Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal 2023, 17(2), 322–329
https://doi.org/10.12913/22998624/161820
ISSN 2299–8624, License CC-BY 4.0

Received: 2022.11.15
Accepted: 2023.03.15
Published: 2023.04.01

Ensemble Model for Network Intrusion Detection System Based on Bagging Using J48

Mohammad Mahmood Otoom[1*], Khalid Nazim Abdul Sattar[1], Mutasim Al Sadig[1]

[1] Department of Computer Science and Information, College of Science in Zulfi, Majmaah University, Al-Majmaah 11952, Saudi Arabia

* Corresponding author's e-mail: m.otoom@mu.edu.sa

ABSTRACT

Technology is rising on daily basis with the advancement in web and artificial intelligence (AI), and big data developed by machines in various industries. All of these provide a gateway for cybercrimes that makes network security a challenging task. There are too many challenges in the development of NID systems. Computer systems are becoming increasingly vulnerable to attack as a result of the rise in cybercrimes, the availability of vast amounts of data on the internet, and increased network connection. This is because creating a system with no vulnerability is not theoretically possible. In the previous studies, various approaches have been developed for the said issue each with its strengths and weaknesses. However, still there is a need for minimal variance and improved accuracy. To this end, this study proposes an ensemble model for the said issue. This model is based on Bagging with J48 Decision Tree. The proposed models outperform other employed models in terms of improving accuracy. The outcomes are assessed via accuracy, recall, precision, and f-measure. The overall average accuracy achieved by the proposed model is 83.73%.

Keywords: cyber security, network intrusion, ensemble learning, machine learning.

INTRODUCTION

With the recent advancement in technology due to the development of artificial intelligence (AI) and the rapid rising of the huge amount of data on the internet as big data technologies, network security challenges are more complex threats than in the past (1). Moreover, with the rising of cybercrimes and large data on the internet, and extended network connectivity, computer systems are turning out to be more susceptible to attack (2). This focuses on the critical need for an efficient and reliable network intrusion detection system (NIDS), which has emerged as a significant research area. This is because it is theoretically impossible to construct a system with no susceptibility. To that goal, throughout the last few decades, researchers have developed a variety of systems, each with its own set of advantages and disadvantages. However, there is still a need for an efficient NIDS with improved accuracy and prediction (3). Recent global technological developments make users vulnerable to several cyber-attacks, where intrusion compromise users' sensitive information. To identify cyber-attacks, there are different social and technology-assessed techniques. Social techniques bring user awareness; however, they are incapable of significantly identifying the cyber-attacks that demand technology-assisted techniques. Over the last decade, especially with the advancement of IoT-based systems, The Internet's population is constantly growing. According to data from Data-Report, there are 4.5 billion Internet users. Many individuals, as well as organizations, are depending on the Internet to facilitate communication, store information, and conduct business (4). Because of the large number of users and applications, there is growing concern about data privacy and security (5). To address these concerns, academics and industry are collaborating in the area of cyber security to offer security and privacy. To protect

Internet-based systems from attacks, various protection tools such as firewalls, user authentication, data encryption, anti-malware, and antivirus software have been proposed (6). These security technologies prevent many attacks but lack in-depth packet analysis due to which they cannot provide security as required to the organization's network (7). Some of the latest ML techniques have been used to overcome these shortcomings (8). Signature-based and anomaly-based are the two categories of IDS. Signature-based detection systems detect attacks by analyzing the previous attack patterns. Because detection is based on data from previous attacks, this technique is vulnerable to novel attack detection (9). An anomaly-based detection system, on the other hand, identifies assaults by detecting conditions or patterns that are not deemed normal, and so these systems successfully identify known and undiscovered attacks (8). There are continuous improvements in the performance of Network Intrusion Detection Systems (NIDS), but still, further, improvement is required (10). To this end, this study aims to propose an ensemble model based on Bagging using J48 for NIDS compared with J48 Decision Tree (J48), Random Forest (RF), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM).

In recent years, the use of machine learning (ML) techniques to NIDS has been a popular study subject. Based on numerous publicly available datasets created plenty of new approaches to the problem of detecting network attacks. There are a variety of ways employed, including anomaly and signature-based IDS, or a combination of both. On the subject of signature-based IDS, there are numerous research proposals are available in the literature (11). Most of the previous work focuses on binary classified datasets and neglect multiclass real-time data sets. This section presents related work on both binary and multiclass datasets. The authors' presents the ability of the ML-based AIDS using the CICIDS2017 dataset for analysis. Supervised and unsupervised ML techniques are applied and tested on 48 different models. Due to poor results about 17 models are excluded from the results and 31 are included. Supervised ML models achieves 99.32% accuracy for ANN, 99.49% accuracy for DT for KNN, 98.86% for NB and 96.72% accuracy for SVM. The unsupervised ML model achieves 60.06% for EM, 23.41% for K-means and 59.06% for SOM. Among all ANN performs best while the performance of K-mean is the poorest one. This study further focuses on the impact of feature selection and deems to use developing a deep learning model for analysis (12). The researchers apply two new ML models: ANN and KNN to the defense data traffic for anomaly detection in the network. In this paper, the researchers select multiples performance metrics such as: accuracy, precision, TPR and FPR for analysis. The results shows that KNN achieves 0.9957 accuracies, 0.9949 precision, 0.9959 TPR and 0.9956 TNR while ANN achieves 0.9923 accuracies, 0.9910 precision, 0.9926 TPR and 0.9920 TNR. The main drawback of this work is the feature selection and the calculation of the distance between new and existing points is so large that it adversely affects the performance. In the case of specific datasets KNN proves to be better for classification (13). For intrusion detection two datasets: CICIDS2017 and ISCXIDS2012 are used for analysis. A hybrid approach based on packed and session classifiers achieve the highest accuracy of 99.9% for CICIDS2017 and accuracy of 97.37% for ISCXIDS2012. Furthermore, the results of the hybrid model are compared to other models such as: RF, Adaboosted DT (ADT), Deep Neural Network (DNN), SMOTE+RF, Support Vector Machine (SVM), DTNB, TSE including Rotational Forest, Extreme Learning Machine (ELM) and Gradient Boosting Tree (GBT). Among all ELM and SVM perform poorly. The main disadvantage of this study is the practical implementation which is too much difficult due to its high cost and complexity (14). Data-driven IDS incorporates several steps: data processing, datasets exploration and ML-based models. The effectiveness of Data-driven analysis by the authors using 10-fold cross-validation and KDDcup99 dataset. This dataset includes 4898432 instances and 41 attributes. In this work, one dataset, two ML models and three performance metrics such as: accuracy, precision and recall are considered. The results show that the RF model achieves 94% accuracy, 99% precision and 93% recall while DT achieves 93% accuracy, 98% precision and 92% recall (15). The authors apply DoS and probe attacks in the NSL-KDD dataset to an IoT network, especially Routing Protocol for Low-Power and Lossy Networks (RPL) and 6LoWPAN networks, utilizing the Contiki-NG operating system. Furthermore, the dataset is fed into machine learning algorithms to examine their capacity to categories various network threats. The findings show that tree-based techniques and ensemble algorithms such as XGBoost, DT, Bagging Trees, and RF outperform and achieve better than 96% accuracy (16). The goal of this study is to design an intrusion detection system to detect the intrusions early and accurately. This goal is achieved using an ensemble machine learning model which is based on bagging and J48.

EXPERIMENTAL SETUP

This study focuses on determining the prediction of intrusion in networks and their impact on systems and data. Most recent intrusion detection models can quantify anomalies in the network data flow, these approaches can discern anomalies; nonetheless, they have limited capacity in preventing these anomalies and intrusions from attacking (17). Intrusion detection is a high-security issue as it can be the cause of data loss and defacement of information. These flaws must be addressed as quickly as feasible to reduce the risk of data loss and defacement. To this end, our research methodology applied is presented in Figure 1. Where, after data collection, ML models are trained and tested. The training and testing criteria are discussed in the subsequent. The models are evaluated with some of the standard assessment measures including accuracy, recall, precision, and f-measure (18). All the experiments are done through the system with specifications containing Microsoft Windows 10-based machine with Intel® Core i5 processor and eight-gigabyte memory. Each result obtained is averaged over 20 simulation runs by changing random seed values and keeping the parameters fixed.

DATASET DISCUSSION AND MODELS TRAINING

For IDS, the dataset focused in this study is NSL-KDD which is last updated in 2019 and is available at https://www.unb.ca/cic/datasets/nsl.html. For the training and testing of ML models, different datasets are used. The dataset used for training contains 125973 instances where 58631 are anomalies while the rest of 67342 are normal records. The testing set consists of 22544 instances. In both the test and train sets, there are 42 features, one of which is a class feature used to determine if a record is abnormal or normal. One of the 42 features is a class attribute, while the remaining 41 features are divided into four separate classes, as detailed below:

- basic (B) characteristics are the characteristics of individual TCP connections;
- content (C) features are the properties inside a domain knowledge-suggested relationship;
- traffic (T) features are qualities calculated using a two-second time frame;
- host (H) Features are qualities meant to evaluate assaults lasting more than two seconds.

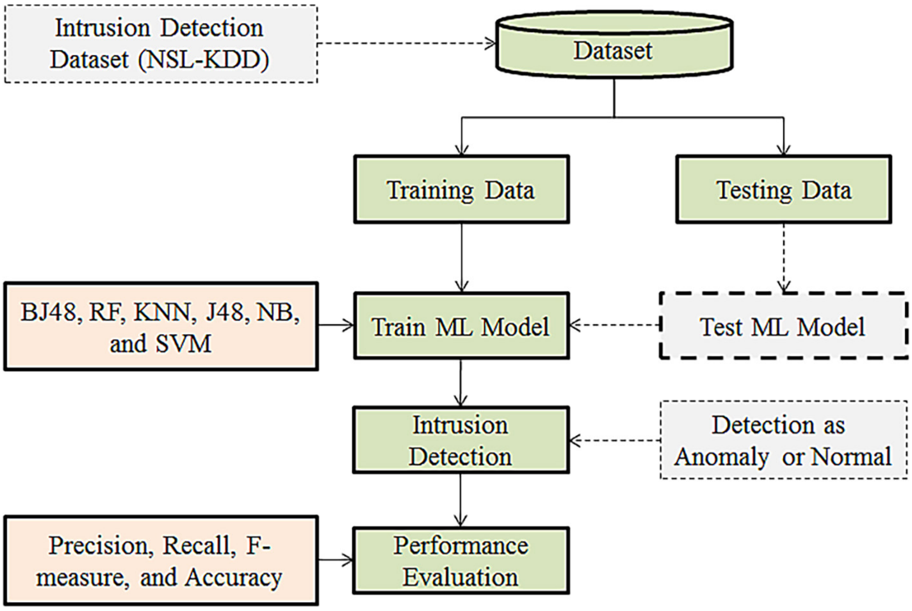

Intrusion Detection Dataset (NSL-KDD)
Dataset
Training Data
Testing Data
BJ48, RF, KNN, J48, NB, and SVM
Train ML Model
Test ML Model
Intrusion Detection
Detection as Anomaly or Normal
Precision, Recall, F-measure, and Accuracy
Performance Evaluation

Fig. 1. Research methodology

PERFORMANCE ASSESSMENT

The core phase of any experimental study is to test the performance of use models (19). Hence, this study focuses on standard assessment measures including accuracy (20), (21), recall, precision, and f-measure (22–24). These measures can be calculated as:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{1}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{2}$$

$$\text{F} - \text{Measure} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where: *TP* – true positive, presents the records that are anomalies and models predicted these as an anomaly;
TN – true negative are those records that are normal and also predicted as normal;
FP – false positive are those records that are normal but predicted as an anomaly;
FN – false negative are those records that are anomaly but predicted as normal.

PROPOSED MODEL

This study uses Bagging with the J48 classifier to design an ensemble model to improve the accuracy of NIDS. Bagging makes decisions from multiple classifiers, and here the classifiers are J48. Bagging generates subsets of training. To create each of the new subgroups, training instances from the initial training data are randomly sampled and replaced. Therefore, certain instances may be chosen time and time again while others may be omitted. All fresh training subsets in bagging have the same number of instances as the data. The J48 is utilized as the basis

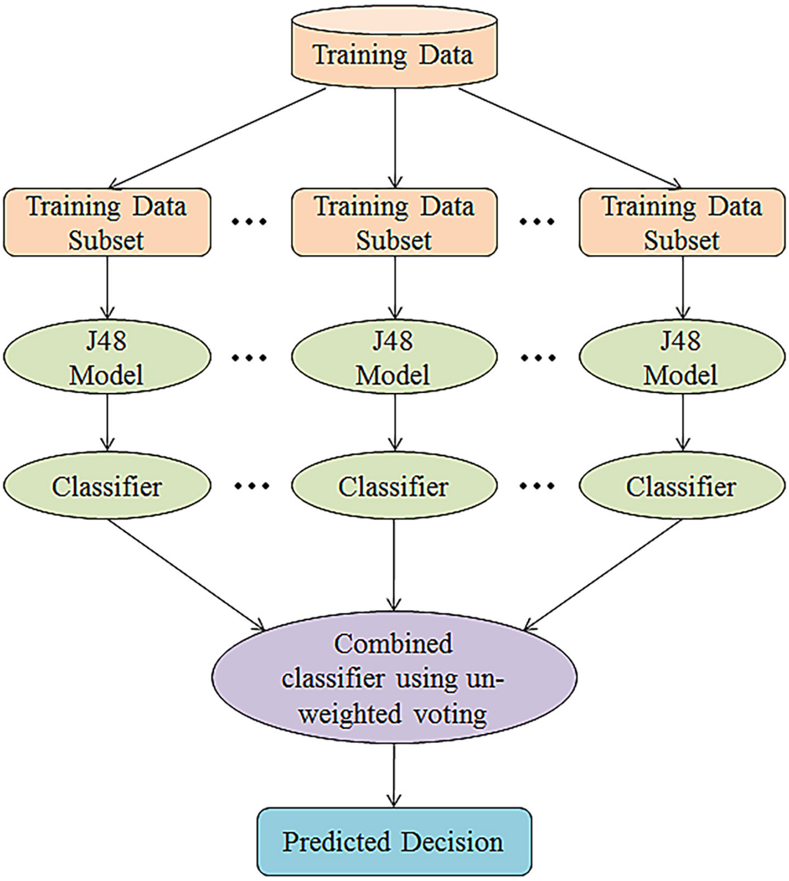


Fig. 2. Process of ensemble model (bagging employing J48)

classifier in the construction of one classifier from each of these subgroups. Thus, the results of various training subset classifiers are combined using un-weighted voting to get the final result from the structured classifier. Each classifier records their vote for a modulation scheme in this case to categorize an instance. The modulation scheme chosen as the winner is the one with the most votes at that point. It's important to note that Bagging improves identification performance primarily by minimizing variance error (25). The complete process is shown in Figure 3 and algorithm 1. The algorithm for the proposed model is:

Algorithm 1 Bagging using J48 Classifier Algorithm:

```
• input:
    − training data
    − base classifier
    − number of training subsets (iterations)
1. For training set 1 to training set n do
2. Generate training subsets
3. Constructed classifier = J48
4. end for
5. Instances constructed by classifier = Sum of all
   largest number of votes.
```

Sum of all largest number of votes are selected as final decision

- Output: constructed classifier (final)

RESULTS

The experimental outcomes achieved through the proposed models and the rest of the employed models including J48 (26), KNN, RF (27), NB (28), and SVM (29). These were trained using the NSL-KDD dataset and evaluated using accuracy, recall, precision, and f-measure. For training and testing, two different datasets are used. The training set consists of 125973 instances and the testing set consists of 22544 instances. Figure 3 illustrates the true positive rate (TPR) and FPR of each employed model. It shows the better performance of the proposed model with 0.837 (83.7%) of TPR. On the other hand, SVM shows the weakest performance with 0.754 (75.4%) of FPR.

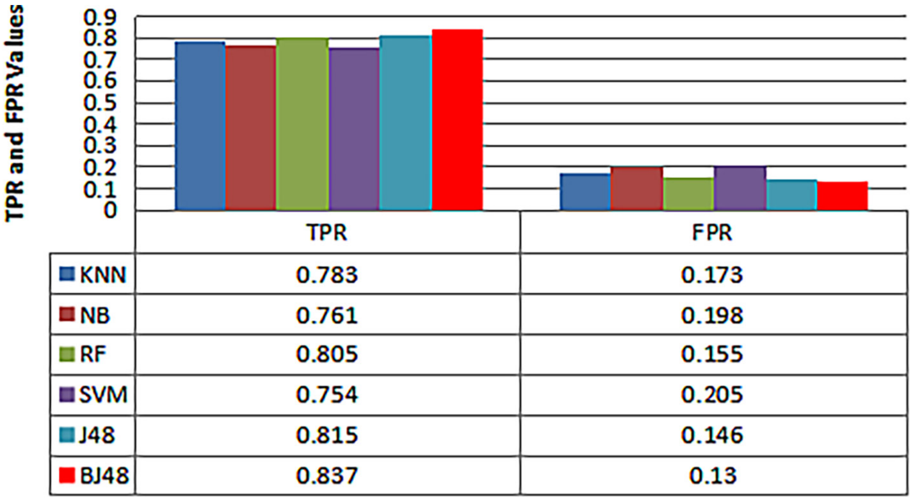
Figure 4 presents the outcomes assessed via precision, recall, and f-measure. The number of positive class forecasts that fall within the positive class is measured by precision. Recall measures how many correct class predictions were produced using all of the successful cases in the dataset. Precision and recall issues are balanced in a single number by F-single Measure's score. As discussed above, these analyses also show the better performance of proposed model as compared with the rest of the employed models. Bagging using J48 aggregates the predictions from various J48 models to determine which prediction is the best. The J48 models specialize in particular regions of the feature space, allowing predictions from all models to be combined to serve the greatest good.

Figure 5 presents the accuracy analysis of the proposed model and other models. When doing scientific studies, it is critical to measure exactly and precisely. Accuracy refers to how closely a measurement approaches its true value. This is critical since substandard hardware, insufficient data processing, or human mistake might result in erroneous and unrealistic results. The accuracy analysis presents that the proposed model outperforms other models with 83.73% accuracy. SVM shows the weakest performance with 75.39%. It also can be demonstrated that there is very little difference between J48 and RF where J48 achieved 81.85% accuracy and RF achieved 80.45% accuracy. The detail of the percentage difference (PD) between the proposed model and the rest of the models is illustrated in Figure 6. The percentage difference is calculated as:

$$PD = \left(\frac{n1 - n2}{\frac{n1 + n2}{2}} \right) * 100 \qquad (5)$$

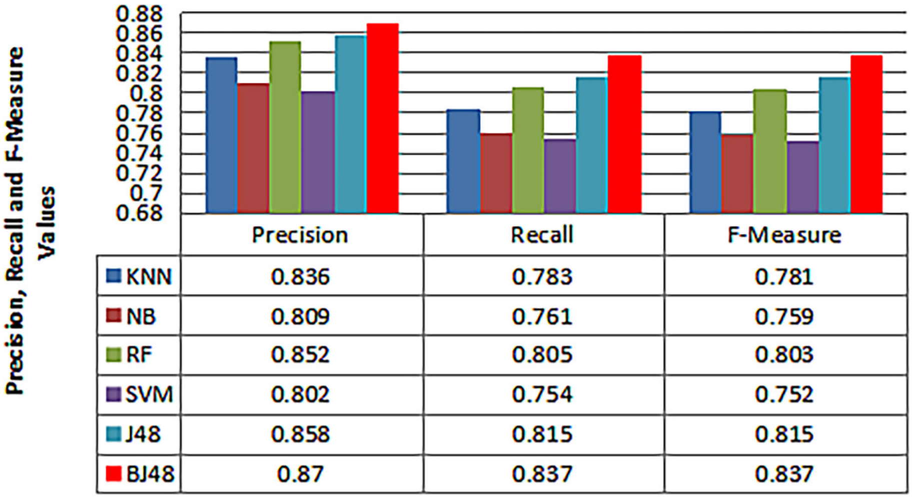
where: $n1$ – shows the values of Bagging using J48,
$n2$ – depicts the values of the rest of the models. SVM's weak performance is due to it does not perform well when the dataset contains a large number of instances. Moreover, if the number of features for each of the data points exceeds the number of training data samples, in this case, SVM will underperform.

The proposed model outperforms well in the current situation, however, there are some threats to the validity. The current study utilized two different sets for training and testing, though, if someone changes the training and testing criteria using any of the methods e.g. percentage splitting, K-fold cross-validation, etc. then the current



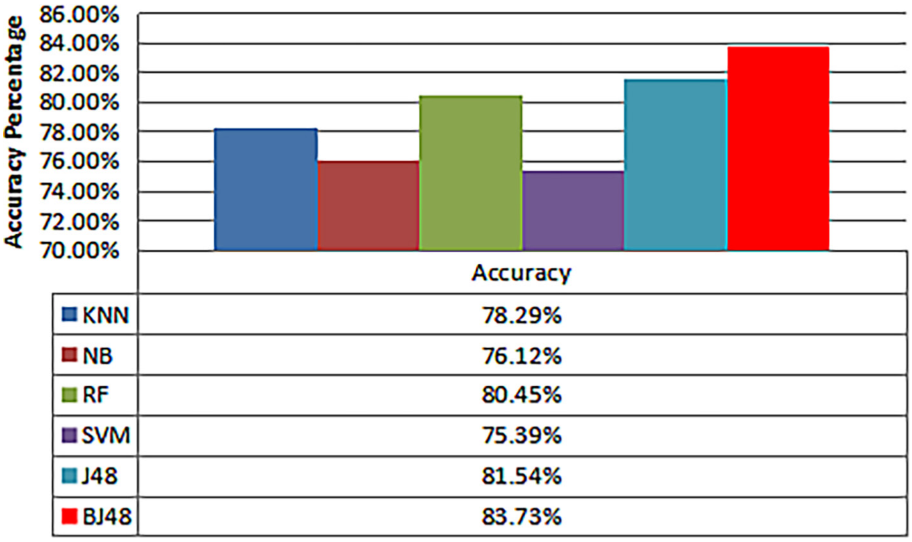
	TPR	FPR
KNN	0.783	0.173
NB	0.761	0.198
RF	0.805	0.155
SVM	0.754	0.205
J48	0.815	0.146
BJ48	0.837	0.13

Fig. 3. TPR and FPR of each employed model



	Precision	Recall	F-Measure
KNN	0.836	0.783	0.781
NB	0.809	0.761	0.759
RF	0.852	0.805	0.803
SVM	0.802	0.754	0.752
J48	0.858	0.815	0.815
BJ48	0.87	0.837	0.837

Fig. 4. Precision, recall and F-measure analysis through each employed model



	Accuracy
KNN	78.29%
NB	76.12%
RF	80.45%
SVM	75.39%
J48	81.54%
BJ48	83.73%

Fig. 5. Analysis of each models through accuracy measurement

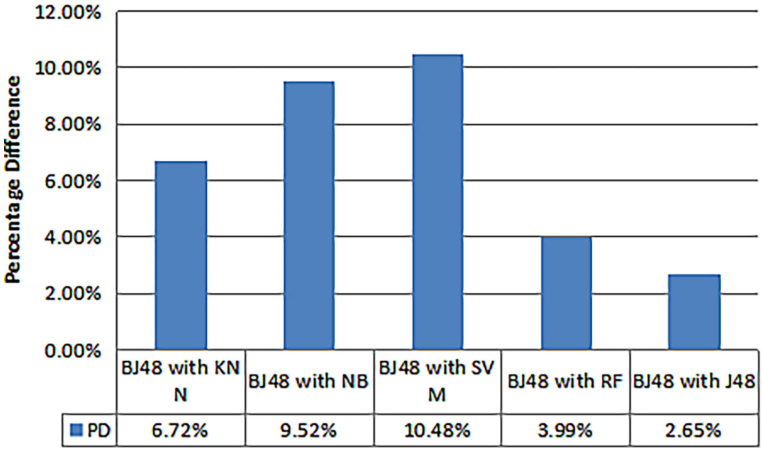


Fig. 6. Performance difference between proposed model and other employed models

outcome may be violated. This study focuses on the NSL-KDD dataset and recall, precision, f-measure, and accuracy as assessment measures, so, any change in the dataset is the selection of some other measurements for assessment that may change the current results.

CONCLUSIONS

In the modern world, the use of technology has become a commodity that, beyond providing several facilities, makes users vulnerable to a variety of cyber-attacks. In this regard, intrusion stands as one of the pivotal attacks to gain unauthorized access to users' sensitive and confidential information. Social techniques are considered the first line of defense, however, these techniques are not effective to prevent and detect these attacks considerably. This demands the use of technology-assisted techniques to overcome the issues on network data and security. To this end, this study proposed an ensemble model based on Bagging using J48. The performance of the proposed is compared with some of the well-known models including KNN, NB, SVM, RF, and J48 based on accuracy, recall, precision, and f-measure. The overall analysis presents the better performance of the proposed ensemble model with 83.73% accuracy.

Acknowledgements

The authors would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under project No. R-2023-50.

REFERENCES

1. Jiang H., He Z., Ye G., Zhang H. Network Intrusion Detection Based on PSO-Xgboost Model. IEEE Access. 2020; 8: 58392–401.
2. Dahiya P., Srivastava D.K. Network Intrusion Detection in Big Dataset Using Spark. Procedia Computer Science [Internet]. 2018; 132: 253–62. https://doi.org/10.1016/j.procs.2018.05.169
3. Selvakumar B., Muneeswaran K. Firefly algorithm based feature selection for network intrusion detection. Computers and Security [Internet]. 2019;81:148–55. https://doi.org/10.1016/j.cose.2018.11.005
4. Sarker I., Abushark Y., Alsolami F. Symmetry AK-, 2020 undefined. Intrudtree: a machine learning based cyber security intrusion detection model. mdpi.com.
5. Rahman M., Asyhari A., Leong L, … GS-SC and, 2020 undefined. Scalable machine learning-based intrusion detection system for IoT-enabled smart cities. Elsevier.
6. Jin D., Lu Y., Qin J., Cheng Z. Security ZM-C&, 2020 undefined. SwiftIDS: Real-time intrusion detection system based on LightGBM and parallel intrusion detection mechanism. Elsevier.
7. Kilincer I., Ertam F. Networks AS-C, 2021 undefined. Machine learning methods for cyber security intrusion detection: Datasets and comparative study. Elsevier.
8. Ashoor A.S., Gore S. Difference between Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). Communications in Computer and Information Science. 2011; 196 CCIS: 497–501.
9. Bamakan S., Amiri B. MM-PC, 2015 undefined. A new intrusion detection approach using PSO based multiple criteria linear programming. Elsevier.

10. Bertoli G., Júnior L. OS-I, 2021 undefined. An end-to-end framework for machine learning-based network intrusion detection system. ieeexplore.ieee.org.
11. Sarnovsky M., Paralic J. Hierarchical intrusion detection using machine learning and knowledge model. Symmetry. 2020; 12(2): 1–14.
12. Maseer Z.K., Yusof R., Bahaman N., Mostafa S.A., Foozy CFM. Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset. IEEE Access. 2021; 9: 22351–70.
13. Dini P., Saponara S. Analysis, design, and comparison of machine-learning techniques for networking intrusion detection. Designs. 2021; 5(1): 1–22.
14. Kim T., Pak W. Hybrid Classification for High-Speed and High-Accuracy Network Intrusion Detection System. IEEE Access. 2021; 9: 83806–17.
15. Alqahtani H., Sarker I.H., Kalim A., Minhaz Hossain S.M., Ikhlaq S., Hossain S. Cyber intrusion detection using machine learning classification techniques. Vol. 1235 CCIS, Communications in Computer and Information Science. Springer Singapore; 2020; 121–131.
16. Liu J., Kantarci B., Adams C. Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. WiseML 2020 – Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning. 2020; 25–30.
17. Choudhury S., Bhowal A. Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection. 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2015 – Proceedings. 2015; May: 89–95.
18. Otoom M.M. ABMJ: An Ensemble Model for Risk Prediction in Software Requirements. IJCSNS. 2022; 22(3): 710.
19. Otoom M.M. Comparing the Performance of 17 Machine Learning Models in Predicting Human Population Growth of Countries. International Journal of Computer Science & Network Security. 2021; 21(1): 220–5.
20. AL-Dreabi E.A., Otoom M.M., Salah B., Hawamdeh Z.M., Alshraideh M. Automated Detection of Breast Cancer Using Artificial Neural Networks and Fuzzy Logic. IJSBAR. 2017; 35: 109–20.
21. Otoom M.M., Jemmali M., Qawqzeh Y., SA K.N., Al Fay F. Comparative Analysis of Different Machine Learning Models for Estimating the Population Growth Rate in Data-Limited Area. IJCSNS. 2019; 19(12): 96.
22. Naseem R., Khan B., Ahmad A., Almogren A., Jabeen S., Hayat B., et al. Investigating Tree Family Machine Learning Techniques for a Predictive System to Unveil Software Defects. Complexity. 2020; 2020: 1–21.
23. Khan B., Naseem R., Shah M.A., Wakil K., Khan A., Uddin M.I., et al. Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021; 2021.
24. Khan B., Naseem R., Muhammad F., Abbas G., Kim S. An empirical evaluation of machine learning techniques for chronic kidney disease prophecy. IEEE Access. 2020; 8: 55012–22.
25. Niranjan A., Prakash A., Veena N., Geetha M., Shenoy P.D., Venugopal K.R. EBJRV: an ensemble of Bagging, J48 and random committee by voting for efficient classification of intrusions. In: 2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE; 2017; 51–4.
26. Bansal D., Chhikara R., Khanna K., Gupta P. Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia. Procedia Computer Science [Internet]. 2018; 132: 1497–502. https://doi.org/10.1016/j.procs.2018.05.102
27. Zamir A., Khan H.U., Iqbal T., Yousaf N., Aslam F., Anjum A., et al. Phishing web site detection using diverse machine learning algorithms. Electronic Library. 2020; 38(1): 65–80.
28. Iqbal A., Aftab S., Ali U., Nawaz Z., Sana L., Ahmad M., et al. Performance analysis of machine learning techniques on software defect prediction using NASA datasets. International Journal of Advanced Computer Science and Applications. 2019; 10(5): 300–8.
29. Babagoli M., Aghababa M.P., Solouk V. Heuristic nonlinear regression strategy for detecting phishing websites. Soft Computing [Internet]. 2019; 23(12): 4315–27. https://doi.org/10.1007/s00500–018–3084–2