Optimization of RBF-SVM Kernel Using Grid Search Algorithm for DDoS Attack Detection in SDN-Based VANET

Goodness Oluchi Anyanwu, Cosmas Ifeanyi Nwakanma, *Member, IEEE*, Jae-Min Lee, *Member, IEEE*, and Dong-Seong Kim, *Senior Member, IEEE*

***Abstract*—The dynamic nature of the vehicular space exposes it to distributed malicious attacks irrespective of the integration of enabling technologies. The software-defined network (SDN) represents one of these enabling technologies, providing an integrated improvement over the traditional vehicular ad-hoc network (VANET). Due to the centralized characteristics of SDN, they are vulnerable to attacks that may result in life-threatening situations. Securing SDN-based VANETs is vital and requires incorporating artificial intelligence (AI) techniques. Hence, this work proposed an intrusion detection model (IDM) to identify Distributed Denial-of-Service (DDoS) attacks in the vehicular space. The proposed solution employs the radial basis function (RBF) kernel of the support vector machine (SVM) classifier and an exhaustive parameter search technique called grid search cross-validation (GSCV). In this framework, the proposed architecture can be deployed on the onboard units (OBUs) of each vehicle, which receive the vehicular data and run intrusion detection tasks to classify a message sequence as a DDoS attack or benign. The performance of the proposed algorithm compared to other ML algorithms using key performance metrics. The proposed framework is validated through experimental simulations to demonstrate its effectiveness in detecting DDoS intrusion. Using the GridSearchCV, optimal values of the RBF-SVM kernel parameters "C" and "gamma" (γ) of 100 and 0.1, respectively, gave the optimal performance. The proposed scheme showed an overall accuracy of 99.33%, a detection rate of 99.22%, and an average squared error of 0.007, outperforming existing benchmarks.**

***Index Terms*—Distributed Denial-of-Service (DDoS) attack, grid search cross-validation (GSCV), hyperparameter optimization, radial basis function (RBF) kernel, software-defined network (SDN)-based vehicular ad-hoc network (VANET), support vector machine (SVM).**

Manuscript received 25 March 2022; revised 13 July 2022; accepted 15 August 2022. Date of publication 18 August 2022; date of current version 9 May 2023. This work was supported in part by the Priority Research Centers Program through NRF funded by MEST under Grant 2018R1A6A1A03024003, and in part by the Grand Information Technology Research Center Support Program under Grant IITP-2022-2020-0-01612 supervised by IITP by MSIT, South Korea. *(Corresponding author: Dong-Seong Kim.)*

Goodness Oluchi Anyanwu, Jae-Min Lee, and Dong-Seong Kim are with the Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea (e-mail: anyanwu.goodnes@kumoh.ac.kr; ljmpaul@kumoh.ac.kr; dskim@kumoh.ac.kr).

Cosmas Ifeanyi Nwakanma is with the ICT Convergence Research Center, Kumoh National Institute of Technology, Gumi 39177, South Korea (e-mail: cosmas.ifeanyi@kumoh.ac.kr).

Digital Object Identifier 10.1109/JIOT.2022.3199712

I. Introduction

THERE is an unprecedented growth of connected and intelligent vehicles and the accompanied demand to guarantee road safety and efficient management as part of a global transportation system upgrade [1]. The gradual shift toward 6G implies a movement to machine-type communications, a platform that would accommodate connection to enormous numbers of devices. These devices will sporadically transmit large amounts of data, enabling cities and infrastructure to communicate in real time [2]. Axiomatically, exposure to threats and vulnerabilities is inevitable. Vehicular ad-hoc networks (VANETs) are the class of ad-hoc networks utilized in intelligent transportation systems. The open-wireless-medium design used in VANET, and its changing architecture and topology, expose this system to several attacks [3]. Given this, secured transportation is a prerequisite for a sustainable smart city to ensure uninterrupted interoperability as fully automated smart vehicles are fast becoming a reality [4]. Acronyms used in this article are listed in Table I.

VANET is accessed by a wide range of devices, incorporating varieties of emerging technologies [5]. Edge computing, cloud computing, and software-defined networking (SDN) are ranges of intelligent technologies incorporated into VANET for improved performance [6]. SDN is an improvement over the traditional VANET and a significant enabler for 5G network deployment, regulating the overall network in a systematic way [7]. SDN, therefore, provides centralized control and innovative scalability for VANET. However, security threats and vulnerabilities are introduced when new entities and architectural components are deployed and integrated into any system [8]. Therefore, SDN-based VANETs are vulnerable to several forms of malicious interference. With the single point of control introduced by the SDN architecture, VANET becomes an easy target for attackers. Furthermore, considering the dynamic properties and architecture of VANET, it becomes difficult to identify and trace these attacks based on frequent handovers and short-lived links.

Conventionally, attacks are aimed at disrupting high-speed networks, flooding the primary control medium with a vast volume of malicious data [9]. In SDN-Based VANETs, attackers attempt to disable the SDN architecture by targeting the SDN controller, resulting in inaccessible services and decreased performance resulting in a Distributed Denial of Service (DDoS) [10]. A DDoS attack on SDN-VANET can

TABLE I
ACRONYMS USED IN THIS ARTICLE AND THEIR MEANING

Acronyms	Meaning
ANOVA	Analysis of Variance
AUC	Area Under Curve
CKS	Cohen Kappa Score
CM	Confusion Matrix
CSV	Comma-separated Values
CV	Cross Validation
DDoS	Distributed Denial of Service
DL	Deep Learning
DPV	Default Parameter Value
GNB	Gaussian Naive Bayes
FN	False Negatives
FP	False Positives
GSCV	Grid Search Cross Validation
IDM	Intrusion Detection Model(s)
IoT	Internet of Things
kNN	k-Nearest Neighbor
KPCA	Kernel Principal Components Analysis
LR	Logistic Regression
MSE	Mean Squared Error
ML	Machine Learning
NCP	Network Control Plane
OBU	On-Board Unit
OPV	Optimized Parameter Value
PC	Principal Component
PCA	Principal Component Analysis
RBF	Radial Basis Function
ROC	Receiver Operating Characteristics
RF	Random Forest
RSUC	Roadside Unit Controller
SDN	Software Defined Network
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VANET	Vehicular Ad-hoc Network

occur due to a compromised controller, with a high impact on a network, affecting its availability. A high volume of illegitimate connection requests from multiple sources may confuse and create problems for legitimate users. In VANET, DDoS attacks are the most difficult-to-handle and consequential type of attack [11]. The most prevalent DDoS attacks occur at various levels, taking advantage of weaknesses in the network connectivity, eliminating chances of information exchange, and rendering the network unresponsive [12]. A basic DDoS attack overwhelms the resources of legitimate nodes, while an extended attack occurs when the whole channel is jammed.

A DDoS attack on the SDN controller targets its process and communication capacities, simultaneously overwhelming the overall network and disrupting the mainstream traffic of the server. Hence, securing VANETs from DDoS intrusions is a challenging task [13]. Since these networks operate in real time, the application of encryption and authentication intrusion detection methods (IDMs) involving general mitigation techniques is unreliable. These forms of IDM only focus on the data's payload, thereby reducing the dimension of the supposed data [14]. However, artificial intelligence (AI) and machine learning (ML) models have been adopted to detect and classify various forms of distributed malicious intent on VANET and its technologies, as these methods are less affected by encryption, thereby preserving the vehicular space. In addition, these approaches aim for high reliability and in-depth analysis of the input data [15], [16].

A vast volume of research has emphasized the importance of securing vehicular communication, presenting IDMs that use either ML or deep learning (DL) methodologies. Random forest (RF), *k*-nearest neighbor (kNN), and support vector machine (SVM) are the most implemented supervised ML approaches employed as these methods guarantee reliable outcomes [17]. However, SVM remains one of the ubiquitous approaches that have found relevance in detecting intrusion for VANET, achieving reliable performances [18]. In this work, for reliable DDoS detection accuracy and a robust learning mechanism, the radial basis function (RBF) kernel of SVM is assisted by an exhaustive grid search cross-validation (GSCV) for optimizing SVM parameters. The RBF kernel is associated with its suitability for nonlinear data. The GSCV, on the other hand, implements an exhaustive exploration over a defined parameter grid to reach cross-validated and optimized parameter values [19]. Research gap from previous literature lies in developing models with optimal points to achieve the tradeoff of computing time, accuracy, and reduction in computational complexity. To the best of our knowledge, this is a pioneering attempt to experimentally show the impact and evaluation of parameterization and optimization of SVM using a grid search algorithm for detecting DDoS attacks in SDN-VANETs.

Consequently, this work ascertained an optimized RBF-SVM model for DDoS detection in VANETs. The proposed IDM was compared with state-of-the-art ML models with similar modeling and theoretical background. These models include kNN, Gaussian naive Bayes (GNB), and logistic regression (LR). Furthermore, a comprehensive and comparative investigation into different kernels, parameter adjustments, and data set properties is provided to appreciate the proposed model. The experimental results obtained using the publicly available DDoS evaluation data set for SDN architectures from IEEE dataport [20] show that the proposed RBF-SVM outperformed the default SVM, other SVM kernel variants, and other random parameter selections in terms of classification accuracy and with better generalization. In addition, the proposed model beats the accuracy of previous works and can be embedded within an SDN-Based VANET to define security rules for preventing DDoS attacks within reasonable execution time. This proposed solution can protect VANET from these types of attacks. The novel contributions of this work are as follows.

1) To propose a novel extension of the regular RBF-SVM kernel-based detection approach which can be defined on each vehicle's onboard unit (OBU) to identify DDoS attacks in an SDN-Based VANET. The proposed RBF-SVM kernel structure is based on parameter tuning of the basic RBF-SVM kernel to achieve an optimal selection toward enhancing the detection rate and accuracy of identifying DDoS attacks.
2) To achieve an optimal objective function, parameter selection, and accurate classification hyperplane, an exhaustive GSCV optimization technique is employed over a fivefold cross-validation (CV).
3) Evaluation using two publicly available DDoS data sets for verification of the experimental approach to access the performance of the proposed model under changing conditions.

4) A comprehensive and comparative investigation into other SVM kernel variants, default and random parameter selections, selected state-of-the-art ML algorithms, and data set properties for better performance analysis and evaluation.

The remainder of this article is structured as follows: Section II reviews current VANET DDoS attack detection and mitigation approaches, while our proposed research methodology and details of our implementation process are presented in Section III. The evaluation metrics are outlined in Section IV. Section V describes obtained results and effectiveness of our proposed approach in detecting and mitigating DDoS attacks over other approaches. Finally, Section VI concludes this article.

II. Related Works

Recently, researchers are exploring the use of ML frameworks for providing and addressing the security challenges of high-mobility VANETs, specifically SDN-based and 5G-based VANETS. For instance, to detect DDoS attacks in an SDN-based network, the ML approach equipped with a neighborhood component analysis (NCA) using SVM as the base classifier was developed and modeled in [21]. Different models were also implemented and compared with the SVM approach, which achieved an accuracy of 97.20% over ten-fold CV. Adhikary *et al.* [22] employed a hybrid algorithm for attack detection, yielding an appreciable result of approximately 98% accuracy. The algorithm adopted by these authors is a combination of two SVM kernel Dot methods (ANOVA and RBF). Similarly, Kaushik *et al.* [23] and Bangui *et al.* [24] implemented an ML-DL hybrid model for mitigating DDoS threats on VANET. The hybrid model in [23], a combination of decision tree and neural network, achieved an accuracy of 96.40% compared to a single implementation of the algorithms at 41.77% and 76.81%, respectively.

Although the hybrid accuracy achieved by these authors is high and may be considered a near-reliable requirement for VANET, their approaches result in increased computational complexity. In addition, there is room for improvement in the accuracy achieved by these works as VANET represents a mission-critical system. Motivated by this drawback, the focus in [25] was on low-rate DDoS attacks in a flexible SDN-based architecture evaluated on an open network operating system controller running on a Mininet virtual machine. The low-rate DDoS attack triggers a timeout on specific protocol control mechanisms. The detection rate of 95% was achieved using a simple SVM model as the base model. Correspondingly, RBF-SVM was implemented to characterize DDoS attacks, achieving accuracies of 94.07% and 97.3% in [26] and [27], respectively. Deka *et al.* [28] achieved accuracies between 92% and 97% by implementing SVM on selected data sets to evaluate DDoS attacks on VANET.

To explore the capability of SVM models with optimization algorithms, Alsarhan *et al.* [29] achieved a detection rate of close to 99% by employing a genetic algorithm (GA) for optimizing SVM parameters. The authors' goal was the accurate classification of benign and attack events. Also, they added a penalty factor to the objective function to control the complexity of the IDM. This factor represents an automated and reliable method for selecting values optimal for the SVM-based IDM. However, training their proposed model on a small subset of network scenarios cast doubt on the robustness of the model. Similarly, a combination of GA and kernel principal component analysis (KPCA) was employed to optimize RBF kernel parameters and the tube size of SVM in [30]. The experimental results obtained by the RBF kernel achieved high-predictive accuracy, faster convergence speed, and better generalization compared with other detection algorithms and SVM kernels modeled.

A simple-structured SVM-based IDM was designed for an ad-hoc network in [31] to detect DoS attacks. The authors attempted to demonstrate the effectiveness of the suggested approach. However, as the number of attackers increased, the IDM experienced more disruption and poor detection ability. The one-class SVM investigated in [32] yielded accurate results despite using only default hyperparameter values for the configuration. As part of the Smart City concept, a hybrid DL approach for detecting DDoS in [33] achieved high accuracy of 98.37%. Although the results demonstrated improved performance of the proposed model over other ML and DL models from the literature, a real-time critical system requires over 99% attack detection performance. Shams *et al.* [34] only presented a comparative analysis and performance evaluation of well-known ML-based anomaly detection models. The authors' result justified the SVM as the best-performing model for classifying the presence or absence of DoS attackers with high precision and recall, averaging about 99%. A trust-aware SVM-based model developed for preventing intrusion in VANETs achieved a precision result of 93.89% in [35].

Similarly, [36] achieved an accuracy of 98.07% by proposing a multilayered SVM and an improved kernel function to reduce the training time of the IDM. Conclusively, from the above works, the design of a comprehensive, reliable, and well-analyzed security framework is required for secure implementation of VANET and its corresponding technologies. Hence, improving the reliability of its information exchange is of immense importance, as this network is dynamic and must respond in real time. From the research work outlined above, the pervasiveness of SVM is also not in doubt. For effective extension and in-depth analysis of data patterns, SVM was selected and considered most appropriate for this study. SVM deals with high-dimensional data while providing reliable accuracy [18]. In addition, SVM possesses a solid theoretical basis with evidence of performing incomparably well in a variety of practical and real-world learning tasks. It works by implementing kernel tricks that map data points into new spaces and establish appropriate decision boundaries [37].

Although previous research efforts have shown tremendous improvements in developing IDM models for VANETs and their integrated technologies, the standard SVM, kernels, and its variants still have some limitations as performance is dependent on the selection of optimal parameters used in the training process. A basic SVM model may produce a large number of support vectors, resulting in higher computational complexity and training time. The two steps taken to improve the

performance of the SVM models include accurate IDM setup and optimal parameters selection. One of the focuses of this work is the relevance of choosing the most suitable kernel and performance parameters. In addition, some of the research works presented above lack detailed analysis of the effects of kernel type, parameter optimization, and the response of the model to data set properties and preprocessing decisions.

Based on these findings, in conjunction with the GSCV, this work utilizes the SVM technique as the prime classifier and the RBF kernel option for investigating and classifying DDoS attacks and benign services in SDN-based VANET. A hard margin SVM was implemented to select the optimal hyperplane with a minimal form of misclassification. This work is the pioneering attempt to use the GSCV in an IDM, especially in the VANET and SDN-based VANET. This work addresses the following drawbacks in related literature.

1) Suitable selection and data set analysis compatible with the right SVM kernel toward providing basis for effective detection and classification.
2) Although the accuracy achieved by surveyed hybrid algorithms is high and may be considered as a near-reliable requirement for VANET; it can, however, result in increased computational complexity.
3) Improvement in the accuracy (>99%) achieved by the aforementioned works as VANET represents a mission-critical system.
4) Creating a reliable method for selecting the optimal hyperparameters for the RBF-SVM-based IDM to enable high and accurate detection of DDoS intrusions.
5) Validating the performance of the proposed RBF-SVM IDS using the CICDDoS data set. Some of the presented IDSs used very old or small-dimensional data sets to validate the detection method.

III. Experimental Methodology

A. Proposed Detection Architecture for SDN-Based VANET

One of the most significant advantages VANET provides is an accurate and rapid assessment of any situation on the road [2]. The most critical and difficult aspect of this provision, and the overall development process of the transportation system, is ensuring its security. As 5G networks necessitate stable and flexible operation of VANET applications, SDN is introduced to support VANET functionalities. An SDN-based VANET architecture is divided into three primary planes: 1) the application; 2) control; and 3) data plane. Vulnerabilities can be exploited in any of these planes [7]. The network control plane (NCP) of SDN-VANET, which is responsible for routing and communication, is majorly susceptible to DDoS attacks [9]. Additionally, with the introduction of SDN, independent deployment of processing entities, control, and traffic forwarding is introduced [38]. However, more security complications are introduced as third-party applications liable to security vulnerabilities are introduced.

AI-integrated security systems promise self-adaptive, cost-effective, and sustainable requirements even in a changing and uncertain environment [39]. In this work, the goal of the proposed AI/ML technique is to detect DDoS traffic using an

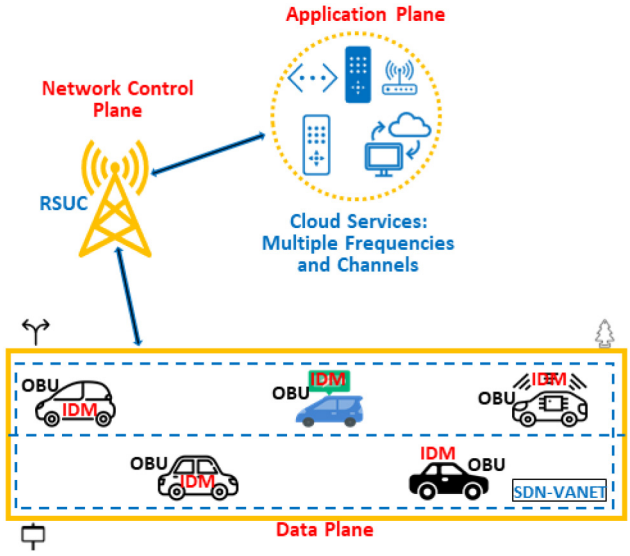

Fig. 1. Proposed DDoS detection architecture for SDN-VANET.

SDN-based VANET as the target system. A traditional VANET network may comprise an IDM device that is often located in a specific area of the network, limiting its detection capabilities to malicious activity, especially in a narrow network region. In this work, we provided an ML approach to represent the security framework of SDN. The solution to the DDoS threat is defined on each node of the system. SDN-based VANET relies heavily on the NCP, which is made up of modules that coordinate the heterogeneous networks of the intelligent infrastructure, ensuring the distribution of VANET policy rules and network behaviors [40]. The network centrality metrics can be used to identify crucial nodes in the system to efficiently evaluate the security of SDN [41].

The self-defence RBF-SVM Kernel-based framework is embedded in the OBU of each vehicle to make decisions to resist a DDOS attack. Each node on the SDN-VANET data plane carries its own local intrusion detection classifier (the proposed RBF-SVM), which uses the ML mechanisms to analyze the data received from the NCP and from other vehicles. As shown in Fig. 1, when an attack is detected, the OBU relies on the IDM integrated into each vehicle to make a decision. The OBU possesses the capability to transmit on one or more radio-frequency channels [42]. If a DDoS attack or jam is encountered on any channel by the ML algorithm implemented, the roadside unit controller (RSUC) in the NCP sends the information received to the application plane. The OBU can successively request a technology or channel switch from the application plane by requesting access from the NCP.

Dedicated short-range communication (DSRC) is a communication enabler and a component of the application plane in SDN-VANET. The Federal Communications Commission allocated 75 MHz of radio spectrum for this purpose; hence, there are several channels and frequencies. Although the control channel is the main channel dedicated to broadcasting data for safety applications, service announcements, and vehicle-to-vehicle messages, it is usually the preferred channel for message dissemination [43]. However, in the case of network unavailability, NCP may request a channel switch.

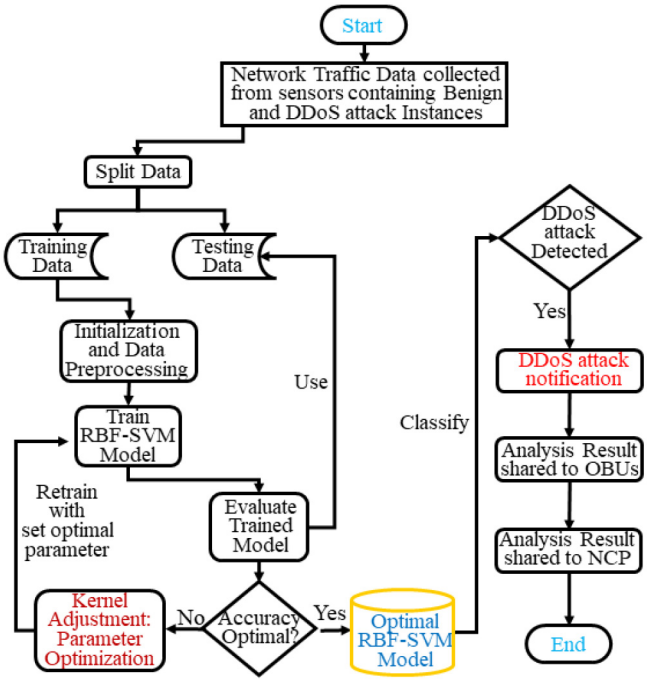

Fig. 2. Flowchart of the proposed RBF-SVM DDoS identification model.

The RSUC serves as a gateway between OBUs and the communications/cloud infrastructure, providing connectivity and support to the SDN-based VANET [42]. The RSUC dynamically implements a fall-back mechanism by switching into the appropriate frequency channels among the available channels for given data transport. This method can be employed to obtain network availability. The data may be forwarded to the next OBU in the network, and network availability is ensured at all times. The flowchart in Fig. 2 summarizes the overall identification and detection flow of the proposed ML model.

B. Proposed Machine Learning Intrusion Detection Model

Amidst the available solutions for intrusion detection in VANETs, ML is well established and is proven to be very beneficial for the prediction, classification, and detection of attacks. ML-based IDMs are based on statistical methods to define typical network behaviors [2], [16]. Any deviation exceeding a certain threshold different from such a pattern is statistically examined, and when it occurs, the security system believes an intrusion is underway. The benefit of this detection technique is that it allows for the identification of previously unknown attacks without the need to update an attack database [44]. This study proposes an ML-based supervised SVM classifier for DDoS detection and classification, which is formulated on the maximal margin classifier. SVM is undoubtedly one of the best and most accurate techniques for classification problems, with an edge in obtaining global minimum optimality rather than a local minimum [45]. With SVM, an IDM can perform well even in high-dimensional spaces and with irregular and unknown data distribution. In addition, with its associated kernel function, complex classification problems can be solved.

An SVM attempts to draw a hyperplane (decision plane) separating two data subsets to create a classification model. A maximum margin is drawn between these two sets of data creating a distance between each nearest point and the hyperplane. The closest data points to the hyperplane of each class are known as the support vectors [46]. SVM separates the training set of vectors that represent the incoming traffic as either DDoS attack or normal. The IDM assigns each event in the SDN-based VANET to a DDoS attack or normal event based on previous observations. Consider an X-dimensional input space of a network connection with n, number of features/predictors and m as number of points/samples usually assumed to be different from each other. The network connection can be denoted as $x = \{x_1, x_2, x_3, \ldots, x_m\}, x \subseteq \mathbb{Z}^n$ where $x_j(j{=}1, 2, 3, \ldots, m)$ representing the jth feature. Given a training set $\{(x_j, y_j)\}, m \leq j \leq m, x_j \epsilon X \subseteq \mathbb{Z}^n$, and $y_j \epsilon Y = \{0, 1\}$. Using an exhaustive grid search implementation denoted by $f(x) : X \Rightarrow Y$ evaluated at each point, an optimal hyperplane can be implemented to classify normal and DDoS attack scenarios. Y represent the predicted or classified values.

1) Proposed RBF-SVM Model and Simulation Parameters: The choice of a kernel function and its parameters is critical because an appropriately constructed kernel results in a model that fits well with the structure's underlying data. In SVM, the kernel function is responsible for transferring the data set to a higher dimensional space [18]. The linear, polynomial, sigmoid, and RBF are SVM kernel functions utilized to generate higher dimensional space. Hence, adjusting a kernel argument allows you to specify the influence of the kernel on the model. The RBF Kernel is one of the most extensively and indefinitely smoothed kernels. Relative to other kernels, RBF kernels can perform complex nonlinear mappings, and parameters are computed with ease, allowing for a rapid and reliable learning mechanism. RBF kernels function by estimating the similarity and closeness between two points [47]. Equation (1) mathematically expresses the RBF kernel, where σ is the variance and hyperparameter, $d_{i,j}$ represents the Euclidean distance $\|d_i - d_j\|^2$ between two points d_i and d_j

$$\mathrm{K_{RBF}}(d_i, d_j) = \exp\left(\frac{-\|d_i - d_j\|^2}{2\sigma^2}\right) \tag{1}$$

$$\gamma \ \alpha \ \frac{1}{\sigma}. \tag{2}$$

SDN-based VANET DDoS detection problem is seen as a complex classification best solved with kernels whose feature space has infinite number of dimensions (k). From (1), and (2), the RBF kernel decreases with distance and ranges between zero and one ($d_i = d_j$). For $\sigma = 1$, and using the multinominal theorem, we have the following:

$$\exp\left(\frac{-\|d_i - d_j\|^2}{2}\right) = \exp\left(\frac{2}{2}d_i^\tau d_j - \frac{1}{2}\|d_i\|^2 - \frac{1}{2}\|d_j\|^2\right) \tag{3}$$

$$= \exp(d_i^\tau d_j)\exp\left(-\frac{1}{2}\|d_i\|^2\right)\exp\left(-\frac{1}{2}\|d_j\|^2\right) \tag{4}$$

$$= \sum_{k=0}^{\infty} \frac{(d_i^\tau d_j)^k}{k^!}\exp\left(-\frac{1}{2}\|d_i\|^2\right)\exp\left(-\frac{1}{2}\|d_j\|^2\right). \tag{5}$$

2) Kernel Function Selection and Parameter of Proposed Model: The kernel function for RBF can be specified optimized by the kernel parameter σ in (1). Proper selection of parameter values is paramount to the SVM's performance [45]. While training an RBF-SVM, it is important to carefully considered and compute the right value for σ as this value is allied with this kernel. The two parameters associated with the RBF Kernel are "gamma" (γ) for the RBF kernel and the penalty parameter ("C") for the SVM model. The RBF kernel "γ" parameter is inversely proportional to σ (2) and indicates how powerful a single training example is. The bigger the γ, the closer the other samples must be to be influenced [48]. The parameter C specifies the influence of the kernel on the model and tells the SVM optimization how much misclassification to avoid in each training example. It is common to all SVM kernels and trades off misclassification of training instances for the decision surface's simplicity [49]. A low C regularizes the decision surface, whereas a high C tries to correctly identify all training samples. The model was optimized using the parameter optimization principle in order to achieve the best bias–variance tradeoff toward enhancing the best generalization performance for the RBF kernel.

3) IDM Optimization Using Exhaustive Grid-Search Technique: Finding the right σ for a given data set is critical and can be accomplished through the use of hyperparameter tuning techniques such as GSCV. The idea behind this technique is geared toward increasing the efficiency of the RBF-SVM kernel. A wrong choice of parameters may result in a poor fit to the data and, in turn, poor model performance [19]. Hyperparameters are not learned directly within estimators. However, with kernel optimization, a model can yield specific and optimal parameters for the classifier, achieving the lowest error rate while keeping the model's complexity under control [50]. For an optimal selection of parameters, a cross-validated grid-search was implemented over a parameter grid of C and γ spread exponentially apart to achieve a more optimal selection of parameter values (see Algorithm 1). GSCV is a comprehensive method that searches all possible preset combinations to discover the best point in the domain. Hence, this optimization method produces a more accurate combination of parameters rather than relying on the randomness implemented in equivalent optimization approaches.

The Algorithm 1 illustrates the procedure of implementing the GSCV for detecting DDoS attacks from normal behavior by using the proposed technique. A *fivefold* CV technique was utilized as a performance metric, dividing the D, training set into mutually exclusive and exhaustive equal-sized subsets. The subsets are fits of K splits, five folds for each of the H_{set} 36 candidates (range of selected C and γ parameter values), totaling 180 fits. For each iteration, the model tries a combination of hyperparameters in a specific order, fits the model on each combination and records its performance. Finally, it returns the best model result with the best hyperparameters combination. Although, the wider the range of parameters, the greater the chances the search mechanism has of finding the ideal combination of parameters [19]. The majority of algorithms have a subset of hyperparameters that have the greatest influence on the search procedure. Invariably, finer tuning can be achieved using a wider range of parameters, but at a much higher computational cost. However, computational complexity was considered to avoid interminable fitting time. Hence, the ranges of C and γ are limited as follows: C: [0.01, 0.1, 1, 10, 100, 1000], while γ: ["scale," 1, 0.1, 0.01, 0.001, 0.0001]. All simulation parameters are summarized in Table II.

Algorithm 1 Fivefold CV With GridSearch

Ensure: Optimal Hyper-tune parameter C and γ
Require: $RBF - SVM - Accuracy \geq 99\%$
Require: $C = 0.01, 0.1, 1, 10, 100, 1000$
Require: γ = Scale, 1, 0.1, 0.01, 0.001, 0.0001 ▷ C and γ account for the grid point in the grid search
Require: $H_{sets} = C$ and γ combination
Require: D; input dataset of DDoS attack features and output binary classification
Require: K_o, K_i, ▷ where K_o is the number of outer folds, and K_i inner folds
for $r = 1 \leftarrow K_o$ splits **do** split D into D_r^{train}, $D_r^{validation}$ for the r'th split
end for
for $g = 1 \leftarrow K_i$ splits **do** split D into D_g^{train}, $D_g^{validation}$ for the g'th split
end for
for each H in H_{set} **do** train RBF-SVM on D_g^{train}
end for
▷ compute validation error $E^{validation}$ for RBF-SVM with $D_g^{validation}$
▷ Select optimal hyper-parameter set H_o from H_{set}, where $E_g^{validation}$ is best
while Train RBF-SVM with D_r^{train}, using H_o **do**
Compute validation error $E_r^{validation}$ for RBF-SVM with $D_r^{validation}$
if $Accuracy \geq 99\%$ **then**
$H_{set} \leftarrow C \times Y$ is optimal
end if
end while

TABLE II
SUMMARY OF SIMULATION PARAMETERS

Simulation Parameters	Values/Ranges
Traffic Type	SDN-Based IoT/IoV Network Traffic
No of Statistical features	21
No of Observations	104,345
Data Visualization	Andrew's Plot showing non-linearity
ML Model/Kernel Type	RBF-SVM
Training Set	73041
Validation Set	31,304
Cross-Validation	5-Folds
Optimization Technique	Grid Search Cross Validation
Range of 'C'	0.01, 0.1, 1, 10, 100, 1000
Range of 'γ'	Scale, 1, 0.1, 0.01, 0.001, 0.0001
Total No of Iteration	180 fits
Feature Analysis	PCA
Evaluation Metrics	Accuracy, Precision, Recall, CKS, MSE, Computational Cost (Training Time)

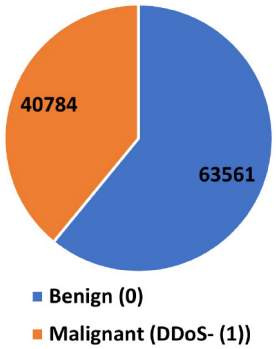

Fig. 3. Distribution of benign and DDoS attack instances of the SDN-DDoS data set.

TABLE III
STATISTICAL FEATURES/PREDICTORS OF THE IDM

dt	*switch*	*src*	*dst*
dur	*pkt-count*	*byte-count*	*dur-n-sec*
flows	*packets-ins*	*tot-dur*	*pkt-per-flow*
byte-per-flow	*pkt-rate*	*pair-flow*	*protocol*
rx-bytes	*port-no*	*tx-bytes*	*tx-kbps*
AI	*A2*	*tot-kbps*	*rx-kbps*

TABLE IV
PERFORMANCE OF VARIOUS SVM-KERNEL CONDITIONS WITHOUT DATA SET SCALING

Kernel Type	Accuracy (%)	Computing Time (seconds)
RBF	67.56	667
Poly	80.60	501
Sigmoid	51.08	414
Linear	-	-

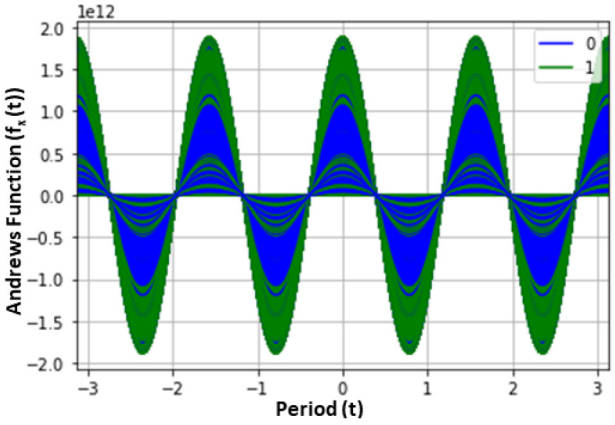

Fig. 4. Andrews plot showing nonlinearity of the DDoS data set.

C. Data Set Description, Properties, and Preprocessing

The data set considered for experimentation and ML evaluation is the SDN DDoS attack data set from IEEE DataPort [20] developed to evaluate DDoS attacks on SDN-based connected networks. The SDN DDoS data set was generated using the Mininet emulator and presents an observation of DDoS traffic for detecting suspicious and distributed intrusions in SDN-integrated networks. Mininet is a helpful tool that enables the creation of virtual network topologies. The data set consists of the Transmission Control Protocol, Internet Control Message Protocol, User Datagram Protocol, and Internet Control Message Protocol benign and DDoS attack traffic. The initial data set consists of network traffic image instances of 5×5 pixels each. For evaluating ML IDMs, the data set is converted into CSV format, consisting of 104 335 samples of attack and benign scenarios with distribution in the pie chart in Fig. 3. The data set contains twenty-four (24) statistical predictors outlined in Table III. The predictors considered for modeling and evaluation effectively capture the properties of SDN-based VANETs as functional components and subsets of IoT.

For further evaluation, the CICDDoS2019 data set was also considered [51]. The CICDDoS2019 data set captures reflection-based and exploitation-based attack types and is a representation of DDoS evaluation in any network. The data set contains raw data, including network traffic (Pcaps) and event logs.CICFlowMeter-V3 was used to extract more than 80 traffic features. In both data sets, each DDoS attack instance is labeled as "1," while benign data communication has its behavior set as "0." The data was preprocessed to handle missing and categorical attributes before training and testing. Disregarding this technique could thwart the ability of the IDM to correctly classify all instances of benign and malignant data [44].

1) Data Set Scaling: SVM is considered a superior classifier in terms of accuracy and generalization capabilities. However, its limitation, particularly when implemented without the right kernel selection, in-depth knowledge, and data structure analysis is the increased computational time and complexity corroborated in the default kernel performance result displayed in Table IV. To formally select a suitable SVM kernel for evaluation, there is a need for precise learnability of the data set toward analyzing the success of the SVM approach. An Andrews plot visualizes clusters and structures in high-dimensional data. An Andrews plots represent observations by a function "*f(t)*," of a continuous variable, "*x*," over an interval, "*t*" [52]. Furthermore, data set scaling also affects the performances of SVM models. Table IV represents poor performance and higher computing time of SVM kernel variants without data set scaling, confirming the relevance of data set scaling for the SVM classifier irrespective of the kernel used.

Moreover, the outputs of RBF-Kernel-based models are inadequately reliable without data set scaling [47]. Consequently, when the differential range of continuous values in a data set is vast, this results in increased classification errors regardless of the kernel type. The linear kernel is not so effective on data sets with overlapping classes, as shown in the overlapping Andrews plot of the data set used for evaluation in Fig. 4, hence it is impossible to linearly separate the classes and generate results using the linear kernel within a reasonable time.

Using the linear kernel, the model trained for days. The RBF kernel consumed the second-top computing resources at a fair accuracy of 67.56% while the sigmoid kernel had the least computing time at a lesser accuracy of 51.08%. The standard scaling procedure was applied, which involves removing

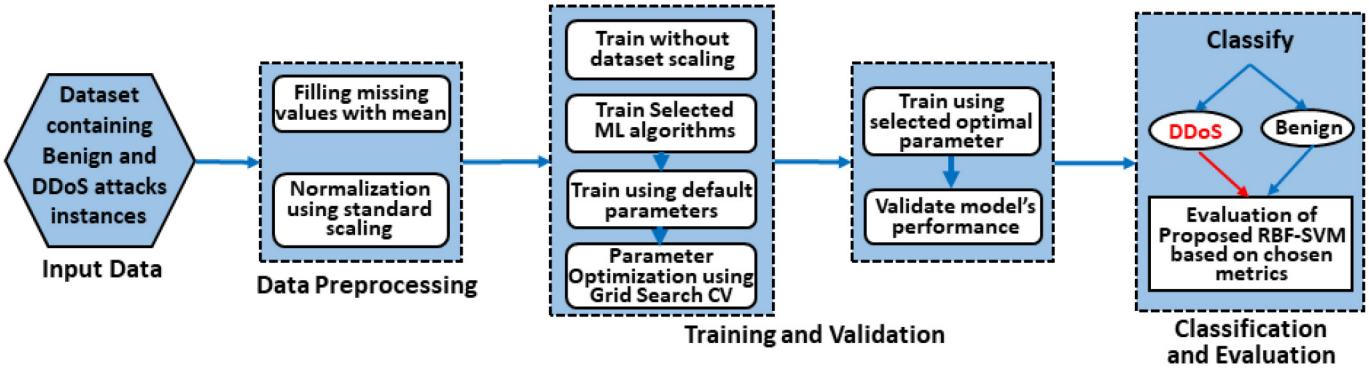

Fig. 5. Overall RBF-SVM kernel-based DDoS detection flow.

the mean of the variables and scaling them to unit variance. This method makes the model less susceptible to outliers and misclassifications. With data set scaling, the RBF-SVM can converge quickly, and classification inaccuracies are reduced

D. Candidate Algorithms Implemented

1) *LR:* LR and SVM generally perform comparably in practice. SVM tries to find the best margin that separates classes, reducing the risk of error. Correspondingly, LR creates decision boundaries with different weights near the optimal point. In addition, both algorithms handle optimization problems indistinguishably [37].
2) *kNN:* In similarity, kNN and RBF kernel-based SVM are nonparametric methods for estimating the probability density of a given data. Both classifiers calculate the distance between instances. A kNN algorithm assigns a new pattern, X, to that category to which the plurality of its k closest neighbors belong. The kNN method can be thought of as estimating the values of the probabilities of the classes given X [44].
3) *GNB:* GNB is a Naive Bayes variant that uses the Gaussian normal distribution and works with continuous data. The conditional probability density of response patterns given a stimulus class is modeled using this multivariate classifier. A GNB utilizes a less flexible distributional model, assuming zero off-diagonal co-variance similar to an RBF-SVM [53].

The overall flow and subdivisions of experimental implementation and detection techniques carried out in this work is summarized in Fig. 5.

IV. Performance Evaluation

The performance of the proposed IDM for SDN-based VANET was evaluated through its ability to correctly classify vehicular communication and events as DDoS attacks and benign behaviors. Optimal parameter combination results are obtained and used for training the RBF-SVM kernel method. The baseline for performance metrics was an accuracy level higher than 99%. All implementations were done using a Windows-10 computer with a Core i5-8500 processor and 8 GB of RAMThe RBF-SVM was coded using Python programming language within the scikit-learn library [48]. To assess and evaluate the effectiveness of the proposed IDM when compared with the existing state-of-the-art and other ML models, seven (7) performance metrics were considered.

1) *Confusion Matrix (CM):* CM is a standard metric used to visually evaluate an IDM's effectiveness and identify misclassifications. CM represents a 2-D array that compares predicted category labels to the actual label in binary classifications. Other essential evaluation metrics derived from it are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The result of the model correctly predicting the positive class is TP. The result of the model correctly predicting the negative category is TN. FP is the result of the misclassified positive class. FN represents the prediction of the negative class that does not meet the standards [37].
2) *Accuracy:* Accuracy is calculated as the ratio of correctly foreseen DDoS and normal flow over total flows [37]. It also provides an estimate of how well the IDM can generalize out-of-sample data (test data). Equation (6) computes the percentage of correct classification

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{6}$$

3) *Recall (Detection Rate):* To compute the proportion of accurately detected attacks in the test data set as a fraction of all attacks [26], (7) is used

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

4) *Precision:* The proportion of accurately categorized DDoS traffic to all attack records is computed using

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{8}$$

5) *Cohen Kappa Score (CKS):* CKS measures how closely the instances classified by the ML classifier agree with the data labeled as the ground truth. It is generally thought to be a more robust evaluation measure than accuracy rate [44]. The CKS is usually a number between -1 and 1. The maximum value signifies perfect agreement, while values less than 1 signify lesser or chance agreement [54]. Values closer to 1 are desirable, and depicts the closeness of predicted values to the

TABLE V
OPTIMAL PARAMETER VALUES OF THE PROPOSED RBF-SVM MODEL

Kernel Type	'C'	gamma
'RBF'	100	0.1

actual values. The formula to calculate Cohen's kappa for binary classification is presented in

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{9}$$

where p_o is the relative observed agreement among observation and p_e is the hypothetical probability of chance agreement.

6) *Mean Squared Error (MSE):* Equation (10) is used to determine the average magnitude of the error by evaluating the difference between the predicted and the corresponding actual values of the data set. The MSE is most beneficial when huge errors are unacceptable [44]. The closer the MSE value is to 0, the better the model's performance

$$\text{MSE} = \frac{1}{n}\sum_{i=1}\left(Y_i - \hat{Y}_i\right)^2 \tag{10}$$

where n represents the number of data points, Y_i and $\hat{Y}_i$ represents observed and predicted values respectively.

7) *Computational Complexity in Terms of the Model's Computing Time:* VANET is time-critical where transmission delays and delivery of safety-related messages are met within 100 ms [5]. To achieve this real-time constraint, the deployment of fast algorithms is a prerequisite, as carrying out message authentication is best done in real time. It is paramount to access an IDM's runtime complexity to ensure reduced latency and meet real-time requirements.

V. RESULTS, DISCUSSION, AND ANALYSIS

A comparative analysis of the proposed RBF-SVM-IDM and other ML approaches in terms of kernel selection, variants, parameter selection, data set property, and scalability will be discussed in this section. The proposed IDM's efficiency is assessed based on the test set of the realistic data set that covers contemporary DDoS attacks in SDN-based scenarios. The overall experimental result shows that, compared with the other schemes and the randomly selected parameter schemes, the GSCV RBF-SVM algorithm can reduce the computational complexity while achieving higher detection accuracy and reliability. As shown in Table V, optimal and ideal values for C and γ were captured at 100 and 0.1 respectively. At these optimal parameter points, the model achieved an appreciable and enhanced detection accuracy of 99.33% in comparison with other state-of-the-art ML approaches, properties, and RBF-SVM structures presented in this work.

A. Comparison Under Default Various Kernel Conditions after Data Set Scaling

In the preliminary stage of the experimental study, using the scaled data set, four variants of default SVM kernels were modeled without optimization. The experimental result and significance of data set scaling are presented in Table VI, showing the accuracy performances rate of initial and default parameter values (DPVs) for all used SVM kernel methods after data set scaling. The impact of data scaling is evident here as this technique resulted in a observable improvement relative to experimental results obtained without data set scaling. The RBF-SVM model was compared with these variants of SVM kernels after data set scaling for initial evaluation. As shown, RBF-SVM had the best initial accuracy result of 97.84% and a detection rate of 98.37% compared to other commonly used kernel methods.

TABLE VI
CLASSIFICATION RESULT AND PERFORMANCE COMPARISON OF "DEFAULT" SVM KERNEL CONDITIONS AFTER DATA SET SCALING

Kernel Type	Accuracy (%)	Recall (%)	Precision (%)	Computing Time (seconds)
'RBF'	97.84	98.37	96.18	71
'Poly'	97.47	95.65	97.98	75
'Linear'	80.64	72.01	76.98	673
'Sigmoid'	50.46	36.51	36.59	364

The "poly" kernel had the second-best classification accuracy, and a lower detection rate while the "linear" kernel had the third-best experimental result after data conversion into less-dimensional space. The poorest accuracy of 50.46% was recorded by using the sigmoid kernel method, which also consumed high computational resources of 364 s and a poor detection rate of 36.51%. The highest computational resource of 673s was consumed by the linear kernel method with an accuracy of 80.64%. Additionally, the data set is not linearly separable as represented in the overlapped Andrews plot in Fig. 4, resulting in a poor "linear-kernel" SVM classification result even after data set scaling. Linear kernel SVM models work best with small-dimensional data elements arranged in a sequential manner [45]. High-dimensional data makes linear computation more complex and takes a long time. Using default settings, it is evident that the RBF kernel performed better; hence, provides justification for its selection. However, the desired accuracy was not yet met. Hence, to enhance its performance and also beat the accuracies from reviewed literatures, the optimization technique was proposed.

B. Comparison Under Various Parameter Conditions

To distinguish default, random, and optimal parameter selection, an additional evaluation methodology was carried out by testing several RBF-SVM variants. Table VII provides a tabular representation of results for the SDN DDoS data set, comparing the proposed model with parameter performances for a variety of random parameter values. In all cases, an instance of $C = 100$ was combined with different and values of γ randomly. The combination of RBF-SVM and GSCV as optimization techniques achieved a better accuracy performance. The accuracy achieved by the proposed IDM confirms the necessity for implementing the hyperparameter optimization and selection technique rather than random parameter selections. To extensively validate the proposed model's results, other performance metrics for evaluating

TABLE VII
IMPACT ASSESSMENT OF SVM-RBF UNDER VARIOUS HYPERPARAMETER CONDITIONS

'RBF' Condition	Accuracy (%)	Recall (%)	Precision (%)	CKS (#)	MSE (#)	Computing Time (seconds)
Without Scaling	67.56	57.08	58.76	0.315	0.324	782
With Scaling (Default Parameter Values)	97.84	98.37	96.18	0.954	0.022	71
Proposed Model Optimal Parameter Values 'C'= 100; 'γ'= 0.1	99.33	99.22	99.08	0.986	0.007	62
Random Parameter Selection 'C'= 100; 'γ' = 1	98.20	97.13	98.22	0.962	0.018	1730
Random Parameter Selection 'C'= 100; 'γ' = 0.01	98.26	98.67	96.93	0.963	0.017	167
Random Parameter Selection 'C'= 100; 'γ' = 0.001	95.85	97.60	92.24	0.913	0.041	155
Random Parameter Selection 'C'= 100; 'γ' = 0.0001	86.63	80.69	84.41	0.716	0.134	319
Random Parameter Selection 'C'= 100; 'γ' = 'scale'	99.31	99.38	98.87	0.985	0.007	94

reliability and error rate introduced in the previous section were also captured in this table.

The proposed model achieve a recall of 99.22%, a precision of 99.08%, a CKS of 0.986, and a minimal error rate of 0.007 at 62 s, a significant improvement in comparison to using DPV. This shows that optimizing parameters values by incorporating GSCV can be significantly beneficial with regards to detection accuracy. Following the proposed model in terms of classification performance was the RBF-kernel variant with γ as "scale," having an accuracy of 99.31%, a closely accurate recall, CKS, and precision although longer computing time of 94 s. The same trend of increase computing resources is observed in other parameter variants. Experimental results without scaling recorded the poorest performance with a very poor CKS of 0.315 and an average squared error of 0.324.

C. Average Squared Error Evaluation for RBF-SVM Variants

As shown, our optimized model improves the IDM system in all performance metrics compared to another standard SVM models which were used on our data set only for comparison. The proposed model also shows a significant improvement in the MSE compared to others. The smaller a model's MSE, the closer the model is to computing the best line of fit. In this work, the estimator was tweaked using the GSCV to get the least classification error. It can be seen that by using the OPV, the average error is closer to zero (0). The lower the value, the better, and 0 means the model is perfect. The proposed RBF-SVM algorithm with optimized parameters used achieved a minimum error of 0.007 for predicting an attack. The highest error of 0.324 was recorded by using unscaled data set followed by the random parameter selection of $C = 100$ and $\gamma = 0.0001$ which recorded 0.134 as its MSE.

D. Training Time Complexity

In this section, the performance of the model was evaluated based on the average time it took to process the whole training data network packet as compared with other models and variants of SVM kernel. It is easy to tell how scalable an algorithm is just by evaluating the time of operation with data size and parameter selection. Training time complexity is always

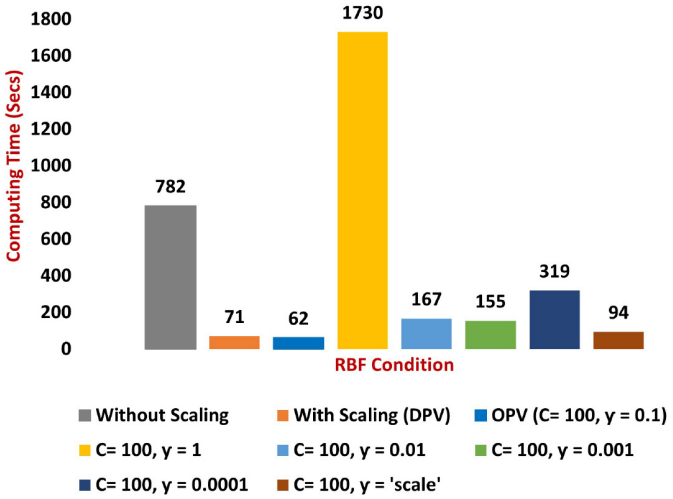

Fig. 6. Computational complexity (time) comparison of various RBF-SVM variants.

expressed in terms of some input size and optimal parameter selection. With the RBF-based kernel, the computing memory is utilized in a very efficient manner. Its effectiveness on nonlinear data is also without doubt. In addition, the complexity associated with this kernel remains the same even with an increase in the input size of the data. The advantage of the RBF-Kernel over other models is its ability to overcome the space complexity problem as RBF-SVM models just needs to store the support vectors during training and not the entire data set.

As shown in Fig. 6, the proposed IDM's computation time (62 s) with the use of optimal parameters was much lower than default parameter selection and other parameter settings. As a result, the optimal parameters aids the ML approach to produce better results for the IDM model. It can be seen that by using the OPV, the computing time of the proposed model could be reduced compared to the computing time of the DPV which is slightly higher and other structures, while also achieving a higher classification accuracy. The RBF variant with parameters at $C = 100$ and at 1 was the slowest of all, taking over 28 min to classify the traffic flow. In real-time environments, where high-speed data analysis and classification are

TABLE VIII
COMPARATIVE ANALYSIS OF PROPOSED SVM-RBF VERSUS APPROACHES FROM RELATED LITERATURE

Ref.	Technique Adopted	Accuracy (%)
[21]	SVM + NCA	97.20
[22]	RBF + ANOVA	98
[23]	Hybrid (DT + NN)	96.40
[25]	SVM	95
[26]	RBF-SVM	94.07
[27]	RBF-SVM	97.3
[29]	SVM + GA	99
[33]	Hybrid Model	98.37
[34]	SVM	99
[35]	SVM	93.89
[36]	Multi- Layered SVM	98.07
Model 1	KNN	97.88
Model 2	LR	78.04
Model 3	GNB	65.21
Proposed Model	**RBF-SVM**	**99.33**

critical, this would be unacceptable. VANETs are time-critical applications, thus demanding fast processing. Conclusively, algorithms that yield accurate results in a minimum amount of time are vital for VANET's deployment.

E. Comparing Performances of State-of-the-Art ML-Based IDM for DDoS Attacks

To validate the effectiveness of the proposed RBF-SVM-based algorithm, we set it side-by-side with comparable supervised ML methods modeled in this work and classical approaches techniques from related works. The selected ML methods (models 1–3) are similar to the proposed model in terms of modeling, structure and theoretical background. These models include model 1 (kNN), model 2 (GNB), and model 3 (LR). As shown in Table VIII, the classification accuracy of the proposed RBF-SVM model is superior to various variants of SVM approaches adopted in [21], [22], [25], [26], [27], [29], [34], [36], and [35], and the hybrid model, implemented in [23] and [33]. The experimental results demonstrate that the use of the GSCV technique can provide more additional discriminatory information for improving classification performance than other parameter optimization algorithm used in previous works. The proposed IDM achieved the best accuracy of 99.33% while the SVM model used by Alsarhan *et al.* [29] and Shams *et al.* [34], achieved the second-best result of 99%. For a mission-critical system, accuracy levels of over 99% are usually required. Meanwhile, kNN achieved an accuracy of 97.88%, GNB and LR were the least accurate of the lot at 65.21% and 78.04%, respectively.

F. Evaluating the Robustness of the Proposed Model

For further evaluation, the influence of the proposed model on a larger volume of DDoS attacks was evaluated using the CICDDoS2019 data set. The CICDDoS2019 data set was developed for evaluation, security testing, and malware prevention in large-scale networks [51]. Computational memory is used extremely efficiently using the proposed RBF-based kernel. Even as the data input size increases, the complexity associated with this proposed kernel remains constant. The

TABLE IX
EVALUATING THE SCALABILITY OF THE PROPOSED MODEL ON THE CICDDOS2019 DATA SET

Accuracy (%)	Precision (%)	Recall (%)	Computing Time (seconds)
99.96	98.46	100	73

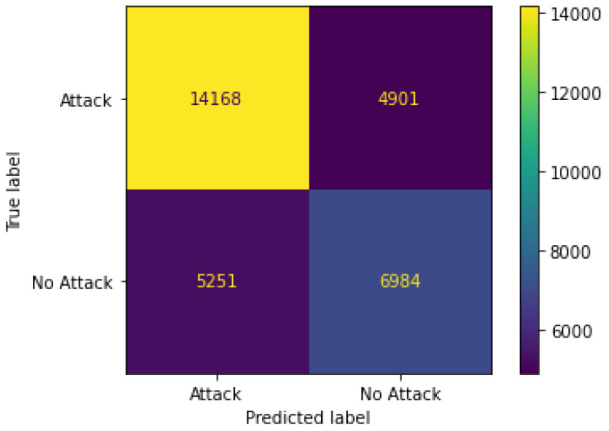

Fig. 7. CM of unscaled RBF-SVM data.

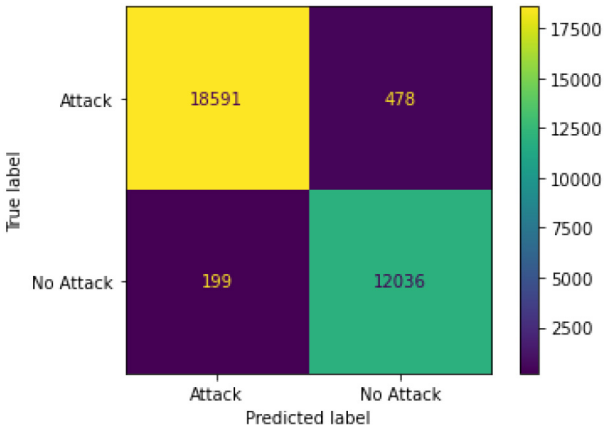

Fig. 8. CM of RBF-SVM using default parameter settings.

experimental result of the proposed model and its performance in handling this data set is shown in Table IX achieving an accuracy of 99.6%. The proposed model achieved a precision of 98.46% and 100% recall at a computing time of 73 s. This also implies that the proposed solution could be applied to any network scenario.

G. Visualizing Misclassification Using Confusion Matrix

The CMs in Figs. 7–9 show the number of correct and incorrect predictions of selected RBF-SVM variant. The selected models and structures are: "without data set scaling," "with default parameter setting using scaled data," and "optimal parameter tuning using GSCV on scaled data," respectively. The CM's left top-to-down diagonal elements reflect correct predictions representing the TP and TN, while the purple diagonals reflect wrong predictions and a representation of FP and FN, respectively. The default variant of the RBF-based SVM without scaling has the highest misclassification. However,

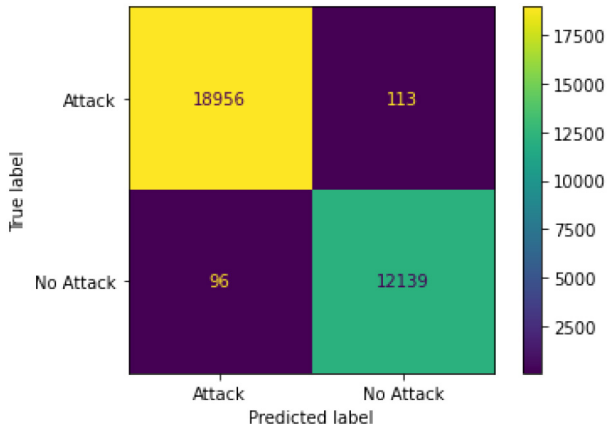

Fig. 9. CM of proposed RBF-SVM with optimal parameter selection using gridsearchcv.

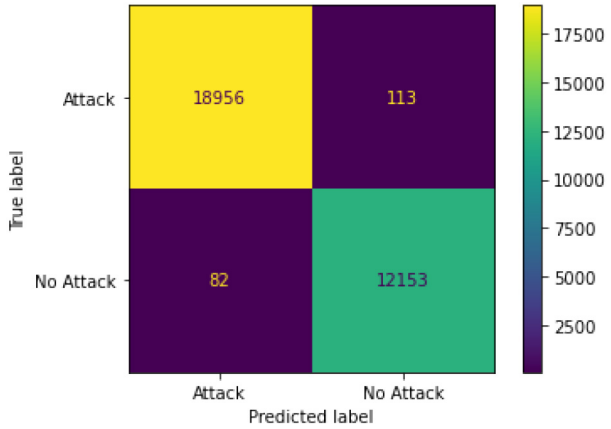

Fig. 10. CM result using different H_{sets}.

using the GSCV, the proposed model achieved the least misclassification (FN and FP) of 96 and 113 samples, respectively, and a minimal error rate of 0.007.

With a further action of optimizing the parameters of the RBF-SVM, using different ranges and sets of C and γ values, a slight increase in accuracy of 99.4% and a lower FN (misclassification) of 82 was achieved, as shown in Fig. 10. However, a computing time of 1617 s (approximately 27 min) was used for fitting and training the RBF-SVM model using the parameters obtained. This 27 min is considered high computing time. Thus, the need for a tradeoff between accuracy and computing time. Therefore, the proposed parameter values C and γ at 100 and 0.1, respectively gave the optimal performance as shown in Fig. 9. As seen in the table, the points gave the best accuracy–time tradeoff. The proposed optimal result achieved an accuracy of 99.33% and a detection rate of 99.22% at a computing time of 64 s.

H. Principal Component Metrics and Analysis

A scree plot is a plot of the eigenvalues of main components in a multidimensional study. The scree plot is evaluation factors in exploratory factor analysis or how many principal components (PCs) to keep in a PC analysis (PCA) [37]. This technique helps in identifying statistically significant

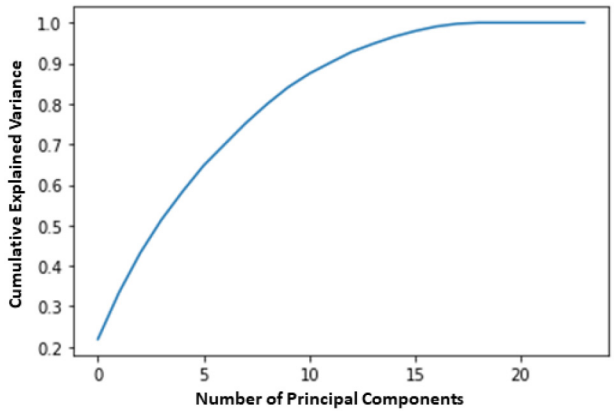

Fig. 11. Scree plot showing cumulative explained variance of PCs.

TABLE X
EIGENVALUES OF THE DATA PREDICTORS

$$\begin{bmatrix} 0.2191385 & 0.3342918 & 0.4316374 & 0.5133736 \\ 0.583036 & 0.6474400 & 0.7000804 & 0.7523807 \\ 0.7990743 & 0.8407513 & 0.8742843 & 0.9014635 \\ 0.9278435 & 0.9475216 & 0.9652817 & 0.9792826 \\ 0.9905486 & 0.9972964 & 0.9999888 & 0.9999992 \\ 0.9999997 & 1. & 1. & 1. \end{bmatrix}$$

predictors or components. Scree plot shows the variance of the features in the data set in descending/descending order, with the eigenvalues ordered from the greatest to smallest or the smallest to the greatest. The scree plot in Fig. 11 represents the significance of the predictors in the SDN-DDoS data set used for modeling. From this plot, the eigenvalue of the last component is higher than that of the second-next. The progression of the values follows the same pattern and so on.

To visualize and extract and information from the high-dimensional and nonlinear SDN-DDoS data set, PCA was utilized to measure the variance of each predictor by employing kernel function. The scree plot in Fig. 11 shows the explained variance of each PCs in order of the variation they cover as a function of the number of components. The last PC represents the eigenvalue that explains most of the information variance. In addition, as shown in screen plot (Fig. 11), the first ten components contain approximately over 80% of the variance, while 15 components and more are needed to describe close to 100% of the variance. Conclusively, the relevance and specific values of the twenty-four (24) predictors is captured in the array (see Table X). As shown in this array, most of the predictors has explained variance of more than 0.5, confirming the significance of the predictors in the data set used for modeling. The least explained variance is captured at 0.2.

Finally, the receiver operating characteristics (ROC) curve in Fig. 12 shows the performance of the proposed model at all classification thresholds plotting the TP rate against the FP rate. The area under the curve (AUC) is the measure of the ability of a classifier to distinguish between classes. It serves as a summary of the ROC curve. The range of the AUC value is from 0 to 1. A model with completely inaccurate predictions has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0. The higher the AUC, the

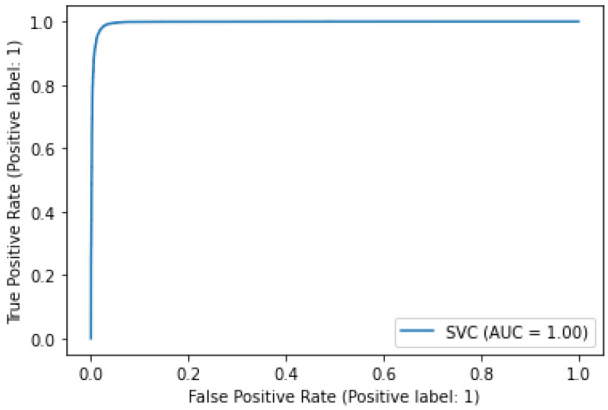

Fig. 12. TP versus FP rate at different classification thresholds.

better the performance of a model at distinguishing between the positive and negative classes. The optimal AUC of the proposed model is 1.

VI. Conclusion

This work presented an IDM for SDN-integrated VANETs using the SDN DDoS attack data set. Our model includes the parameter optimization elements to reduce error to the minimum resulting in better accuracy while achieving lesser complexity. The experimental results obtained with the publicly available data sets show that the proposed scheme can detect attacks with very high accuracy. The proposed model had the best performance of 99.33% accuracy and 99.22% detection rate in binary classification problems compared to all other ML methods. An ML framework that benefits from different types of attack detection using more complex data at a higher level can be a promising future research issue. In addition, ML approaches that considers space and memory computational analysis for SDN-VANETs can also be a research direction for future work.

References

[1] T.-T. Ngo, T. Huynh-The, and D.-S. Kim, "A novel VANETs-based traffic light scheduling scheme for greener planet and safer road intersections," *IEEE Access*, vol. 7, pp. 22175–22185, 2019.

[2] A. Talpur and M. Gurusamy, "Machine learning for security in vehicular networks: A comprehensive survey," 2021, *arXiv:2105.15035*.

[3] A. R. Gad, A. A. Nashat, and T. M. Barkat, "Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset," *IEEE Access*, vol. 9, pp. 142206–142217, 2021.

[4] M. M. Elahi, M. M. Rahman, and M. M. Islam, "An efficient authentication scheme for secured service provisioning in edge-enabled vehicular cloud networks towards sustainable smart cities," *Sustain. Cities Soc.*, vol. 76, Jan. 2022, Art. no. 103384.

[5] M. Safwat, A. Elgammal, E. G. AbdAllah, and M. A. Azer, "Survey and taxonomy of information-centric vehicular networking security attacks," *Ad Hoc Netw.*, vol. 124, Jan. 2022, Art. no. 102696.

[6] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: A survey," *Mobile Netw. Appl.*, vol. 26, pp. 1145–1168, Jun. 2021.

[7] M. Chahal, S. Harit, K. K. Mishra, A. K. Sangaiah, and Z. Zheng, "A survey on software-defined networking in vehicular ad hoc networks: Challenges, applications and use cases," *Sustain. Cities Soc.*, vol. 35, pp. 830–840, Nov. 2017.

[8] M. Arif *et al.*, "SDN-based VANETs, security attacks, applications, and challenges," *Appl. Sci.*, vol. 10, no. 9, p. 3217, 2020.

[9] G. C. Amaizu, C. I. Nwakanma, S. Bhardwaj, J.-M. Lee, and D.-S. Kim, "Composite and efficient DDoS attack detection framework for B5G networks," *Comput. Netw.*, vol. 188, Apr. 2021, Art. no. 107871.

[10] H. Shafiq, R. A. Rehman, and B.-S. Kim, "Services and security threats in SDN based VANETs: A survey," *Wireless Commun. Mobile Comput.*, vol. 2018, Apr. 2018, Art. no. 8631851. [Online]. Available: https://doi.org/10.1155/2018/8631851

[11] K. M. A. Alheeti, A. Gruebler, and K. McDonald-Maier, "Intelligent intrusion detection of Grey hole and rushing attacks in self-driving vehicular networks," *Computers*, vol. 5, no. 3, p. 16, 2016.

[12] J. F. Balarezo, S. Wang, K. G. Chavez, A. Al-Hourani, and S. Kandeepan, "A survey on DoS/DDoS attacks mathematical modelling for traditional, SDN and virtual networks," *Eng. Sci. Technol. Int. J.*, vol. 31, Jul. 2022, Art. no. 101065.

[13] M. Dibaei *et al.*, "Attacks and defences on intelligent connected vehicles: A survey," *Digit. Commun. Netw.*, vol. 6, no. 4, pp. 399–421, 2020.

[14] A. Verma, R. Saha, G. Kumar, and T.-H. Kim, "The security perspectives of vehicular networks: A taxonomical analysis of attacks and solutions," *Appl. Sci.*, vol. 11, no. 10, p. 4682, 2021.

[15] G. O. Anyanwu, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Real-time position falsification attack detection system for Internet of Vehicles," in *Proc. 26th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, 2021, pp. 1–4.

[16] P. Sharma, S. Jain, S. Gupta, and V. Chamola, "Role of machine learning and deep learning in securing 5G-driven industrial IoT applications," *Ad Hoc Netw.*, vol. 123, Dec. 2021, Art. no. 102685.

[17] M. Dibaei *et al.*, "Investigating the prospect of leveraging blockchain and machine learning to secure vehicular networks: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 683–700, Feb. 2022.

[18] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Mar. 2002.

[19] I. Syarif, A. Prügel-Bennett, and G. B. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 14, no. 4, pp. 1502–1509, 2016.

[20] S. Sambangi, L. Gondi, S. Aljawarneh, and S. R. Annaluri, 2021, "SDN DDOS attack image Dataset," IEEEDataPort. [Online]. Available: https://dx.doi.org/10.21227/k06q-3t33

[21] Z. Tonkal, H. Polat, E. Başaran, Z. Cömert, and R. Kocaoğlu, "Machine learning approach equipped with neighbourhood component analysis for DDoS attack detection in software-defined networking," *Electronics*, vol. 10, no. 11, p. 1227, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/11/1227

[22] K. Adhikary, S. Bhushan, S. Kumar, and K. Dutta, "Hybrid algorithm to detect DDoS attacks in VANETs," *Wireless Pers. Commun.*, vol. 114, pp. 3613–3634, Jun. 2020.

[23] A. Kaushik, B. Shashi, K. Sunil, and D. Kamlesh, "Decision tree and neural network based hybrid algorithm for detecting and preventing DDoS attacks in VANETS," *Int. J. Innov. Technol. Exploring Eng.*, vol. 9, pp. 669–675, Mar. 2020.

[24] H. Bangui, M. Ge, and B. Buhnova, "A hybrid data-driven model for intrusion detection in VANET," *Procedia Comput. Sci.*, vol. 184, pp. 516–523, Jan. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921006967

[25] J. A. Pérez-Díaz, I. A. Valdovinos, K.-K. R. Choo, and D. Zhu, "A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning," *IEEE Access*, vol. 8, pp. 155859–155872, 2020.

[26] Y. Gao, H. Wu, B. Song, Y. Jin, X. Luo, and X. Zeng, "A distributed network intrusion detection system for distributed denial of service attacks in vehicular ad hoc network," *IEEE Access*, vol. 7, pp. 154560–154571, 2019.

[27] F. A. Alhaidari and A. M. Alrehan, "A simulation work for generating a novel Dataset to detect distributed denial of service attacks on vehicular ad hoc NETwork systems," *Int. J. Distrib. Sensor Netw.*, vol. 17, no. 3, 2021, Art. no. 15501477211000287.

[28] R. K. Deka, D. K. Bhattacharyya, and J. K. Kalita, "Active learning to detect DDoS attack using ranked features," *Comput. Commun.*, vol. 145, pp. 203–222, Sep. 2019.

[29] A. Alsarhan, M. Alauthman, E. Alshdaifat, A.-R. Al-Ghuwairi, and A. Al-Dubai, "Machine learning-driven optimization for SVM-based intrusion detection system in vehicular ad hoc networks," *J. Ambient Intell. Humanized Comput.*, to be published.

[30] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Appl. Soft Comput.*, vol. 18, pp. 178–184, May 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494614000477

[31] E. A. Shams and A. Rizaner, "A novel support vector machine based intrusion detection system for mobile ad hoc networks," *Wireless Netw.*, vol. 24, pp. 1821–1829, Jul. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11276-016-1439-0
[32] S. Zavrak and M. İskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108346–108358, 2020.
[33] A. A. Elsaeidy, A. Jamalipour, and K. S. Munasinghe, "A hybrid deep learning approach for replay and DDoS attack detection in a smart city," *IEEE Access*, vol. 9, pp. 154864–154875, 2021.
[34] E. A. Shams, A. H. Ulusoy, and A. Rizaner, "Performance analysis and comparison of anomaly-based intrusion detection in vehicular ad hoc networks," *Radio Eng.*, vol. 29, no. 4, pp. 664–671, 2020.
[35] E. A. Shams, A. Rizaner, and A. H. Ulusoy, "Trust aware support vector machine intrusion detection and prevention system in vehicular ad hoc networks," *Comput. Security*, vol. 78, pp. 245–254, Sep. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404818307569
[36] K. S. Sahoo *et al.*, "An evolutionary SVM model for DDOS attack detection in software defined networks," *IEEE Access*, vol. 8, pp. 132502–132513, 2020.
[37] A. Subasi, "Chapter 3—Machine learning techniques," in *Practical Machine Learning for Data Analysis Using Python*, A. Subasi, Ed. London, U.K.: Academic, 2020, pp. 91–202.
[38] P. Manso, J. Moura, and C. Serrão, "SDN-based intrusion detection system for early detection and mitigation of DDoS attacks," *Information*, vol. 10, no. 3, p. 106, 2019.
[39] S. S. Gill *et al.*, "AI for next generation computing: Emerging trends and future directions," *Internet Things*, vol. 19, Aug. 2022, Art. no. 100514.
[40] M. Adnan *et al.*, "Towards the design of efficient and secure architecture for software-defined vehicular networks," *Sensors*, vol. 21, no. 11, p. 3902, 2021.
[41] T. Eom, J. B. Hong, S. An, J. S. Park, and D. S. Kim, "A systematic approach to threat modeling and security analysis for software defined networking," *IEEE Access*, vol. 7, pp. 137432–137445, 2019.
[42] R. G. Engoulou, M. Bellaïche, S. Pierre, and A. Quintero, "VANET security surveys," *Comput. Commun.*, vol. 44, pp. 1–13, May 2014.
[43] S. Malik and P. K. Sahu, "A comparative study on routing protocols for VANETs," *Heliyon*, vol. 5, no. 8, 2019, Art. no. e02340. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844019360001
[44] A. Subasi, "Chapter 2—Data preprocessing," in *Practical Machine Learning for Data Analysis Using Python*, A. Subasi, Ed. London, U.K.: Academic, 2020, pp. 27–89.
[45] K. El Bouchefry and R. S. de Souza, "Chapter 12—Learning in big data: Introduction to machine learning," in *Knowledge Discovery in Big Data from Astronomy and Earth Observation*, P. Škoda and F. Adam, Eds. Amsterdam, The Netherlands: Elsevier, 2020, pp. 225–249.
[46] S. M. Basha and D. S. Rajput, "Chapter 9—Survey on evaluating the performance of machine learning algorithms: Past contributions and future roadmap," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, A. K. Sangaiah, Ed. Amsterdam, The Netherlands: Academic, 2019, pp. 153–164.
[47] P. Chudzian, "Radial basis function kernel optimization for pattern classification," in *Computer Recognition Systems. Advances in Intelligent and Soft Computing*, vol. 95, R. Burduk, M. Kurzyński, A. Woźniak, and M. Żołnierek, Eds. Berlin, Germany: Springer, 2011, pp. 99–108.
[48] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[49] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems*, vol. 20, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Red Hook, NY, USA: Curran Assoc., Inc., 2007, pp. 1177–1184.
[50] T. Yu and H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications," 2020, *arXiv:2003.05689*.
[51] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," in *Proc. Int. Carnahan Conf. Security Technol. (ICCST)*, 2019, pp. 1–8.
[52] V. Grinshpun, "Application of Andrews plots to visualization of multidimensional data," *Int. J. Environ. Sci. Educ.*, vol. 11, no. 17, pp. 10539–10551, 2016.
[53] M. Misaki, Y. Kim, P. A. Bandettini, and N. Kriegeskorte, "Comparison of multivariate classifiers and response normalizations for pattern-information fMRI," *NeuroImage*, vol. 53, no. 1, pp. 103–118, 2010.
[54] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochem Med*, vol. 22, no. 3, pp. 276–282, 2012.

Goodness Oluchi Anyanwu received the B.Tech. degree in information management technology from the Federal University of Technology, Owerri, Nigeria, in 2018. She is currently pursuing the master's degree in information technology convergence engineering with Kumoh National Institute of Technology, (KIT), Gumi, South Korea.

She is a full time Researcher with the Networked System Laboratory, KIT. Her research interests are real-time systems, artificial intelligence, Internet of Things/Vehicles, data analytics, and security of vehicular networks.

Cosmas Ifeanyi Nwakanma (Member, IEEE) received the Diploma degree (Distinction) in electrical/electronics engineering from the Federal Polytechnic Nekede Imo State Nigeria, Nekede, Nigeria, in 1999, the Bachelor of Engineering (B.Eng.) degree in communication engineering, the master's (M.Sc.) degree in information technology, and the Master of Business Administration (MBA) degree in project management technology from the Federal University of Technology, Owerri, Nigeria in 2004, 2012, and 2016, respectively, and the Ph.D. degree in IT-convergence engineering from the Networked System Laboratory, IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea, in 2022.

He has been a Lecturer and Researcher with the Federal University of Technology since 2009. He was an Intern with Asea Brown Boveri, Lagos, Nigeria, in 2003. He is also a Postdoctoral Researcher with the ICT Convergence Research Center, Kumoh National Institute of Technology. His research interests include the reliability of artificial intelligence application to the Internet of Things for smart factories, homes, vehicles, and Metaverse.

Dr. Nwakanma is a member of the Computer Professionals Registration Council of Nigeria and Nigeria Society of Engineers, and registered by the Council for the Regulation of Engineering in Nigeria.

Jae-Min Lee (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2005.

He was a Senior Engineer with Samsung Electronics, Suwon, South Korea, from 2005 to 2014, where he was a Principal Engineer from 2015 to 2016. He has been an Assistant Professor with the Department of IT-Convergence Engineering, School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, Gyeongbuk, South Korea, since 2017. His current research interests include industrial wireless control networks, performance analysis of wireless networks, and TRIZ.

Dong-Seong Kim (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2003.

He worked as a full time Researcher with ERC-ACI, Seoul National University from 1994 to 2003. He worked as a Postdoctoral Researcher with the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, from March 2003 to February 2005. He was a Visiting Professor with the Department of Computer Science, University of California at Davis, Davis, CA, USA, from 2007 to 2009. He is currently the Director of the KIT Convergence Research Institute and ICT Convergence Research Center (ITRC and NRF Advanced Research Center Program) supported by the Korean Government, Kumoh National Institute of Technology, Gumi, South Korea. His research interests include real-time IoT and smart platform, industrial wireless control networks, and networked embedded systems.

Dr. Kim is a Senior Member of ACM.