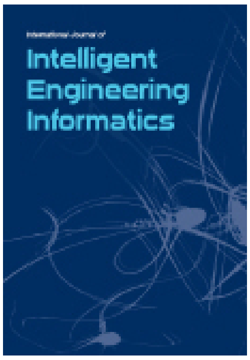


International Journal of Intelligent Engineering Informatics

ISSN online: 1758-8723 - ISSN print: 1758-8715
https://www.inderscience.com/ijiei

An investigation of machine learning-based intrusion detection system in mobile ad hoc network

C. Edwin Singh, S. Maria Celestin Vigila

DOI: 10.1504/IJIEI.2023.10054577

Article History:

Received:	16 November 2022
Last revised:	30 January 2023
Accepted:	30 January 2023
Published online:	03 May 2023

An investigation of machine learning-based intrusion detection system in mobile ad hoc network

C. Edwin Singh*

Department of CSE,
Noorul Islam Centre for Higher Education,
Nagercoil, Tamil Nadu-629180, India
Email: cedwinsingh@gmail.com
*Corresponding author

S. Maria Celestin Vigila

Department of Information Technology,
Noorul Islam Centre for Higher Education,
Tamil Nadu-629180, India
Email: celestinvigila@gmail.com

Abstract: Building stable networks is one of the most demanding issues in the current era, as the world is increasingly reliant on computers and technology. The standard MANET protocols, software, and facilities presume a collaborative and networking atmosphere that does not consider protection. Intrusion detection systems (IDS) that track centralised network operations and detect malicious nodes are often used to supplement certain security because mitigation strategies are never sufficient. This study describes ML techniques for distributing valuable properties to IDS for green smart transportation on MANET. The performance of ML-IDSs and a review of their adequacy in MANETs help the users determine intrusion when learning about the MANET context. ML optimised KDD IDS. Ensemble learning in this IDS process gave anomaly scores to controlled packets. Our solution to actual MANET dataset shortages is this ML technique. ML techniques, simulation, and a functioning prototype had created a more resilient IDS for green smart transportation. ML-enhanced IDS detected and reduced MANET harmful activity. This research expanded IDS knowledge through ubiquitous learning.

Keywords: mobile ad hoc network; MANET; IDS; QoS; machine learning; ML; security; attacks.

Reference to this paper should be made as follows: Singh, C.E. and Vigila, S.M.C. (2023) ‘An investigation of machine learning-based intrusion detection system in mobile ad hoc network’, *Int. J. Intelligent Engineering Informatics*, Vol. 11, No. 1, pp.54–70.

Biographical notes: C. Edwin Singh is pursuing his PhD at Noorul Islam Centre for Higher Education, Kumaracoil, with the Department of Computer Science and Engineering. He received his Bachelor’s and Master’s in Computer Science and Engineering in 2008 and 2012, respectively, from Anna University, Chennai. His research interest includes network security and intrusion detection, mobile ad hoc networks and wireless networks.

S. Maria Celestin Vigila is working as an Associate Professor in the Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil. She received her BE and ME in Computer Science and Engineering in 1996 and 1999, respectively. She received her PhD in Data Security from Anna University, Chennai, in 2013. She is an active member of ISTE and IET. She is the reviewer for quite a few peer-reviewed international journals. Her research interest includes cryptography and data security, wireless networks and information hiding.

1 Introduction

For many research findings, wireless communication for green smart transportation has become the method of optimal. With the support of green manoeuvrability and intelligent transportation technologies, connectivity, teamwork, adaptability, and profitability will all be boosted. Additionally, new trend approaches make it possible to preserve cutting-edge technology with every application, increasing its accessibility. Due to their limited-range wireless communication and elevated network node mobility, mobile ad hoc networks (MANET) are wireless technologies with a high node mobility level. Mobile routers and wireless networks form the MANET, a self-configuring network lacking an access point. Due to nodes' freedom to migrate, communication links in MANET frequently break. MANETs do not require a permanent framework; in contrast to other cellular networks and offering necessary communication, nodes should cooperate with the constantly changing connection to fulfil MANET requirements. The complexity of the protocols enabling MANET operations makes them ideally suited to harsh or unpredictably changing green smart transportation environments. With applications in many areas, such as Green Smart Transportation, emergencies, tactical operations, environmental management, and military services, MANETs are now a widely studied topic.

Methods of machine learning (ML) have been used to detect maliciousness and anomalies. ML techniques focus on algorithms that can learn from data without being expressly built. ML is helpful because there are so many types of network traffic. Despite ML characteristics, harmful detection still rules in the real world, and IDS techniques are rarely applied. IDS adoption is typically attributed to the significant Fake Positive Rate (FPR) problem. Even an FPR of 1.08 % on a network with high traffic levels could result in so many false alarms that MANET would be unable to process them. This study suggests employing ML techniques to deploy actual MANETs for green smart mobility in order to enhance IDS. The conclusions of this study can not be completed by improving IDS accuracy on existing training datasets because they do not apply to real-world MANETs. ML approaches learn the traffic in a dataset rather than watching it. Due to the lack of labelled train datasets for MANET assaults, they must be retrained on the network element, which is impossible. Using ML models, we suggested an effective architecture for IDS on KDD. The following optimisation methods are used in this framework:

a data augmentation

b parameter estimation

c supervised learning.

This method obtained a maximum prediction performance of 94.48 % on the KDD test data collection with a low probability of 1.98 %. Consider simple metrics like Precision, recall, and F-score when assessing particular ML effectiveness for IDS types. The paper's organisation is as follows: Section 2 is based on related works. The proposed ML-based green smart transportation IDS is described in Section 3. In Section 4 mathematical model of the ML the process is explained. Important results are given in Section 5. Performance analysis is shown in Section 6. The conclusion is given in Section 7.

2 Related works

Massive IDS methods have been anticipated to minimise security issues in MANET to improve effective IDS performance. However, these approaches have challenges in increasing prediction performance, and presently MANET remains challenging due to the security challenges such as energy usage and packet delivery ratio. A penetration testing system is a computing device that detects the system's software behaviour in the MANET to identify intruders. Since information technology can be vulnerable to security flaws, it is cost-effective and challenging to network setup that is not resistant to threats. IDSs accurately find and defend against attacks by monitoring nodes and user misbehaviour. IDS have been considered an integral part of solutions for preventing recent trends. Mishra et al. (2004) developed a systematic IDS and solution framework for MANETs in innovative research in wireless IDS study. MANET concentrated on policies for creating IDS that can identify attacks instantly. The mobile node has an administrator connected to it, and every node in the system interacts with the IDS. To promote collaboration of neighbouring nodes, shared IDS is recommended. To develop the standardised profile, Huang et al. (2004) introduce an innovative aspect of the data processing approach that uses 'hybrid-feature' analytics to obtain the intelligent system standard forms of MANET traffic.

For information mining jobs, Weka is a suite of ML algorithms. It has tools for discovering association rules, grouping, analysis, categorisation, and visualisation of data. Object model detection, by contrast, hand, picks up on intruders if they deviate from a routed protocol's established behaviour reference. This method can directly identify when the adversarial nodes violate the traffic restrictions, which makes it useful in identifying topological attacks. Weka is a species of bird that can only be found in Both the moniker and the birds have the same etymology. Features recognition has thus far been used in comparable circumstances, such as protecting various sensors and ad hoc network protocols. In such a setting, relays are typically explicitly rated by specialists using their theoretical requirements. There are several other specifications as well, but none of them was able to identify topological attacks.

Standard of care is a network and switched function that optimises traffic so that the most relevant traffic gets through first. The performance of critical systems has improved. The switch's QoS changes depending on its level; the greater the platform's level, the better the internet access layer it supports. The network connection's QoS manages packet loss and decreases latency and jitters. Service quality is a collection of capabilities that function on a network to ensure that the high programs and traffic run reliably even when bandwidth is constrained. Based on the QoS, anomaly-based IDS can be divided into sub-categories (Kim et al., 2006). These are mathematical and ML-based types. Nonparametric, multi-variable, and time sequence statistics are used in statistical data. Case and N-based, expert classifications, and ML are examples of knowledge-based devices that use finite-state models and laws. Buczak and Guven (2016) recommend DL algorithm selection based on problem-solving (Rahul and Shah, 2002). ANNs, grouping, genetic algorithms, and other algorithms are examples of algorithms. Specification-based is a hybrid model that combines the strengths of both trademark and anomaly-based models. Yin et al. (2017). used a recurrent neural network (RNN) to evaluate the NSL-KDD datasets, and the RNN's output correlated to that of other ML classification models J48, SVM, and RF. It is used to classify single-and multi-data and has higher IDS performance. Since RNN processing takes longer, researchers have investigated, so in the coming years, long short-term memory (LSTM)/gated recurrent unit (GRU) will be used to solve the problem (Ma et al., 2016).

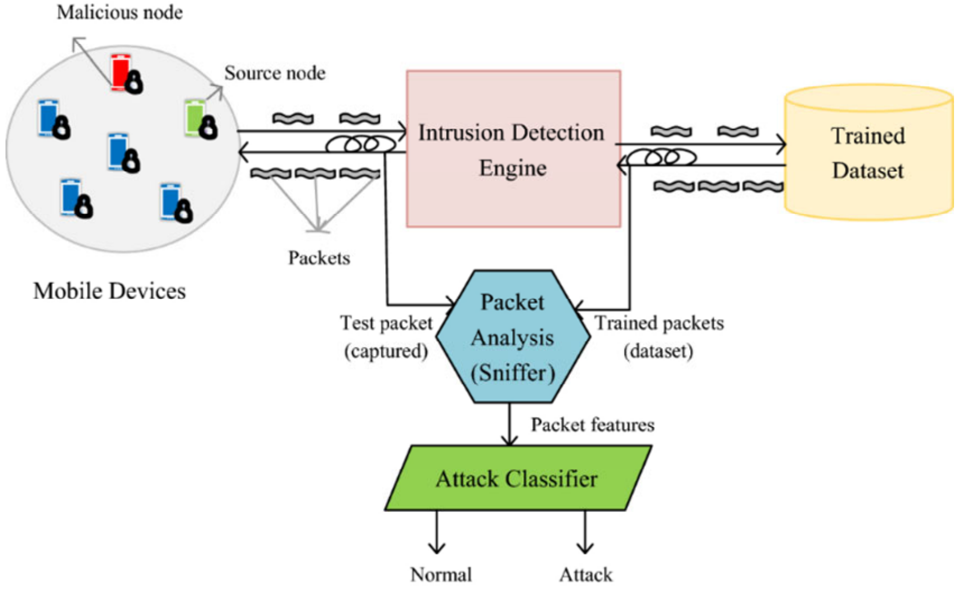
ML is a sub-domain of AI that exploits data's inherent information to identify ML methods and help with decision-making automatically (Mejía Quiroga et al., 2020). Various algorithms and techniques synthesise a powerful tool for designing intelligent predictive systems for a broad spectrum of business tasks. In cyber security, specifically IDS, we use classification algorithms to discover anomalies and offensive incidents in a secured network (Bhuyan et al., 2014). The strength of ML depends on the ability to recognise patterns based on many features within vast quantities of network traffic data, which would otherwise necessitate manual mining on the part of a skilled (Tavallaee et al., 2009). The extracted patterns are used to build generalised models, predicting the legitimacy of future network traffic instances automatically/determining their proximity to anomalous/legitimate behaviours. Hence, ML-based IDSs go beyond standard-setting thresholds, monitor specific metrics, and adapt to demanding MANET environments (Sharma and Kumar, 2022).

Ad hoc network service security is a difficult problem to solve. The first step in building robust cybersecurity solutions is always understanding probable attackers. The secure transfer of information in MANET relies heavily on security. MANETs are more sensitive to virtual than wired networks due to lacking a central planning tool and a wireless transmission medium. Several attacks affect MANET. In MANET security, assault on the protocol stack, malicious code, and repudiation are all possible. Threats on the Packet Filtering Session include theft and SYN overflow. Drowning, black hole, and grey hole are all used to describe the in-network layer. Many algorithms like Institutional fuzzy, genetic algorithms, relevance vector machines, zone sampling based trace back algorithms, rough set theory, support vector machines and neural networks are suitable for MANET. Attacks such as worm hole and link spoofing occur. Deep packet inspection and surveillance take place at the Data Link Layer. In the physical layer, traffic jamming, eavesdropping occur. IDS detects the system's software behaviour in the MANET to identify intruders.

3 Proposed ML-based green smart transportation IDS

ML methods create an explicit or implicit framework that categorises the patterns analysed. The requirement for class labels to exercise the behavioural model, a technique that applies reflective training on resources, is a unique feature of these strategies (Abbas et al., 2021). While the former centred on creating ML models that enhance efficiency based on existing research investigations, ML ideas' efficacy also correlates with statistical approaches (Pahadiya et al., 2021). As an outcome, ML-IDS can adapt its methodology as new knowledge is acquired (Palenzuela et al., 2016). While this feature can make it more attractive to use such strategies in all Green Smart Transportation network environments, MANET resource-intensive scheduling is a key drawback. IDS have been subjected to many ML-based approaches. The most significant are presented in this research article, and the key benefits and disadvantages of ML on MANET (Figure 1).

Figure 1 Architecture of ML-based IDS (see online version for colours)



The primary investigation of ML-IDS is to recognise and classify the nodes' observable misbehaviour patterns using IDS and the necessary ML features as inputs. As a result, effective IDS system design and implementation following the application's problems is another significant issue. MANET architecture necessitates maximum energy efficiency (Taherdoost, 2019). As a result, IDS processes can use less energy consumption while achieving the desired outcomes (Thombre et al., 2016; Buczak and Guven, 2016). Comparing all the most commonly utilised ML models concerning the critical key factors when determining the IDS for MANETs is summarised in Table 1.

In addition to the regular mentioned above, specific IDS could help with the volume of trained datasets. Analysing the primary components is a strategy for reducing a dataset's difficulty. PCA is an intermediate detecting system rather than a detection scheme itself. PCA is a translation method that uses 'n' correlated variables to describe PCA (Jindal and Singh, 2019). PCA fetches the total factors down to DS > n. And next step is to find associations between different extracting features from the trained dataset using association-rule mining. For example, internal links among data consistent with a specific relationship can be open using ML association rules (Sharma et al., 2020). knowledge discovery in databases (KDD) hit the scene in the 1990s, promising to ‘‘define fresh, true, particularly effective, and clear and understandable patterns for data.’ (Tavallaee et al., 2009). Data exploration methods emerged as a subset of KDD, consisting of ‘‘applying DL to massive databases to explore valuable knowledge immediately’.

Locating valuable trends in information is referred to by many names in diverse regions. Most frequently, statistics and databases researchers, as well as, more lately, the accounting and business groups, use the phrase ‘data mining.’ Here, extracting relevant knowledge from data is referred to as ‘KDD’, or information retrieval from data. A key step in this procedure is data collection, which involves applying specialised algorithms to data in order to achieve a good detection rate. The additional processes in the Process model, such as information extraction, data identification, and data cleaning, as well as the assimilation of pertinent background experience and accurate analysis of the processing findings, guarantee that knowledgeable conclusions are drawn from the facts. Database dredging, rightfully criticised as a harmful practice that can lead to the finding of nonsensical patterns, is one of the criticisms against the blind application of data mining techniques. Research in areas like datasets, computer vision, pattern matching, statistics, intelligent systems, justification with confusion, effective learning for intelligent systems, data visualisation, computer discovery, scientific innovation, information retrieval, and rising computer science has led to the development of KDD and is still leading to it. Each of these domains’ ideas, techniques, and techniques are incorporated into KDD systems.

KDD and ML are widely utilised to compare network traffic instances in network-related repositories as a particular application. IDS sensors can be integrated into the network using one of three approaches: venue, infrastructure, or hybrid. Every access point has the proximity sensor implemented thanks to presenter IDS. Each node will take on the role of a surveillance node, keeping an eye on its neighbours’ operations. This approach uses a significant amount of the node’s memory and computational power, creating significant overhead. The detection module is implemented at the sink by network-based IDS, however. All system components will be notified and asked for the information required for the identification decision before it is gathered and sent to the sink. Additionally, this approach adds to the transmission rate and does not ensure that the source has access to all the required data. The phrase ‘data mining’ is commonly used to categorise and relate various IDS decision bases as a standard wildcard assessment method (Sommer and Paxson, 2010).

The malevolent external nodes that snoop on messaging services and take private information for their gain pose a hazard to the internet nodes. Additionally, the nodes engage in malicious behaviour while the network is routing data and launching the grey-hole assault. The authors provide a paradigm to achieve internet authentication that comprises two blockchains, encryption techniques, and digital signatures to overcome the problem. However, the utilisation of two blockchains results in increased connectivity and processing overheads. Additionally, the authors use Merkle trees and blockchain to track malicious activity and overcome the specific point of failure problem. The author uses a secure hashing methodology based on a blockchain for node identification. The authors provide a secure data storage system based on a blockchain that uses the algorithm to detect minutes at least in the network. However, compared to other centralised platforms, blockchain-based data storage is more expensive.

Table 1 Taxonomy of machine learning

ML types	*Suitability in MANETs*		
	Accuracy	*Training time*	*Resource constraint*
ANN	Strong	Strong	Good
Bayesian network	Good	Weak	Strong
Markov model	Strong	Good	Good
SVM	Strong	Weak	Weak

3.1 Pre-processing and data augmentation

The research has continuously criticised the KDD for its flaws and fixed redundant information and inconsistencies. The threats and links in these databases do not accurately reflect the assaults and fraud committed by a network. The suggested architecture covers the entire MANET. According to the article, comparing the results of our method to instances from the cutting edge will show how practical it is. The researchers conclude that the detection rates in these datasets cannot be extrapolated to the actual MANET (Tracy Camp and Boleng, 2002).

3.2 Parameters optimisation

The neural network parameters are the variables chosen before training. The number of nodes, packet size, optimisation method, learning rate, and features is examples of simulation parameters. Each hyperparameter is altered separately, or every potential combination of a value set is examined. This strategy has the advantage of object planetary optimal values and is glorious as facility seeking. H. The parameter combination results in the best sorting outcome. However, it is difficult to apply in implementation because a large investigative squad identified several parameters. random look and tree-structured Parzen estimator are the two most efficient and state-of-the-art algorithms we opt for. A creation machine acquisition issue to be investigated is an uneven collection. In our test situation, fewer linkages exist in the regular and DoS groups. R2L and U2R classes have two main methods for class reorganisation: under- and oversampling.

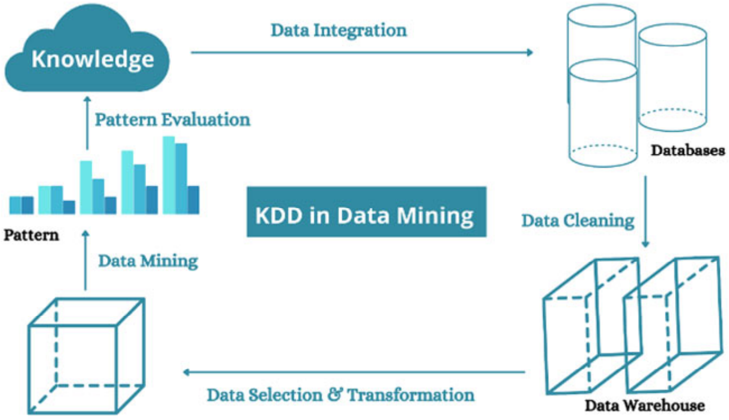
3.3 Ensemble learning

By combining many frameworks, ensemble learning is an approach for increasing overall predictive efficiency. The fundamental idea is that a group of base learners performs better than a single dominant learner, increasing the accuracy of the prediction. The Ensembling learning strategy observed results with many ML techniques, such as KDD.

3.4 Flow chart and algorithms for ML classification for malicious node detection

The constituent flow chart is given in Figure 2.

Figure 2 Flow chart for ml classification for malicious (see online version for colours)



Algorithm 1 for ML classification for malicious node detection

```
Step 1    Initialise vector v, b = 0
Step 2    Set dataset = ($D_1$, $D_2$), …, ($D_n$, $D_n$)
              Train SVM to learn decision function
Step 3    For each dataset is D Do
            Classify $D_i$ using decision function $f_x(D_i)$
Step 4    If (function margin < 1) Then
Step 5    Compute
Step 6    For sample data for reducing errors to predict Do
            $f_x = (HyperPlane^T D^i + Length\ of\ Intercept)$
            $Functon\ Margin\ (FM) = (HyperPlane^T D^i + Length\ of\ Intercept)$
Step 7    If (Prediction is Accurate) Then
            Do next
Step 8    Else
            Train SVM
Step 9    End If
Step 10   End If
            Classify $D_i$ as malicious nodes
Step 11   End
```

Algorithm 2 for optimal solution for acceptable error rate

```
Step 1    Begin
Step 2    Set network nodes
Step 3    Assess appropriateness for discrete nodes
Step 4    While (! check state) Do
Step 5    Proceed the cross over process
Step 6    Proceed the mutation process
Step 7     Evaluate the new node
Step 8     Select individual to substitute nodes
Step 9     Update node state
Step 10    End
Step 11    Return optimal solution for path selection
Step 12    End
```

Algorithm 3 For packet anomaly prediction

```
Step 1     The input of the network packet
Step 2     Output for packet anomaly
Step 3     Extract protocol headers from data packet
Step 4     For network protocol header Do
               Choice the parameters
               Translate unconditional into mathematical features
               Normalise features is {0 , 1}
Step 5     End for
Step 6     If CNN protocol header is trained then
               Set anomaly scores
Step 7     Else
               Training dataset on neural network
               Set anomaly scores
Step 8     End If
Step 9     End
```

The SVM algorithm trains the dataset to detect the malicious node in ML classification. SVM trains on a set of label data. SVM is used for both classification and regression problems. SVM draws a decision boundary that is a hyperplane between two classes to separate or classify them. It imports the dataset and then splits it into training and test samples. It classifies the target and initialises the Support Vector Machine to fit the training data. The classes for the test set have been predicted, and then the prediction is attached to the test set for comparison; thus, the prediction result is determined, and the malicious node is detected.

Train historical data to recognise outlier data and predict packet anomalies. Anomaly detection finds unusual events or observations that deviate statistically from the rest. SVM anomaly detection clusters normal data behaviour using a learning area. The testing example finds abnormalities outside the learning area. Fitting a neural network requires updating model weights with a training dataset to relate inputs to outputs. An optimisation technique finds weights that optimally relate inputs to outputs when training a neural network. Neural networks are trained iteratively to identify parameters that minimise error or loss when evaluating the training dataset.

4 Mathematical model of ML process

For the proposed ML method, we formulate a statistical approach. Each intelligent metre contributes to the prevalence of unique characteristics viewed for this time-space in a 3D matrix, $M_i = (M_{i1}, M_{i2}, M_{i3})$. Here, the certain effective rutting period and a 'HELLO' data packet connection to the 1-hop neighbourhood are encompassed in each channel. We measure the centre matrix of M with N time frames using equation (1).

$$\overline{M} = \frac{1}{N}\sum_{i=0}^{N}[M_{i_1,\ldots}M_{i_3}] \tag{1}$$

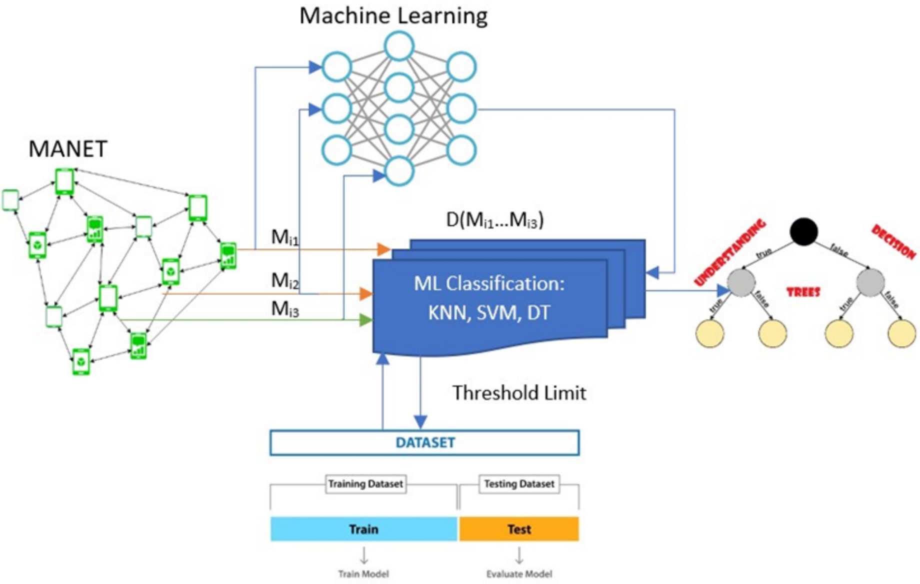
Then researchers measured the error to the mean vector from input sample data M, equation (2).

$$D(M) = [M - \overline{M}]^2 \tag{2}$$

If the routing time is significantly higher than the $[D(M) > T]$, this implies that it is not inside the acceptable range traffic and is therefore considered a threat. Here, the optimum projection path from the ML dataset is processed, equation (3)

$$\begin{aligned} &T = D(M_i) \\ &where, i = ArgMax[D(M_i], M_i \in D \end{aligned} \tag{3}$$

Figure 3 ML-IDS decision flow (see online version for colours)



ML assesses the 'T' threshold and constructs training data for our ML models using this mathematical formula in MANET routing protocol for green smart transportation. This 'T' extends only to N slots, and the 'T' will then be updated according to network security settings on MANET. Each intelligent metre can therefore perform within an adaptive 'T' range. Hello, packet, RREQ and RREP route optimisation data are now evaluated for every successive congestion control in an ML framework to estimate D(M). Classifiers, including the ML algorithm, are considered test data D(M) and its three comparative distinct features for our ML model. Our proposed model's complete flow charts are exposed in Figure 3.

5 Result and discussion

5.1 Simulations

The simulation is conducted on the open-source Waikato Environment for Knowledge Analysis (WEKA), a data analysis platform that offers a simple GUI for training and testing ML with different parametric settings. It contains features for information pre-processing, correlation, grouping, association rules, and other ML implementation tools. It's ideal for creating advanced ML models.

On the contrary hand, the specification-based MANET can identify assailants if they deviate from a transportation protocol's established conduct reference. This method can directly identify when the adversarial nodes violate the protocol restrictions, which makes it effective in identifying topological attacks. Features monitoring has thus far been used in comparable circumstances, such as protecting various sensors and ad hoc network protocols. In such a setting, specialists typically manually analyse data packets using their theoretical requirements. There are several other RPL specifications, but none successfully stopped RPL topological assaults. In particular, we simply provided a MANET prototype, which lacked evaluation and functionality.

5.1.1 Findings of KDD

For evaluating Dendron using the KDD, this research adopted a set of pre-processing steps broadly used in the literature study. We randomised the datasets to make them smaller while maintaining their characteristic state. Furthermore, all redundancy is avoided from the incomplete dataset. Table 2 describes the occurrences used in our findings' research training datasets. Except for the U2R class, 51.45% of the test cases are considered, and the training dataset includes 11.34 % of each type's instances. Dendron was evaluated using the KDD dataset to construct a training dataset comprising 10.12% of the preliminary cases, with the remaining 90.86% used in the testing process. Table 2 shows instances used in the training and test set included in the simulation results.

Since the optimisation process is not probabilistic, these two variables challenge the testing. As a result, the ML method would not produce the same result for each variable, regardless of the discrepancies in the outcomes between the different data distributions. The appropriate optimisation process was designed to obtain the end's ID rules. The parameters and values used to evaluate the design for each dataset are given in Table 3. We suggested our model's appropriateness using the Mean-F-Measure (σ-FM) method. The aim is to optimise this standard metric so that Dendron considers all the dataset's classes evenly. FM was designed to calculate the correlation among Recall and Precision measures amongst all dataset groups in the sense of multi-classification issues. The Recall and Precision measures take the False Negative Rate and Positive Rate of a class.

As a result, FM is a good measure for balancing detection and false reports. Also, under the measures, equate Dendron's efficiency to that of 'Rule of Thumb' algorithms. We may confidently assert that our methodology can provide adequate and consistent intrusion prevention rules to help a violation detection method using case-based analysis. Dendron achieved high efficiency in testing data, demonstrating that FM is a proper investigation metric. The increased values of the mean metrics of FM, maximum precision, and threat accuracy show that Dendron achieved a high detection rate for all dataset categories.

Table 2 Training and Test Dataset using KDD

Class	*Training dataset*	*Test dataset*
Normal	9,192	8,1292
DoS	5,819	5,1345
PRB	250	2,241
R2L	123	928
U2R	39	49

Table 3 Values for each dataset's test parameters

Symbol	*Parameters*	*KDD dataset*
Class	List of class	8
F	List of features	27
SD	4	
I	No. of Iterations	1,500
β	Beta value	45

6 Performance analysis

The performance measures are considered when assessing and comparing the QoS model.

- Precision (P) = TP/(TP + FP), which is the percentage of classification intrusions that indeed occur
- Recall (R) = TP/(TP + FN), which would be the proportion of correctly predicted intrusions to the total intrusions
- F-score(F) = (2/(1/P) + (1/R)), to determine the prediction performance, a trade-off regarding classification accuracy was achieved.

We used different ML algorithms with many evaluation strategies to analyse the KDD dataset and the appropriate subgroups related to performance measures. The records are paradigmatised using a safety subset analyser with the best search. The development and analysis of ML methodologies use ML environments. The prediction measures for testing datasets with unknown characteristics are Expertise, Recall, and F-Score, as illustrated in Table 4 and Figure 4.

Table 4 Simulation test results

ML classification	*Precision*	*Recall*	*Accuracy*	*Test time in sec.*
Naive Bayes	0.718	0.682	0.929	37.18
Multilayer perception	0.827	0.729	0.937	50.18
NB tree	0.891	0.791	0.948	102.34
Random forest	0.918	0.9387	0.993	367.238

Figure 4 Analysis of simulation (see online version for colours)

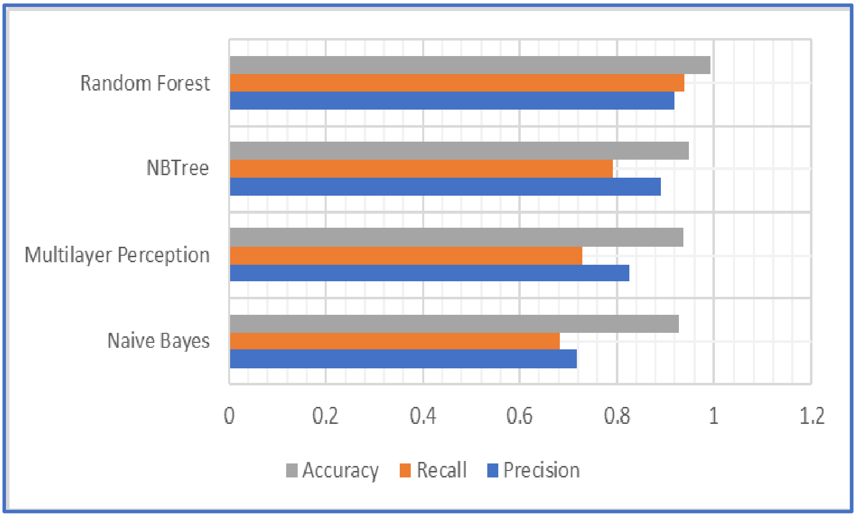


Table 5 Dendron Metrics of training dataset in %

Training dataset	*Accuracy*	*µ-FM*	*Average accuracy*	*Attacks accuracy*	*Attack detection rate*	*False attack rate*
KDD	98.12	87.19	91.21	89.36	98.91	0.89
NB15	91.16	67.18	71.28	67.19	89.10	0.68

The technique mentioned in the preceding section combines the features of ML to generate semantically understandable and reliable IDS rules that can predict known attacks on green smart transportation. Furthermore, Dendron can address network traffic problems according to ML methodologies in MANET. When mixed with the α, β, and γ weights, the distribution possibility function is balanced ML algorithms' propensity to be partial toward the dataset's significant elements of threats–the overall detection measure of σ-FM, which was used as an optimal function in this ML method. Dendron improves the DTs to improve the detection ratio across all attack classes in the datasets by optimising this standard measure. Our plan outperformed its rivals on standardised metrics such as FM, average, attacks accuracy, a high detection accuracy rate, and a low FPR (Table 5). On the other hand, Dendron lacks the precision metric, and Dendron emphasises all dataset groups relatively. Consequently, standard deviation measures become much more significant than precision for algorithms such as Dendron. To put it differently, prediction precision in multi-classification issues with asymmetry datasets from an ML algorithm's a greater tendency to support the dataset's important groups while ignoring the simple issues.

Figure 5 σ-FM of IDS accuracy (see online version for colours)

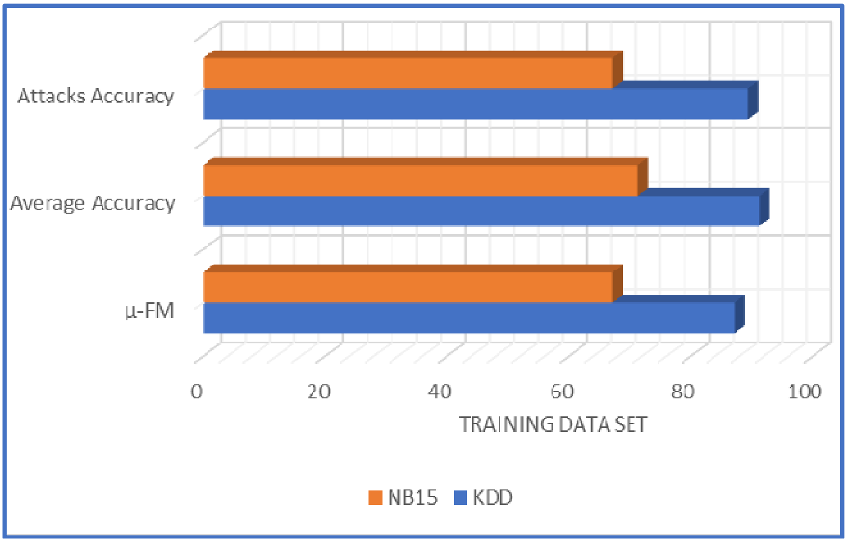
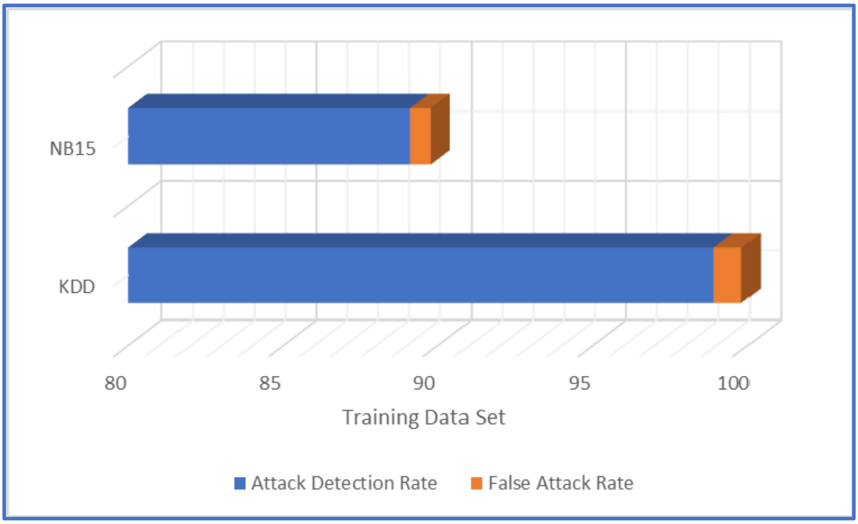


Figure 6 Attack detection vs. false attack rate (see online version for colours)



In the MANET environment, referred to in Figure 5, our proposed methodology outperforms the standard ML approach. In 98.12% of the regions, the proposed method exceeded the passive approach in terms of reliability. The static method has an accuracy rate of 91.21%, and the σ error rate is 87.19%.

Figure 6 shows the efficiency of the attack detection rate in every 100 nodes. Remember that the attack detection rate demonstrates harmful incidence detectability by explicitly measuring the precision of identifying threat cases. Overall, the adaptive solution achieves a 98.91% of attack detection efficiency.

Figure 7 Performance analysis of recall (see online version for colours)

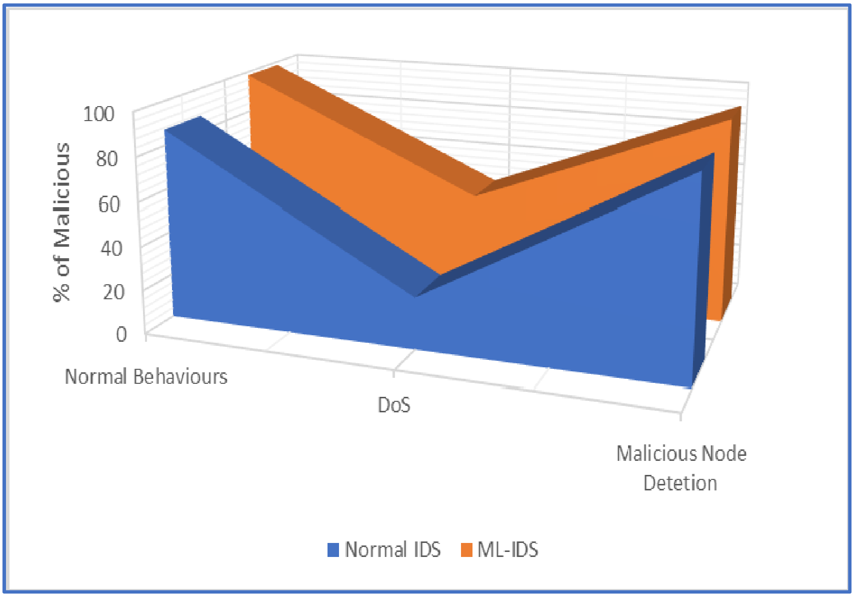
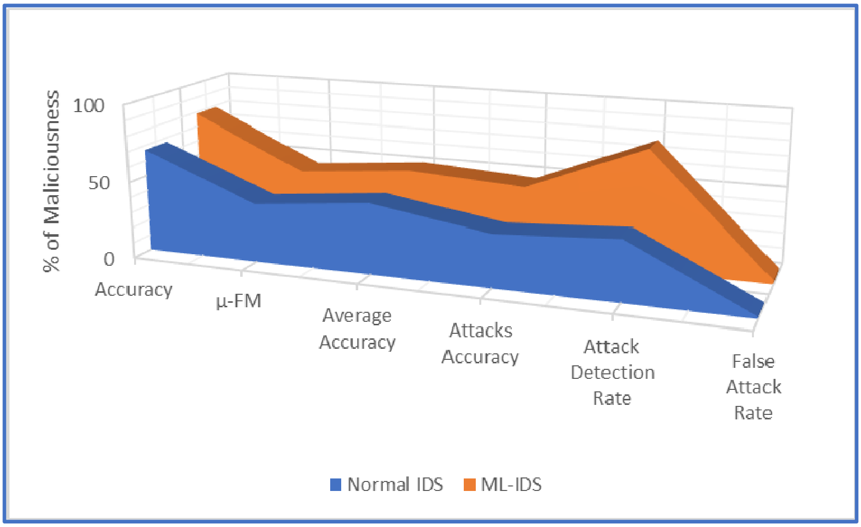


Figure 8 Performance analysis % of maliciousness (see online version for colours)



Besides, Figure 7 illustrates the total Recall for each class used in the trained datasets. We provide an adjacent comparison of the results of fixed and adaptive approaches to every type. And for the Normal level, the precision of the ML techniques is virtually equal. There are indeed 0.45% variance outcomes in dynamic analysis, and this variation is defined as small as if this precision is the mean of the Recall throughout 100 nodes. On the other hand, there is a considerable contrast in the precision of the DoS attack class. The proposed methodology, in particular, achieved an overall Precision rate of 91.21%.

Figure 8 depicts the cumulative performance of both the adaptive and static processes. All prevention measures support the adaptive strategy's reign. Except for the precision and attack detection ratio parameters thoroughly discussed above, the remaining metrics show that our ML-IDS approach is outstanding. The proposed ML-IDS technique effectively coordinates recall and precision in all dataset classes according to the statistic.

7 Conclusions

We required the ML-IDS detection method to use behavioural features to classify faulty events in a green smart transportation MANET environment. Albeit briefly, in this study, the primary IDS and its key architecture enable identification for each targeted 'Maliciousness' system. The evolutionary computing organisation has successfully provided ML-IDS on different test cases. MANET investigates the current state-of-the-art in ML classifiers of KDD. A structured cascade design used to lower the FPR and boost TPR accuracy is 88.39 % and 1.94 % FPR on KDD by building a series of meta-specialists. Also, ML techniques incorporate modified techniques and are merged to generate meta-specialists. Our research findings ensure our concept's ability to adapt to the new Green Smart Transportation network MANET environment, obtaining ratings that exceed fixed IDS by increasing by 73.37%. We claim that our MANET analysis provides new ADR guidelines as an automatic retraining mechanism that can substantially reduce human interference.

References

Abbas, M., Dwivedy, S.K. and Narayan, J. (2021) 'Adaptive iterative learning-based gait tracking control for paediatric exoskeleton during passive-assist rehabilitation', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 6, p.507, DOI:10.1504/ijiei.2021.10046403.

Bhuyan, M.H., Bhattacharyya, D.K. and Kalita, J.K. (2014) 'Network anomaly detection: methods, systems and tools', *IEEE Communications Surveys and Tutorials*, Vol. 16, No. 1, pp.303–336, DOI:10.1109/surv.2013.052213.00046.

Buczak, A.L. and Guven, E. (2016) 'A survey of data mining and machine learning methods for cyber security intrusion detection', *IEEE Communications Surveys and Tutorials*, Vol. 18, No. 2, pp.1153–1176, DOI:10.1109/comst.2015.2494502.

Huang, Y.-A., Fan, W., Lee, W. and Yu, P.S. (2004) 'Cross-feature analysis for detecting ad hoc routing anomalies', *23rd International Conference on Distributed Computing Systems, 2003, Proceedings*, IEEE.

Jindal, R. and Singh, I. (2019) 'A survey on database intrusion detection: approaches, challenges and application', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 6, p.559, DOI:10.1504/ijiei.2019.104565.

Kim, D.S., Shazzad K.M. and Park, J.S. (2006) 'A framework of survivability model for wireless sensor network', *First International Conference on Availability, Reliability and Security (ARES'06)*, IEEE.

Ma, T., Wang, F., Cheng, J., Yu, Y. and Chen, X. (2016) 'A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks', *Sensors* (Basel, Switzerland), Vol. 16, No. 10, p.1701, DOI:10.3390/s16101701.

Mejía Quiroga, J., Herrera, M.M. and Méndez Morales, A. (2020) 'Exploring the linkages between the patent applications and energy transitions: a system dynamics perspective', *International Journal of Intelligent Engineering Informatics,* Vol. 8, Nos. 5–6, p.526, doi:10.1504/ijiei.2020.10038545.

Mishra, A., Nadkarni, K. and Patcha, A. (2004) 'Intrusion detection in wireless ad hoc networks', *IEEE Wireless Communications*, Vol. 11, No. 1, pp.48–60, DOI:10.1109/mwc.2004.1269717.

Pahadiya, P., Saxena, S. and Vijay, R. (2021) 'Optimisation of thresholding techniques in de-noising of ECG signals', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 5, p.487, DOI:10.1504/ijiei.2021.10044780.

Palenzuela, F., Shaffer, M., Ennis, M., Gorski, J., McGrew, D., Yowler, D. and Taha, T.M. (2016) 'Multilayer perceptron algorithms for cyberattack detection', *2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS)*, IEEE.

Rahul, C. and Shah, J.M. (2002) 'Energy-aware routing for low energy ad hoc sensor networks', *IEEE Wireless Communications and Networking Conference*, Vol. 1, pp.350–355.

Sharma, I. and Kumar, V. (2022) 'Multi-objective tunicate search optimisation algorithm for numerical problems', *International Journal of Intelligent Engineering Informatics*, Vol. 10, No. 2, p.119, DOI:10.1504/ijiei.2022.125859.

Sharma, N., Sharma, H., Sharma, A. and Bansal, J.C. (2020) 'Dung beetle inspired local search in artificial bee colony algorithm for unconstrained and constrained numerical optimisation', *International Journal of Intelligent Engineering Informatics*, Vol. 8, No. 4, p.268, DOI:10.1504/ijiei.2020.112030.

Sommer, R. and Paxson, V. (2010) 'Outside the closed world: on using machine learning for network intrusion detection', *2010 IEEE Symposium on Security and Privacy*, IEEE, Canada.

Taherdoost, H. (2019) 'Electronic service quality measurement: development of a survey instrument to measure the quality of e-service', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 6, p.491, DOI:10.1504/ijiei.2019.10026271.

Tavallaee, M., Bagheri, E., Lu, W. and Ghorbani, A.A. (2009) 'A detailed analysis of the KDD CUP 99 dataset', *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, IEEE, Piscataway, NJ, USA.

Thombre, R.S., Islam, K. and Andersson, M.S. (2016) 'IP based wireless sensor networks: performance analysis using simulations and experiments', *Ubiquitous Computing, and Dependable Applications*, Vol. 7, No. 3, pp.53–76.

Tracy Camp, J. and Boleng, V. (2002) 'Wireless communication and mobile computing', *IEEE International Conference on Distributed Computing Systems*, Vol. 2, No. 5, pp.483–502.

Yin, C., Zhu, Y., Fei, J. and He, X. (2017) 'A deep learning approach for intrusion detection using recurrent neural networks', *IEEE Access: Practical Innovations, Open Solutions*, Vol. 5, pp.21954–21961, DOI:10.1109/access.2017.2762418.