International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I3P131

Volume 72 Issue 3, 358-369, March 2024
© 2024 Seventh Sense Research Group®

Original Article

Enhancing Secure Communication in IoT-Based Automated Vehicle Systems through Accurate Prediction of Abnormal Traffic Data

Kesava Reddy Jangam[1], R. Kumudham[2], V. Rajendran[3], M. Ramkumar Prabhu[4]

[1,2,3]*Electronics and Communication Engineering, Vels Institute of Science Technology and Advanced Studies, Chennai, India.*
[4]*Electronics and Communication Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.*

[1]*Corresponding Author : kesava464@gmail.com*

Received: 20 November 2023 Revised: 20 February 2024 Accepted: 29 February 2024 Published: 17 March 2024

Abstract *- Ensuring secure communication among Automated Vehicles (AV) on the Internet of Things (IoT) applications requires accurate prediction of abnormal traffic data. Information security within vehicles is crucial for transmitting traffic-related information reliably and guiding vehicles along the correct path. Existing algorithms for classification and prediction have aimed to provide clear predictions of malicious data. However, traditional techniques have faced challenges in achieving both accuracy and computational efficiency. This study proposes a three-stage implementation to address these issues. The first phase involves pre-processing to reduce computational complexities. This initial step streamlines the data for further analysis. The processed data then undergoes feature selection in the second phase, employing a multi-GGA (Greedy Genetic Algorithm) approach to identify the most relevant features. By utilizing this algorithm, the system can detect significant information even in the presence of misleading data. Finally, the third phase involves classification using a combination of Random Forest (RF) and AdaBoost algorithms. This integrated approach enables the system to distinguish between normal and abnormal traffic data in vehicle-to-vehicle datasets. Through experimental evaluations and comparative analysis, the efficiency of the proposed system is demonstrated in terms of accuracy, precision, recall, and F1-score, outperforming several existing algorithms. Overall, this proposed prediction system shows great potential in effectively classifying misleading and normal data with high accuracy. Addressing the limitations of traditional techniques offers a reliable solution for secure communication and decision-making in AV systems.*

Keywords *- Autonomous Vehicle, Internet of Vehicle, MGGA, Hybrid RF, and Adaboost.*

1. Introduction

1.1. Background and Motivation

Integrating IoT (Internet of Things) with smart devices has revolutionized various aspects of human life, offering increased flexibility and numerous benefits through its wide range of applications. Among these applications, autonomous vehicular systems have emerged as a promising technology capable of sensing and navigating users to their desired destinations without human intervention [1]. However, as autonomous vehicles interact with remote infrastructures and services provided by different service providers in Cyber-Physical Systems (CPS) [2], concerns regarding the security and privacy of sensitive information, such as location data, have become apparent.The advancement of connected automated vehicles (CAVs) and their integration into mixed-traffic environments alongside traditional vehicles has become increasingly prevalent. The driving techniques of CAVs employ a cyber-physical approach that combines cyber factors related to traffic data with vehicles' physical and internal components. Previous research [3] has presented a mixed-traffic framework within the context of CPS, focusing on the data collected from various vehicles and the delayed reactions of drivers. Stability conditions in the proposed model have been derived through linear stability analysis. Additionally, the study investigates fuel consumption and carbon dioxide emissions, highlighting the impact of CAV penetration rates.

Experimental evaluations demonstrate the potential for stability in mixed-traffic scenarios, which are influenced by driver reaction delays and the penetration frequency of CAVs. Moreover, the analysis reveals that the information collected by CAVs plays a crucial role. Simulation results confirm the system's enhanced stability and reduced CO2 emissions and fuel consumption [4]. The integration of IoT with CPS has greatly benefited autonomous vehicle systems, enabling users to submit queries related to locations and services of interest to service providers. However, concerns arise regarding the privacy and security of query contents, as their exposure may compromise user privacy. Existing privacy-preserving studies have relied on location perturbation or k-level anonymities. However, they often suffer from inadequate privacy protection or reduced query utility when processing multiple forms of queries for a single query content. To address these challenges, a recommended research study [5] proposes using Client-based Personalized k-anonymities (CPkA) for managing privacy and query options in autonomous systems. The study measures the performance of CPkA in terms of utility and privacy metrics. The study introduces two essential modules: the first module establishes mechanisms to achieve optimal privacy within each content group. In contrast, the second module employs linear programming models to compute optimal grouping techniques. By combining in-group methods and grouping formulations, CPkA achieves optimal privacy with high utility. The study employs a real-time dataset for simulation, evaluating the efficiency of the proposed approach.

1.2. Significance of the Study

To address safety concerns in CPS, another research study [6] proposes the use of forward reachability analysis techniques and reinforcement-based learning controllers. These techniques rely on efficient and accurate reachability methods to ensure the system's safety. The study focuses on aeronautical systems, utilizing intelligent tools such as sensors and wireless information analysis equipment to enhance efficiency, safety, and reliability.[7] [8] Privacy and security aspects of vehicles are also considered, particularly the vulnerability of unprotected systems [9]. Cybersecurity techniques based on Machine Learning (ML) often require access to attack data. Simulators generate high-fidelity data in the context of IoT, enabling accurate detection rates. Experimental evaluations consider factors such as detection rate, energy consumption, and memory usage in controlled environments.

Intelligent autonomous vehicles are seamlessly integrated into Cyber-Physical Systems (CPS) networks, facilitating secure wireless communication with other smart vehicles and devices, which in turn ensures a safer planning strategy. However, wireless communications can be unreliable, making vehicles an easy target for attacks. Such compromises can compromise vehicle autonomy, increase the latency between vehicles, and drain power significantly [10]. These compromises contribute to traffic congestion and jeopardize passenger safety, resulting in financial losses. Real-time attack detection is crucial for ensuring the safety of smart transportation systems. In this regard, a recommended approach [11] combines static and dynamic analysis methods for real-time attack detection. The proposed methodology effectively detects and preserves privacy by leveraging the strengths of combined analysis.CPS is designed to enable collaboration and communication among physical components, such as actuators, control systems, and sensors, within a communication framework.

A recent study [12] employs transfer learning techniques to identify cyber-attacks against connected physical elements. The study combines the Controller Area Network (CAN) with a simulation framework, generating datasets using pre-processing methods [13]. Despite the plethora of conventional methodologies and techniques available for attack prediction and the classification of normal and abnormal data, there is still room for improvement in terms of prediction accuracy. Furthermore, CPS mechanisms alone are insufficient for effectively predicting the normal transformation of information. Therefore, there is a need for ML-based algorithms that incorporate feature selection, classification, and effective pre-processing to enhance accuracy.

1.3. Research Gaps

The study was made on existing research work and identified few gaps. They have used intelligent prediction algorithms in autonomous vehicle systems and artificial intelligence-based classifiers to detect malware information from autonomous vehicle systems. To ensure the quality of the prediction algorithms, it is necessary to perform validation of abnormal vehicular data patterns. It is identified that there are a few limitations in the validation process.

- To detect distributed denial-of-service attacks and compute network traffic datasets, the Adaboost algorithm is a suitable algorithm for better accuracy.
- The prediction rate in the existing systems is insufficient for a better correlation of temporal features.
- The existing classification algorithms are not focused enough to detect and classify the higher-priority attacks in a faster time [15].
- Predicting normal and abnormal data accuracy is insufficient for highly secured automotive vehicular systems.
- The performance measure metrics need to be improved for better-performed systems.

1.4. Research Objects

This research aims to address the challenges faced by existing studies in achieving security in autonomous vehicle applications. Based on an analysis of various approaches, the objectives of this research are as follows:

- Perform pre-processing on the vehicle-to-vehicle (v2v) dataset to reduce computational complexity and improve the model's accuracy in terms of security in autonomous vehicles.
- Achieve accurate prediction by utilizing a hybridization of Multi Greedy-based Genetic Algorithms (MGGA) for feature selection.
- Implement classification using ML algorithms such as Random Forest (RF) and AdaBoost to accurately differentiate between malware information.

- Evaluate the system's performance using metrics such as precision, recall, F1 score, and accuracy to demonstrate improved accuracy in predicting normal and abnormal data.

2. Literature Review

2.1. Overview of IoT-based Automated Vehicle Systems

Privacy-preserving techniques in Connected Automated Vehicles (CAVs) have gained significant attention due to the increasing connectivity and reliance on AI-based systems. Integrating smart systems and connectivity has enabled CAVs to perform complex tasks more efficiently [14].

However, this connectivity has also introduced cybersecurity threats, such as malware attacks, phishing, and botnets, which can disrupt communication and pose risks to the connected systems. A recommended research study focuses on detecting botnet attacks in CAVs by utilizing network traffic analysis and temporal features [15]. The study compares the efficiency and accuracy of the proposed approach with existing techniques, demonstrating its effectiveness in early attack detection.

2.2. Importance of Secure Communication in AV Systems

Ensuring a secure communication infrastructure is crucial for deploying CAVs to enhance road safety. A recommended approach focuses on a cooperative, trust-aware Tolerant Misbehavior Detection System (CT2-MDS) to exchange sensitive information securely within CAVs [18].

The study constructs a Vehicle-to-Everything (V2X) framework and simulates various attacks, evaluating the cooperative trust factor and improving precision through uncertainty measurement in sensor observations. The proposed multi-modal fusion detection mechanism and cross-validation method outperform traditional classifiers in predicting attack densities.[17]

The Internet of Vehicles (IoV), which connects vehicles through IoT, has emerged as a solution for road safety and efficient transportation. ML algorithms have been employed to predict and mitigate threats and attacks in the network data of IoV [19]. Integrating IoT technologies in AVs enhances their capabilities in sensing and collecting data from the environment. [21]

However, this reliance on technology also makes AVs vulnerable to cyber-attacks [20]. Fault diagnosis systems utilizing SVM-based classification techniques have been proposed to enhance AV security [22]. The under-sampling technique combined with the Grey Wolf Optimizer (GWO) improves classification performance, and the SVM algorithm aids in fault diagnosis. Risk assessment plays a vital role in AV safety and collision avoidance. A suggested approach identifies risk status during critical traffic scenarios by selecting risk factors using Random Forest (RF) and the Pearson correlation coefficient method [23]. The model accurately predicts risk status, and a comparative analysis with other classification techniques, such as Classification and Regression Tree (CART) and SVM, demonstrates the superiority of RF. In the context of IoV, safety is enhanced through the connectivity of vehicles and the application of intelligent transportation systems [24]. Quality management systems have been employed to enhance AV safety by mitigating risks. The perception of customer requirements is crucial in this context, and Natural Language Processing (NLP) techniques have been employed to collect and classify user-related requirements [25].

Information exchange within vehicles is essential but requires validation to ensure privacy and security [26]. A recommended strategy utilizes supervised ML algorithms to detect attacks, categorize misbehavior information, and achieve efficient detection in IoV [27]. Another research study analyzes the performance and effectiveness of the VeReMi position attack detection model using precision, recall, and ROC analysis [28].

Deploying connected physical elements in Cyber-Physical Systems (CPS) has improved AV control and safety. An optimal forward distribution application is presented to ensure the safety of CAVs in smart transportation systems [29]. The CPS-based application model enhances vehicle safety by utilizing historical information for predictive analysis and integrating different behaviors into the control system. The application of CPS mechanisms improves the safety parameters of the vehicle [30].

2.3. Existing Techniques for Abnormal Traffic Data Prediction

In the context of CAVs, Machine Learning (ML) algorithms have played a significant role in various applications, including image processing. A suggested research study utilizes Particle Swarm Optimization (PSO) for classification in AVs, training the model using a vehicular dataset from Kaggle [16]. The model is evaluated using the Cubic Support Vector Machine (CSVM), outperforming traditional accuracy and time consumption techniques.

2.4. Limitations of Traditional Techniques

Existing studies have focused on intelligent prediction methods and AI-based classifiers for malware detection in AVs. However, some limitations remain in the validation of abnormal vehicular data patterns:

The Adaboost algorithm has shown suitability for detecting specific DDoS attacks but requires validation with other network traffic datasets to increase accuracy. Additionally, stronger correlations between temporal features can enhance prediction rates [15].

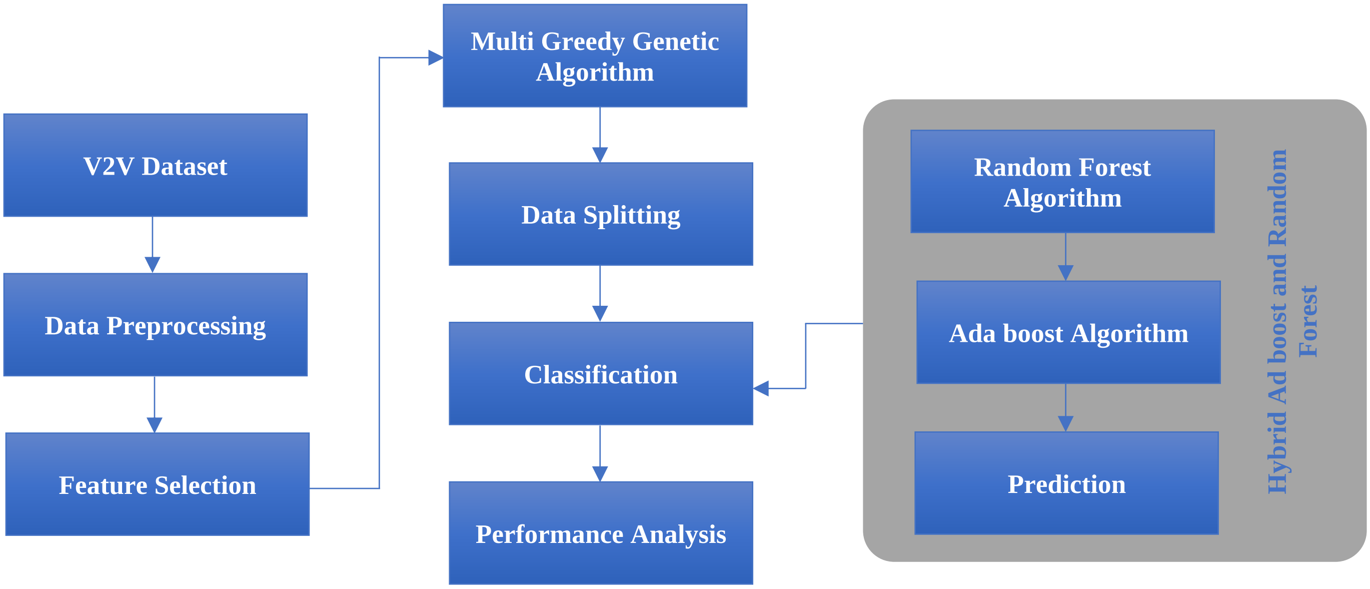

Fig. 1 Architecture of the proposed system

While the proposed approach for securing AV systems and automated CPS has demonstrated adaptiveness, future research should focus on developing classification algorithms capable of detecting and classifying higher-priority attacks more efficiently [29].

3. Methodology

Ensuring the security of Autonomous Vehicles (AVs) is paramount in preventing unauthorized control and malicious interference. The prediction of malware-infected data transmitted within AVs is particularly important, as it can safeguard against misleading instructions that may compromise the entire smart transportation system.

While conventional approaches have addressed the prediction of misleading data in Cyber-Physical Systems (CPS), they face challenges in heavy traffic scenarios, leading to reduced classification accuracy and increased computational complexities. To address these issues, a proposed system is designed to enhance accuracy by integrating feature selection and classification algorithms. The comprehensive functionality of the proposed system is illustrated in Figure 1.

The data input is derived from the v2v (vehicle to vehicle) dataset and undergoes pre-processing, facilitating faster computation and significantly reducing computational complexities. As depicted in Figure 1, following the pre-processing stage, the data goes through two implementation phases: feature selection and classification. Feature selection is initially performed using the Multi Greedy Genetic Algorithm (MGGA), employing a greedy approach to select the most optimal features. Population parameters are initialized through genetic operations like crossover, mutation, and fitness evaluation, which identify the best-fitting individual. The selected features are then fed into the classification phase, where hybridization of ML algorithms, namely Random Forest (RF) and Adaboost, is employed. RF is preferred due to its ability to handle larger datasets and mitigate overfitting issues, while Adaboost complements it. This combined approach ensures improved accuracy and addresses the optimization of input parameters. The classification algorithms utilized in this research effectively predict misleading information in Av, thereby enhancing security measures. Finally, the system's performance is assessed through metrics such as accuracy, precision, f1-score, and recall, determining its overall efficiency.

3.1. Feature Selection with MGGA (Multi Greedy Genetic Algorithm)

The initial phase of the proposed study involves the implementation of MGGA, which combines genetic algorithms with a greedy-based approach to select the most optimal features. The greedy technique plays a crucial role in identifying the best choices at each stage, collecting the most relevant features from the data sequence to effectively predict changes and distinguish between normal and abnormal patterns. By integrating GA with the greedy sequential technique, the problem is approached with the maximum search space exploration. The greedy approach acts as an operator within the genetic algorithm, enhancing search capabilities and improving chromosome fitness. The crossover operation between individual chromosomes facilitates exploitation between two parent solutions, modifying genes and maintaining population diversity. The following diagram illustrates the flow of MGGA:

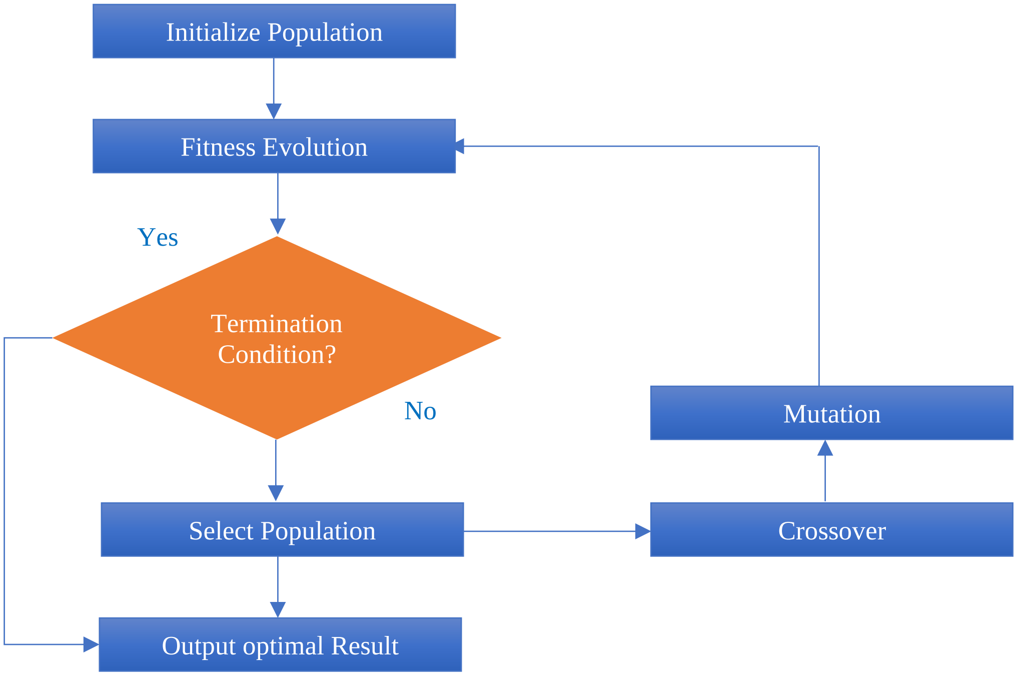

Fig. 2 Flow chart of MGGA

Figure 2 presents the flowchart diagram depicting the Genetic Algorithm (GA) operation. A greedy approach has been incorporated into the GA framework to obtain the global optimum value, allowing for the acquisition of local optimum values at each stage. The GA is recognized as an evolutionary algorithm that utilizes genetic operators to simulate the survival of fitness. Each genetic solution is represented as a chromosome, with each parameter being denoted as genes within that chromosome. The fitness function evaluates the fitness objective for each individual in the population. A random selection operation is performed using an operator that employs a random selection technique to enhance the obtained solution.

The operator selects the best solution based on the proportional fitness value. The superior solution obtained in each iteration is captured as the local solution, providing an estimate for the global optimum. This process of obtaining the best solution from every iteration improves the subsequently generated solutions, resulting in the overall population improving at each generation level. One of the primary advantages of the greedy approach is its ease of implementation and its guarantee of finding effective and optimal solutions at each step. It greatly assists in achieving the global optimal value.

```
Input: v2v information
Output: Best multiple features selected from the data
Step 1: Begin
Step 2: Initial population Generation with POP_ii (0).
Step 3: Estimation of POP_ii (0).
Step 4: Repeat
Step 5: Selection of parents.
Step 6: Crossover operation for Generation of new chromosomes.
Step 7: Involving mutation on the new chromosomes.
Step 8: Applied the greedy-based sequential function.
Step 9: Estimate POP_ii (t_i).
Step 10: Until (Terminating condition is reached).
Step 11: End.
```

Fig. 3 Pseudo Code1: Greedy-Based Genetic Algorithm Pseudo Code

The initial population is randomly comprised of both feasible and non-feasible solutions, which is followed by the selection of parents using the tournament-selection operation. Subsequently, crossover and mutation operations are performed, along with a sequential-based correction module. Finally, the entire set of chromosomes is evaluated to select the new parent population for the next life cycle. The following pseudo-code, denoted as Pseudo Code 1, illustrates the function of the greedy approach combined with the GA technique. The algorithm commences by generating a population with a random distribution, aiming to enhance diversity. This population consists of multiple solutions representing individual chromosomes. Each chromosome contains a set of variables that emulate genes. By spreading the solutions across the search space, diversity is fostered, leading to more accurate predictions of promising segments. The selection process favors the fittest individuals, allowing their genes to contribute to the production of the next generation. The algorithm continues to improve the population using three operators until the specified criterion is met. Ultimately, the best solution from the resulting population yields the optimal global solution for the problem at hand. The algorithm's control parameters can be adjusted to optimize performance. The crossover rate plays a crucial role in achieving efficient outcomes. For each graph G, the chromatic number x(G) represents the minimum number of colors required to color G, a known NP-complete problem. The proposed technique initializes the population P with feasible coloring graphs.

The population is obtained through a graph-coloring technique, which produces suitable colorings for the graph. The initial population function randomly generates various chromosomes, each with a different number of colors. The tournament-selection method serves as the selection mechanism, randomly choosing two significant individuals from the population group and selecting the one with the best features as the parent.By employing the crossover operator, two parents produce two children, inheriting genetic elements from their parents. The three-point operator is utilized in this process. Additionally, a uniform level of mutation introduces and maintains genetic diversity but with a lower probability. The algorithm employs a greedy approach where, for each child node, constraint verification is not necessary for all genes. In cases where a distinguished color from a neighbor is affected, it must be one of the colors utilized in the chromosomes. If no such occurrences are found, modifications are made accordingly.

3.2. Classification with Hybrid RF and Adaboost

After the feature selection process, the classification stage is carried out to enhance the system's accuracy. To maximize efficiency, the proposed system incorporates two distinct classifiers into a unified module. By harnessing the strengths of both RF and Adaboost classifiers, this research adopts a hybrid approach, capitalizing on the advantages offered by each algorithm. This integration aims to optimize the system's performance and achieve superior results.

3.2.1. Random Forest Algorithm

The RF-based ML mechanism employs the creation of multiple trees, which are later combined to determine more accurate predictions. A unified result is achieved by aggregating the outputs from various decision trees. This approach offers enhanced usability and flexibility, making it highly suitable for classification tasks. Through supervised learning, the classification algorithm utilizes training data to categorize new observations. The dataset is employed to train the model and classify input data into distinct classes. Below is a pseudo-code outlining the step-by-step functioning of the Random Forest algorithm: The algorithm starts by randomly selecting a set of records from the v2v dataset to form a random forest. Each record is used to construct a decision tree, which produces an individual output. The average value is computed across all the decision trees to improve accuracy. Instead of relying on a single decision tree, the random forest predicts the output based on the majority vote of all the trees.

```
Input: Training data samples
Output: Data Classified with labels
Step 1: $W_{ii}$
$= Value_{random}$  /
/ Initialization of  Weights by Random Values
Step 2:  $S_{hi} := \sum_{i=1}^{n} X_{ii}.W_{ii} + b_i$ // calculate the sum of the inputs $(X_{i1}, X_{i2}, X_{i3} \ldots \ldots \ldots \ldots .. X_{in})$ multiplied with their $(W_{i1}, W_{i2}, W_{i3} \ldots \ldots \ldots \ldots .. W_{in})$ for hidden layer node
Step 3: output hidden $\leftarrow \emptyset(s_{hi})$ output for hidden layer node
Step 4: $S_{out} := \sum_{i=1}^{n} X_{ii}.W_{ii} + b_i$ // input for output layer nodes
Step 5: $output_{predict} \leftarrow \emptyset(S_{out})$  // At output nodes
Step 6:Compare Error $\left(output_{predict} - output\ Actual\right)$
Step 7: Find Error Rate // from hidden layer Hl to output layer nodes.
Step 8: Find Error Rate // from hidden layer Il to output layer nodes.
Step 9: Updating the network weight.
Step 10: $\left(output_{predict} - output\ Actual\right)$
```

Fig. 4 Pseudo Code2: Random Forest Pseudo Code

The more trees in the forest, the higher the classification level, and the random forest outperforms other classification algorithms in terms of prediction. However, the random forest classifier is susceptible to noise in the data, which can lead to overfitting and reduced accuracy when testing the model. Additionally, without pruning, tree growth can become complex. To address these issues, the random forest algorithm is combined with Adaboost. This combination aims to achieve maximum accuracy and mitigate overfitting problems.

```
Given $(x_{i1}, y_{i1}) \ldots \ldots \ldots \ldots \ldots \ldots \ldots (x_{im}, y_{im}), x_{ii} \in X_i, y_{ii} \in Y_i = \{-1,1\}$
Initialize$D_{ii}(ii) = 1/m_i$
For $t_i = 1 \ldots \ldots \ldots . T_i$
    1.  Train Weak classifier using distribution $D_i$
    2.  Get a Weak hypothesis $h_{ii}: X_i \to \{-1,1\}$With error $\sum_{ii:h_{it}(x_{ii}) \neq y_{ii}} D_{it}(x_{ii})$
    3.  Choose $\alpha_{it} = \frac{1}{2} \log \left(1 - \frac{\varepsilon_{ti}}{\varepsilon_{ti}}\right)$
    4.  Update
        $D_{it+1}(ii) = \frac{D_{it}(ii)}{Z_t}$
        $= \{\frac{e^{-\propto t} \text{ if instance ii is correctly classified}}{e^{\propto t} \text{if instance ii is not correctly classified}}$
Where
$Z_{if}$ is a normalization factor (chosen so that $\sum_{ii=1}^{m_i} D_{it+1}$
1
Output the Final hypothesis $H_i(x_i) =$
$sign(\sum_{t_i-1}^{T} \propto_{it} h_{it}(x_{ii}))$
```

Fig. 5 Pseudo Code3: Adaboost Pseudo Code

3.2.2. Adaboost Algorithm

Boosting, a fundamental iterative technique, plays a crucial role in enhancing the accuracy of prediction rules by combining weak classifiers. In the boosting process, the training set is utilized, consisting of pairs (u1, v1), ..., (um, vm), where ui belongs to the X domain and each variable ri varies in R={-1, 1}. At each iteration T, a weight distribution St is assigned to the training samples, and a weak classifier is constructed based on St. Initially, all weights are set to the same value, and subsequent rounds focus on increasing the weights of misclassified samples to prioritize the learning of weak classifiers for accurate classification. The iterative process continues for a predetermined number of iterations. The ensemble of weak classifiers collectively contributes to predicting newly labeled samples, aiming to maximize accuracy. The pseudo-code for the Adaboost algorithm encapsulates the entire procedure.

Based on the given pseudo-code, it is evident that three types of boosting methods exist that are specifically designed for dealing with imbalanced classes in problem scenarios. The first category of algorithms addresses correct and incorrect classifications by adjusting the weights of false positive and negative samples in equal proportions. The prediction for unlabeled samples is determined by averaging individual classifications, and the weighting is based on the accuracy of the individual classifiers.

The second category, suitable for handling rare classes, employs ad cost, which assigns varying costs to misclassifications in the training data. This technique assigns higher misclassification costs to the rare classes, addressing the class imbalance. The third category of boosting techniques focuses on rare classes and involves modifying the weights of training data based on specific treatment criteria for true or false predictions. The final classification stage considers both single classifiers' positive and negative accuracy levels. The effectiveness of the current boosting approach relies on the weighted accuracy of the classifier, which surpasses the threshold of half of the classification. When dealing with classification problems, weak classifiers outperform random options, particularly when the occurrence of a class is close to zero, thereby improving prediction accuracy for that class. The accuracy of the class ensures a secure system that can avoid weak classification issues.

3.2.3. Integration of Adaboost with Random Forest

The present system employs a combination of Random Forest (RF) and AdaBoost techniques, utilizing RF as the weaker learner to generate a prediction framework with minimal error frequency. While boosting algorithms generally perform efficiently with weaker learners, RF is specifically chosen in this case due to its formulation with a vehicle-to-vehicle dataset used for predicting malware information. The hybridization of machine learning algorithms in Pseudo code 4 follows the subsequent steps:

From the below pseudo-code, it is evident that the combined technique offers significant advantages in terms of improved prediction and performance capabilities within v2v datasets. To effectively predict misleading traffic flow in CPS as AV, the proposed system utilizes hybridized techniques as the base learning algorithms, aiming to minimize the rate of error frequency, which serves as a fundamental measurement strategy. This approach allows for a comprehensive evaluation of the strengths and weaknesses of different algorithms in various assessment systems, incorporating inputs such as X_i1, X_i2, X_i3, and so on. By integrating these algorithms, the prediction performance in terms of accuracy and precision can be enhanced using the data gathered from the dataset.

The accuracy of the classifier, which measures the correctness of outcomes achieved in the test set, is determined as a percentage. The classifier is trained using D_ii as the distribution, with weights assigned as W_i1, W_i2, W_i3, and so forth for all layers. The final output is computed using the function H_i(x_i). The Adaboost algorithm plays a significant role in prediction tasks involving classification, providing better prediction reliability. This classifier generates multiple classifiers and selects the most effective one. It offers flexibility in combining classification methods and requires fewer input variables to enhance model accuracy. Additionally, Random Forest (RF) is a powerful classification technique widely used in pattern recognition and machine learning for high-dimensional data. The recursive partitioning method employed by RF performs splitting based on class 0 and 1 predictor, generating rules that maximize class purity within subsets. The bagging technique in the random forest classifier allows for the creation of multiple classifiers for high-dimensional information in a faster manner, using a voting mechanism with random vectors. By combining the advantages of both algorithms, the proposed system accurately classifies misleading information from normal traffic flow data in AV.

4. Results and Analysis

The proposed model's performance has been thoroughly evaluated through the computation of prediction rates using feature selection and classification accuracy derived from the algorithm. In the subsequent section, the computation results and comparative analysis with other algorithms are presented to assess the effectiveness of the proposed structure in AV applications, particularly in terms of accuracy.

4.1. Experimental Results

The vehicle-to-vehicle dataset was utilized in the proposed work, and this section presents the computational results obtained from the algorithm. The simulation outcomes, including the confusion matrix and correlation matrix, were measured for the classification problems at various threshold values. These results demonstrate the model's ability to effectively distinguish between different classes, showcasing its capability to accurately classify the data.

```
Given $(x_{i1}, y_{i1}) \dots \dots \dots \dots \dots \dots \dots (x_{im}, y_{im}), x_{ii} \in X_i, y_{ii} \in Y_i = \{-1,1\}$
Initialization of $D_{ii}(ii) = 1/m_i$
For $t_i = 1 \dots \dots \dots . T_i$
Training the Weak classifier with the distribution $D_{ii}$
Acquire Weak hypothesis using $h_{ii}: X_i \to \{-1,1\}$ along with minimum error $\sum_{ii:h_{it}(x_{ii}) \neq y_{ii}} D_{it}(x_{ii})$
Choosing $\alpha_{it} = \frac{1}{2}\log\left(1 - \frac{\varepsilon_{ti}}{\varepsilon_{ti}}\right)$
Updating
$S_{hi} := \sum_{i=1}^{n} X_{ii}.W_{ii} + b_i$ // calculating the sum of inputs with $(X_{i1}, X_{i2}, X_{i3} \dots \dots \dots \dots .. X_{in})$ and multiplied with the weights $(W_{i1}, W_{i2}, W_{i3} \dots \dots \dots \dots .. W_{in})$ for all layer node
output $\leftarrow \emptyset(s_{hi})$ output for layer node
$S_{out} := \sum_{i=1}^{n} X_{ii}.W_{ii} + b_i$ // input for output layer nodes
Formulate $D_{it+1}(ii) = \frac{S_{out}(ii)}{Z_t} = \{\frac{e^{-\propto t} \text{ if instance ii is correctly classified}}{e^{\propto t} \text{if instance ii is not correctly classified}}$
Where $Z_{if}$ is referred as the normalization factor (chosen so that $\sum_{ii=1}^{m_i} D_{it+1} = 1$
The output of the Final hypothesis is indicated as $H_i(x_i) = \text{sign}(\sum_{t_i-1}^{T} \propto_{it} h_{it}(x_{ii}))$
```

Fig. 6 Pseudo Code4: Integration of Adaboost with Random Forest Pseudo Code

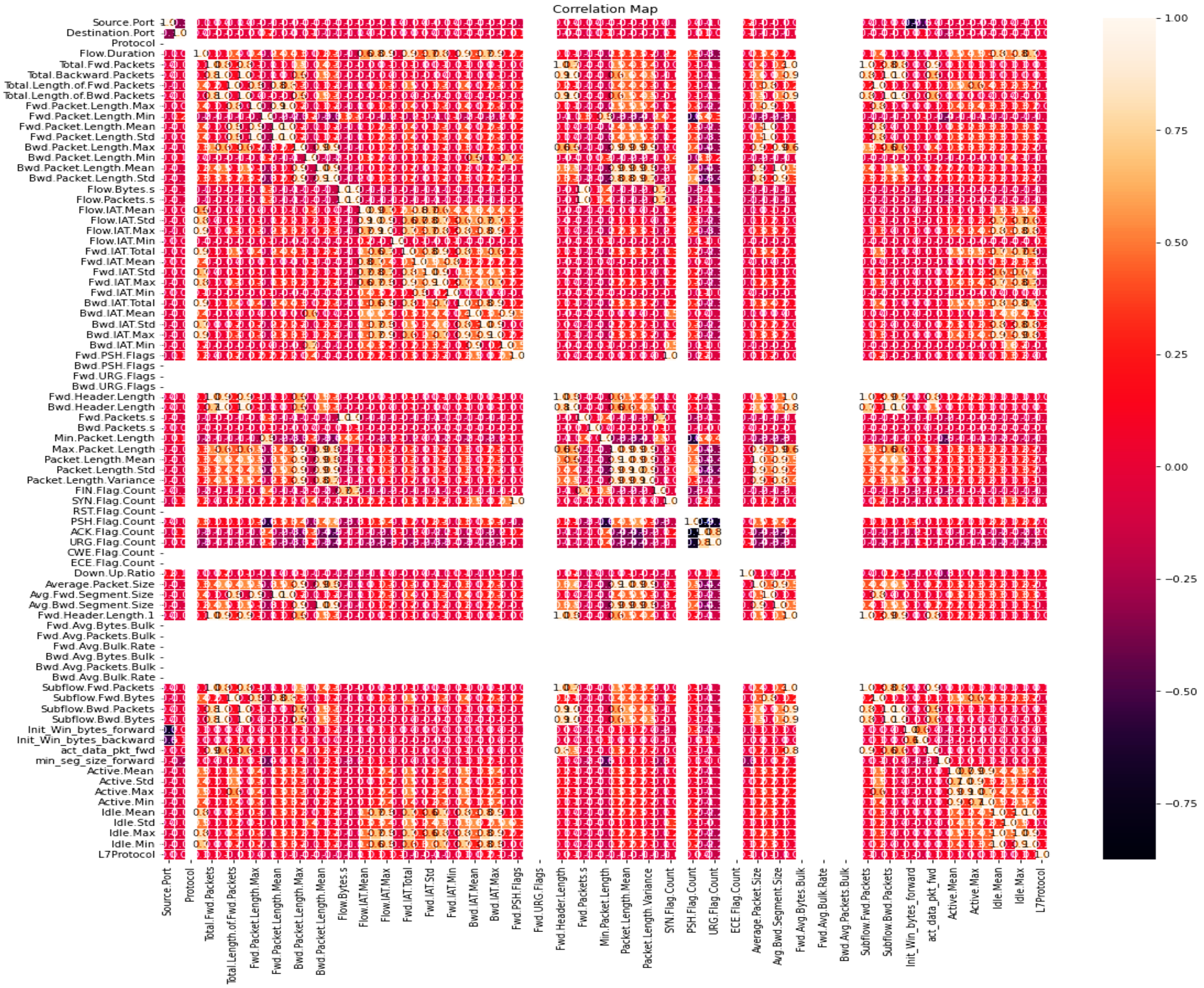

Fig. 7 Correlation matrix of the proposed system

The automotive vehicle systems dataset is considered input to the algorithms for the experiments to validate the normal and abnormal data. By using intelligent simulation tools, classification analysis is performed, and corresponding outcomes are recorded in multiple iterations. It helps to identify the gaps in implementation and optimize the accuracy for better results.

Figure 7 illustrates the correlation coefficients among the variables. The darker portions indicate a higher level of correlation, confirming the prediction of false information. Conversely, the lighter portions indicate a lower correlation, indicating the absence of misleading malware data occurrences in automated vehicles.

Figure 8 presents the rate of prediction and classification using the confusion matrix. This visualization provides insights into the classification process's accuracy, highlighting the proposed model's effectiveness in accurately predicting and classifying data. Figure 8 showcases the combination of actual and predicted values. It is observed that the number of correctly predicted values from both the positive and negative classes is high, indicating the system's superior accuracy. The efficiency and performance of the system are further assessed and calculated in the subsequent section.

4.1.1. Comparative Analysis with Existing Algorithms

The proposed work utilizes a v2v dataset, and this section provides an overview of the computational results obtained from each algorithm. The accuracy, precision, and detection rate outcomes derived from the proposed approach are compared with those of other algorithms in Table 3. This comparison allows for an assessment of the effectiveness and performance of the proposed work with alternative methods.

Table 1. Comparative prediction assessment of the proposed system with existing algorithms [31]

Conventional Methods	Accuracy
Without Ml	78
Convolutional Ml Scheme	82
Proposed ML SVM	94.2
Proposed Method	96.2

Table 1 provides a clear comparative analysis between the proposed algorithm and existing ML techniques, as well as methods without ML and ML with SVM, in terms of accuracy. The present work achieves a significantly higher accuracy value of 96.2, surpassing the accuracy values of 82, 78, and 94.2 obtained by the other methods.

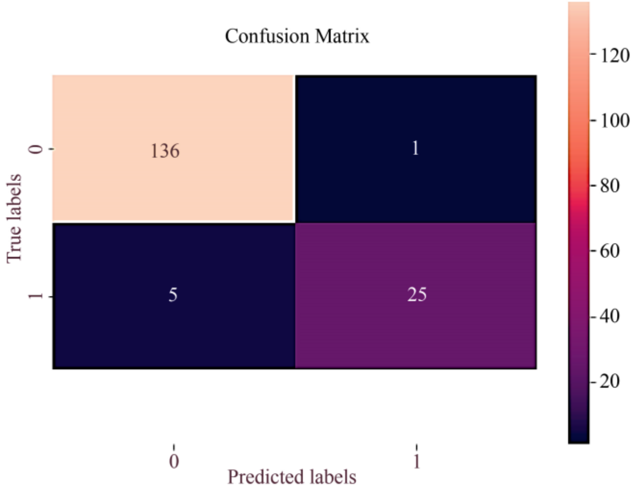

Fig. 8 Confusion matrix

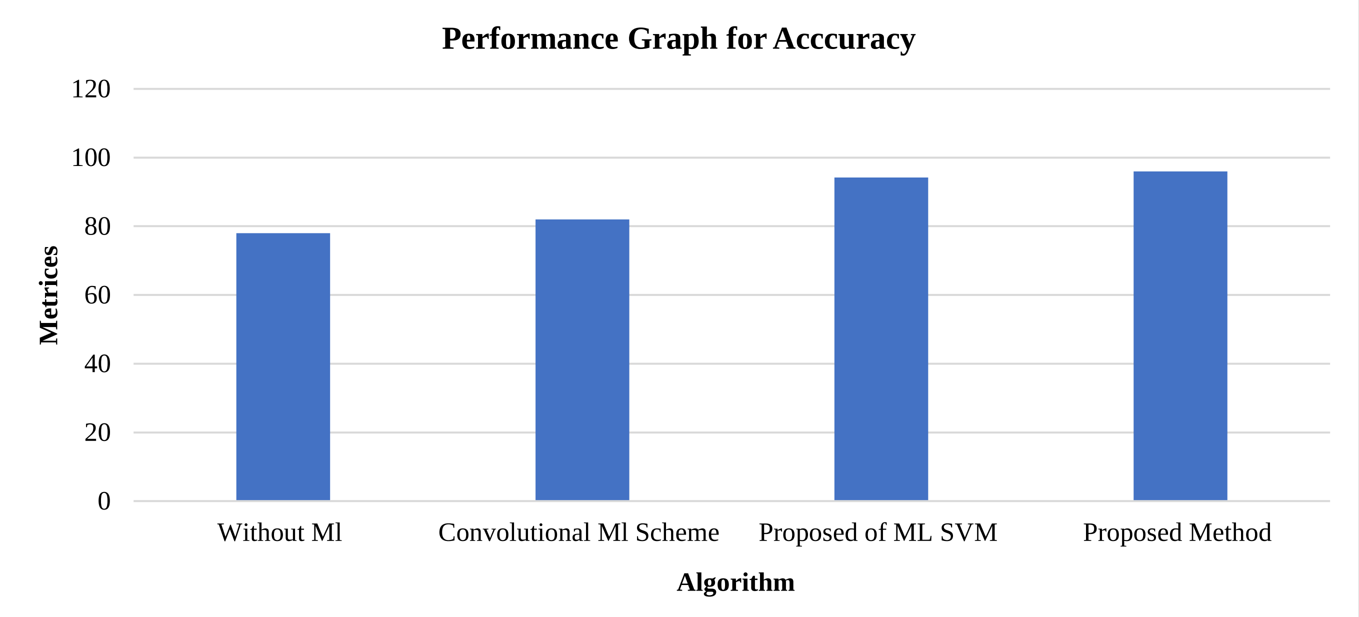

Fig. 9 Graphical Comparative Assessment on Accuracy [31]

Table 2. Accuracy comparison of conventional technique with the proposed work [32]

Method	Accuracy
Existing Method	78
Proposed Method	96.2

Table 3. Comparison of traditional classification algorithms with the proposed method [33]

Methods	Precision	Recall	F1 score
CNN	0.889	0.821	0.854
RegionNet	0.899	0.91	0.904
RF	0.746	0.852	0.795
TFL-CNN	0.906	0.968	0.936
Proposed Method	0.962	0.991	0.984

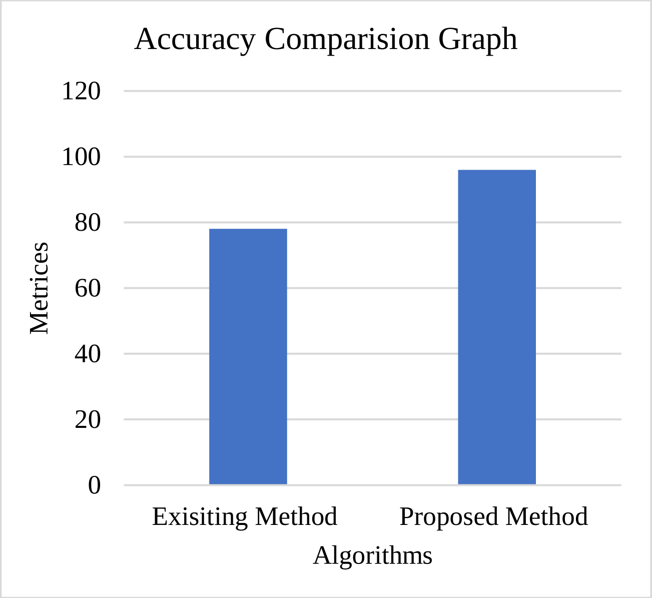

Fig. 10 Accuracy comparison of proposed method with existing method

Figure 9 presents a graphical representation of the performance assessment to further evaluate the performance. This visualization allows for a comprehensive understanding of how the proposed algorithm outperforms the alternative methods in terms of accuracy. Figure 9 demonstrates that the proposed work outperforms other traditional techniques regarding accuracy. The higher accuracy values achieved by the proposed model indicate its efficiency and effectiveness. To further illustrate this, Table 2 compares the proposed model and other conventional algorithms with a detailed accuracy comparison.

According to the findings presented in Table 2, the proposed approach outperforms existing methods in accurately classifying malicious and normal information, achieving an impressive accuracy of 96.2%. These results are visually depicted in Figure 10, providing a graphical representation of the comparative analysis.

Based on the insights gleaned from Figure 10, it is evident that the proposed study excels in terms of accuracy when compared to existing methodologies. This efficiency in accuracy underscores the system's enhanced capability in detecting abnormal events. To provide a comprehensive overview of this comparison, a detailed tabulation of the proposed work's performance against the existing system can be found in Figure 10. The comparative evaluation conducted on existing approaches, as presented in Table 3, showcases the performance metrics of precision, recall, and f1-score. Traditional CNN achieved a precision of 0.889, recall of 0.821, and f1-score of 0.854. RegionNet surpassed these values with a precision of 0.899, recall of 0.91, and f1-score of 0.904. RF yielded a precision of 0.746, recall of 0.852, and f1-score of 0.795. TFL with CNN demonstrated superior performance with a precision of 0.906, recall of 0.968, and f1-score of 0.936. Notably, the proposed prediction model outperformed all conventional approaches, achieving significantly higher values of 0.962 for precision, 0.991 for recall, and 0.984 for the f1-score. A visual representation of this comparative evaluation can be observed in Figure 11.

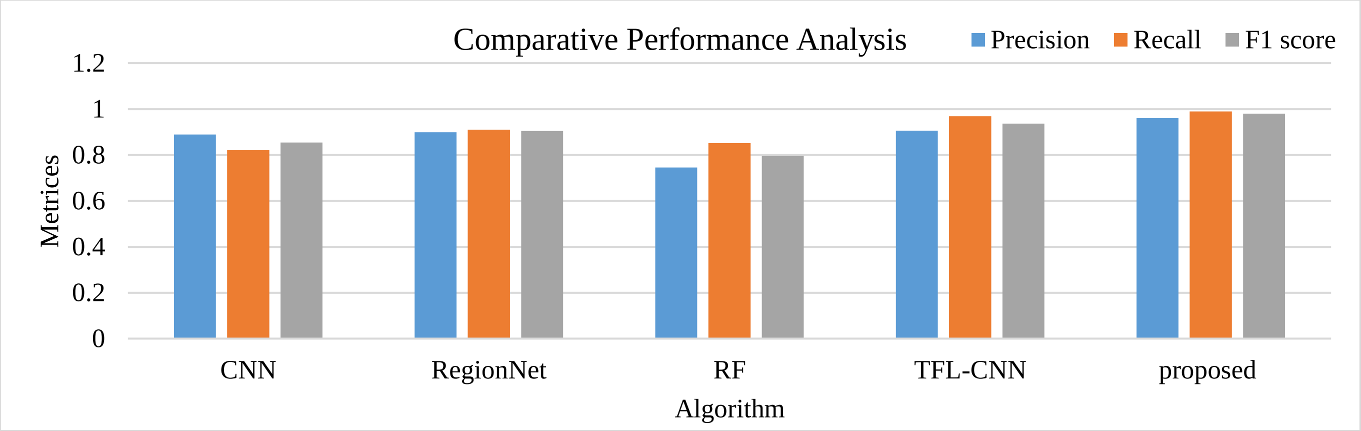

Fig. 11 Graphical analysis of comparison of conventional methods with the proposed method

The comparative analysis of accuracy between the proposed work and existing methodologies highlights the remarkable efficiency of the present system in classifying malware data. Through the integration of RF with Adaboost, the proposed classification approach achieves significantly higher prediction accuracy in distinguishing between normal and abnormal data. The key advantage of this research lies in its ability to deliver precise and accurate predictions, enabling effective decision-making based on the quality of the results. These results showcase the efficiency of the implemented algorithms and provide better security to the automotive vehicular systems. It enhances secure communication in the system by accurately predicting abnormal data.

5. Conclusion

The proposed detection model in the Cyber-Physical System employed a hybridization technique, combining GGA with RF and Adaboost, to enhance accuracy and efficiency. This hybrid methodology was evaluated and compared against traditional algorithms, yielding promising results. The proposed technique, incorporating ML algorithms, demonstrated notable improvements in terms of accuracy, precision, recall, and F1 score. Among the various existing techniques, the classification analysis utilizing GGA-RF-Adaboost on traffic data showcased remarkable efficiency, surpassing the accuracy values achieved by other methods. Pre-processing the input vehicle data and applying feature selection and classification played a crucial role in achieving a high accuracy rate of 96.2% in this research. Experimental evaluations emphasized that integrating the detection algorithm resulted in superior performance in discerning normal and abnormal traffic information from the v2v dataset. The proposed security framework exhibited remarkable efficiency when implemented in surveillance facilities for automated vehicle applications, thanks to its heightened accuracy in detection. Future research endeavors could focus on minimizing the execution time required for the prediction process, thus enabling effective real-time implementation.

Acknowledgments

The authors would like to express our sincere gratitude to the Velse Institute of Science Technology and Advanced Studies for their guidance in successfully completing this research study.

References

[1] Hui Yang et al., “BLCS: Brain-Like Distributed Control Security in Cyber-Physical Systems,” *IEEE Network*, vol. 34, no. 3, pp. 8-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Saraju P. Mohanty, “Advances in Transportation Cyber-Physical System (T-CPS),” *IEEE Consumer Electronics Magazine*, vol. 9, no. 4, pp. 4-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Shuang Jin et al., “Modeling and Stability Analysis of Mixed Traffic with Conventional and Connected Automated Vehicles from Cyber-Physical Perspective,” *Physica A: Statistical Mechanics and its Applications*, vol. 551, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ao Li, Jinwen Wang, and Ning Zhang, “Chronos: Timing Interference as a New Attack Vector on Autonomous Cyber-Physical Systems,” *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2426-2428, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jinbao Wang, Zhipeng Cai, and Jiguo Yu, “Achieving Personalized K-Anonymity-Based Content Privacy for Autonomous Vehicles in CPS,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4242-4251, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Hoang-Dung Tran et al., “Safety Verification of Cyber-Physical Systems with Reinforcement Learning Control,” *ACM Transactions on Embedded Computing Systems*, vol. 18, no. 55, pp. 1-22, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yassine Maleh, *Machine Learning Techniques for IoT Intrusions Detection in Aerospace Cyber-Physical Systems*, Machine Learning and Data Mining in Aerospace Technology, vol. 836, pp. 205-232, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Franco Van Wyk et al., “Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1264-1276, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Nelson H. Carreras Guzman, and Adam Gergo Mezovari, “Design of IoT-Based Cyber-Physical Systems: A Driverless Bulldozer Prototype,” *Information*, vol. 10, no. 11, pp. 1-15, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Khan Sakib Mahmud, “*Connected and Automated Vehicles in Urban Transportation Cyber-Physical Systems*,” Clemson University, pp. 1-141, 2019. [Google Scholar] [Publisher Link]

[11] Bradley Potteiger, Zhenkai Zhang, and Xenofon Koutsoukos, “Integrated Moving Target Defense and Control Reconfiguration for Securing Cyber-Physical Systems,” *Microprocessors and Microsystems*, vol. 73, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Denise Ratasich et al., “A Roadmap toward the Resilient Internet of Things for Cyber-Physical Systems,” *IEEE Access*, vol. 7, pp. 13260-13283, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Sebastian Paul, and Patrik Scheible, “Towards Post-Quantum Security for Cyber-Physical Systems: Integrating PQC into Industrial M2M Communication,” *25th European Symposium on Research in Computer Security, ESORICS 2020*, Guildford, UK, vol. 12309, pp. 295-316, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Johan Olstam et al., “An Approach for Handling Uncertainties Related to Behavior and Vehicle Mixes in Traffic Simulation Experiments with Automated Vehicles,” *Journal of Advanced Transportation*, vol. 2020, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Abdul Rehman Javed et al., “Ensemble AdaBoost Classifier for Accurate and Fast Detection of Botnet Attacks in Connected Vehicles,” *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Adi Alhudhaif et al., “A Particle Swarm Optimization Based Deep Learning Model for Vehicle Classification,” *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 223-235, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yee Mun Lee et al., “Road Users Rarely Use Explicit Communication When Interacting in Today’s Traffic: Implications for Automated Vehicles,” *Cognition, Technology & Work*, vol. 23, pp. 367-380, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ying Liu et al., “CT2-MDS: Cooperative Trust-Aware Tolerant Misbehavior Detection System for Connected and Automated Vehicles,” *IET Intelligent Transport Systems*, vol. 16, no. 2, pp. 218-231, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Sina Nordhoff et al., “A Multi-Level Model on Automated Vehicle Acceptance (MAVA): A Review-Based Study,” *Theoretical Issues in Ergonomics Science*, vol. 20, no. 6, pp. 682-710, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Yiyang Wang, Neda Masoud, and Anahita Khojandi, “Real-Time Sensor Anomaly Detection and Recovery in Connected Automated Vehicle Sensors,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1411-1421, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Manuel Dietrich, “Addressing Unequal Risk Exposure in the Development of Automated Vehicles,” *Ethics and Information Technology*, vol. 23, pp. 727-738, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Malik Mustafa et al., *Multitask Learning for Security and Privacy in IoV (Internet of Vehicles)*, Chapter 12, Autonomous Vehicles Volume 1: Using Machine Intelligence, Wiley Online Library, vol. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Wang Hao, Adam Asrul, and Han Fengrong, “Improving the Efficiency of Customer Requirements Classification on the Autonomous Vehicle by Natural Language Processing,” *International Journal of Computing and Digital Systems*, vol. 9, no. 6, pp. 1213-1219, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Xiaohua Ge et al., “Dynamic Event-Triggered Scheduling and Platooning Control Co-Design for Automated Vehicles Over Vehicular Ad-Hoc Networks,” *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 1, pp. 31-46, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Muammer Eren Sahin, Lo’ai Tawalbeh, and Fadi Muheidat, “The Security Concerns on Cyber-Physical Systems and Potential Risks Analysis Using Machine Learning,” *Procedia Computer Science*, vol. 201, pp. 527-534, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] R.J. Kavitha et al., “Automated Vehicle Theft Prevention Using Atmega Embedded Systems,” *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 5, pp. 901-904, 2020. [Google Scholar] [Publisher Link]

[27] Prinkle Sharma, and Hong Liu, “A Machine-Learning-Based Data-Centric Misbehavior Detection Model for the Internet of Vehicles,” *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4991-4999, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sana Aurangzeb et al., “CyberSecurity for Autonomous Vehicles against Malware Attacks in Smart-Cities,” *Cluster Computing*, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] Lulu Guo et al., “Systematic Assessment of Cyber-Physical Security of Energy Management System for Connected and Automated Electric Vehicles,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3335-3347, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[30] Nelson H. Carreras Guzman et al., “Conceptualizing the Key Features of Cyber-Physical Systems in a Multi-Layered Representation for Safety and Security Analysis,” *The Journal of the International Council on System Engineering*, vol. 23, no. 25, pp. 189-210, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[31] Muhammad Zohaib Anwar, Zeeshan Kaleem, and Abbas Jamalipour, “Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2526-2534, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[32] Rajesh Gupta et al., “Machine Learning Models for Secure Data Analytics: A Taxonomy and Threat Model,” *Computer Communications*, vol. 153, pp. 406-440, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Xiaokang Zhou et al., “Two-Layer Federated Learning with Heterogeneous Model Aggregation for 6g Supported Internet of Vehicles,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5308-5317, 2021. [CrossRef] [Google Scholar] [Publisher Link]