8]. There are various risks of attacks on the WSN networks, intrusion detection has proved to be the best defense technique against WSN attacks in the past [9].

The Intrusion detection techniques are divided into two types: Anomaly-based and Signature-based [10], in the Anomaly-based pattern technique, needs to check the network connectivity at regular intervals and also compare the ongoing WSN network activities with the current normal behavior.

The type of data which is used in the process of Machine Learning plays a significant role in the successful Detection of anomalies in the WSN environment. Data is mostly downloaded from the internet and contains a lot of abnormalities, Machine Learning provides a solution by providing tools like MinMaxScaler to scale the data, Normalizer to normalize the data, and Standard Scaler to standardize the data. Scaling the data makes the the dimensionality of the data more linear, whereas normalizing the data rounds off the data into a much similar and simpler format [11].

Feature selection techniques also play a major role in increasing the performances of the Machine learning algorithms. In the past, Feature selection techniques such As Recursive feature Elimination, Univariate Statistical method using SelectKBest and Feature Importance have been used for the detection of anomalies in the IoT and WSN. It is considered that feature selection is one of the important steps in the process of Machine learning. The features that contribute most towards the detection the accuracy is extracted through these feature selection techniques, another advantage of using Feature selection for the Dimensionality reduction of the data.

Machine Learning has been on the rise in detecting anomalies and attacks. Various Machine Learning techniques have been used to improve the performances of attacks such as DoS attacks in the WSN environment. Supervised Machine learning approaches are one of the techniques which are used to classify WSN attacks. Machine learning intends to enable machines to learn by themselves using the provided data and making accurate predictions. Machine Learning has been used in various fields such as Virtual Personal Assistants, Video Surveillance, Email spam, Malware filtering, Online Fraud Detection, speech recognition, medical services, Online service, Online customer support, and Image recognition, etc. [12].

Aruhaily et al. proposed a multi-layer intrusion detection method using two protection layers. The first layer used the Naïve Bayes-based method and utilized the binary classification method for classifying normal or malicious traffic whereas the second layer was allocated to the cloud that only managed the legitimate traffic using a multi class Random Forest classifier [13].

Carswell et al. contributed to the performance improvements in the field of the Hidden Naive Bayes Binary classifier model for intrusion detection and used the concept of multi-classification. Their experimental setup showed better performance than the traditional NB techniques [14].

Ibrahim et al. proposed a multi-class classification framework for intrusion detection in WSN. The framework contained 3 levels the first two levels were situated in WSN and the third level resided on the cloud. With the help of the OPNET simulation tool traffic was generated and performance was analyzed [15].

Sun et al. based on the enhanced V-detector technique, presented a WSN-NSA model to detect intrusions for WSN. The principal component analysis is utilized by authors to minimize detection features, and the V-detector technique is changed by altering detector generation rules [16].

Xia et al. in their experiment extracted 19 feature sets from the KDDcup99 dataset and created a classifier to detect the legitimate entry in the traffic using SVM and a combination of the clustering algorithm [17].

Nancy et al. proposed an algorithm for dynamic recursive feature selection (DRFSA) that automatically generates optimal features. Moreover, convolution neural networks have been used for classification. All the mechanism has been tested on the KDD cup dataset [18].

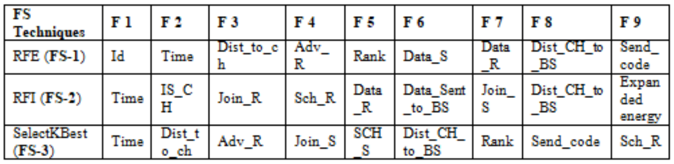
III. RESEARCH METHODOLOGY

Dataset: The dataset which is used in this research is the WSN-DS dataset which was downloaded from the Kaggle website [19]. The dataset contains four different types of attacks on the WSN environment Grayhole attacks, Blackhole attacks, TDMA, and Flooding attacks. The original dataset contains 23 features which were reduced after the process of Feature selection. There are more than 3 lakh rows in the dataset and 23 different features.

Pre-processing: The data preprocessing is the next step that was followed, as the Machine only understands the numerical data the data was converted into the numerical form using one-hot encoding. After converting the data into numerical form, the Scaling, Normalization, and Standardization of the data was done. For Scaling the data MinMaxScaler was used from the Sklearn library, after scaling the data Normalizer function was used to normalize and simplify the data. The standardization of data was done by the StandardScaler from the Sklearn library, and finally, the data was prepared for the feature selection.

Feature selection: Feature selection is the technique that's used for extracting the necessary feature from the data, the advantage, and importance of using the Feature selection technique is to reduce the dimensionality of the data and increase the chance of getting better detection accuracy [20]. Three different feature selection techniques were used on the WSN dataset which gave three different feature sets as the output. Two wrapper methods and one filter method were used to analyze the impact of both methods on detection accuracy. All the feature selection techniques returned 9 different features which were then used differently for multi-classification

Table 3.1: Different feature sets obtained from feature selection techniques



FS Techniques	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9
RFE (FS-1)	Id	Time	Dist_to_ch	Adv_R	Rank	Data_S	Data_R	Dist_CH_to_BS	Send_code
RFI (FS-2)	Time	IS_CH	Join_R	Sch_R	Data_R	Data_Sent_to_BS	Join_S	Dist_CH_to_BS	Expanded energy
SelectKBest (FS-3)	Time	Dist_to_ch	Adv_R	Join_S	SCH_S	Dist_CH_to_BS	Rank	Send_code	Sch_R

Table 3.0 displays the different feature sets that are obtained after using three different feature selection techniques. Recursive feature elimination (RFE) gave 9 different features out of 23 features, RFE is a wrapper method, and Logistic Regression was used as a wrapper method to automatically extract the most important features from the data, for linearity and fair evaluation of only the top 9 features were considered from all the feature selection techniques. Random feature importance uses a Random Forest algorithm (RFI) to assert different weights on the features based on their importance, it's also a wrapper method. Lastly, the Univariate statistical method was used which is a filter method to get different feature sets based on the weights of the features,

The Univariate statistical method uses the chi^2 method to calculate the results, K best features were extracted where k being the value 9.

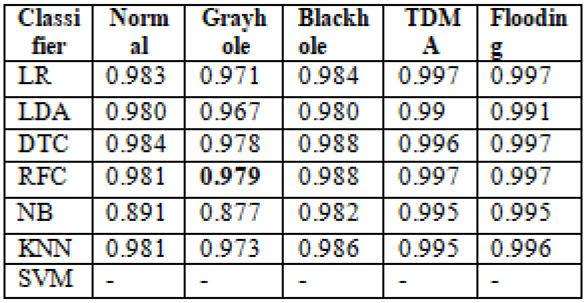
Multi-class classification: Multi-class classification [21] was done on the different feature sets using 7 different supervised machine learning algorithms. A confusion matrix was used to evaluate the performance of the different supervised machine learning algorithms, it is to be noted that to the best of our knowledge. Classification of all four attacks on the WSN environment was done on all 7 classifiers in all the 3 different feature sets to compare and evaluate the performances of the Machine Learning models.

Train-test-split: Train test-split method was used to divide the data into training testing and validation sets. The ratio of 80:20 is used for the training, and testing. The data was trained on all seven classifiers and then 30% of reserved data was tested by the train-test-split method.

IV. RESULTS

Three different feature selection techniques were used on the WSN-DS dataset and three different feature sets were obtained after feature selection. Seven Supervised Machine Learning Algorithms Logistic Regression, Linear Discriminant Analysis, K-nearest neighbor (KNN), Random Forest (RF), Decision Tree (DT), Naïve Bayes, and Support Vector Machine (SVM) were implemented on all three feature sets. Parameters such as Precision, Accuracy, Recall, and F1-support was used to evaluate and compare results.

Table 3.0: Accuracy of all 7 classifiers in detecting the gray hole attack on the FS-1 (RFE).



Classi fier	Norm al	Grayh ole	Blackh ole	TDM A	Floodin g
LR	0.983	0.971	0.984	0.997	0.997
LDA	0.980	0.967	0.980	0.99	0.991
DTC	0.984	0.978	0.988	0.996	0.997
RFC	0.981	**0.979**	0.988	0.997	0.997
NB	0.891	0.877	0.982	0.995	0.995
KNN	0.981	0.973	0.986	0.995	0.996
SVM	-	-	-	-	-

Table 3.0 represents the detection accuracy of all the implemented classifiers on Feature set 1, it can be observed from the table that the detection accuracy of the Normal class in LR was 98%, 97% in detecting gray hole attacks, 98% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting the flooding attacks. The detection accuracy of the Normal class in LDA was 98%, 96% in detecting grayhole attacks, 98% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting flooding attacks. The detection accuracy of the Normal class in DTC was 98%, 97% in detecting gray hole attacks, 98% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting the flooding attacks. The RF classifier gave the best accuracy in detecting grayhole attacks with 97.9% accuracy, while 98% on the black hole attack, 98% on the Normal class, and 99% in detecting both TDMA and flooding attacks. The Naïve Bayes classifier gave a detection accuracy of 87% in grayhole attacks which are average as compared to other classifiers, it detected black holeattack with 98.9% accuracy, while 89%was on the Normal class, and 99% in detecting both TDMA and flooding attacks. The KNN classifier gave a detection accuracy of 97% in gray hole attacks which is encouraging, it detected black hole attacks with 98% accuracy, while 98%was on the Normal class, and 99% in detecting both TDMA and flooding attacks. The results of SVM are awaited because of the limitations of the infrastructure required for viewing the algorithm’s performance. It can be concluded after observing the results that the RF classifier performed best out of all the other classifiers in detecting the gray hole attack with 97.9% detection accuracy, other classifiers also gave encouraging results but the RF classifier performed the best.

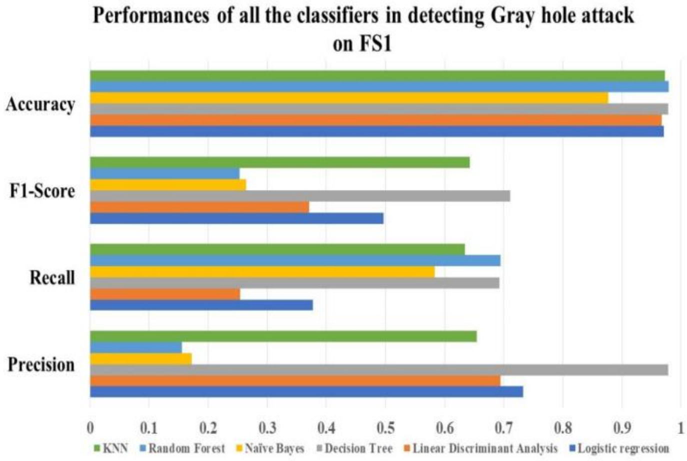
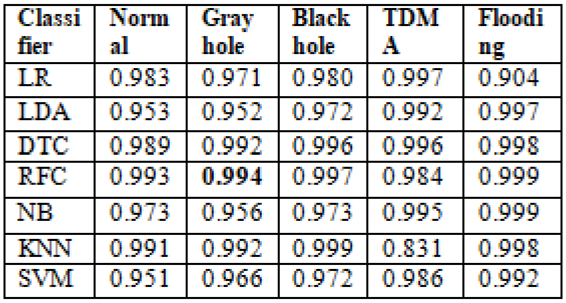


Figure 3.0: Precision, Recall, and F1-score of all seven classifiers on detecting gray hole attack in the FS-1 (RFE)

Figure 3.0 represents the Precision, Recall, and F1 support of the implemented classifiers on the feature set 1, it can be observed from figure 3.0 that Logistic regression gives the precision in detecting the gray hole was below average 63%, 37% recall, and F1-score of 49%. LDA performed poorly with 69% in detecting gray hole attacks, recall of 25%, and F1-score of 37%. DTC gave 97% in detecting the gray hole attack, recall of 69%, and F1-score of 71%.

Naïve Bayes performed poorly giving only 17% precision on the gray hole attack, recall of 58%, and F1-score of 26%. KNN classifier gave 65% precision on the gray hole attack, recall of 63%, and F1-score of 64%. RF classifier gave 15% precision in the gray hole attack, recall of 69%, and F1-score of 25%. It can be concluded from observing figure 3.1 that DTC gave the best precision of 98% in the gray hole attack on the FS-1. Scores of SVM classifiers could not be displayed because of the lack of the infrastructure required for evaluating the performance of the algorithm.

Table 3.1: Accuracy of all 7 classifiers in detecting the gray hole attack on the FS-2 (Univariate statistical method).



Classifier	Normal	Gray hole	Black hole	TDMA	Flooding
LR	0.983	0.971	0.980	0.997	0.904
LDA	0.953	0.952	0.972	0.992	0.997
DTC	0.989	0.992	0.996	0.996	0.998
RFC	0.993	**0.994**	0.997	0.984	0.999
NB	0.973	0.956	0.973	0.995	0.999
KNN	0.991	0.992	0.999	0.831	0.998
SVM	0.951	0.966	0.972	0.986	0.992

Table 3.1 represents the detection accuracy of all the implemented classifiers on Feature set 2, it can be observed from the table that the detection accuracy of the Normal class in LR was 98%, 97% in detecting gray hole attacks, 98% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 94% in detecting the flooding attacks. The detection accuracy of the Normal class in LDA was 95%, 95% in detecting gray hole attacks, 97% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting the flooding attacks.

The detection accuracy of the Normal class in DTC was 98%, 99% in detecting gray hole attacks, 99% in detecting the black hole attack, 99% in detecting TDMA attacks, and 99.4% in detecting flooding attacks. The RF classifier gave the best accuracy in detecting gray hole attacks with 99.4% accuracy, while 99% on the black hole attack, 99% on the Normal class, 98% in detecting TDMA, and 99% in detecting the flooding attacks.

The Naïve Bayes classifier gave a detection accuracy of 95% in grayhole attacks, it detected black hole attacks with 97% accuracy, while 97% was on the Normal class, and 99% in detecting both TDMA and flooding attacks. The KNN classifier gave a detection accuracy of 99% in gray hole attacks, it also detected black hole attacks with 99% accuracy, while 99% on the Normal class, 83% in detecting

TDMA, and 99% detection accuracy in flooding attacks. The results of SVM were extracted from the confusion matrix on the FS2 as the infrastructure allowed, gray hole was detected with 96% accuracy, 97% accuracy in detecting the black hole attack, 95% detection accuracy in Normal class, 98% detection accuracy in TDMA attacks, and 99% detection accuracy on the Flooding attacks. It can be concluded after observing the results that the RF classifier performed best out of all the other classifiers in detecting the gray hole attack with 99.4% detection accuracy, other classifiers also gave encouraging results but the RF classifier performed the best.

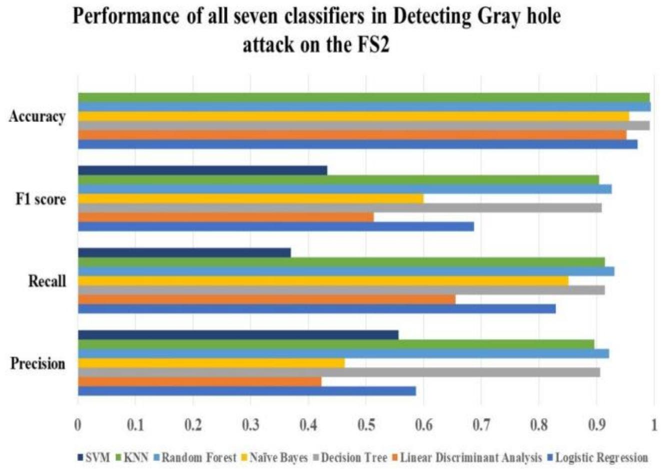
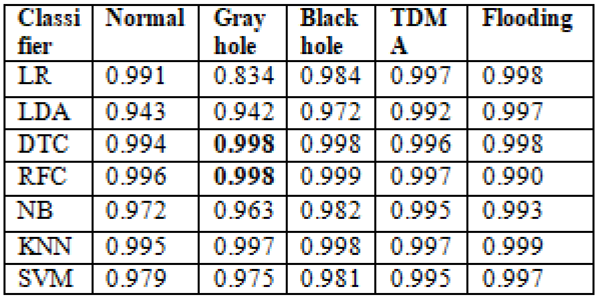


Figure 3.1: Precision, Recall, and F1-score of all seven classifiers on detecting gray hole attack in the FS-2 (Univariate statistical method)

Figure 3.1 represents the Precision, Recall, and F1 score in detecting gray hole attacks after implementing all seven classifiers on the feature set 3, it can be observed from figure 3.1 that Logistic regression gave the precision in detecting the gray hole was below average %, recall 82%, F1-score of 68%. LDA performed poorly with only 69% precision in detecting gray hole attacks, recall of 65%, and F1-score of 51%. DTC gave encouraging results of 90% precision in detecting the gray hole attack, recall of 91%, and F1-score of 90%, it can be observed that the precision went slightly higher in DTC as compared to LR and LDA. Naïve Bayes performed poorly giving only 17% precision on the gray hole attack, recall of 85%, and F1-score of 59%. KNN classifier gave 65% precision on the gray hole attack, recall of 91%, and F1-score of 90%. RF classifier gave 92% precision in the gray hole attack, recall of 93%, and F1-score of 92%. KNN classifier gave 65% precision on the gray hole attack, recall of 91%, and F1-score of 90%.SVM classifier gave 55% precision in the gray hole attack, recall of 35%, and F1-score of 43%. It can be concluded from observing figure 3.1 that the RF classifier gave the best performance on the FS2 as compared to all other classifiers.

Table 3.2: Accuracy of all 7 classifiers in detecting the gray hole attack on the FS-3 (Random Feature Importance).



Classifier	Normal	Gray hole	Black hole	TDMA	Flooding
LR	0.991	0.834	0.984	0.997	0.998
LDA	0.943	0.942	0.972	0.992	0.997
DTC	0.994	**0.998**	0.998	0.996	0.998
RFC	0.996	**0.998**	0.999	0.997	0.990
NB	0.972	0.963	0.982	0.995	0.993
KNN	0.995	0.997	0.998	0.997	0.999
SVM	0.979	0.975	0.981	0.995	0.997

Table 3.2 represents the detection accuracy of all the implemented classifiers on Feature Set 3, it can be observed from the table that the detection accuracy of the Normal class in LR was 99%, 83% in detecting gray hole attacks, 98% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting the flooding attacks.

The detection accuracy of the Normal class in LDA was 94%, 94% in detecting gray hole attacks, 97% in detecting the black hole attack, 99% in detecting the TDMA attacks, and 99% in detecting the flooding attacks. The detection accuracy of the Normal class in DTC was 99.4%, 99.8% in detecting gray hole attacks, 99.8% in detecting the black hole attack, 99.6% in detecting the TDMA attacks, and 99.8% in detecting the flooding attacks.

DTC and RF classifiers gave the best accuracy in detecting gray hole attacks with 99.8% accuracy, while 99% on the black hole attack, 99% on the Normal class, 98% in detecting TDMA, and 99% in detecting the flooding attacks. The Naïve Bayes classifier gave a detection accuracy of 96% in the gray hole attack, it detected the black hole attack with 98% accuracy, while 97% was on the Normal class, and 99% in detecting both TDMA and flooding attacks.

The KNN classifier gave a detection accuracy of 99% in gray hole attacks, it also detected black hole attacks with 99% accuracy, while 99% on the Normal class, 99% in detecting TDMA, and 99% detection accuracy in flooding attacks.

The results of SVM were extracted from the confusion matrix on the F3 as the infrastructure allowed, gray hole was detected with 97% accuracy, 98% accuracy in detecting the black hole attack, 97% detection accuracy in Normal class, 99% detection accuracy in TDMA attacks, and 99% detection accuracy on the Flooding attacks.

It can be concluded after observing the results that the RF and DTC classifier performed best out of all the other classifiers in detecting the gray hole attack with 99.8% detection accuracy, other classifiers also gave encouraging results but the RF classifier performed the best.

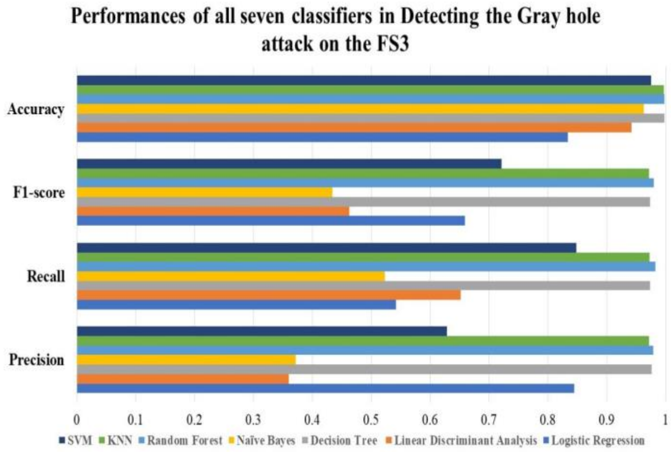


Figure 3.2: Precision, Recall, and F1-score of all seven classifiers on detecting gray hole attack in the FS-3 (Random Feature Importance).

Figure 3.2 represents the Precision, Recall, and F1 score in detecting gray hole attacks after implementing all seven classifiers on the feature set 3, it can be observed from figure 3.2 that Logistic regression gives the precision in detecting the gray hole was below average 84%, recall 54%, F1-score of 65%. LDA performed poorly with only 36% precision in detecting gray hole attacks, recall of 65%, and F1-score of 46%. DTC gave encouraging results of 97% precision in detecting the gray hole attack, recall of 97%, and F1-score of 71%, it can be observed that the precision went slightly higher in DTC as compared to LR and LDA. Naïve Bayes performed poorly giving only 37% precision on the gray hole attack, recall of 52%, and F1-score of 43%. KNN classifier gave 97% precision on the grayhole attack, recall of 97%, and F1-score of 97%. RF classifier gave 97% precision in the gray hole attack, recall of 97%, and F1-score of 98%. SVM classifier gave 62% precision in the gray hole attack, recall of 84%, and an F1-score of 72%. It can be concluded from observing figure 3.1 that the RF and KNN classifiers gave the best performance on the FS3 as compared to all other classifiers.

V. ANALYSIS

The Analysis from the results section can be drawn that the results in the case of Feature set 1, the RF classifier performed the best in terms of detection accuracy out of all other classifiers in detecting the grayhole attack, though other classifiers also gave a high performance, RF performed the best. In terms of Precision, the DTC classifier gave the best precision on the FS-1 of 98%, while giving 69% recall. On Feature set 2, the RF classifier performed the best in terms of detection accuracy out of all other classifiers with 94% accuracy in detecting the grayhole attack, though other classifiers also gave a high performance, RF performed the best. In terms of Precision, the again RF classifier gave the best precision on the FS-2 of 91% while giving 93% recall. On Feature set 3, the RF and DTC classifier performed the best in terms of detection accuracy out of all other classifiers with 98% accuracy in detecting the grayhole attack. In terms of Precision, again RF and KNN classifiers gave the best precision on the FS-3 of 97% while giving the best recall of 97% in DTC, RF, and KNN.

It can be observed that different feature selection techniques affect the performances of the models and the detection accuracy, out of all the feature selection techniques Univariate Statistical method performed the best with the highest accuracy of 99.8% in the RFC and DTC model, other two techniques also gave encouraging results. It can also be observed that the performances of Linear models i.e., Logistic Regression, and LDA are relatively low compared to the other supervised Machine learning classifiers, though RFE and Univariate Statistical method techniques performed better than Random Feature importance in Linear Models, while Tree-based models i.e DTC, and RF performed better in Random Feature importance. KNN also performed very well under all the feature selection techniques.

VI. CONCLUSION & FUTURE SCOPE

WSN introduces numerous security threads as a result of its widespread use. The most frequent attack that can harm WSNs traffic is a DOS attack. Gray hole is the popular attack among them against WSN. The WSN-DS dataset used in this experiment consists of various types of DOS attacks that have been used for training purposes. Three different feature selection techniques i.e., Recursive Feature Elimination, Univariate statistical method (SelectKBest), and Random Feature Importance have been used. These three-feature selection techniques gave 3 feature sets as output. Supervised machine learning algorithms were implemented on each feature set. Out of all the feature selection techniques, the Univariate Statistical method performed the best with the highest accuracy of 99.8% in the RFC and DTC model. It can be concluded that feature selection techniques affect the performance of the learning models. In the Future, more feature selection techniques can be applied to WSN-based datasets and can evaluate the behavior of these techniques on different Machine Learning Algorithms.

REFERENCES

[1] M. Carlos-Manicilla, "Wireless Sensor Networks Formation: Approaches and Techniques, " Journal of Sensors, 2016.

[2] P. Langley, "Applications of machine learning and rule induction," Communications of the ACM, 1995.

[3] M. . S. Sabri, "Security Issues in Wireless Sensor Networks (WSN),"SSRN Electronic Journal, 2015.

[4] S. Patil, "DoS Attack Prevention Technique in Wireless Sensor Networks, "ELSEVIER, 2016.

[5] L. Alsulaiman, "Performance Evaluation of Machine Learning Techniques for DOS Detection in Wireless Sensor Network, "IJNSA, 2021.

[6] V. Shanmuganathan, "A Survey on Gray Hole Attack in MANET, "International Journal of Computer Networks and Wireless Communications, 2012.

[7] H. Yang, "Combined Wireless Network Intrusion Detection Model Based on Deep Learning, "IEEE, 2019.

[8] Z. Rehena, "Application of Wireless Sensor Network in Forest Fire Detection," in Second India Disaster Management Congress, New Delhi, 2009.

[9] L. Han, "Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model,"Science Direct, 2019.

[10] A. Khraisat, "Survey of intrusion detection systems: techniques, datasets and challenges,"Springer, 2019.

[11] C.-F. Tsai, "Intrusion detection by machine learning: A review,"ELSEVIER, 2009.

[12] A. A. Iorkaa, "Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights,"International Journal of Engineering Research and Applications, 2021.

[13] N. . M. Alruhaily, "A Multi-layer Machine Learning-based Intrusion Detection System for Wireless Sensor Networks,"International Journal of Advanced Computer Science and Applications, 2021.

[14] A. . D. Carswell , "Network Intrusion Detection Using a Hidden Naïve Bayes Binary Classifier,"International Journal of Simulation, Systems, Science and Technology, 2020.

[15] D. Ibrahim, "Anomaly detection in wireless sensor networks:A proposed Framework,"International association of online engeneering, 2022.

[16] Z. Sun, "An Intrusion Detection Modelfor Wireless Sensor Networks with an Improved V-Detector Algorithm,"IEEE, 2018.

[17] J. Xia, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," ELSEVIER, 2011.

[18] N. Periasamy , "Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks,"IET Communications, 2019.

[19] I. Almomani, "WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks,"Journal of sensors, 2016.

[20] G. Chandrashekar , "A survey on feature selection methods," Elsevier, 2014.

[21] W. Kim, "A multi-class classification approach for target localization in wireless sensor networks," Journal of Mechanical Science and Technology, 2014.