applied sciences

MDPI

Article

FSCB-IDS: Feature Selection and Minority Class Balancing for Attacks Detection in VANETs

Sara Amaouche [1], Azidine Guezzaz [1], Said Benkirane [1], Mourade Azrour [2,*], Sohaib Bin Altaf Khattak [3], Haleem Farman [3] and Moustafa M. Nasralla [3,*]

[1] Technology Higher School Essaouira, Cadi Ayyad University, Essaouira 40000, Morocco; saraamaouche232@gmail.com (S.A.); a.guezzaz@uca.ma (A.G.); said.benkirane@uca.ma (S.B.)
[2] STI Laboratory, IDMS Team, Faculty of Sciences and Techniques, Moulay Ismail University of Meknes, Errachidia 52000, Morocco
[3] Smart Systems Engineering Lab, Department of Communications and Networks, College of Engineering, Prince Sultan University, Riyadh 11586, Saudi Arabia; skhattak@psu.edu.sa (S.B.A.K.); hfarman@psu.edu.sa (H.F.)
* Correspondence: mo.azrour@umi.ac.ma (M.A.); mnasralla@psu.edu.sa (M.M.N.)

Abstract: Vehicular ad hoc networks (VANETs) are used for vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communications. They are a special type of mobile ad hoc networks (MANETs) that can share useful information to improve road traffic and safety. In VANETs, vehicles are interconnected through a wireless medium, making the network susceptible to various attacks, such as Denial of Service (DoS), Distributed Denial of Service (DDoS), or even black hole attacks that exploit the wireless medium to disrupt the network. These attacks degrade the network performance of VANETs and prevent legitimate users from accessing resources. VANETs face unique challenges due to the fast mobility of vehicles and dynamic changes in network topology. The high-speed movement of vehicles results in frequent alterations in the network structure, posing difficulties in establishing and maintaining stable communication. Moreover, the dynamic nature of VANETs, with vehicles joining and leaving the network regularly, adds complexity to implementing effective security measures. These inherent constraints necessitate the development of robust and efficient solutions tailored to VANETs, ensuring secure and reliable communication in dynamic and rapidly evolving environments. Therefore, securing communication in VANETs is a crucial requirement. Traditional security countermeasures are not pertinent to autonomous vehicles. However, many machine learning (ML) technologies are being utilized to classify malicious packet information and a variety of solutions have been suggested to improve security in VANETs. In this paper, we propose an enhanced intrusion detection framework for VANETs that leverages mutual information to select the most relevant features for building an effective model and synthetic minority oversampling (SMOTE) to deal with the class imbalance problem. Random Forest (RF) is applied as our classifier, and the proposed method is compared with different ML techniques such as logistic regression (LR), K-Nearest Neighbor (KNN), decision tree (DT), and Support Vector Machine (SVM). The model is tested on three datasets, namely ToN-IoT, NSL-KDD, and CICIDS2017, addressing challenges such as missing values, unbalanced data, and categorical values. Our model demonstrated great performance in comparison to other models. It achieved high accuracy, precision, recall, and f1 score, with a 100% accuracy rate on the ToN-IoT dataset and 99.9% on both NSL-KDD and CICIDS2017 datasets. Furthermore, the ROC curve analysis demonstrated our model's exceptional performance, achieving a 100% AUC score.

Keywords: attacks; intrusion detection; machine learning; security; VANET

Citation: Amaouche, S.; Guezzaz, A.; Benkirane, S.; Azrour, M.; Khattak, S.B.A.; Farman, H.; Nasralla, M.M. FSCB-IDS: Feature Selection and Minority Class Balancing for Attacks Detection in VANETs. *Appl. Sci.* **2023**, *13*, 7488. https://doi.org/10.3390/app13137488

Academic Editor: Luis Javier Garcia Villalba

Received: 27 May 2023
Revised: 19 June 2023
Accepted: 21 June 2023
Published: 25 June 2023

1. Introduction

VANET is a subclass of MANET that enables wireless communication among vehicles and infrastructure devices to improve traffic safety and efficiency [1,2]. VANET comprises

two types of nodes: on-board units (OBUs) that are installed in vehicles; and roadside units (RSUs) that are deployed along the roads. The vehicular communication module (VCM) is responsible for routing the traffic data from other VANET nodes to the vehicle devices, facilitating effective vehicle monitoring and management [3].

Even though VANET has achieved significant improvements in its domain, the security challenge still remains the biggest challenge. This problem is a direct threat to the quality of human life and constitutes one of the major difficulties faced by VANETs. Various types of attacks may impact VANETs as the vehicles are linked over a wireless medium. There are many dangers, such as flooding, SQL injection, DoS, DDOS, Sybil, and Jamming attack [4,5], which injects undesired traffic to a network and blocks the legitimate users from accessing the resources. Furthermore, malicious software is a kind of code that takes control by exploiting vulnerabilities in the network system [6]. Consequently, the security of vehicles against intrusion became an important point, since traditional models of vehicles do not include this aspect of safety. Since VANETs provide different forms of communication, as V2V and V2I are linked to the Internet, they are vulnerable to different types of attacks, which not only degrade the quality of traffic but also threaten human lives. Consequently, the use of classical security mechanisms, such as encryption techniques or access control, are not relevant to modern vehicles. Reactive systems, in particular intrusion detection systems (IDSs), have been the subject of much attention these days, as they are able to detect any possibility of cyberattacks on VANETs in order to improve their usability.

The objective of the present work is to develop and implement an intrusion detection model for VANETs. To realize this model, we will use the RF classifier and several techniques to improve the quality of the dataset usage and to have better results. Mainly, the mutual information is used for the selection of the group of best features to build an efficient model, the Synthetic Minority Over-sampling Technique (SMOTE) which is an oversampling technique that helps to overcome the overfitting problem due to the random oversampling. This technique is used in our case to deal with the imbalance problem between the classes and the one-hot encoding technique to convert the categorical information into numerical data. We evaluated and compared the results obtained with the following ML techniques: DT, KNN, LR, and SVM. The model was evaluated and compared using three datasets—NSL-KDD, TON-IOT, and CICID2017—in order to validate the accuracy and the efficiency of our system. These results confirm the performance and accuracy of the proposed model.

The rest of the paper is structured as follows: Section 2 presents the background and related work of VANETs architecture, services, network security, and intrusion detection methods. It also covers the relevant literature related to intrusion detection based on machine learning, deep learning, and ensemble learning techniques. Section 3 details the design of the proposed model. A discussion of the results and evaluation study of the system is presented in Section 4. The conclusion and future work are presented in Section 5.

2. Background and Related works

2.1. Background

This section provides background and recent work that has focused on intrusion detection approaches incorporating ML techniques for securing VANETs.

In VANET, vehicles are connected together using both V2V and V2I communication through RSUs and mobile broadband such as 4G/LTE [7]. VANET services comprise vehicular and road security services, efficiency and traffic control services, as well as info services. The safety services for vehicles and roads are to reduce traffic accidents and the death of vehicle users [8]. Efficiency and traffic control services focus on the improvement of traffic flow, traffic management, as well as providing location and mapping information. The objective of information services is to ensure the accessibility of the data, including the transfer of multimedia data and the connection to the Internet [9,10]. The VANET communication architecture may be categorized as Wi-Fi-based ad hoc and hybrid wireless access in vehicle environments (WAVE). In the Wi-Fi-based WAVE architecture, RSUs located

along roadways are used as wireless access nodes which provide communication service to vehicles in their area of coverage. In ad hoc architecture, a group of vehicles are forming ad hoc networks by utilizing WAVE. Those networks work in an autonomous way, with no infrastructure. For hybrid architecture, the cellular and ad hoc architectures employing WAVE execute tasks in collaboration [11]. Initially, VANETs utilized Dedicated Short Range Communications (DSRC), using 75 MHz of bandwidth. The previous standards were not adapted to the expected functions and uses of VANETs. Additional enabling communication is improving the reach of VANETs and enhancing connectivity to more types of supporting technologies such as the Internet of Things (IoT) [12] and cloud computing [13]. Vehicular networks share the same characteristics as predictable ad hoc sensor networks, such as self-structuring and lack of central control [14]. VANETs have some exclusive challenges that force the design of the communication system and the security of its protocol [15]. VANETs constitute the technical basis for the visualization of Intelligent Transport Systems (ITS). The vehicles are characterized by their V2V and V2I communication capabilities [16]. The large number of nodes circulating inside the VANET network, with a high degree of mobility and a high frequency of topological changes, constitute a major challenge for this type of network as the nodes potentially move at high speed. In addition, the challenge of the complexity of keeping the confidentiality of the information related to the security is also a critical challenge, as the data contained in the messages exchanged in the network are broadcasted to all the users of this network and must not be confidential. Finally, vehicle communication capabilities can expose driver or user information such as ID, speed, location, and mobility patterns, which poses a privacy issue. Regardless of the need for message authentication and non-repudiation of safety messages, user and driver privacy must be valued, especially location privacy and ambiguity [15]. Figure 1 presents the architecture of VANET network, illustrating the different forms of communication.

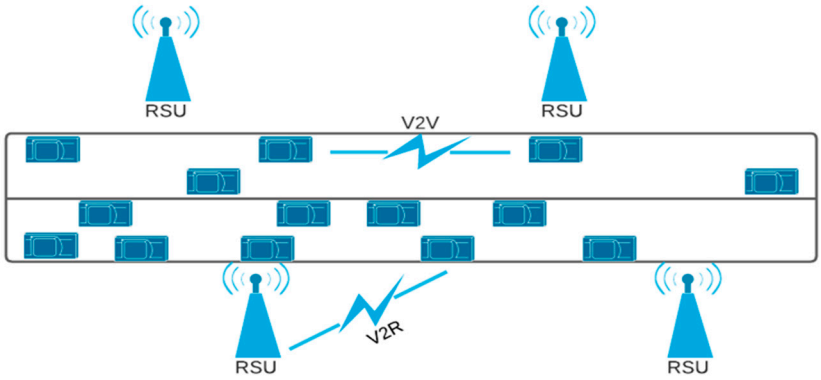


Figure 1. VANET Architecture.

VANETs have a direct connection to the Internet, which makes them susceptible to many security problems. As a result, Intrusion Detection Systems IDSs are identified as one of the leading strategies to address the challenges of VANET security. Specifically, IDSs might be used to protect any network against attacks.

IDS is used to detect abnormal activities over a network [17–19]. IDSs are suggested to detect attacks that are internal and not detectable by cryptographic systems. An IDS is generally implemented as the second layer of security following cryptography methods [20,21]. There are two different approaches of intrusion detection: the anomaly detection and the misuse detection. The anomaly detection consists of defining and characterizing the right static form or the correct dynamic behavior for the system and then the detection of changes or illegal behaviors. It depends on being able to determine the wanted form or pattern of the system and then discriminate whether that form or behavior are undesirable changes or anomalies. The frontier between normal and abnormal form is definable because only one different bit signals a failure. The frontier between normal and abnormal patterns of behavior is harder to identify [22]. The second approach, misuse detection, consists

of the identification of possible ways to access a system. Every one of these elements is usually expressed as a model. The abuse detection system tracks the explicit patterns, where the pattern could be either a string of bits or a sequence of suspicious actions [22]. In various cases, both detection types are integrated in a combined and complementary form in one system. Although, novel forms of pattern specification for abuse detection have been created, and techniques designed for simple systems have been modified and adapted to deal with intrusions in networks and distributed systems. System effectiveness and monitoring have been enhanced, as have user interfaces, including those for specification of new abuse patterns and interaction with the security administrator of the system. A number of researchers have proposed IDS based on ML or DL methods. In vehicular networks, IDSs examine traffic packets in the network to identify malicious activities. Two different approaches to IDSs can be distinguished: for the first, the IDS is installed inside the vehicle, and for the second, the IDS is installed on the RSU. LR [23] which is a modified version of the linear regression method, is commonly used for classification issues. It is helpful in applications like filtering spam and detecting intrusions. The probability of LR is calculated according to Equation (1):

$$h_\theta = \sigma\left(\theta^T X\right) \tag{1}$$

h_θ: the hypothesis function
X: the vector of feature input
θ: the parameters of the LR
σ: stands for the sigmoid function that defines the threshold $\sigma(r) = \frac{1}{1+ e^{-1}}$

DT [24], consisting of a tree graph composed of internal nodes representing the test on an element, of branches showing the result of the test and of leaf nodes representing the class label.

The Gini impurity is used as a dividing criterion, as presented in Equation (2):

$$G(D) = \sum_{I=1}^{C}(P(i) * (1 - P(i)) \tag{2}$$

D: training dataset
C: class label collection
P(i): proportion with the label category I in C

KNN [25] is a non-parametric approach that does not anticipate the distribution of the underlying data. It finds a group of k observations in the training ensemble that are most similar to testing observation and attributes a label according to the common class of the k neighbors.

Equation (3) represents the Euclidian distance which measures the difference between two samples:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

d(x,y): Euclidian difference of two samples
x_i: first observation
y_i: second sampling
N: the observation

Reference [26] is a supervised ML approach that may be applied to solve classification and regression problems. It belongs to the ensemble learning that couples multiple decision trees to create a model to predict the class of a data with higher accuracy. SVM [27,28] is a classification technique that deals with linear and non-liners datasets. This algorithm consists in the identification of a hyperplane which enhances the distinction between the classes.

$$K(x, y) = e^{-\frac{||x-y||^2}{2\sigma^2}} \quad (4)$$

σ: variance
$||x - y||$: Euclidean distance of two points.

2.2. Related Works

A number of researchers have proposed detection methods for IDS in VANET networks. Al-Jarrah et al. [29] combined the RF technique with a bidirectional feature selection process of classification. The KDD-CUP99 was cleaned; duplicate data were eliminated; and then, different techniques of preprocessing were applied, such as normalization and discretization. RF-FSR gave excellent results of 99.90% classification, and RF/BER of 99.88%. Ahmed [30] proposed an IDS for VANET using a dataset called ToN-IoT. They used several techniques to resolve class imbalance, missing data, and encoding of categorical values. The proposed model showed the following results: LR 60.9% of accuracy, NB 51.3%, DT 87%, RF 87%, Adaboost 43.6%, KNN 97.1%, SVM 75.1%, and XGBoost 98%. In this paper, a ML-based IDS was developed by the authors [31] in order to identify the intrusions in VANETs. To guarantee the security of group leaders, they applied an artificial neural network (ANN), and to detect abusive multipoint relays, they employed SVM. Compared with the preceding ML-based approaches, the findings result indicated that the implemented approach had higher robustness and reliability. Wahab et al. [32] provided a model for multi-decisional intelligent detection named CEAP that conforms to one of the most important characteristics of VANETs—that they are highly mobile with a high detection rate and low costs. They used SVM for classifying intelligent vehicles as cooperating or misbehaving. The results obtained by the model based on the density scenario of the network are 99.13% for linear kernel, 99.04% for multilayer perception kernel, 99.13% for quadratic kernel, 99.35% for polynomial kernel, and 99.67% gaussian radial basis function kernel. H. Bangui, M. Ge, and B. Buhnova [33] have developed an application based on the RF to identify possible intrusions on the network. It is integrated with the post-detection phase by utilizing the benefits of the coresets as well as the clustering algorithms. The proposed method had a maximum accuracy of 96.93%, while the smallest accuracy was obtained with Bayesian-Coresets of 82.4%, CNN 95.14%, and SVM 85.2%. The authors in [34] developed a proposed framework using a VANET system network and a multiple agent system of communication which contributes to smart traffic management, helping to make it efficient. In this study [35], an ML-based method is introduced to classify misbehaviors in VANETs. The authors propose a security framework that utilizes packet manipulation to distinguish between malicious and legitimate nodes. The classification process incorporates features like node speed deviation, RSS, and packet delivery and drop counts. The evaluation includes both binary and multi-class approaches, and experiments are conducted in the NCTUns-5.0 simulator using various scenarios. The results demonstrate the effectiveness of the proposed framework in accurately classifying misbehaviors in VANETs, with Random Forest and J-48 classifiers exhibiting superior performance compared to other methods. A novel ECRDP approach of routing has been put forward in [36]. The authors of [37] proposed a new routing protocol based on clustering. Khattab M et al. [32], propose CEAP, a multi-decision intelligent detection model tailored for VANETs. The model utilizes cooperative monitoring and SVM learning to achieve higher detection rates with minimal overhead. To address the highly mobile nature of VANETs, CEAP is implemented on top of the VANET QoS-OLSR protocol, which considers vehicle mobility metrics for stable clustering and extended network lifetime. Additionally, a propagation algorithm is introduced to reduce the exchange of redundant information and repeated detection steps. Simulation results demonstrate that CEAP outperforms other detection techniques in terms of accuracy, attack detection rate, false positive rate, and packet delivery ratio in high mobility scenarios. Omar A et al. [38] introduce a probabilistic cross-layer Intrusion IDS that incorporates ML techniques. The IDS demonstrates a detection accuracy exceeding 90% for spoofing attacks. The authors

introduce a new metric named Position Verification using Relative Speed (PVRS), which significantly influences the classification results. PVRS compares the observed distance between communicating nodes, as captured by On-Board Units (OBU), with the estimated distance obtained through the calculation of relative speed values using exchanged signals in the Physical (PHY) layer. The proposed IDS presents an effective approach to detect spoofing attacks in VANETs. D. Kosmanos et al. [39] emphasizes the implementation of Software-Defined Network (SDN) in IoT networks to provide a more cost-effective solution compared to traditional hardware components. The paper delves into the application of IDS within SDN-based IoT networks, presenting a comprehensive analysis of various studies and conducting comparisons. The primary objective of this survey is to introduce innovative research avenues for SDN-based IoT networks. Additionally, the paper explores the integration of blockchain technology as a means to enhance security in SDN-based IoT networks. Abro et al. [40] examine recent advancements in electric vehicles (EVs) and emerging technologies for e-mobility by 2030. It highlights the benefits of integrating autonomous features in EVs to improve safety and fuel efficiency. The research addresses research gaps and proposes solutions for intelligent vehicles, emphasizing the importance of real-world data and exploring various EV-related fields. In this article [41], a novel approach is introduced for the early and accurate detection of botnet attacks, utilizing common network traffic patterns and temporal features. The study evaluates and compares different ML algorithms, including DT, probabilistic neural network, sequential minimal optimization, and Adaboost classifiers, against existing research. The findings demonstrate the effectiveness of the proposed approach, achieving an impressive true positive rate of 99.7%, and emphasize the significance of incorporating temporal features to enhance botnet detection efficiency. This article [42] introduces a unique simulation technique to create the VANET network distributed DOS dataset, specifically designed for VANET. The dataset captures essential aspects such as network architecture, traffic density, attack intensity, and node mobility, offering comprehensive information on distributed denial of service attacks. Comparisons with existing studies confirm the novelty of the dataset, while the evaluation of various ML models demonstrates their high accuracy rates exceeding 99.5%, except for the SVM achieving 97.3%.

After examining the recent work outlined in Table 1, we have identified numerous challenges and limitations. VANETs, characterized by high vehicle mobility, shared network medium, and the absence of centralized security devices like firewalls and authentication servers, are highly susceptible to attacks on wired networks. Managing the growing volume of vehicular traffic in urban environments poses a significant challenge for IDS. Additionally, VANET networks and their critical infrastructure interface face ongoing security challenges in terms of rapidly transmitting, receiving, archiving, and retrieving information across the network. To address these concerns, we propose an IDS based on an RF classifier, aiming to enhance precision and accuracy. Our model's robustness is demonstrated by comparing it with various ML—DT, KNN, LR, and SVM techniques—and testing it on different datasets—NSL-KDD, TON-IOT, and CICID2017.

Table 1. Classification and comparison studies of selected recent work.

Author	Year	Used Learning Method	Accuracy (%)	Attacks Covered
Grover et al. [19]	2011	IBK	56.0	Multiple Misbehaviors
		Random Forest	92.0	
		J−48	42.0	
		Naive Bayes	92.0	
		AdaBoost1	92.0	
Wahab et al. [16]	2016	SVM		Multiple Misbehaviors

Table 1. *Cont.*

Author	Year	Used Learning Method	Accuracy (%)	Attacks Covered
Abdalla and Ahmed [14]	2021	Logistic regression Naive Bayes K-nearest Neighbor Decision tree XGBoost Support vector machine Random forest AdaBoost	60.9 51.3 97.1 87.0 98.0 87.0 97.1	Multiple Misbehaviors
Al-Jarrah et al. [13]	2014	RF-FSR RF-BER	99.90 99.88	Multiple Misbehaviors
H. Bangui et al. [17]	2021	Random Forest Bayesian-Coresets SVM CNN	96.93 82.40 85.20 95.14	Multiple Misbehaviors
Khattab M et al. [40]	2018	SVM Neural network	99.92	Grey hole Rushing attack
Omar A et al. [41]	2016	SVM	99.67	Multiple Misbehaviors
D. Kosmanos et al. [42]	2020	RF KNN	90	Multiple Misbehaviors

3. Proposed Methodology

This section enlightens the proposed work using RF classifier. We optimize the quality of the data using mutual information for feature selection to reduce the consummation of resources. To solve the problem of class imbalance, we apply the SMOTE technique. Then, we train the RF classifier by considering fifteen features to create the proposed intrusion detection system. The proposed model has three steps as shown in Figure 2: data preprocessing, feature selection, and classification.

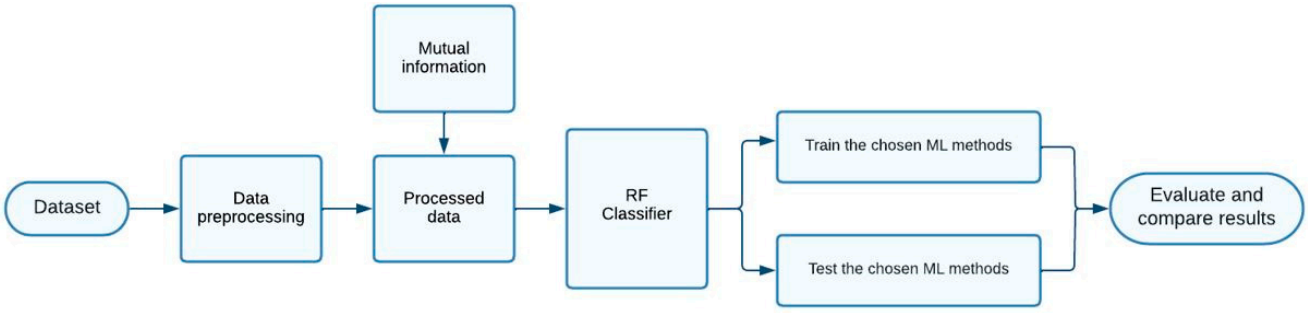


Figure 2. The components detail of our approach.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1. Data Preprocessing

The most important step in the ML method is the preparation of the data to obtain high performance. Initially, the data needs to be cleaned as it can have problems such as missing values, conversion of categorical data to numerical, and imbalance classes. Moreover, the unnecessary data must be removed, which can affect the performance and results of using the ML method. The datasets must be processed properly in order to create a useful analysis. In our case, our datasets contain many missing values, so we will replace them with the most common values. Additionally, we have to convert the categorical data of our datasets into numerical to better use our model. For this purpose, we used a one-shot encoding. And finally, to address the issue of imbalance class, we use the SMOTE

technique [16] which improves random oversampling by the supply of synthetic samples of minority classes and fixes the overfitting issue which may result from simple random oversampling since SMOTE generates new data items in place of duplicating the existing data items.

Steps of oversampling method:

For every motif X in the minority class A:

Select one of its K nearest neighbors x^1 (also belong to the minority class).

Create a new motif S on a random point on the line segment linking the motif and the selected neighbor, as shown in Equation (5):

The following formula is utilized to generate a new example for each $x_i) \in A$ (i = 1, 2, 3 . . . N):

$$S = x + \text{rand}\ (0,\ 1)\ \times\ (x\_\ x_i) \tag{5}$$

3.2. Feature Selection

The objective of feature selection is to eliminate unimportant and/or duplicate features so as to select the optimal set of features to build efficient models of the studied phenomena. The feature selection workflow consists of obtaining a score for each possible feature and selecting the better features. A feature's frequency is calculated in the training process for each instance, both positive and negative individually, and a function of these two is calculated. It is important to verify features in intrusion detection to identify the most crucial ones that will help in eliminating unnecessary data. In this paper, we have used mutual information [43] filtering method that measures the decrease in variance among two variables, as presented in Equation (6).

$$I(X;Y) = H(X) - H(X \mid Y) \tag{6}$$

H(X): X's entropy

H(X | Y): X's conditional entropy given Y

3.3. Data Normalization and Classification

Normalization minimizes the complexity of models as it has a significant role in reducing the overweight features having higher values in comparison with ones having lower values. The normalization method used is Min-Max to scale the feature values between [0:1] as defined in Equation (7).

$$x_{norm}\ = \frac{x - x_{min}}{x_{max} - x_{min}}\ \in [0,\ 1] \tag{7}$$

For the classification process, we will use the RF classifier, and we will compare the results with the following ML techniques: LR, KNN, DT, and SVM. We will perform the tests on the three datasets ToN-IoT, NSL-KDD, and CICIDS2017.

4. Results and Discussion

This section analyzes and compares the efficiency of our employed ML methods for VANET by comparing the results of the three datasets ToN-IoT, NSL-KDD, and CICIDS2017.

4.1. Environment Description

NSL-KDD is a newer version of the existing KDD'99 dataset. It is an efficient reference dataset that helps researchers in comparing various methods of intrusion detection. Every single set of records has different characteristics of the flux with a label indicating whether it is an attack or normal behavior. Because of the limited open data sets for the use of IDS based on networks, NSL-KDD is actually the most effective dataset that can be used to evaluate various intrusion detection approaches [44]. The CICIDS2017 dataset includes benign attacks and the latest current attacks, resembling real-world data (PCAP). In addition, it contains analysis results from network flow based on CICFlowMeter, having

flows that are labeled by timestamp, the source and destination IPs, the protocols, and the attacks (CSV files) [45]. ToN-IoT holds a collection of different information sources taken from the whole IIoT system, as well as telemetry data, linux records, and also IoT system network traffic. The ToN-IoT network is accessible on ToN-IoT repository [46]. In addition, the datasets were displayed in CSV format with a column labeled that represents the attack or the normal behavior, indicating the types of attacks.

We run our tests on an Intel(R) Core(TM) i7-8650U CPU @ 1.90 GHz 2.11 GHz, and 16 GB in RAM, with Windows 10 × 64-bit. We implemented our model and feature engineering using Python v3.10.6.

- TP: True positives define the amount of well identified intrusion.
- FP: False positives represent the amount of badly classified intrusion.
- TN: True negatives define the amount of correctly well normal occurrence.
- FN: False negatives represent the amount of misclassified normal occurrence.

To evaluate our model we used the following metrics: accuracy, precision, recall, and f1-score, and the equations to calculate these metrics are respectively (8), (9), and (10). Those metrics are calculated based on the elements of the confusion matrix represented in the Table 2.

$$\text{Accuracy}: \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{8}$$

$$\text{Precision}: \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{9}$$

$$\text{Recall (True Positive Rate TPR)}: \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{10}$$

$$\text{False Positive Rate (FPR)}: \frac{\text{FP}}{\text{FP}+\text{TN}} \tag{11}$$

$$\text{F1-score}: 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

Table 2. The confusion matrix.

	Predicted Normal	**Predicted Attack**
Actual Normal	TN	FP
Actual Attack	FN	TP

4.2. Result Discussion

In this section, we first present the results for our three datasets NSl-KDD, ToN-IoT network, and CICIDS2017. A cross validation with 10 factors was used for all ML methods. In summary, all the applied techniques obtained significant results due to the different feature engineering methods that were employed on the datasets. First, we imputed missing values and converted categorical values, solving the imbalanced data problem with the SMOTE technique, and then applying the Min-Max normalization technique.

4.2.1. Using NSL-KDD Dataset

Figures 3 and 4 present the outcomes of applying SMOTE on the NSL-KDD dataset to address the issue of imbalanced classes. Figure 3 displays the distribution percentages of each class prior to employing the SMOTE technique, illustrating the initial class imbalance. In contrast, Figure 4 showcases the outcomes after the implementation of SMOTE, revealing a noticeable resolution of the imbalanced classes. The results demonstrate a more balanced distribution among the classes, thereby indicating the effectiveness of the SMOTE technique in mitigating class imbalance.

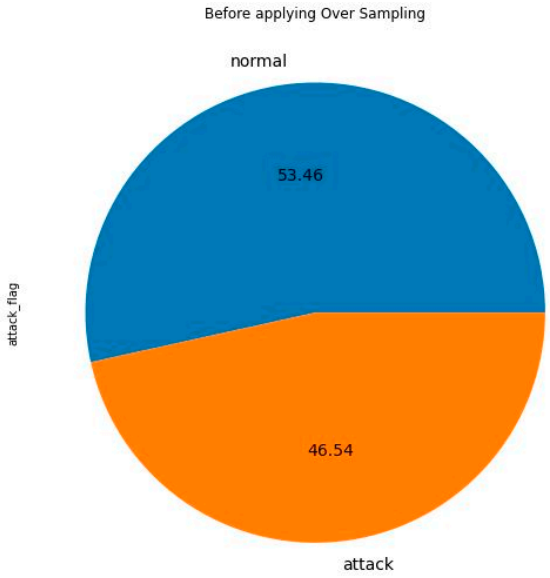


Figure 3. The amount of information in the attack class before applying the over sampling technique (NSL-KDD dataset).

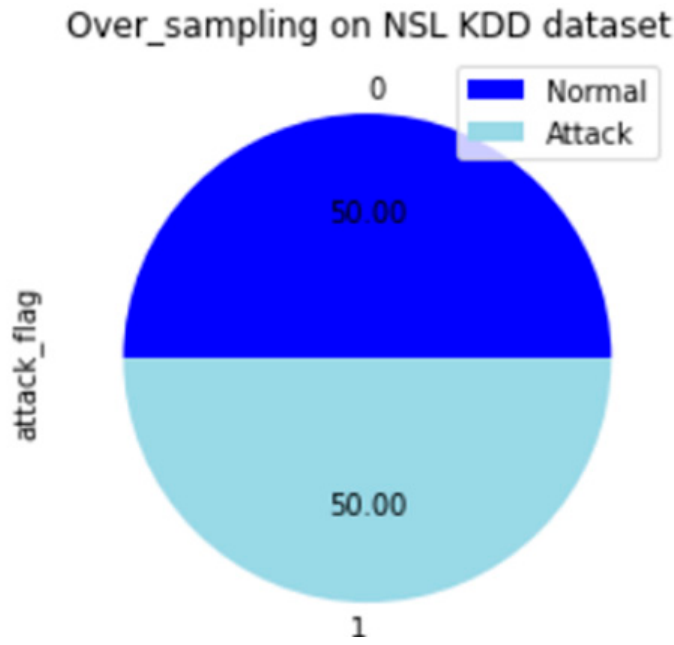


Figure 4. Attack class using SMOTE (NSL-KDD dataset).

The mutual information technique reveals the most influential features in the NSL-KDD dataset, which we use for dimensionality reduction by selecting the columns with the highest variance for improved prediction accuracy. Figure 5 displays the result of this technique, where we choose the top 15 features for our model.

We utilized mutual information as a feature selection method, resulting in the identification of 15 optimal features, as shown in Figure 5. To address the class imbalance issue in the dataset, as depicted in Figure 3, we applied the SMOTE technique, which significantly improved the distribution of the classes, as illustrated in Figure 4.

For the purpose of comparison, we employed the RF classifier alongside LR, DT, KNN, and SVM techniques. The experiments were conducted using the NSL-KDD dataset. Figure 6 displays the performance of our proposed model, exhibiting a satisfactory outcome with a minimal number of FPR and FNR.

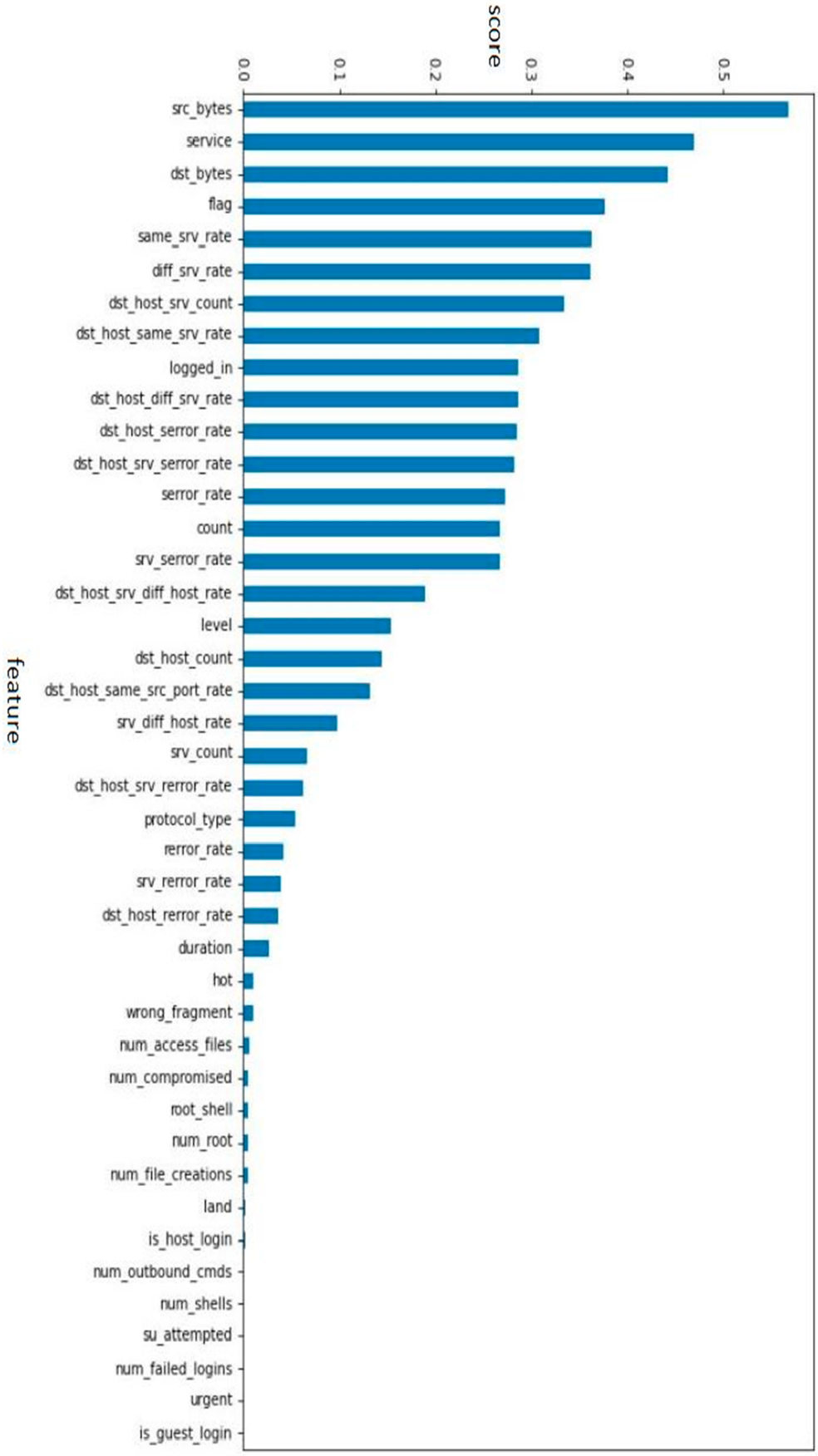
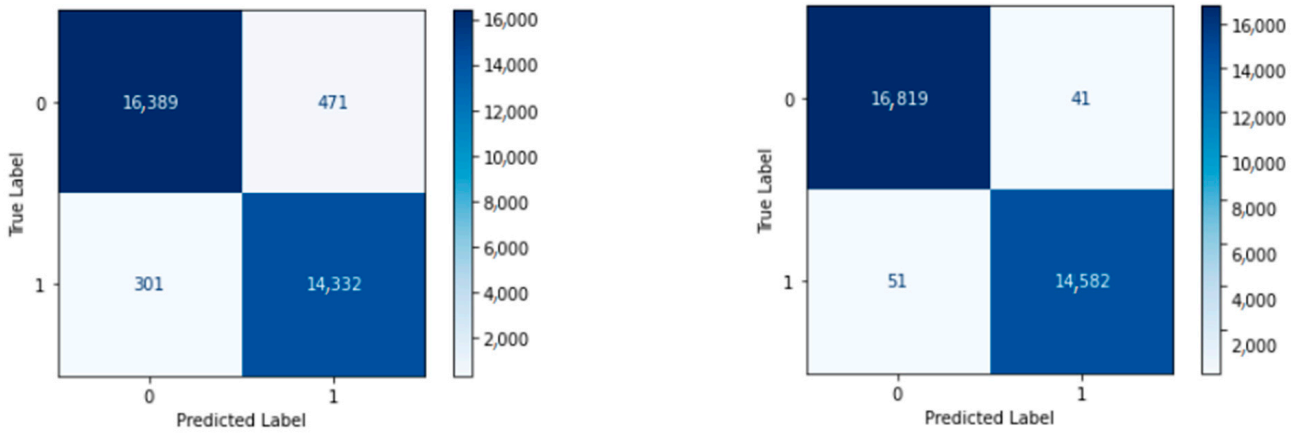


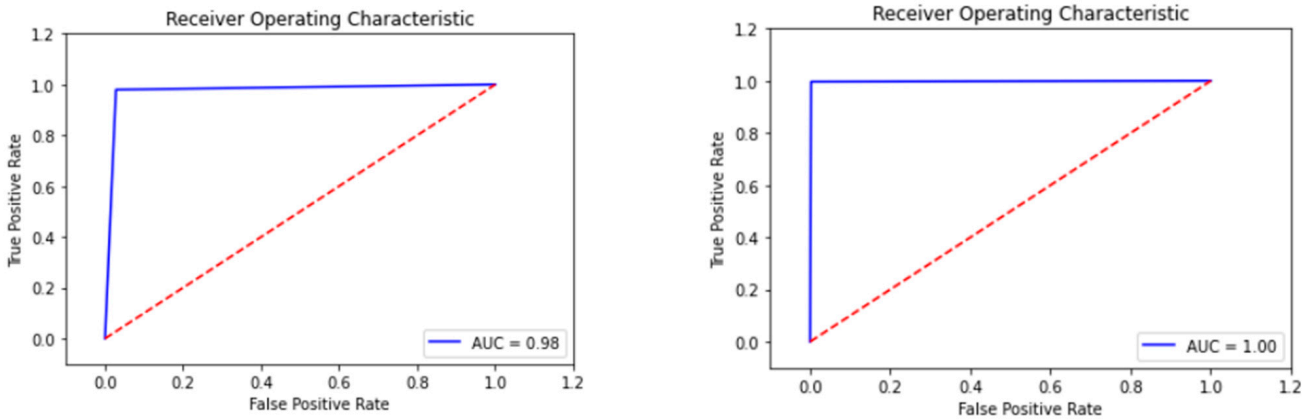
Figure 5. Influential features of detection attack on NSL-KDD dataset.



(**a**) Logistic Regression (**b**) Random Forest

Figure 6. Confusion matrix of prediction for NSL-KDD dataset.

To further assess the performance of the techniques, we examined the ROC curve, as shown in Figures 7 and 8. Our analysis reveals that we achieved good results, particularly with a 99% accuracy for KNN. Moreover, Table 3 provides a comprehensive overview, indicating that our model exhibits high accuracy (0.99), recall (0.97), precision (0.97), and F1 score (0.97) in terms of intrusion detection, with DT and KNN methods following closely.



(**a**) Logistic regression (**b**) Random Forest

Figure 7. ROC curve of the applied techniques on the 15 best features (NSL-KDD Dataset).

Table 3. Experimentation result on NSL-KDD dataset.

Models	Accuracy	Precision	Recall	F1-Score
LR	97.6	97.5	97.5	97.5
KNN	99.8	99.8	99.8	99.8
DT	99.8	99.8	99.8	99.8
SVM	99.6	99.6	99.6	99.6
Proposed Model	99.9	99.7	99.7	99.7

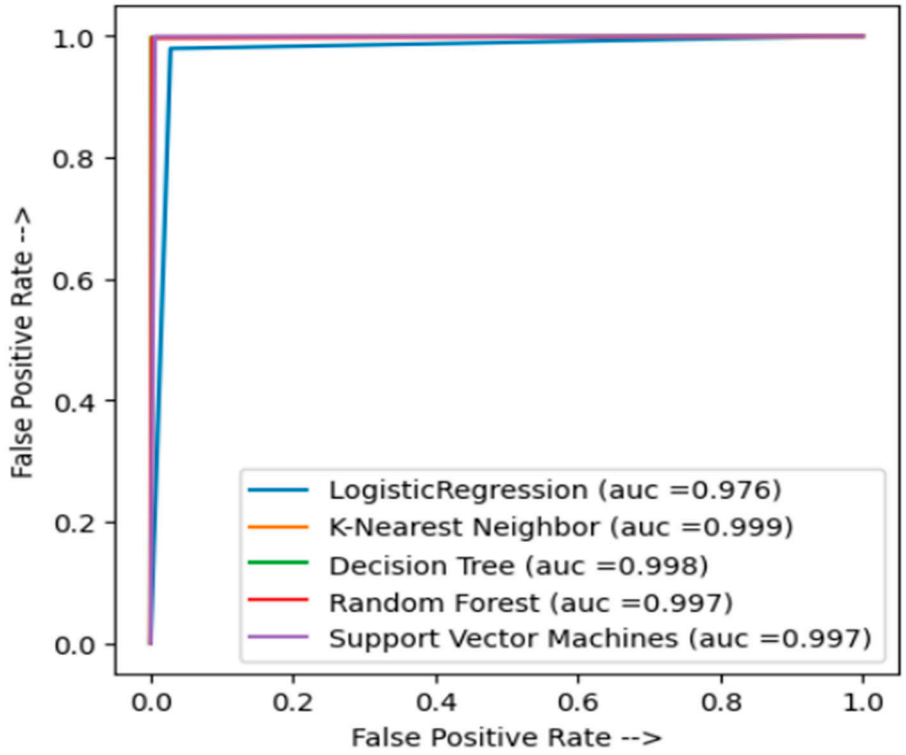


Figure 8. ROC curve for all applied techniques on the 15 best features (NSL-KDD).

4.2.2. Using TON-IOT Dataset

TON-IOT dataset has also the problem of class imbalance which is shown in Figure 9. After using the SMOTE technique, this problem is solved as shown in Figure 10. The influence features present in our TON-IOT dataset are shown in Figure 11. We have selected the 15 best features to obtain better results.

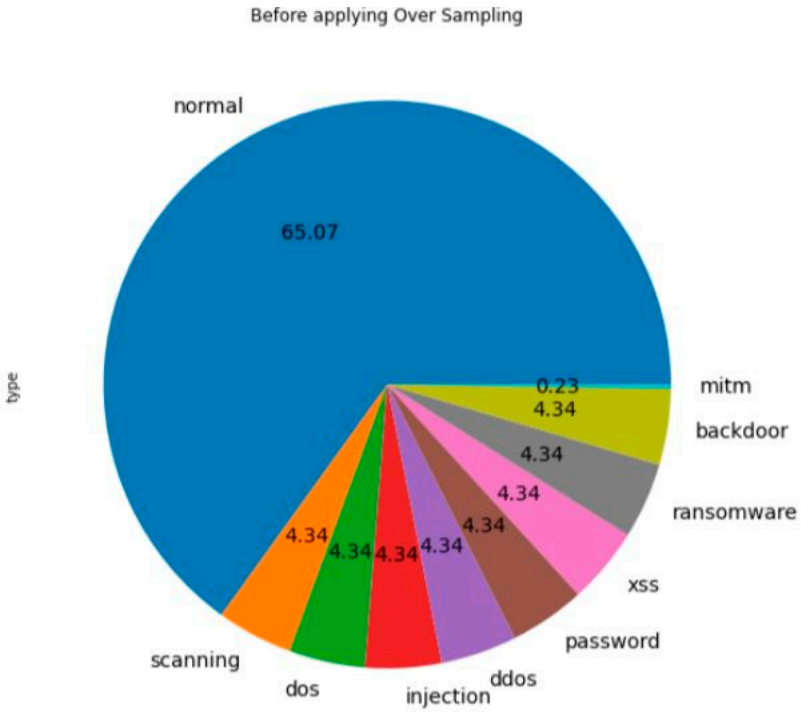


Figure 9. Attack class before applying the over sampling technique by using TON-IOT dataset.

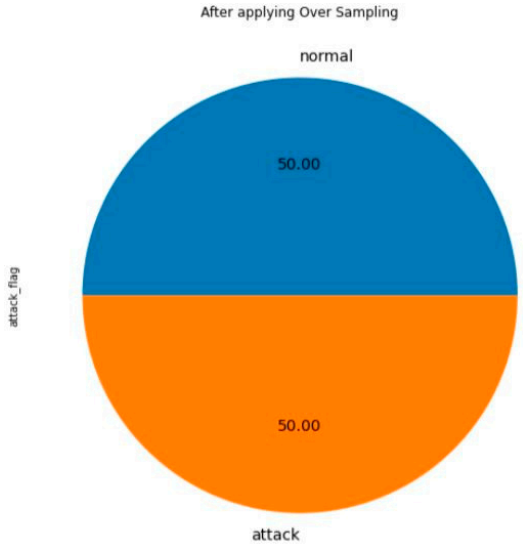


Figure 10. Attack class after applying the over sampling technique by using TON-IOT dataset.

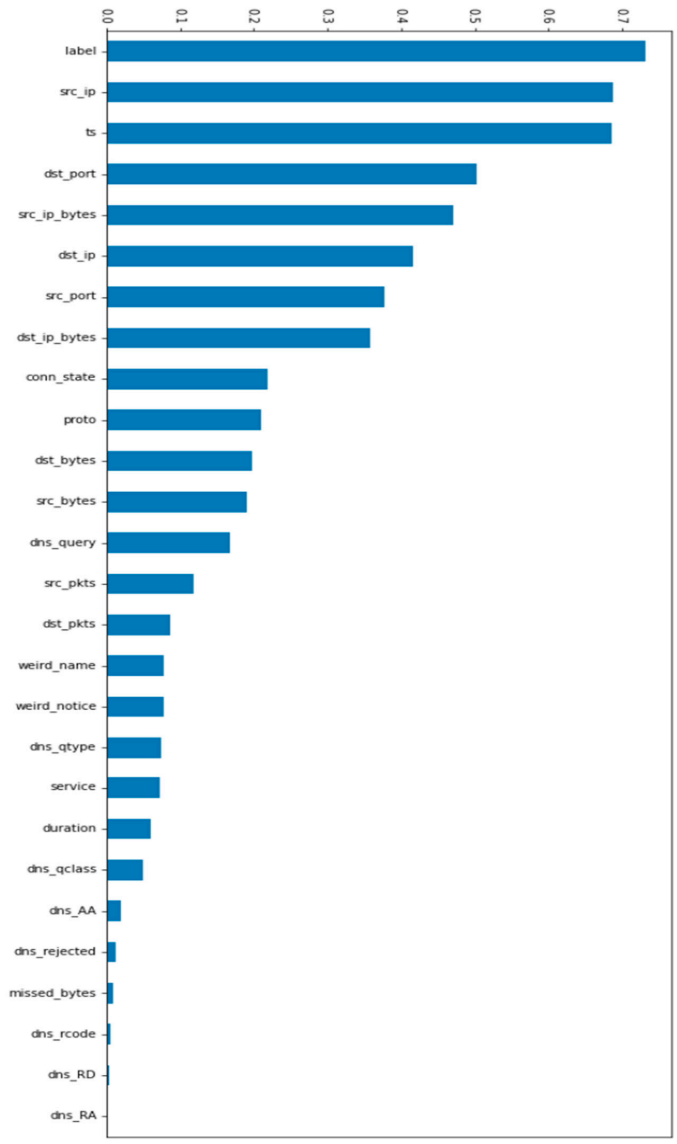
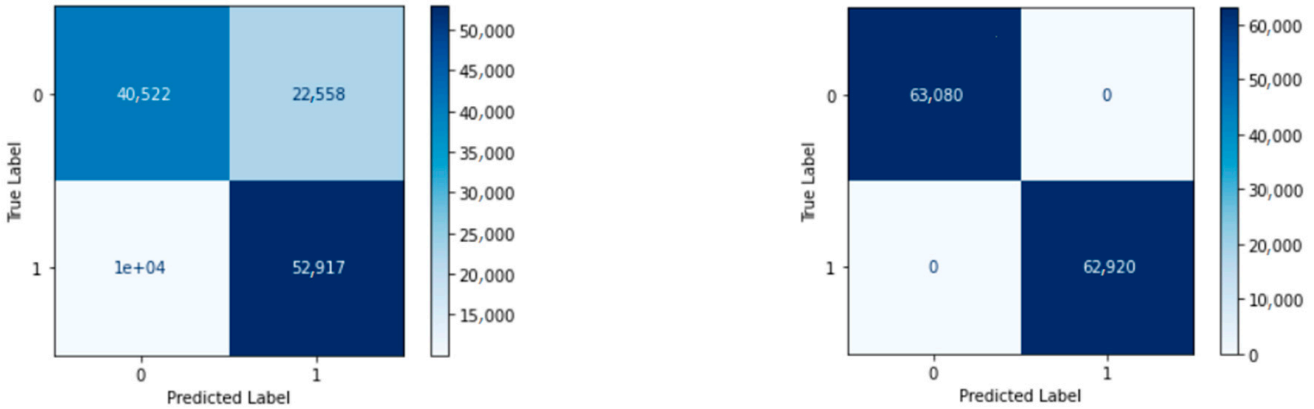


Figure 11. Influential features of detection attack on TON-IOT dataset.

We first applied mutual information to select the 15 best features, as depicted in Figure 5. These features, including ts, src_ip, src_port, dst_ip, dst_port, service, duration, src_bytes, dst_bytes, conn_state, src_pkts, src_ip_bytes, and dst_pkts, were used in subsequent analyses.

Furthermore, Figures 9 and 10 demonstrate the significant impact of employing the SMOTE technique to address class imbalance in the dataset. The improvement in class balance is evident, indicating the effectiveness of SMOTE in mitigating the issue.

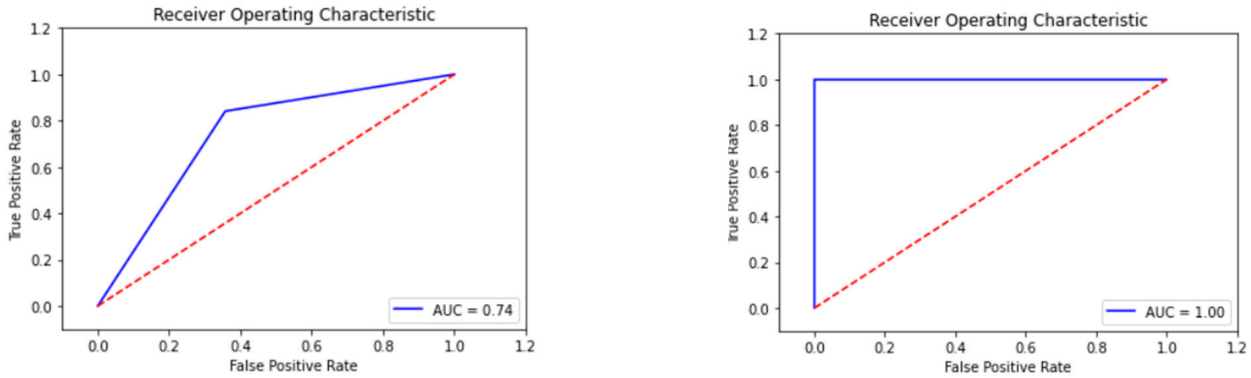
For evaluating the performance of our RF classifier, we compared it with LR, DT, KNN, and SVM techniques using the TON-IOT dataset. To provide a more comprehensive analysis, we examined the confusion matrices in Figure 12. Notably, our model exhibited outstanding performance with 0 FPR and FNR.



(**a**) Logistic Regression (**b**) Random Forest

Figure 12. Confusion matrix of prediction using TON-IOT dataset.

Additionally, Figures 13 and 14 illustrates the ROC curve for all the applied techniques on the dataset. Our model achieved excellent results, attaining 100% AUC, which signifies its robustness in distinguishing between normal and intrusive instances.



(**a**) Logistic Regression (**b**) Random Forest

Figure 13. ROC curve on the 15 best features using TON-IOT Dataset.

To provide a comprehensive assessment, we present the results in Table 4, which showcases the performance metrics of accuracy, recall, precision, and F1 score. Our model outperformed all other techniques, attaining 100% for each of these metrics. The second-best technique, decision tree (DT), also achieved 100% in all metrics.

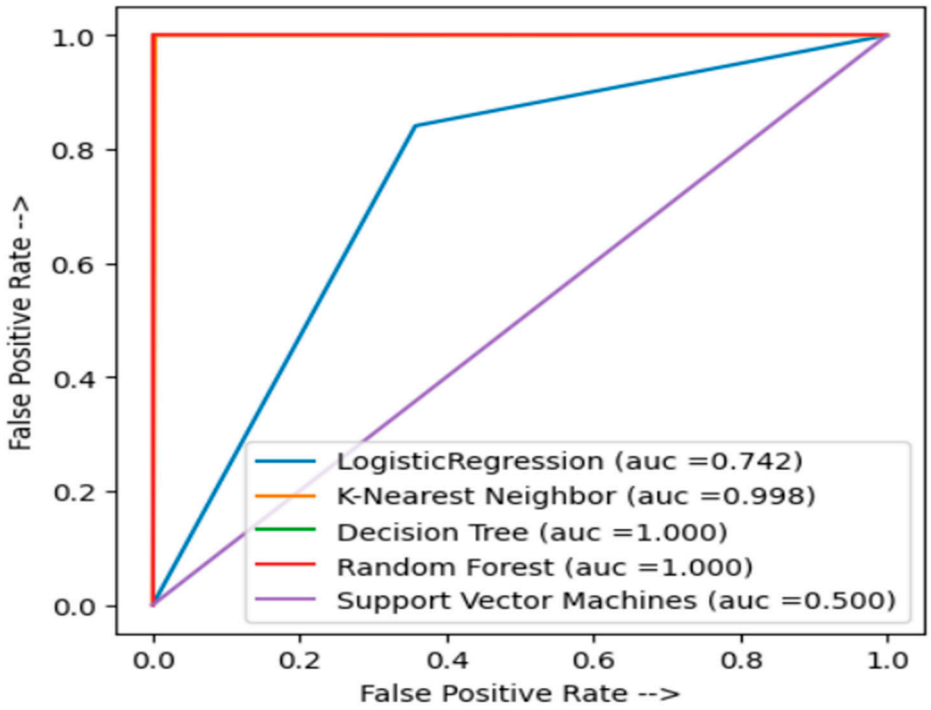


Figure 14. ROC curve of all techniques by considering 15 best features.

Table 4. Experimentation result on TON-IOT dataset.

Classifier	Accuracy	Precision	Recall	F1-Score
LR	73.9	75.1	74.1	73.9
KNN	99.8	99.8	99.8	99.8
DT	100	100	100	100
SVM	50.0	24.9	49.9	66.6
Proposed Model	100	100	100	100

The SMOTE technique is also used on CICIDS2017 dataset to face the problem of class imbalance. Figures 15 and 16 show the percentages before and after the use of SMOTE.

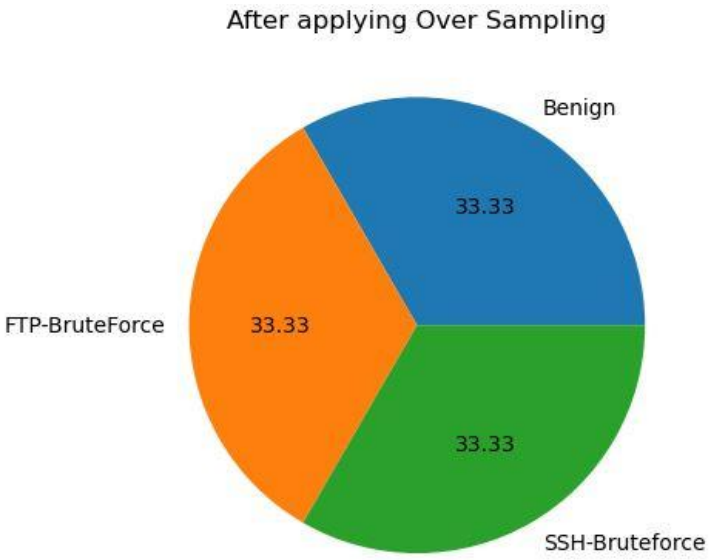


Figure 15. Attack class before applying the over sampling technique using CICIDS2017 dataset.

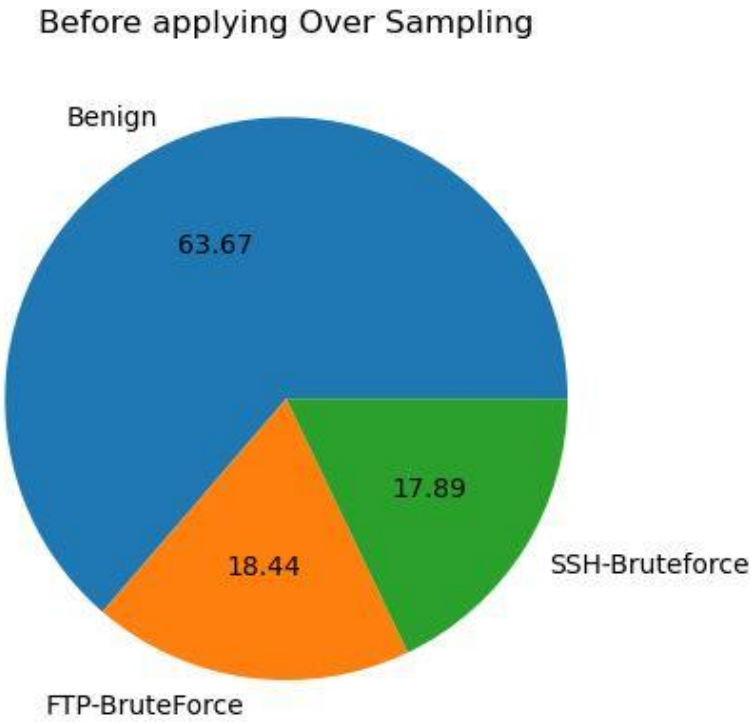


Figure 16. Attack class after applying the over sampling technique using CICIDS2017 dataset.

Figures 15 and 16 display the partitions of the attack class before and after applying the SMOTE technique. As depicted in Figure 16, the problem of class imbalance is effectively resolved, and the partitions are more balanced.

Figure 17 showcases the influential features on our dataset, specifically the CICIDS2017 dataset. These features play a significant role in the performance of our model.

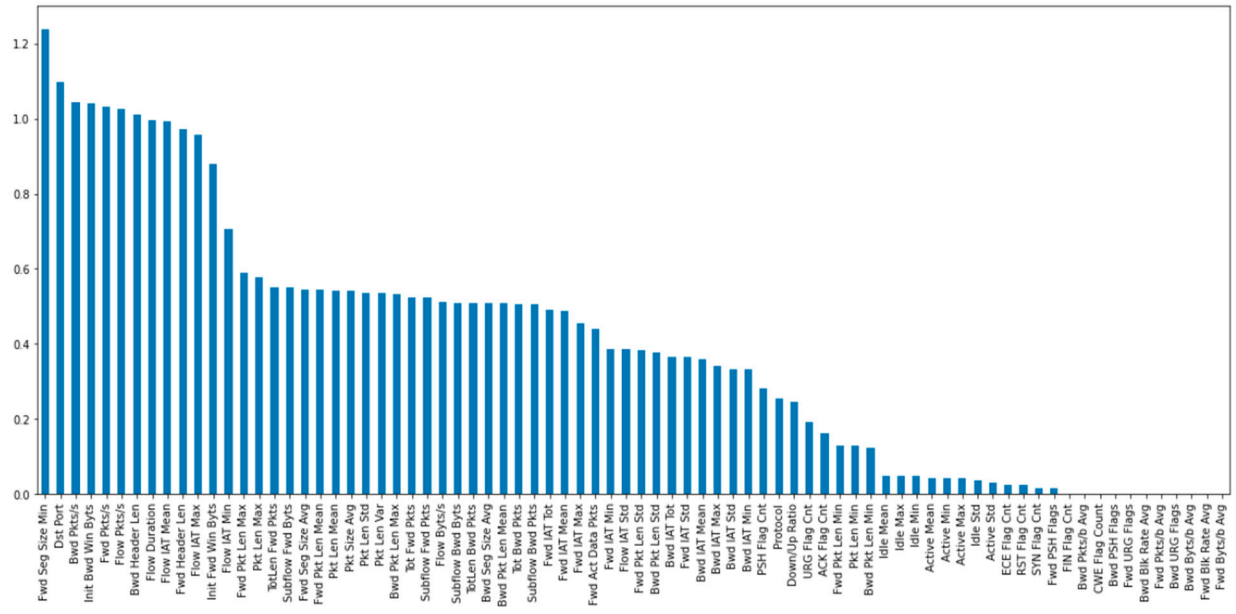
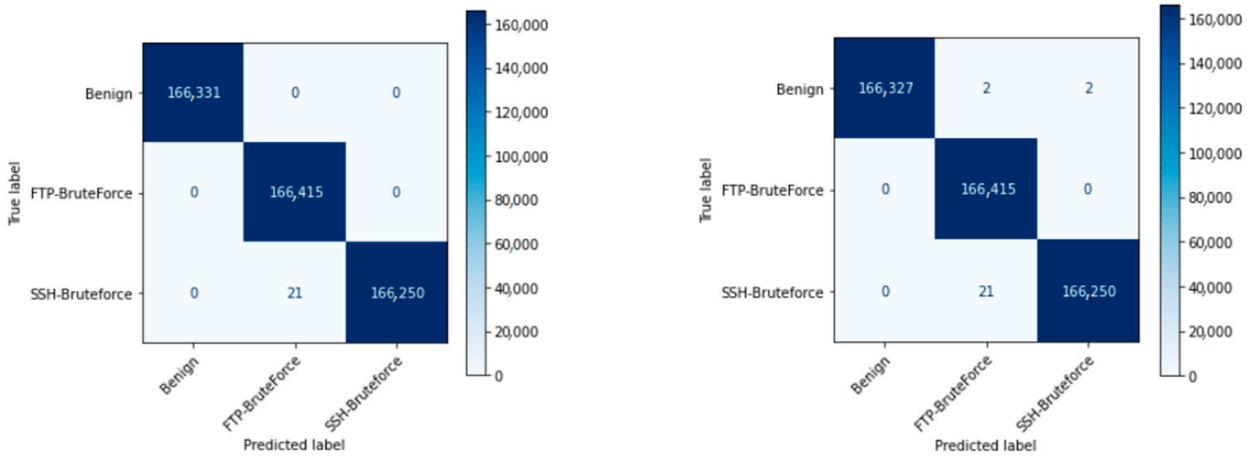


Figure 17. Influential features of detection attack on CICIDS2017 dataset.

Furthermore, the confusion matrices presented in Figure 18 demonstrate the correct implementation of our model, as indicated by the absence of FPR and FNR.

In terms of performance evaluation, Table 5 presents the results obtained by the RF and SVM techniques. The RF technique yields satisfactory outcomes, with an accuracy, recall, precision, and F1 score of 0.999. On the other hand, the SVM technique exhibits

lower performance, with an accuracy of 0.180, recall of 0.340, precision of 0.160, and F1 score of 0.340. The results obtained allow us to conclude that the proposed framework is relevant and achieves significant performance.



(**a**) Logistic Regression (**b**) Random Forest

Figure 18. Confusion matrix of prediction for CICIDS2017 dataset.

Table 5. Experimentation result using CICIDS2017 dataset.

Classifier	Accuracy	Precision	Recall	F1-Score
LR	99.9	99.9	99.9	99.9
KNN	99.9	99.9	99.9	99.9
DT	99.9	99.9	99.9	99.9
SVM	18.0	16.0	34.0	34.0
Proposed Model	99.9	99.9	99.9	99.9

For the TON-IOT dataset, the performance evaluation results indicate that our RF classifier achieves exceptional results. The model exhibits 0 FPR and FNR as shown in the confusion matrices (Figure 12). The ROC curve (Figure 14) demonstrates that our model achieves 100% AUC, indicating its strong discriminatory power. Moreover, the results presented in Table 4 indicate that our RF classifier achieves the highest intrusion detection performance, with 100% accuracy, recall, precision, and F1 score. Moving on to the NSL-KDD dataset, the performance comparison reveals that our RF classifier continues to excel. Figure 3 shows the class imbalance issue in the original dataset, which is effectively addressed by applying the SMOTE technique (Figure 4). The confusion matrices (Figure 6) demonstrate the satisfactory performance of our model, with a low number of FPR and FNR. The ROC curve (Figure 8) exhibits good results, particularly with 99% accuracy for the KNN technique. Table 3 presents the evaluation metrics, indicating that our RF classifier achieves high accuracy (0.99), recall (0.97), precision (0.97), and F1 score (0.97) for intrusion detection, surpassing the performance of other methods such as DT and KNN. Regarding the CICIDS2017 dataset, our RF classifier once again demonstrates superior performance. Figures 15 and 16 illustrate the difference between the partitions of the attack class before and after applying the SMOTE technique, effectively resolving the class imbalance issue. The confusion matrices (Figure 18) highlight the correct implementation of our model, with 0 FPR and 0 FNR. In terms of evaluation metrics (Table 5), the RF technique achieves outstanding results, with 0.999 accuracy, recall, precision, and F1 score. In contrast, the SVM technique lags behind, with significantly lower values for these metrics. Overall, the comparison of results across the TON-IOT, NSL-KDD, and CICIDS2017 datasets consistently demonstrates the superiority of our model based on the RF classifier. It consistently achieves higher performance in terms of accuracy, recall, precision, and

F1 score, outperforming other techniques such as SVM, DT, and KNN. These findings highlight the robustness and effectiveness of our proposed model in intrusion detection tasks across different datasets.

5. Conclusions

VANETs play a crucial role in ensuring the safety, security, and efficiency of transportation systems. However, the interconnected and dynamic nature of VANETs presents significant challenges and risks that require rigorous assessment and mitigation. In this study, we addressed the problem of intrusion detection in VANETs by proposing a model based on mutual information for feature selection and utilizing the SMOTE technique to handle class imbalance. Our experimental evaluations on multiple datasets demonstrated the effectiveness of the proposed model in enhancing the security of VANETs. The model achieved commendable performance in terms of accuracy, precision, and F1 score, highlighting its potential for intrusion detection in VANETs. Additionally, we compared our model with other machine learning techniques, further validating its superiority. However, it is important to acknowledge certain limitations. The fast mobility of nodes in VANETs and the dynamic changes in network typology pose challenges for intrusion detection. The rapid movement of nodes and dynamic network conditions require adaptive and dynamic intrusion detection mechanisms to ensure accurate and up-to-date threat detection. In conclusion, our study contributes to the field of VANET security by proposing an effective intrusion detection model. The use of mutual information and the SMOTE technique demonstrated promising results. Looking ahead, future research should focus on developing adaptive intrusion detection approaches that can address the challenges posed by the fast mobility of nodes and the dynamic network typology in VANETs. By doing so, we can enhance the security and resilience of VANETs, ensuring safer and more reliable transportation systems in the future.

Author Contributions: Conceptualization, S.A. and A.G.; methodology M.M.N., H.F. and A.G.; software, M.A.; validation, S.B., S.B.A.K. and A.G.; formal analysis, M.M.N.; investigation, M.A.; resources, A.G.; data curation, S.A.; writing—original draft preparation, S.A.; writing—review and editing, A.G., S.B., M.A. and M.M.N. visualization, S.B.A.K. and H.F.; supervision, S.B.; project administration, S.B.; funding acquisition, M.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was not funded and did not receive financial support. We did this research work as academic researchers of computer science at the university.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Available on request.

Acknowledgments: The authors would like to acknowledge Cadi Ayyad University, Moulay Ismail University of Meknes, Prince Sultan University (PSU) and Smart Systems Engineering lab for their valuable support. Additionally, the authors would like to acknowledge the support of PSU for paying the Article Processing Charges (APC) of this publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yousefi, S.; Mousavi, M.S.; Fathy, M. Vehicular ad hoc networks (VANETs): Challenges and perspectives. In Proceedings of the 2006 6th International Conference on ITS Telecommunications, Chengdu, China, 21–23 June 2006; IEEE: Piscateville, NJ, USA, 2006.
2. Biswas, S.; Mišić, J.; Mišić, V. DDoS attack on WAVE-enabled VANET through synchronization. In Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA, 3–7 December 2012; IEEE: Piscateville, NJ, USA, 2012.

3. Zeng, Y.; Qiu, M.; Zhu, D.; Xue, Z.; Xiong, J.; Liu, M. A Deepvcm: A deep learning based intrusion detection method in vanet. In Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019; pp. 288–293.
4. Zaidi, T.; Faisal, S. An overview: Various attacks in VANET. In Proceedings of the 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 14–15 December 2018; IEEE: Piscateville, NJ, USA, 2018.
5. Zeadally, S.; Hunt, R.; Chen, Y.S.; Irwin, A.; Hassan, A. Vehicular ad hoc networks (VANETS): Status, results, and challenges. *Telecommun. Syst.* **2012**, *50*, 217–241. [CrossRef]
6. Choudhury, N.; Nasralla, M.M. A Proposed Resource-Aware Time-Constrained Scheduling Mechanism for DSME based IoV Networks. In Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 27–30 September 2021; pp. 1–7. [CrossRef]
7. Shah, Y.A.; Aadil, F.; Khalil, A.; Assam, M.; Abunadi, I.; Alluhaidan, A.S.; Al-Wesabi, F.N. An Evolutionary Algorithm-Based Vehicular Clustering Technique for VANETs. *IEEE Access* **2022**, *10*, 14368–14385. [CrossRef]
8. Khan, M.A.; Nasralla, M.M.; Umar, M.M.; Iqbal, Z.; Rehman, G.U.; Sarfraz, M.S.; Choudhury, N. A Survey on the Noncooperative Environment in Smart Nodes-Based Ad Hoc Networks: Motivations and Solutions. *Secur. Commun. Netw.* **2021**, *2021*, 9921826. [CrossRef]
9. Ku, I.; Lu, Y.; Gerla, M.; Gomes, R.L.; Ongaro, F.; Cerqueira, E. Towards software-defined VANET: Architecture and services. In Proceedings of the 2014 13th Annual Mediterranean ad hoc Networking Workshop (MED-HOC-NET), Piran, Slovenia, 2–4 June 2014; IEEE: Piscateville, NJ, USA, 2014; pp. 103–110.
10. Tang, Y.; Cheng, N.; Wu, W.; Wang, M.; Dai, Y.; Shen, X. Delay-Minimization Routing for Heterogeneous VANETs with Machine Learning Based Mobility Prediction. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3967–3979. [CrossRef]
11. Kaiwartya, O.; Abdullah, A.H.; Cao, Y.; Altameem, A.; Prasad, M.; Lin, C.T.; Liu, X. Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects. *IEEE Access* **2016**, *4*, 5356–5373. [CrossRef]
12. Azrour, M.; Mabrouki, J.; Guezzaz, A.; Kanwal, A. Internet of Things Security: Challenges and Key Issues. *Secur. Commun. Netw.* **2021**, *2021*, 5533843. [CrossRef]
13. Guezzaz, A.; Asimi, A.; Asimi, Y.; Azrour, M.; Benkirane, S. A Distributed Intrusion Detection Approach Based on Machine Leaning Techniques for a Cloud Security. In *Intelligent Systems in Big Data, Semantic Web and Machine Learning*; Springer: Cham, Switzerland, 2021; Volume 1344, pp. 85–94.
14. Nasralla, M.M.; Garcia-Magarino, I.; Lloret, J. Defenses against Perception-Layer Attacks on IoT Smart Furniture for Impaired People. *IEEE Access* **2020**, *8*, 119795–119805. [CrossRef]
15. Nirmala, R.; Sudha, R. A Relativity Cram between MANET and VANET Background along Routing Protocols. *Int. J. Adv. Inf. Sci. Technol.* **2014**, *26*, 153–157.
16. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
17. Guezzaz, A.; Asimi, A.; Asimi, Y.; Tbatou, Z.; Sadqi, Y. A Lightweight Neural Classifier for Intrusion Detection. *Gen. Lett. Math.* **2017**, *2*, 57–66. [CrossRef]
18. Alheeti, K.M.A.; Gruebler, A.; McDonald-Maier, K. Intelligent Intrusion Detection of Grey Hole and Rushing Attacks in Self-Driving Vehicular Networks. *Computers* **2016**, *5*, 16. [CrossRef]
19. Khan, A.S.; Balan, K.; Javed, Y.; Tarmizi, S.; Abdullah, J. Secure Trust-Based Blockchain Architecture to Prevent Attacks in VANET. *Sensors* **2019**, *19*, 4954. [CrossRef] [PubMed]
20. Guezzaz, A.; Benkirane, S.; Azrour, M.; Khurram, S. A Reliable Network Intrusion Detection Approach Using Decision Tree with Enhanced Data Quality. *Secur. Commun. Netw.* **2021**, *2021*, 1–8. [CrossRef]
21. Guezzaz, A.; Azrour, M.; Benkirane, S.; Mohy-Eddine, M.; Attou, H.; Douiba, M. A Lightweight Hybrid Intrusion Detection Framework using Machine Learning for Edge-Based IIoT Security. *Int. Arab. J. Inf. Technol.* **2022**, *19*, 5. [CrossRef]
22. Jones, A.K.; Sielken, R.S. Computer system intrusion detection: A survey. *Comput. Sci. Tech. Rep.* **2000**, 1–25.
23. Ioannou, C.; Vassiliou, V. An Intrusion Detection System for Constrained WSN and IoT Nodes Based on Binary Logistic Regression. In Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Suzhou, China, 25 October 2018; pp. 259–263.
24. Rai, K.; Guleria, A.; Devi, M.S. Decision Tree Based Algorithm for Intrusion Detection. *Int. J. Adv. Netw. Appl.* **2016**, *7*, 2828.
25. Almseidin, M.; Alzubi, M.; Kovacs, S.; Alkasassbeh, M. Evaluation of machine learning algorithms for intrusion detection system. In Proceedings of the 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 14–16 September 2017.
26. Mahesh, B. Machine learning algorithms–A review. *Int. J. Sci. Res.* **2020**, *9*, 381–386.
27. Amiri, F.; Yousefi, M.R.; Lucas, C.; Shakery, A.; Yazdani, N. Mutual information-based feature selection for intrusion detection systems. *J. Netw. Comput. Appl.* **2011**, *34*, 1184–1199. [CrossRef]
28. Guezzaz, S.; Azrour, B.M. A Novel Anomaly Network Intrusion Detection System for Internet of Things Security. In *IoT and Smart Devices for Sustainable Environment. EAI/Springer Innovations in Communication and Computing*; Springer: Berlin/Heidelberg, Germany, 2022.

29. Al-Jarrah, O.Y.; Siddiqui, A.; Elsalamouny, M.; Yoo, P.D.; Muhaidat, S.; Kim, K. Machine-learning-based feature selection techniques for large-scale network intrusion detection. In Proceedings of the 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW), Madrid, Spain, 30 June–3 July 2014; IEEE: Piscateville, NJ, USA, 2014; pp. 177–181.
30. Gad, A.R.; Nashat, A.A.; Barkat, T.M. Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset. *IEEE Access* **2021**, *9*, 142206–142217. [CrossRef]
31. Zeng, Y.; Qiu, M.; Ming, Z.; Liu, M. Senior2local: A machine learning based intrusion detection method for vanets. In Proceedings of the International Conference on Smart Computing and Communication, Tokyo, Japan, 10–12 December 2018; Springer: Cham, Switzerland, 2018; pp. 417–426.
32. Wahab, O.A.; Mourad, A.; Otrok, H.; Bentahar, J. CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks. *Expert Syst. Appl.* **2016**, *50*, 40–54. [CrossRef]
33. Bangui, H.; Ge, M.; Buhnova, B. A hybrid data-driven model for intrusion detection in VANET. *Procedia Comput. Sci.* **2021**, *184*, 516–523. [CrossRef]
34. Benkirane, S.; Guezzaz, A.; Azrour, M.; Gardezi, A.A.; Ahmad, S.; Sayed, A.E.; Naseer, S.; Shafiq, M. Adapted Speed System in a Road Bend Situation in VANET Environment. *Comput. Mater. Contin.* **2022**, *74*, 3781–3794. [CrossRef]
35. Grover, J.; Prajapati, N.K.; Laxmi, V.; Gaur, M.S. Machine learning approach for multiple misbehavior detection in VANET. In *International Conference on Advances in Computing and Communications*; Springer: Berlin/Heidelberg, Germany; pp. 644–653.
36. Kandali, K.; Bennis, L.; El Bannay, O.; Bennis, H. An Intelligent Machine Learning Based Routing Scheme for VANET. *IEEE Access* **2022**, *10*, 74318–74333. [CrossRef]
37. Kandali, K.; Bennis, L.; Bennis, H. A New Hybrid Routing Protocol Using a Modified K-Means Clustering Algorithm and Continuous Hopfield Network for VANET. *IEEE Access* **2021**, *9*, 47169–47183. [CrossRef]
38. Kosmanos, D.; Pappas, A.; Maglaras, L.; Moschoyiannis, S.; Aparicio-Navarro, F.J.; Argyriou, A.; Janicke, H. A novel Intrusion Detection System against spoofing attacks in connected Electric Vehicles. *Array* **2020**, *5*, 100013. [CrossRef]
39. Hassan, H.A.; Hemdan, E.E.; El-Shafai, W.; Shokair, M.; El-Samie, F.E.A. Intrusion Detection Systems for the Internet of Thing: A Survey Study. *Wirel. Pers. Commun.* **2022**, *128*, 2753–2778. [CrossRef]
40. Abro, G.E.M.; Zulkifli, S.A.B.M.; Kumar, K.; El Ouanjli, N.; Asirvadam, V.S.; Mossa, M.A. Comprehensive Review of Recent Advancements in Battery Technology, Propulsion, Power Interfaces, and Vehicle Network Systems for Intelligent Autonomous and Connected Electric Vehicles. *Energies* **2023**, *16*, 2925. [CrossRef]
41. Javed, A.R.; Jalil, Z.; Moqurrab, S.A.; Abbas, S.; Liu, X. Ensemble Adaboost classifier for accurate and fast detection of botnet attacks in connected vehicles. *Trans. Emerg. Telecommun. Technol.* **2020**, *33*, e4088.
42. Alhaidari, A.F.; Alrehan, A.M. A simulation work for generating a novel dataset to detect distributed denial of service attacks on Vehicular Ad hoc NETwork systems. *Int. J. Distrib. Sens. Netw.* **2021**, *17*, 15501477211000287. [CrossRef]
43. Guezzaz, A.; Asimi, A.; Azrour, M.; Batou, Z.; Asimi, Y. A Multilayer Perceptron Classifier for Monitoring Network Traffic. In *Big Data and Networks Technologies*; Springer: Berlin/Heidelberg, Germany, 2020.
44. NSL-KDD Dataset. Available online: https://web.archive.org/web/20150205070216/http://nsl.cs.unb.ca/NSL-KDD/ (accessed on 11 January 2023).
45. CICIDS2017 Dataset. Available online: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 11 January 2023).
46. Moustafa, N. ToN-IoT Dataset. Available online: https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i (accessed on 11 January 2023).