SETSCI conference proceedings
SETSCI Conference Proceedings
17, 13-19, 2024

Cognitive Models and Artificial Intelligence Conference
May 25-26, 2024, İstanbul-Türkiye

AICCONF 2024

https://doi.org/10.36287/setsci.17.1.0013

Intrusion detection system using Optimized Machine Learning Algorithms for cyberattacks in the Internet of Vehicles (IoV)

Mamadou Korka DIALLO*, Dr Oğuzhan KARAHAN [2]

[1]Electronic and Communication Engineering, Kocaeli University, Kocaeli, Türkiye
Electronic and Communication Engineering, Kocaeli University, Kocaeli, Türkiye
fafayakadiallo@gmail.com, oguzhan.karahan@kocaeli.edu.tr

Abstract – The Internet of Vehicles (IoV) is a branch of the Internet of Things that deals with vehicle-to-vehicle communication and intelligent transport systems (ITS). But this connection is not without consequences, because the more a system exchanges information, the more vulnerable it is to various attacks from malicious actors. (hackers). Vehicle Internet security is a great challenge that security professionals face every day. Moreover, despite the deployment of diverse technologies by smart cities to obtain varied, high-performance cloud services, security concerns continue to appear in communications entities that share information. In this article, an intrusion detection system (IDS) based on machine learning is proposed to improve safety in vehicle Internet systems (IoV). The IDS uses random forest (RF) algorithms, decision Tree, Adaboost, and gradient boost on an IoV traffic data set. The hyperparameters of machine learning models are optimized using a meta-heuristic optimization algorithm called the CDO (Chemotactic Differential Evolution) algorithm. The proposed IDS achieved high performance in terms of up to 99.91% accuracy for the Adaboost algorithm in the binary case and 9.81% accuracy in the case of the decimal dataset. High performance of precision, recall, and F1 score were also observed in this study. The optimization has significantly improved the performance of the models by optimizing their hyperparameters. The study was conducted using data sets built by CICIS (Canadian Institute of Cybersecurity) with a real vehicle to evaluate the proposed detection system. The experimental results show that the proposed IDS has significantly higher detection performance.

Keywords – Internet of Vehicle, Cybersecurity, Intrusion Detection System, Machine learning

I. INTRODUCTION

The rapid evolution of communication technologies and their integration into all areas of everyday life have enabled new sectors of activity to emerge for more smooth exchange. It is in this sense that the transport system has undergone a much faster transformation in recent decades from a purely mechanical system to a system integrating electronics and computing for connected transport. Today everything is connected to the Internet, which is why we are talking about the Internet of Things, one of the specifications of the internet of things can also be identified in transportation that has experienced a great change in its recent years. Starting from the Ad-hoc Network of Vehicles (VANet), which only allowed the exchange of data between vehicles and Roadside unit to the Internet of vehicles (IoV) which will have allowed the different vehicles to exchange information with all its surroundings (pedestrians, motorcyclists, vehicles, database, lights, etc...). Many steps have been taken to this goal, but with the combination of the Internet of Things and artificial intelligence in the field of transport this goal is achieved. Nowadays intelligent transportation systems can exchange various types of information, such as road conditions, traffic, entertainment, and weather. However, these advantages are not without consequences, as malicious individuals take advantage of these opportunities to attack systems, thus compromising communications and regular updates. Unfortunately, these exchanges present security challenges because the more data is shared via communication systems, the more it will also be exposed to theft and various attacks. The main attacks listed today on the Internet of vehicles are, among other things, denial of service, GPS hijacking, identity usurpation, black hole, search, Sybil, malware etc... These various known and unknown future attacks expose the Internet of Vehicles to a very high security challenge It is in this sense that to deal with the various security problems, several security systems are put in place, among which we often talk about intrusion detection systems for an improvement in security, but some of them may not be sufficient because they only know the attacks that already exist. Based on this reality, for new attacks developed recently that may not be detected, we propose an IDS combining metaheuristic optimization and some machine learning models like random forest, Adaboost, gradient boost, and decision tree for the implementation of an intrusion detection system to secure the internal network against cyberattacks such as denial of service (DoS), identity theft, etc. This article is structured as follows, after a brief introduction in the first section, the second part focuses literature review, then the third part materials and method, in the fourth part, the result will be show and discuss and the last part a conclusion will be given.

II. LITERATURE REVIEW

Internet of Vehicle security has attracted researchers' attention in recent years, given its crucial importance in saving lives. In this section we will review some studies carried out in this direction in order to reach a good conclusion in the end.

In [1], Sun and others provide a comprehensive overview of the security and privacy issues in the Internet of Vehicles

(IoV). They discus about the characteristics of IoV systems, types of attacks, existing countermeasures, and future research trends. They studied contributes to further study and understanding of IoV security and privacy challenges.

Song and others propose in [2], an Intrusion Detection System (IDS) based on a Deep Convolutional Neural Network (DCNN) to protect the Controller Area Network (CAN) bus of a vehicle from cyber-attacks. The system aims to detect malicious traffic without the need for hand-designed features, by learning the network traffic patterns. The paper also evaluates the IDS using datasets built with a real vehicle, demonstrating its effectiveness in achieving low false negative rates and error rates compared to conventional machine-learning algorithms.

Hafiz and others address security and privacy concerns in Internet of Vehicles (IoV) applications by proposing an Intrusion Prevention System (IPS) based on Fuzzy Logic and Q-Learning in [3]. Their study permits to improve security in next-generation complex heterogeneous networking against sophisticated attacks and to implement a sustainable vehicular network. It presents a mechanism against attacks on IoV and similar dynamic systems, analyses the interaction between attackers and IPS, and evaluates the performance of the proposed solution.

In [4], the authors present GIDS (GAN based Intrusion Detection System), a novel IDS model for in-vehicle networks that uses deep-learning model Generative Adversarial Nets (GAN) to detect unknown attacks using only normal data.

An intrusion detection system (IDS) is included in the automated framework for secure continuous availability of cloud services for connected vehicles in smart cities, as proposed by the authors [5]. This framework protects against security breaches and ensures that services meet users' Quality of Experience (QoE) and Quality of Service (QoS) requirements. To detect and stop intrusion threats, the framework groups smart cars into service-specific clusters and applies machine learning algorithms for data reduction and classification. The goal of the suggested remedy is to improve the reliability and security of cloud services for vehicles within the framework of smart city transportation networks.

In the same study area, recently in 2023, Bifta and others develop in [6] a machine learning-based Intrusion Detection System (IDS) for Vehicle Controller Area Networks (CAN) using Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbour (KNN) classifiers. The IDS aims to detect and classify cyber-attacks on vehicles with high accuracy and minimal error rate, addressing the security vulnerabilities of the CAN protocol. The study also involves testing the IDS on multiple real-world vehicular datasets to ensure its reliability and efficiency.

Hyun and others propose an Intrusion Detection System (IDS) based on a deep convolutional neural network (DCNN) to protect the Controller Area Network (CAN) bus of a vehicle from cyber-attacks in [7]. The study also includes experimental findings that show how the IDS outperforms traditional machine-learning algorithms in terms of false negative and error rates.

In 2021, Li and others address the challenges of updating machine learning-based intrusion detection models in the Internet of Vehicles (IoV) to cope with new and constantly evolving types of attacks in [8]. Their study permit to propose two model update schemes using transfer learning, one assisted by the IoV cloud and the other a local update scheme, to improve detection accuracy and reduce the need for large amounts of labelled data. This aims to allow vehicles to update their detection models independently and respond more effectively to new attacks.

In [9], the authors proposed the use of four new machine learning algorithms for intrusion detection with the KIA Soul dataset, which is an open real-world dataset available on the internet. To improve the performance of studies carried out in the recent past on the same set, they propose the use of Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Multi-Layer Perceptron (MLP) in order to overcome the existing safety issues with on-board CAN Bus.

The issues related to the Internet of Vehicles are multiple. This model allows vehicles to communicate constantly, which improves traffic management. The data used allows in-depth analysis to take necessary measures during peak hours to minimize congestion. Thus, the optimization of journeys to the different destinations is ensured. This reduction also has positive consequences, such as reduced expenses and reduced carbon dioxide emissions. However, the communicating systems of this network are exposed to security problems, like any communication network. In this context, intrusion detection systems are valuable assets for anticipating security problems and sending alerts to authorized entities so that they can take the necessary measures. Not all IDSs are effective in any system, and it is in this context that the proposal for an IDS based on the combination of machine learning and metaphysical optimization seems to bring a plus to the different systems currently existing.

III. MATERIALS AND METHOD

A. Materials:

The use of the Internet of Vehicles is very varied in the intelligent transport system, ranging from congestion detection, information entertainment, travel time optimization, and cost reduction. Vehicles can communicate with road infrastructure to identify the most fuel-efficient routes.

Collision Alert: Intelligent vehicles exchange speed and location information to alert the driver if they are too close to avoid collisions. And if the driver does not react, the vehicle can even brake on its own to avoid accidents.

Weather sharing: Vehicles moving in opposite directions can share the weather conditions of places already crossed to allow others to know the nature of the environment.

Real-time information sharing: Smart vehicles are equipped with all kinds of sensors that allow them to share information in real time, from congestion to accident updates.

Reducing carbon emissions: Since vehicles are in constant communication, shared data helps to update traffic conditions, enabling vehicles to prevent traffic jams and reduce CO2 emissions. Some vehicles even use electricity, which is a great benefit for reducing carbon emissions.

As part of our comparative study, we used the CICIoV2024 database detailed in [10]. Several models based on different learning algorithms have been trained combined with meta-heuristic optimization, with the aim of choosing which is most suitable for our work in terms of accuracy so that we can put in place a robust system of intrusion detection to minimize the security problems that IoV may face. The selected data set is one of the few publicly available datasets on the subject of IoV. It consists of two different sets; the first is the binary version, which contains 156 characteristics of the system as well as 3 labelling characteristics that take into account various

classification situations. And the second one consists of 13 characteristics, three of which are labelling. The labelling attributes are: label, category, and specific category.

In the label attribute, two main elements should be noted. Useful, i.e., data without attack provided by the various components of the system, and attacks are data injected by malicious persons to divert vehicles from their main targets

Table 1. Number of instances benign and attack present in the CICIoV2024 dataset

Label	Count
BENIGN	1223737
ATTACK	184482

The category attributes consist of three main elements: useful data, identity usurpation, and denial of service attacks.

Table 2. Number of instances for benign, spoofing and DoS present in the CICIoV2024 dataset

Category	Count
BENIGN	1223737
SPOOFING	109819
DoS	74663

The third categorical attribute consists of five elements that can be listed as useful data, denial of service attack, speed attack, fuel attack, steering wheel attack.

Table 3. Number of instances for each specific class present in the CICIoV2024 dataset

Specific_class	Count
BENIGN	1223737
DoS	74663
RPM	54900
SPEED	24951
STEERING_WHEEL	19977
GAS	9991

For our study, we will focus exclusively on binary classification, which means that we will only consider the label column for the tests that we are going to perform.
As is often the case in machine learning algorithms studied by everything, the usual pre-treatment must be performed on the data set first, such as the removal of duplications, zero values, etc., then the given set is encoded binary, then subdivided into training and test data to allow classification with the algorithms that will be implemented.

B. Proposed Approach Developing

1. Data pre-processing:

Before any performance study of machine learning algorithms, the data set must first be passed to pre-treatment to avoid any errors in the implementation of the models that will then be studied. In order for machine learning models to exploit data effectively, they need to be cleaned and converted into quantitative values. For this reason, the initial processing of the data plays an essential role in the whole procedure. The quality of the input data is essential to ensure the accuracy and efficiency of the final model, and data pre-processing ensures the consistency, accuracy and usefulness of the input data. Inconsistencies and errors in the data can be removed, making it easier to compare and analyse. It also facilitates the management and use of data for machine learning models [11].

Pre-processing is the step of preparing the data before providing it to the machine for learning. The goal is to put the data in a format conducive to the development of machine learning models but also to have the cleanest dataset possible in order to improve the learning performance of the models to be implemented. To achieve this goal, several operations can be used, among which, in this article, it will be a question of encoding, normalization, cleaning of missing and aberrant values, and the extraction of certain characteristics.

First, the "isduplicate" function of the Pandas library is used to check whether the data set contains duplicate lines. If duplicates are detected, the function returns True for these lines. Next, the "drop_duplicates" function from the same library is used to remove duplicate rows from the database [12]. In the same logic of cleaning the dataset, the "isna" function of the Pandas library is used to check the existence of missing values. In the event that the result of the function also returns True, the "dropna" function of the same library is also used to delete the missing values. When all pre-processing is done, the data set will have to be subdivided into training and test set; in this case, 30% of the data set is used as a test and the remaining 70% as model training.

As already mentioned above, several algorithms have been implemented in order to find the best for the models of the data set studied in combination with CDO optimization. A detailed study of these algorithms will be given in the following sub-section:

2. Algorithms:

In this article, four machine learning algorithms were combined with a meta-heuristic optimization algorithm to conduct a comparative study of the performances achieved. Improving learning performance remains the main objective, as do those with the aim of 'increasing the security of the various constituents. The algorithms studied are detailed in this section.

a) Random Forest: Random Forest is a machine learning method that combines many decision trees to make a final classification. Each tree is constructed using a random subset of data and characteristics, which helps reduce over-adjustment and improve model generalization. The final result is determined by a majority vote of the different trees, where each tree votes for the projected class of the input data point. The class with the highest number of votes is then predicted as the final class of the random forest model.
b) Decision Tree: Decision trees are a highly effective and widely used tool for classification and prediction. They are an organizational chart-like tree structure, where each internal node represents a test conducted on an attribute, each branch represents the outcome of the test, and each leaf node (terminal node) contains a class label. The decision tree is considered to be one of the most powerful methods for making predictions and decisions in various fields [13],[14].
c) Ada Boosting is a learning technique that combines several weak classifiers, such as simple decision trees, to form a robust classifier. It assigns and varies weights to each data point of the training set gradually

with iteration to focus more on the badly classified points at each step. This improves the accuracy of the model by focusing on previous errors.

d) Gradient Boosting: This is a boostage technique that allows you to sequentially combine several weak models to create a strong predictive model. It works by optimizing the loss function by adjusting iteratively to the different successive weak patterns.

Gradient boosting is an ensemble learning technique that aggregates multiple weak models to form a robust model. This method differs from Ada boost in that it utilizes decision trees as its weak models and iteratively refines the model's performance by minimizing the loss function [15].

e) Support Vector Machine (SVM): It is an algorithm for supervised machine learning that has various applications. The algorithm is capable of performing a range of tasks, such as text and picture categorization, spam detection, handwriting recognition, gene expression analysis, face detection, and anomaly detection. The primary objective of this algorithm is to find the optimal hyperplane in an N-dimensional space that separates data points into different classes in the feature space. The hyperplane aims to maximize the separation between the closest points in each class. The size of the hyperplane is determined by the number of entities, and when there are more than three entities, its definition becomes more complex [16].

3. Optimization model:

In the field of network and system security detection, a diverse range of optimization algorithms is used to optimize detection models and improve their ability to identify anomalous behaviour. These include bio-inspired approaches such as Ant Colony Optimization (ACO), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Bacterial Colony Optimization (BCO) and CDO (Chernobyl disaster optimizer). These algorithms draw their inspiration from the collective and adaptive behaviour of living organisms to explore the solution space, optimize the parameters of detection models, and improve the sensitivity of traffic anomaly detection. By combining these bio-inspired approaches with deep learning and machine learning techniques, researchers are developing more robust and effective intrusion detection systems, helping to strengthen network security against cyber threats.

CDO Optimization: It is one of many meta-heuristic optimizations that uses the random aspect in the research process and addresses difficult instances of problems by providing satisfactory solutions (optimal solution or close to optimal solution) within a reasonable timeframe. They provide no guarantee of finding optimal overall solutions. Their uses in many applications show their power and effectiveness in solving important and complex problems. They are usually slightly faster than comprehensive research and are strategies that guide research processes. They are generally used for difficult optimization problems and are not specific to any particular type of problem.

The Chernobyl Disaster Optimizer (CDO) [17] employed in this study is an innovative optimization algorithm that was inspired by the nuclear reactor core explosion at Chernobyl. This method mimics nuclear radiation to efficiently explore and exploit the search space and find global optima in optimization problems. The CDO algorithm uses radiation from gamma, beta, and alpha particles emitted during the Chernobyl disaster to optimize solution search spaces. Each particle has a defined role: gamma particles are responsible for space exploration, beta particles for exploitation, and alpha particles for balancing between the first two. The CDO algorithm effectively evaluates and explores the solution search area by simulating the movement of the various particles. It was evaluated using benchmark functions from the Congress on Evolutionary Computation (CEC 2017) testbed suites and was compared to established optimization methods such as Sperm Swarm Optimization (SSO) and the Gravitational Search Algorithm (GSA). Although recently discovered, CDO has demonstrated an improvement in the field of metaheuristic optimization, making it a viable alternative in this field for new studies. Overall, the CDO algorithm, inspired by the Chernobyl disaster, offers impressive performance and should be considered in optimization processes.

Tables 4, 5, 6, 7 and 8 below show the hyperparameters that will be used in the CDO model in order to be optimized for study implementation. In the next part, these values will be used to train and test the different machine learning algorithms, then the performance will be analysed and a conclusion will be drawn.

Table 4. Random Forest used hyperparameters

Hyperparameters	Range
max_depth	(1, 10)
min_samples_split	(1, 20)
n_estimators	(1, 200)

Table 5. Decision Tree used hyperparameters

Hyperparameters	range
max_depth	(1, 100)
min_weight_fraction_leaf	(0.0, 0.5)
min_samples_split	(2, 50)

Table 6. Gradient boost used hyperparameters

Hyperparameters	Range
learning_rate	(1, 10)
n_estimators	(10, 100)
max_depth	(1, 50)

Table 7. Adaboost used hyperparameters

Hyperparameters	Range
learning_rate	(0.01, 1.00)
n_estimators	(10, 100)
max_depth	(1, 100)
min_samples_split	(0.1, 1.0)

Table 8. Support Vector Machine used hyperparameters

Hyperparameters	Range
C	(1, 100)
gamma	(0.1, 10)

4. Evaluation metrics:

The primary objective of machine learning is to build a model that is capable of accurately predicting outcomes using new data. Evaluation metrics play a crucial role in determining

how well a model is achieving this objective. By assessing a model's performance using metrics, we can pinpoint areas where the model is not performing as expected and implement fixes to enhance its accuracy and usefulness [18].

a) **Accuracy:** It is generally used to measure the performance of the ratio of correctly predicted values to the total number of elements evaluated. [18]. The accuracy equation is present in the equation 1

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

b) **Precision:** Precision is used to find the ratio of the number of true positive values that are correctly predicted to the total predicted positive models in a classification [18]. The precision equation is as follows 2:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

c) **Recall:** It is used to determine the correlation between the number of true positive predictions and the total number of true prediction samples. The recall equation is 3:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN}}$$

d) **F1-score:** It's the harmonic mean of precision and recall. The F1 score equation is 4:

$$\mathrm{F1-Score} = 2\mathrm{x}\frac{Precision\ x\ Recall}{Precision + Recall}$$

IV. RESULT AND ANALYSE:

The optimized values of the hyperparameters defined in the previous sub-section are shown in table 7 and 8.

The hyperparameters optimized in Table 7 are used to train and test the performance of the different machine learning algorithms studied in this article on the binary dataset.

Table 9. Optimized hyperparameters for binary dataset train and testing

Methods	Optimized params
Random Forest	n_estimators=31, max_depth=9, min_samples_leaf=1
Decision Tree	max_depth= 82, min_weight_fraction_leaf = 0.42 min_samples_split = 48
Adaboost	n_estimators= 15, learning_rate= 0.25, algorithm=SAMME.R, max_depth=1, min_samples_split=17
GradientBoost	n_estimators= 29, learning_rate= 1., max_depth=4
Support Vector Machien	C=10.9803, kernel=linear, gamma = 0.42614294402582253

The hyperparameters optimized in Table 8 are used to train and test the performance of the different machine learning algorithms studied in this article on the decimal dataset.

Table 10. Optimized hyperparameters for decimal dataset train and testing

Methods	Optimized params
Random Forest	n_estimators=144, max_depth=2, min_samples_leaf=4
Decision Tree	max_depth=90, min_weight_fraction_leaf = 0.37 min_samples_split = 28
Adaboost	max_depth=1, min_samples_split=0.8242999706920595), learning_rate=0.019398274717060408, n_estimators=14
Gradient Boost	n_estimators= 54, learning_rate= 6, max_depth=11
Support Vector machine	C=24.614912110178434, gamma=3.4971471626898953e-05

The table 8 provides a comparative analysis of the performance of different machine learning algorithms implemented on the dataset used. This comparative study of the performance of machine learning algorithms evaluates the results of different models in different localization contexts.

Algorithm performance was measured using the following metrics: accuracy, precision, recall, and F1 score. These results form the basis for a detailed analysis aimed at identifying the strengths and weaknesses of each algorithm in different application contexts.

After careful examination of the results presented in this table, it was found that the use of optimization can increase the learning performance of IDS systems based on informed decisions about the choice of the best algorithm for a particular machine learning task.

Table 11. Performance metrics

Type	ML Techniques	Accuracy	Precision	Recall	F1 Score
Binary	**Random Forest**	99.53%	99.54%	99.54%	99.47%
	Decision Tree	95.07%	99.15%	95.08%	96.75%%
	Adaboost	99. 91%	99. 16%	99.71%	99.43%
	Gradiantbosst	99. 44%	99.50%	99.44%	99.47%
	Support Vector Machine	98.89%	98.40%	98.89%	99.57%
Decimal	**Random Forest**	97.67%	99.29%	97.68%	98.28%
	Decision Tree	99.63%	99.67%	99.63%	99.64%
	Adaboost	99.81%	99.81%	99.81%	99.81%
	Gradiantbosst	99.54%	99.68%	99.53%	99.76%
	Support Vector Machine	99.44%	99.45%	99.44%	99.34%

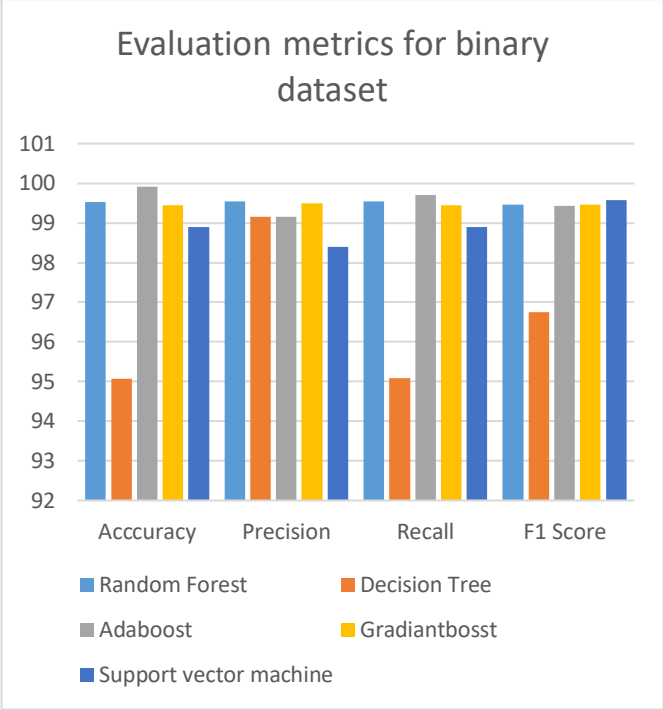

Fig. 1 Metrics performance graph for binary classification

This first graph shows the analysis performance of the algorithms implemented in this study, as shown in the first part of the table. For binary prediction, the Adaboost algorithm obtains the best overall performance with the highest values of accuracy but performance in terms of precision, recall and F1 score is still a little weak compared to other models. The Random Forest and Gradient Boost algorithms achieve slightly lower performance, while the Decision Tree algorithm obtains the lowest performance. These results suggest that the Adaboost algorithm is the most suitable with this optimization for setting up the desired IDS.

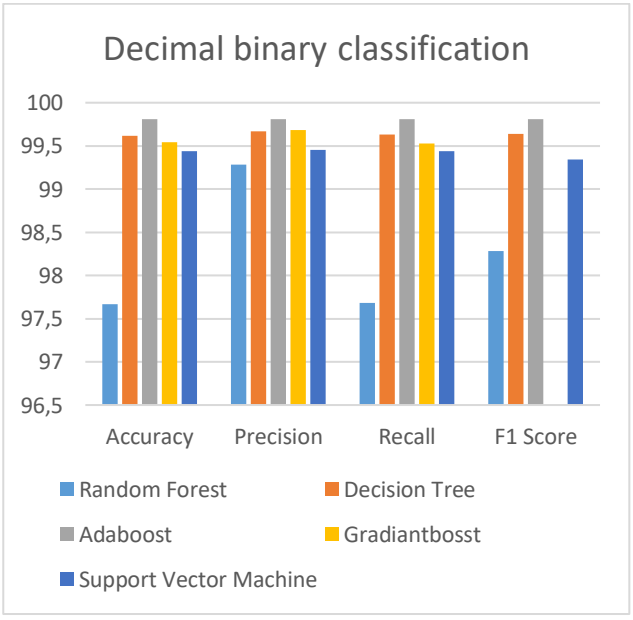

Fig. 2 Metrics performance graph for decimal dataset

The second graph shows the analysis performance of the algorithms implemented in this study, as shown in the second part of the table. For decimal prediction, the Adaboost, and Decision Tree algorithm obtains the best overall performance with the highest values of precision, recall, and F1 score. The Decision Tree and Support Vector Machine algorithms achieve slightly lower performance, while the Random Forest algorithm obtains the lowest performance. These results suggest that for decimal classification Random Forest algorithm is the most suitable with this optimization for setting up the desired IDS.

V. CONCLUSION

Based on all the differences, it can be concluded that security in the Internet of Vehicles (IoV) domain is of crucial importance due to the potential risks associated with cyber threats and attacks. The increasing connectivity of transportation systems and the massive exchange of information make these systems vulnerable to various attacks from malicious actors. The security of IoV systems is a major challenge for security professionals, as security breaches can not only lead to data theft but also endanger human lives and damage physical assets.

The need to protect today's vehicles, which have become sophisticated electronic and computer systems containing millions of lines of code, cannot be overstated. The use of security mechanisms such as intrusion detection systems, security policies and user awareness are crucial to protecting interconnected vehicle networks against cyber threats. The integration of artificial intelligence, in particular machine learning, into intrusion detection systems for the Internet of Vehicles is an effective way of enhancing transport system security. Leveraging techniques such as metaheuristic optimization and machine learning models, including boosting methods (Adaboost, Gradient Boosting), Random Forest, Decision Tree and Support Vector Machine, this paper focuses on the development of an intrusion detection system capable of improving the effectiveness of detection and prevention of cyber-attacks, such as denial of service, identity theft, hijacking and the sharing of false weather updates, in the IoV environment.

In summary, compared with simple intrusion detection systems, the implementation of artificial intelligence-based systems combined with metaheuristic optimization represents a promising approach to enhancing the security of IoV systems and defending users against cyberthreats. The performance evaluations in this study demonstrate that these methods have high levels of accuracy, precision, recall and F1 score, underlining the effectiveness of these strategies in ensuring the safety of interconnected vehicle networks.

REFERENCES

[1] Md. Alamgir Hossain, Md. Saiful Islam, *Ensuring network security with a robust intrusion detection system using ensemble-based machine learning*, Array, 2023

[2] Song, Hyun Min and Woo, Jiyoung and Kim, Huy Kang, *In-vehicle network intrusion detection using deep convolutional neural network*, Vehicular Communications, Elsevier, 2020

[3] Sherazi, Hafiz Husnain Raza, Iqbal, Razi, Ahmad, Farooq, Khan, Zuhaib Ashfaq, Chaudary, Muhammad Hasanain, *DDoS attack detection: A key enabler for sustainable communication in internet of vehicles*, Sustainable Computing: Informatics and Systems, Elsevier, 2019

[4] Seo, Eunbi, Hyun Min Song, Huy Kang Kim, GIDS: *GAN based intrusion detection system for in-vehicle network* 2018 16th Annual Conference on Privacy, Security and Trust (PST), IEEE, 2018

[5] Moayad Aloqaily, Safa Otoum, Al Ridhawi, Ismaeel, Yaser Jararweh, *An intrusion detection system for connected vehicles in smart cities* Ad Hoc Networks, Elsevier, 2019

[6] Bifta Sama Bari, Kumar Yelamarthi, Sheikh Ghafoor, *Intrusion detection in vehicle controller area network (can) bus using machine learning: A comparative performance study*, Sensors, 2023

[7] Hyun Min Song, Jiyoung Woo, Huy Kang Kim, *In-vehicle network intrusion detection using deep convolutional neural network*, Vehicular Communications, Elsevier 2020.

[8] Xinghua Li, Zhongyuan Hu, Mengfan Xu, Yunwei Wang, Jianfeng Ma, *Transfer learning based intrusion detection scheme for Internet of vehicles*, Information Sciences, Elsevier, 2021

[9] M. A. Hossain, & M. S. Islam, *Ensuring network security with a robust intrusion detection system using ensemble-based machine learning.* Array, 2023.

[10] Hamideh and Dadkhah, Sajjad and Iqbal, Shahrear and Xiong, Pulei and Rahman, Taufiq and Ghorbani, Ali, *Ciciov2024: Advancing Realistic Ids Approaches Against Dos and Spoofing Attack in Iov Can Bus*

[11] Subasi, A. (2020). Chapter 2-data preprocessing. Practical Machine Learning for Data Analysis Using Python, 27-89.

[12] Md. Alamgir Hossain, Md. Saiful Islam, *Ensuring network security with a robust intrusion detection system using ensemble-based machine learning*, Array, 2023.

[13] MOHAMMED AMIN BOUKERTOUTA, "*Detection des intrusions basée sur l'apprentissage automatique dans les systèmes IdO (Internet des Objets)*." (2022).

[14] Martindale, Nathan, Ismail, Muhammad, Talbert, A Douglas *Ensemble-based online machine learning algorithms for network intrusion detection systems using streaming data* Information, 2020

[15] Jason Brownlee, *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery, 2020.

[16] https://www.geeksforgeeks.org/support-vector-machine-algorithm/.

[17] Shehadeh,A.Hisham, *Chernobyl disaster optimizer (CDO): a novel meta-heuristic method for global optimization*, Neural Computing and Applications, 2023

[18] HOSSIN, Mohammad et SULAIMAN, Md Nasir. *A review on evaluation metrics for data classification evaluations.* International journal of data mining & knowledge management process, 2015, vol. 5.