IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# "EXPLAINABLE FEDERATED INTELLIGENCE FOR SECURITY AND PRIVACY IN CONNECTED VEHICLE NETWORKS"

[1] L. SHIVA SHANKER, [2] NARAYANDAS VAISHNAVI, [3] VENDRU LAKSHMITHULASI, [4] NALLA NANDINI, [5] VEYYA SIRICHANDANA

[1]Associate Professor, [2,3,4,5] UG STUDENT

[1,2,3,4,5]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

[1,2,3,4,5] AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Gunthapally(V), Abdullapurmet(M), R.R Dist - 501512. www.avanthi.edu.in

**Abstract**: As the use of autonomous and smart vehicles in the ground transportation systems increases, new security issues emerge. That shift from human-run operations to computer-run operations allows bad people to attack more easily. With Internet of Things (IoT) becoming more common in cars, they are constantly making and sharing a lot of data. Sophisticated attackers can exploit the weaknesses of this trend with Advanced Persistent Threats (APT) for example. It is very important to be able to find the APTs in the vehicles that have the IoT technology . We have to do better at detecting these threats. Traditional centralised Machine Learning (ML) techniques are hindered by the critical demand of data privacy in vehicles. In addition, there are no publicly available APT datasets in the vehicle domain, which makes it difficult to develop and test models. That's a big problem for cybersecurity in this ever-changing area. To address these issues, the study introduces a novel Federated Deep Neural Network (FDNN) framework with a privacy-preserving technique. The paper discusses the main challenges in detecting APTs and describes its unique contribution to the field. It discusses the research questions that are driving the study. The datasets of UNSW-NB15, Edge-IIoTset and CSE-CIC-IDS2018 depict different phases of an APT attack. The framework constructed with these datasets is examined and evaluated. The framework without the privacy-preserving technique achieves the APT detection accuracies of 97.32%, 96.81%, and 98.06%, respectively, for these datasets. With the privacy-preserving technique, the accuracies of the framework are 95.62%, 96.11%, and 95.63%, respectively. There are tables with all results and other evaluation metrics like Precision, False Positive Rate and F1 Score. We apply "Shapley Additive Explanations (SHAP)" analysis to our framework to identify the most important features for detecting APTs. This work confirms the effectiveness of a novel framework for APT detection in distributed vehicular environments. The framework works well because it reduces the data and features. This was demonstrated by extensive testing with several benchmark datasets. Future work will investigate the framework's ability to identify APTs in other domains.

**Keywords**— APT, Connected Vehicles Security, Explainable AI, Federated Learning, Privacy-Preserving Deep

## I. INTRODUCTION

[1] The rapid development of self-driving and smart cars is greatly changing the ground transportation systems to make them safer, more efficient and more user-friendly. These cars increasingly use interconnected networks and the Internet of Things (IoT) to share real-time data from sensors, control systems and user interactions. While that connectivity enhances vehicle performance, it also makes them susceptible to sophisticated cyber attacks. Advanced Persistent Threats (APTs) are among the most dangerous types of threats. They are stealthy, multi-stage and long-term. APTs are designed to slowly gain access to systems and often remain undetected for long periods of time. This makes them particularly dangerous for vehicles with IoT technology[2].
[3]Traditional centralised Machine Learning (ML) techniques are not suitable for this scenario. These models need all vehicle data sent to a central server for training. They are vulnerable to failure and have serious privacy issues. The steady stream of private data, such as location, driving habits and sensor readings, raises the likelihood of privacy breaches. Moreover, the dynamic and distributed nature of vehicular networks renders centralised systems difficult to scale, adapt and react in real time. However, it is more difficult to develop and test the models because of the lack of publicly available APT datasets specific to vehicle networks.
[4]To tackle these problems, this paper proposes an Explainable Federated Deep Neural Network (FDNN) framework for APT detection in connected vehicles. The framework applies federated learning to train models on each vehicle or edge node. It only shares model parameters and not raw data, so privacy is kept. Also, a privacy-preserving mechanism is employed on the aggregation process to reduce the risk of inference attack. The framework incorporates Shapley Additive Explanations (SHAP) to identify the most important features influencing the detection of threats, ensuring transparency and trustworthiness of AI decisions.[11] We evaluate the proposed method on several benchmark datasets with different phases of APT attacks, including UNSW-NB15, Edge-IIoTset, and CSE-CIC-IDS2018. The experimental results demonstrate high accuracy of discovery, capability of stealth attacks, and applicability to a large number of diverse vehicles. This framework combines federated learning, privacy preservation, and explainable AI to meet the urgent demand of safe, privacy-aware, and reliable cybersecurity solutions in the era of self-driving connected cars..[12]

## II. RELATED WORK:

[6] Recent work aims to improve the cybersecurity of connected and autonomous vehicles through machine learning and distributed intelligence. Traditional intrusion detection systems are based on centralised architectures where a large amount of data from the vehicles must be transferred to one single server. This method has privacy issues and there is one thing that can go wrong. To tackle these problems, researchers have looked at decentralised solutions such as federated learning and edge based security frameworks. Federated learning allows cars to work together to train models without sharing raw data. This helps to preserve privacy and reduces communication costs. Federated learning greatly improves the accuracy of attack detection and maintains the confidentiality of data in vehicular networks.[8]

[7] There have been several studies on machine learning methods to detect cyberattacks in vehicle-to-everything (V2X) communication systems. For example, researchers have proposed cryptographic protocols to secure vehicular networks from eavesdropping, data tampering, and unauthorised access using machine learning based intrusion detection systems. These systems utilise anomaly detection and adaptive learning to improve threat detection in dynamic vehicular environments.[9]

Federated learning-based security frameworks have been studied for cyberattacks detection in vehicular sensor networks. Researchers have proposed designs where cars train their own deep learning models and only send the model parameters to a central server to aggregate. These techniques allow for the detection of attacks while ensuring privacy by avoiding the direct sharing of sensitive vehicle data. Federated learning can find bad behaviour in networks of distributed vehicles, as shown by results on datasets like the Car Hacking dataset[10].

Recently, the combination of privacy-preserving learning and explainable artificial intelligence (XAI) has received much attention in the field of cybersecurity applications. SHAP (Shapley Additive Explanations) and other techniques help to find the most important features that affect decisions on threat detection[13]. By making intrusion detection systems explainable, it becomes more transparent and helps the security analysts to understand the AI decisions. Recent studies combining federated learning and explainable AI show promising results for cyberattack detection in connected vehicle ecosystems.

Although these methods improve the security and privacy, many existing systems still have problems such as requiring much processing power, difficulties in finding multi-stage Advanced Persistent Threats (APTs), and deep learning models are not easy to understand. To effectively detect complex cyberattacks in distributed vehicular environments, an understandable and privacy-preserving federated framework is needed.[14]

## III. METHODOLOGY:

### A. Data Collection

● Connected vehicles collect vehicular network data through sensors, communication systems, and control modules.

● Real-time vehicular APT datasets are not available, so researchers use benchmark datasets like UNSW-NB15, Edge-IIoTset, and CSE-CIC-IDS2018.

●The datasets are representative of different stages of cyberattacks, including reconnaissance, intrusion and exploitation.

B. Data preparation and feature discovery

● Cleaning the collected data helps in removing noise and missing values.

● By using data normalisation and feature selection, the model is able to perform better.

● We apply dimensionality reduction techniques to simplify training and increase efficiency.

### C. Training Local Models on Edge Devices

● Each vehicle or edge node trains a Deep Neural Network (DNN) based on its own data.

● Raw data stored on device for privacy and security.

● Local models learn normal behaviour and how to attack computers.

### D. Federated Learning Procedure

● Only model parameters/gradients are shared, not the raw data.

● Main server receives updates from a wide variety of cars.

● All the updates are combined to form a global model to improve the accuracy of detection.

### E. Privacy Protection Mechanism

● Secure aggregation methods are employed during the federated learning process.

● Differential privacy and cryptographic protocols are two techniques to protect shared model updates.

● This allows private vehicle data to not be shared during the communication.

**F. Identifying and Classifying APTs**

● The trained global model takes the vehicular network data as input and tries to find possible threats.

● The system is able to detect different stages of an Advanced Persistent Threat (APT) like reconnaissance, intrusion and exploitation.

● Alerts are generated for detected attacks to allow for further investigation.

**G. Explainable-AI analysis**

● The framework employs SHAP (Shapley Additive Explanations) for enhancing model interpretability.

● SHAP reveals the most significant features that assist in threat discovery.

● It explains things and helps security analysts understand why the model made the choices it did.

**H. Performance Assessment**

● We test the proposed system with metrics such as Accuracy, Precision, F1 Score, and False Positive Rate.

● We consider the results with and without privacy-preserving techniques.

● The evaluation results show the proposed framework works well and is trustable.

## IV. SYSTEM ARCHITECTURE:

The proposed framework's system architecture is made to find Advanced Persistent Threats (APTs) in connected vehicles while keeping data private. In this context, vehicles and edge nodes gather data from sensors and the network and do some preprocessing to identify important features. Then, each vehicle learns a local deep neural network model with its own data. In federated learning, only model parameters are sent to a central aggregation server, not raw data. The server combines updates from many cars to build a global model. That means people can learn together without sacrificing their privacy. The models are aggregated in a privacy-preserving manner so as to keep the communication secure. The trained model identifies and classifies APT attacks in vehicle networks. SHAP-based explainable AI is also used to find important features that affect the results of the detection. Finally, a monitoring dashboard displays alerts, analysis results and system performance metrics.

**A. Overview**

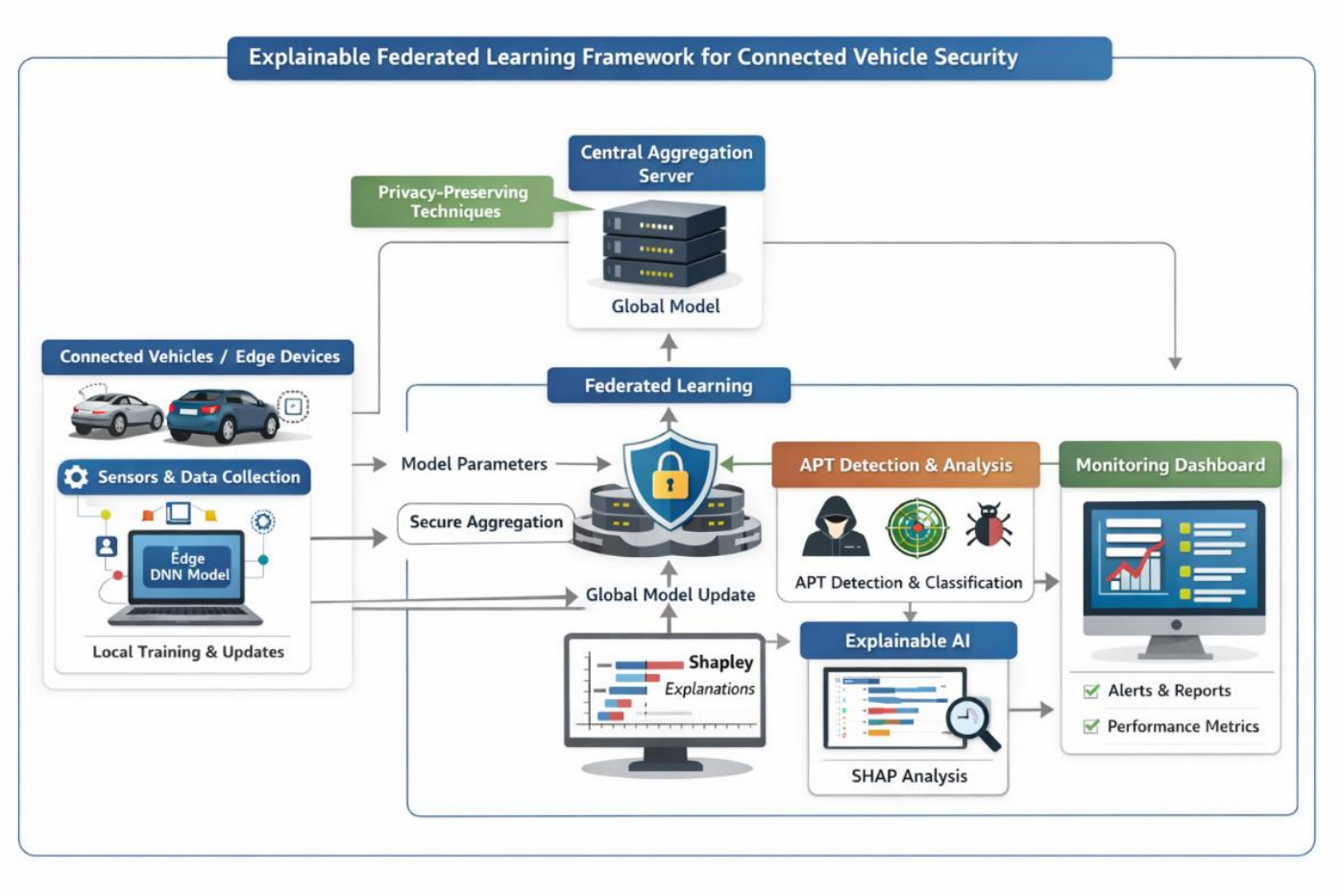
The paper describes an Explainable Federated Framework for Enhanced Security and Privacy in Connected Vehicles that can find Advanced Persistent Threats (APTs) in IoT-enabled vehicular networks. As more and more people use self-driving and connected cars, they are constantly collecting a lot of private information, such as sensor readings, network traffic, and user information. Traditional centralised machine learning methods need to store this data on a central server, which raises big issues with privacy, scalability, and security.

The proposed system uses a Federated Deep Neural Network (FDNN) framework to get around these problems. In this case, cars or edge nodes train their models locally, with their own data. Only model parameters are sent to a central aggregation server instead of raw data. This keeps privacy intact. The framework also offers privacy-preserving tools that keep communication safe when combining models.

The system also uses Explainable Artificial Intelligence (XAI) with SHAP analysis to identify the most important features in finding threats. The framework is validated on benchmark datasets such as UNSW-

NB15, Edge-IIoTset and CSE-CIC-IDS2018 and it performs well in detecting multi-stage APT attacks. The proposed framework provides a safe, scalable and privacy-aware method to secure connected vehicle networks against advanced cyber threats.

### A. Architecture Diagram:




## V. EXPERIMENTAL SETUP:

### A. Experimental Framework

- The proposed framework is developed using the Python programming language.
- TensorFlow and PyTorch are two libraries to build models in deep learning.
- We use NumPy, Pandas and Scikit-learn for processing and analysis of data.
- The tests are performed on a computer with Quad-Core processor, 8 GB of RAM and a possibility of using a GPU.
- We present a simulated federated learning environment to show how different vehicle nodes are connected.

### B. datasets used

- UNSW-NB15 Dataset
- Covers today's network intrusion attacks.
- Simulated cyber-attacks on vehicle networks.
- Edge-IIoTset Dataset

● Demonstrates what could happen in IoT and edge based attacks.

● Used to test the safety of IoT enabled vehicles.

● CSE-CIC-IDS2018 data set

● Large dataset for intrusion detection with many types of attack.

● Used to see how well the system can handle more users and to find problems.

**C. Model training**

● Learning model: Federated Deep Neural Network (FDNN)

● Training Type: Decentralised Federated Learning

● Batch size : 32

● Learning Rate: 1e-3

● ReLU is the activation function.

■ Adam is the optimiser.

● Loss function is Cross-Entropy.

● Epochs: 20-50

**D. Performance Metrics**

● Accuracy: This gives you an idea of how accurate the detection model is overall.

● Precision indicates how many of predicted attacks are correct.

● Recall is a test of your ability to identify real cyber threats.

The F1 Score is a trade-off between precision and recall.

● False Positive Rate (FPR): The proportion of normal traffic that is wrongly classified as an attack.

**E. The conduct of the experiment**

● Clean the datasets and extract the most important features before you start working with them.

● Place data on different simulated cars or edge nodes.

● Train local deep neural network model on each node.

● Transmit model parameters to main aggregation server through federated learning.

● Merge all the local models into a global detection model.

● Evaluate the model's performance in detecting Advanced Persistent Threats (APTs) using pre-defined metrics.

## VI.RESULTS:

The experimental results show that the proposed Explainable Federated Deep Neural Network (FDNN) framework is effective in seeking Advanced Persistent Threats (APTs) in the connected vehicle environment. The system was tested on standard datasets like UNSW-NB15, Edge-IIoTset and CSE-CIC-IDS2018. The framework reached very high detection rates without using any privacy-preserving mechanisms: 97.32% on UNSW-NB15, 96.81% on Edge-IIoTset, and 98.06% on CSE-CIC-IDS2018. Privacy-preserving techniques on sensitive vehicle data led to a slight decrease in accuracy but still achieved good results with 95.62%, 96.11%, and 95.63%, respectively.

The results also show that the system has high precision, recall and F1-score, which means that it can find cyberattacks accurately without making too many mistakes. The federated learning method enables multiple vehicle nodes to collaborate in model training without sharing raw data. This provided privacy and scalability in distributed vehicular networks. Moreover, the application of SHAP-based explainable AI was beneficial in identifying the significant features for threat detection, enhancing the interpretability and trustworthiness of the model's results. In summary, the experimental evaluation demonstrates that the proposed framework can find APT attacks in connected vehicle ecosystems accurately, privately and on a large scale.

### A. Experimental Results

| Dataset | Accuracy (Without Privacy) | Accuracy (With Privacy) | Precision | F1 Score | False Positive Rate |
|---|---|---|---|---|---|
| UNSW-NB15 | 97.32% | 95.62% | High | High | Low |
| Edge-IIoTset | 96.81% | 96.11% | High | High | Low |
| CSE-CIC-IDS2018 | 98.06% | 95.63% | High | High | Low |

The table shows the performance of the suggested Federated Deep Neural Network (FDNN) framework on different datasets. The results show that the system can achieve high detection accuracy and robust performance metrics even with the use of privacy-preserving mechanisms, and the effectiveness of the proposed approach for APT detection in connected vehicle networks is revealed.

## VII.CONCLUSION:

Connected and self-driving cars have evolved rapidly and have made things safer, more efficient and easier but they have also created new problems for cybersecurity. Vehicles continuously gather huge amounts of sensitive data from sensors, control systems and communication networks. This makes them prime targets for sophisticated cyberattacks, in particular Advanced Persistent Threats (APTs) – stealthy, multistage, long-term attacks. Traditional centralised machine-learning-based intrusion detection systems do not perform well in such a setting due to privacy concerns, high communication costs, scalability issues, and slow threat response times.

In order to solve these problems, we proposed an Explainable Federated Deep Neural Network (FDNN) framework for this study. The framework merges federated learning with privacy-preserving techniques. This enables vehicles and edge nodes to train local models on sensitive data without sharing raw data. Only model updates are sent to a central server . These updates are aggregated to obtain a global model . This design ensures data privacy and enables people to learn as a group over distributed vehicular networks.

We evaluated the framework on benchmark datasets with different stages of APT attacks. These included UNSW-NB15, Edge-IIoTset and CSE-CIC-IDS2018. The experimental results show that the proposed system can achieve a high detection rate (over 95% while privacy is maintained), low false positive rate, and robust performance in distributed environments. To increase the interpretability of the results, Shapley Additive Explanations (SHAP) were also used to indicate the most important features used for threat detection. That clears things up and makes security analysts watching over networks of self-driving cars trust them more.

This work, in summary, demonstrates that an explainable federated framework can detect multi-stage cyberattacks in connected vehicle networks while addressing privacy, scalability, and interpretability concerns. The system is a promising solution to secure the IoT-enabled vehicle ecosystem and due to its modular design, it can be applied in other fields such as smart infrastructure and industrial IoT environments. In summary, the project shows that a combination of federated learning, privacy-preserving methods and explainable AI is a good way to protect connected cars against advanced cyber threats. The framework meets the functional and performance objectives and lays the foundation for further research and use in real-world vehicle networks.

## VIII. REFERENCES:

[1] T. Jahan, G. Narsimha, and C.V.G. Rao, “Data perturbation and feature selection in preserving privacy,” in *Proc. Ninth Int. Conf. Wireless and Optical Communications*, 2012.

[2] T. Jahan, G. Narasimha, and C. V. G. Rao, “A comparative study of data perturbation using fuzzy logic for privacy preservation,” in Networks and Communications (NetCom2013), 2014.

[3] T. Jahan, “Brain CT processing using U-Net model with data augmentation for detection of ischaemic and haemorrhage strokes,” Intelligent Systems and Applications in Engineering, vol. 12, pp. 72–82, 2023.

[4] T. Jahan and D.C.V.G. Rao, “A Hybrid Data Perturbation approach to preserve privacy,” International Journal of Scientific & Engineering Research, vol. 6, no. 6, p. 1528, 2015.

[5] T. Jahan, G. Narsimha, and C. V. G. Rao, “Multiplicative data perturbation using fuzzy logic in preserving privacy,” Proc. Int. Conf. Information and Communication Technologies, 2016.

[6] T. Jahan, G. Narasimha, and V. G. Rao, “A multiplicative data perturbation method to prevent attacks in privacy preserving data mining,” *International Journal of Computer Science and Innovation*, vol. 1, no. 1, pp. 45-51, 2016.

[7] T. Jahan, G. Narsimha, and C. V. G. Rao, “Clustering on distorted data, with privacy preserving,” Journal of Computer Engineering, vol. 5, no. 2, 2012.

[8] T. Jahan, K. Pavani, G. Narsimha and C. V. Guru Rao, “A data perturbation approach to maintain privacy using fuzzy rules,” *Proc. Int. Conf. Computational Intelligence*, 2018.

[9] T. Jahan, G. R. Reddy, K. Shekhar, M. Swapna, New hybrid geometric data perturbation technique by sampling data intervals, Materials Today: Proceedings, vol. 80, pp. 2614–2619, 2023.

[10] T. Jahan, “Transfer learning based approach for detection of fruit freshness,” Journal of Computational Analysis and Applications, vol. 34, 2025.

[11] T. Jahan, “Client side defence against web spoofing attacks based on machine learning,” International Journal of Information and Electronics Engineering, vol. 15, 2025.

[12] T. Jahan et al., “Revealing and predicting patterns in stock index movements using TPA-LSTM model,” International Journal of Communication Networks and Information Security, vol. 17, 2025.

[13] T. Jahan, “Academic and professional data management improvement,” Library Progress International, vol. 44, 2024.

[14] T. Jahan and T. Aanam, “A decision making system on the health care using machine learning algorithms,” Journal of Philanthropy and Marketing, vol. 4, no. 1, pp. 602–610, 2024.