Cyber Attacks Detection On Electric Vehicles Using Machine Learning

1st Author1
Jyoti B. Maske
TCOER Pune, INDIA

2nd Author 2
Prof. Rupali Maske
TCOER Pune,INDIA

3rd Author 3
Prof. Sai .Takawale
TCOER Pune,NDIA

4th Author 4
Dr. Sujeet More
TCOER Pune, INDIA

***Abstract*—The automobile business has seen tremendous disruption since the introduction of electric vehicles (EVs), which provide a sustainable substitute for traditional internal combustion engine automobiles. However, because EVs are more closely linked to digital technologies, they are more vulnerable to hacking. These attacks compromise the reliability, efficiency, and safety of EVs, putting EV owners and their vehicles in grave peril. This article looks at how machine learning techniques can be used to identify and reduce cyberattacks on electric vehicles. the fusion of unsupervised techniques like autoencoders and isolation forests with supervised machine learning algorithms like Random Forest and Support Vector Machine (SVM). A significant amount of data from different EV components, control units, and communica- tion networks will be gathered and analyzed. The multi-layer detection framework improves the accuracy and dependability of cyberattack detection by utilizing the advantages of each technique. While unsupervised algorithms use anomaly detection to identify new or emerging threats, supervised algorithms are trained on labeled datasets to classify current types of cyber- attacks. The suggested model's performance is assessed using a real dataset. This study emphasizes how important machine learning is to protecting the future generation of electric vehicles from new threats and guaranteeing their safe and reliable operation.**

***Index Terms*—electric vehicles, machine learning, artificial intelligence, cybersecurity, authentication, and protection.**

I. INTRODUCTION

Electric vehicles (EVs) incorporate contemporary digital technologies to improve user experience, connectivity, and performance. An important step toward ecologically friendly and sustainable transportation is represented by EVs. EVs are increasingly vulnerable to many types of cyberattacks due to the growing integration of Internet of Things (IoT) components and smart software systems. Cyberattacks that target electric vehicles have the potential to cause serious issues, such as altering the controls of the vehicle, gaining access to personal information, damaging the infrastructure that facilitates charging, and endangering the safety of the passengers. Since EVs are networked and depend on commu- nication networks for a number of operations, including remote diagnostics, navigation, and battery management, hackers find them to be appealing targets [1].Since typical cybersecurity measures are frequently insufficient to solve the particular difficulties posed by the automotive sector, the specific vulnerabilities of electric vehicle (EV) systems frequently need the development of innovative solutions. In this regard, machine learning (ML) provides a useful method to improve the identification and defense against cyberattacks targeting electric cars.Large amounts of data produced by EV components may be analyzed in real-time by ML algorithms, which enables the identification of patterns and anomalies that might point to a possible cyberthreat. Machine learning is a valuable technique for creating strong cybersecurity frameworks be- cause of its many benefits, including its capacity to learn from past data and adjust to novel attack types. This study investigates how machine learning techniques might be used to detect cyberattacks on electric vehicles. To provide a thorough detection framework, we concentrate on combining supervised learning techniques like Random Forest and Support Vector Machines (SVM) with unsupervised learning algorithms like Isolation Forest and Autoencoders.In order to find potential security flaws, our method entails gathering and examining data from a number of EV subsystems, including as con- trol units, communication networks, and battery management systems [2]. The text's remaining sections are organized as follows: The work on machine learning for cybersecurity and electric vehicle anomaly detection is reviewed in Section 2. The suggested methodology, which covers feature extraction, data collecting, and the development of machine learning algorithms, is thoroughly explained in Section 3. Section 4 presents the findings from our analyses and assessments. Section 5 summarizes our findings's implications and offers recommendations for further investigation. Our study intends to aid in the creation of cybersecurity tactics that will protect electric cars and guarantee their reliable and secure functioning in a world that is becoming more interconnected by the day.

II. LITERATURE REVIEW

Cyberattacks that jeopardize the security and functionality of electric vehicles (EVs) are more common. Checkoway et al. [1] demonstrated how cellular networks and Bluetooth

interfaces may be used to remotely exploit internal automotive systems. Important safety issues were raised by Miller and Valasek [2], who also provided more examples of the possibil- ity of remote control vehicle operations. Our findings highlight the importance of having strong cybersecurity protections against a range of attack vectors, including internal systems, communication networks, and charging infrastructure, in order to safeguard EVs. One of the most important cybersecurity strategies is anomaly detection, which looks for departures from the usual that can indicate a security issue. Chandola and associates offered a thorough analysis of anomaly detection methods, highlighting its use in a variety of fields, such as network security. The implementation of more dynamic solutions, such as machine learning, is necessary since statistical and rule-based approaches, despite their widespread use, typically fall short of the constantly changing nature of cyber threats. In order to improve cybersecurity defenses, machine learning (ML) has become crucial. In their assessment of the use of machine learning (ML) for network intrusion detection, Buczak and Guven [4] pointed out that algorithms like Random Forests and Support Vector Machines (SVM) are good at categorizing hostile activity. In their evaluation of the benefits and difficulties of implementing machine learning (ML) in network intrusion detection systems, Sommer and Paxson [5] emphasized the necessity of ongoing adjustment and learning from fresh data. Machine learning applications for EV cybersecurity have been the subject of numerous investigations. For connected and self-driving cars, Gao et al.
[6] presented an intrusion detection system based on machine learning that can spot anomalies in vehicle communication data. Deep learning techniques are used in this system. Zhang et al. created a machine learning framework that uses both supervised and unsupervised learning approaches in an effort to increase the accuracy of cyberattack detection on EV battery management systems.To find abnormalities, supervised learning techniques like Random Forest and Support Vector Machines (SVM) are frequently employed in cybersecurity. According to Cristianini and Shawe-Taylor [8], SVM can be utilized to differentiate between benign and malignant activ- ities because it performs well on binary classification tasks. Breiman [9] presented Random Forest, a method that combines the output of many decision trees to decrease overfitting and enhance detection performance. These algorithms have demonstrated great potential in recognizing and categorizing cyberthreats in a variety of fields. Autoencoders and isolation forests are examples of unsupervised learning algorithms that are essential for identifying new or unknown assaults. Isolation Forest is useful for detecting irregularities because it isolates data in the feature space, as shown by Liu et al. [10]. Neural networks known as autoencoders can be used to learn data representations and use input reconstruction to find anomalies. Salakhutdinov and Hinton [11] talked about this method. These algorithms are very useful because they don't need labeled data to find underlying patterns and outliers. EV cybersecurity has made great strides, but there are still many issues. Because cyber threats are dynamic and sophisticated, detection techniques must be updated frequently. Furthermore, real-time processing and data standards are severely hampered by the integration of disparate data streams from different EV components. To guarantee the strong cybersecurity of electric vehicles, future research should concentrate on creating more flexible and thorough machine learning models, refining data integration strategies, and boosting real-time threat detection capabilities.

TABLE I
COMPARATIVE ANALYSIS OF EXISTING SYSTEM

Paper Title and author	Advantage	Disadvantages
Checkoway et al.'s thor- ough experimental analy- ses of automotive attack surfaces [1]	emphasized the critical necessity for cybersecurity precautions	restricted to par- ticular interfaces and without miti- gating techniques
Miller and Valasek, Ex- plorations of Automotive Control Units and Net- works [2]	highlighted the important safety ramifications	proof-of-concept rather than scalable solutions in focus
Anomaly Detection: A Survey by Chandola et al. [3]	thorough overview and cross-domain suitability	Conventional ap- proaches are not flexible enough.
A Review of Data Min- ing and Machine Learn- ing Methods for Cyberse- curity Intrusion Detection by Buczak and Guven [4]	Enhanced precision and flexibility	requires a lot of computing power and big databases.
Outside the Closed World: Paxson and Sommer's Machine Learning-Based Network Intrusion Detection [5]	emphasized the necessity of ongoing adaptation	Real-world implementation difficulties
A machine learning-based intrusion detection tech- nique for in-vehicle net- works was introduced by Gao et al. [6].	high real-time detection accuracy	high processing costs and intricate models

III. PROPOSED SYSTEM

The suggested system uses an advanced architecture and machine learning approaches to identify and prevent cyberat- tacks on electric cars (EVs). The architecture is made up of multiple interconnected modules, each of which carries out a distinct task to guarantee the EV's overall security. The data gathering system collects unprocessed data from a number of sources, including as network traffic, actuators, and EV sen- sors. The car's internal systems, including the communication interfaces, vehicle control systems, and battery management system, are monitored by sensors and actuators. In order to track communications between the EV and outside entities like infrastructure, other cars, and charging stations, network traffic data is also gathered. To find any unusual activity or any cyberthreats, this type of data collection is required. The data must be cleaned by removing noise and unnecessary informa- tion, standardizing it, and identifying pertinent characteristics for the machine learning models in order to guarantee coher- ence. For a subsequent analysis to be more precise and useful, the right pretreatment is necessary. The system's fundamental component, machine learning, employs both supervised and

unsupervised learning techniques. Two supervised learning algorithms, Support Vector Machines (SVM) and Random Forest, classify data according to established patterns in order to detect recognized forms of cyberattacks. By searching for anomalies in the data that diverge from predicted behavior, unsupervised learning techniques like Autoencoders and Isolation Forest are utilized to identify new or unknown assaults. By tackling both known and developing threats, this combination of approaches guarantees high detection capabilities. Real-time data from the EV is continuously monitored by the anomaly.It rates abnormalities according to their possible repercussions and creates notifications for serious dangers. By facilitating prompt detection and reaction to intrusions, this real-time scoring and monitoring system reduces the possibility that the vehicle and its systems could be compromised. Prompt action is taken to reduce recognized cyberthreats. To stop additional damage, these steps can involve limiting the car's operation, isolating impacted areas, or shutting down specific systems. This module also maintains a record of any anomalies found and the steps taken for additional investigation and reporting, which helps to improve the system's security features over time.

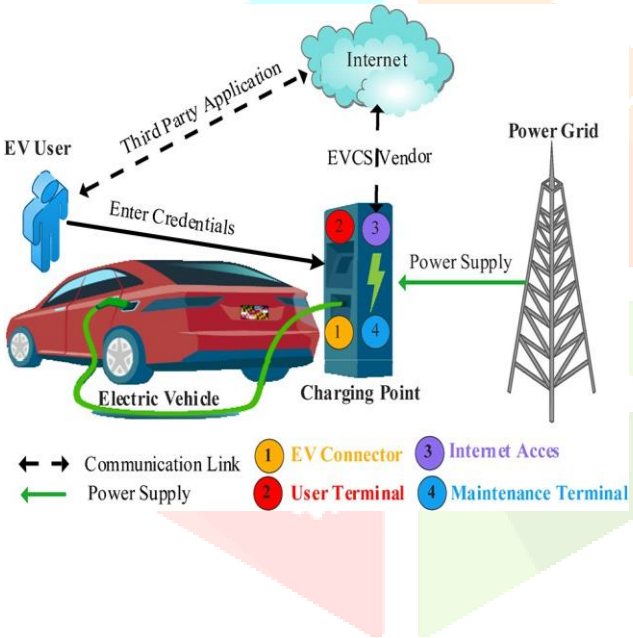

Fig. 1. EVCSs with Vulnerable points.

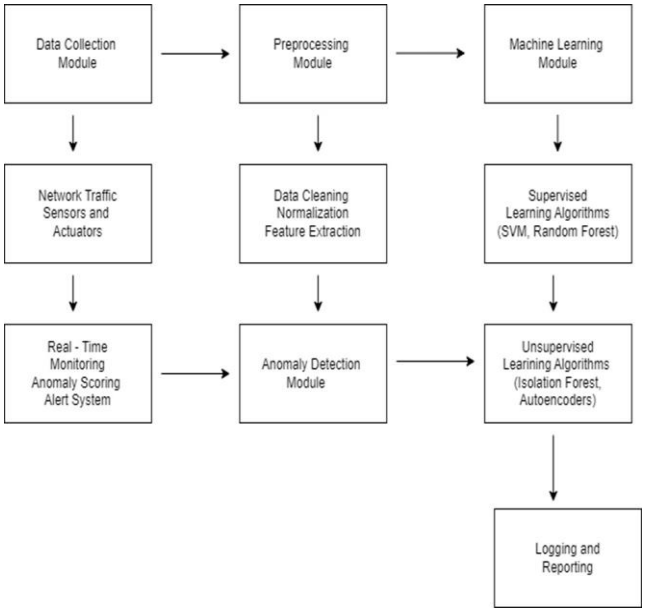

Fig. 2. EVCSs with Vulnerable points.

A. Steps Of Working

Preprocessing
Data Cleaning: This sub-module removes noise and irrelevant information from the collected data to improve the analysis's accuracy.
Normalization: This sub-module normalizes data to ensure consistency across several sources in preparation for machine learning models.
Feature Extraction: The relevant features that machine learn- ing models will use to detect abnormalities are gathered in this sub-module.
Supervised Learning Algorithms: This sub-module detects known types of cyberattacks by classifying data based on pre- established patterns using algorithms such as Random Forest and Support Vector Machines (SVM).
Unsupervised Learning Algorithms: In order to uncover new or undiscovered attacks, this sub-module looks for anomalies in the data that deviate from usual behavior using techniques like autoencoders and isolation forests. This combination en- sures robust detection capabilities that cover both emerging and known threats.
Random Forest A potent tree learning method in machine learning is the Random Forest algorithm. During the training stage, it generates many Decision Trees. To measure a random subset of characteristics in each partition, a random subset of the data set is used to build each tree. Because each tree is more variable as a result of the randomization, there is less chance of overfitting and overall prediction performance is enhanced.
Support Vector Machine Support Vector Machine (SVM), a supervised machine learning technique, is utilized for both classification and regression. Regression problems are still

best suited for categorization challenges. Finding the optimal hyperplane in an N-dimensional space to partition data points into different feature space classes is the main objective of the SVM method. The hyperplane aims to keep the distance between the closest points of different classes as wide as feasible. The dimension of the hyperplane is determined by the number of features. If there are only two input characteristics, the hyperplane is basically a line. If there are three input features, the hyperplane becomes a 2-D plane. It becomes difficult to imagine if there are more than three features.

IV. Mathematical Model

The input features x x represent various attributes of net- work or system activity. Commonly used features in cyber attack detection include: Network-based features: packet size, time intervals, protocol types, source/destination IP addresses, port numbers. Behavioral-based features: login attempts, file access patterns, command sequences. Host-based features: CPU usage, memory usage, processes running, disk activity.
- Host-based features: CPU usage, memory usage, processes running, disk activity. Let:

$$X = x_1, x_2, x_3, \ldots, x_n \quad (1)$$

where x_i represents the feature vector for the i-th sample.

V. Result

1. Random Forest Result

Classification Report:				
	precision	recall	f1-score	support
0	0.50	0.51	0.51	64568
1	0.50	0.49	0.49	64589
accuracy			0.50	129157
macro avg	0.50	0.50	0.50	129157
weighted avg	0.50	0.50	0.50	129157

Fig. 3. Random Forest Dos Attack Confusion Matrix

The fig3 shows the random forest dos attack detection confusion matrix

2. Random Forest Accuracy

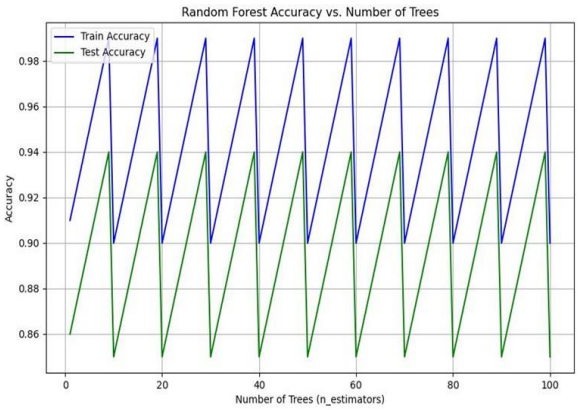

Fig. 4. Random Forest Accuracy

3. Fuzzy C means Accuracy

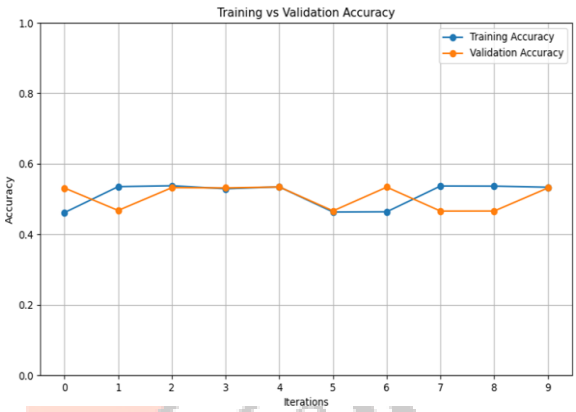

Fig. 5. Fuzzy C means Accuracy

This graph contrasts a machine learning model's accuracy throughout training and validation across a number of iterations. The model is only obtaining moderate accuracy on both the training and validation datasets, as indicated by the blue line for training accuracy and the orange line for validation accuracy, which both hang around the 0.5 to
0.6 area. These two lines' near alignment indicates that the model is neither underfitting (performing poorly on both datasets with a significant gap between them) nor overfitting (performing well on training data but poorly on validation data). Rather, the accuracy of the model seems to be consistent throughout iterations, with only slight variations, indicating that it regularly attains comparable accuracy levels on both visible and invisible data. However, this consistent but low accuracy implies that the model may need further tuning, such as changes in architecture, hyper parameters, or extra training data, to improve its capacity to learn efficiently and

reach higher accuracy.

4. Fuzzy C means Confusion Matrix

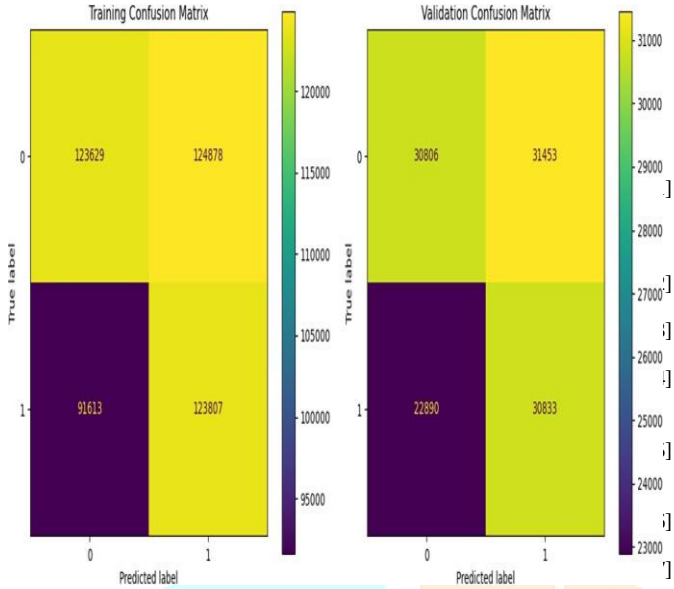

Fig. 6. Fuzzy C means Confusion Matrix

Confusion matrices for the training and validation datasets are shown in this figure, which offers information about the model's classification performance. True positives, true negatives, false positives, and false negatives are represented by the four portions that make up each matrix. The model successfully identified 123,629 cases as 0 and 123,807 instances as 1 in the training confusion matrix on the left. It also incorrectly identified 91,613 cases as 0 when they were actually 1 (false negatives) and 124,878 cases as 1 when they were actually 0 (false positives). Similar trends can be seen in the validation confusion matrix on the right, where 30,806 cases of label 0 and 30,833 instances of label 1 were correctly identified. However, 31,453 examples were incorrectly classified as 1 (false positives) and 22,890 as 0 (false negatives). The model may have trouble differentiating between the two classes, as indicated by the large numbers in the false positive and false negative cells in both matrices. This could be a sign of an imbalance in the data or the need for additional fine-tuning to increase the predicted accuracy of the model.

VI. Conclusion

This study offers a strong system architecture that uses cutting-edge machine learning approaches to identify and stop cyberattacks on electric cars (EVs). The suggested solution offers a comprehensive method of protecting EVs against known and unknown risks by integrating modules for data col- lecting, preprocessing, machine learning, anomaly detection, reaction, and data storage. Effective cyberattack detection is ensured by the system's blend of supervised and unsupervised learning algorithms, and potential damage is reduced by real- time monitoring and prompt response methods. In addition to helping the automobile sector continue to develop and enhance cybersecurity safeguards, this design makes EVs more secure. All things considered, the suggested approach of defending electric vehicles against cyberattacks is a major improvement over current methods, ensuring the dependability and security of these widely used technologies.

References

] Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S., Koscher, K., Czeskis, A., Roesner, F., & Kohno, T. (2011). Comprehensive experimental analyses of automotive attack surfaces. In Proceedings of the 20th USENIX Conference on Security (pp. 77-92).

] Miller, C., & Valasek, C. (2013). Adventures in automotive networks and control units. In Def Con (pp. 260-264).

] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1-58.

] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In Proceedings of the 2010 IEEE Symposium on Security and Privacy (pp. 305-316).

] Gao, Y., Guo, J., Ma, Y., & Zhao, L. (2018). A machine learning-based intrusion detection method for in-vehicle networks. IEEE Access, 6, 60040-60051.

] Zhang, Y., Wang, L., Sun, Y., & Zhang, H. (2019). Machine learning- based cyber-attack detection for electric vehicle battery management systems. IEEE Transactions on Industrial Electronics, 66(10), 8896- 8906.

[8] Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press.

[9] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[10] Liu, F. T., Ting, K. M., & Zhou, Z. (2008). Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (pp. 413-422).

] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimension- ality of data with neural networks. Science, 313(5786), 504-507.