Journal of Information Hiding and Multimedia Signal Processing ©2023 ISSN 2073-4212
Ubiquitous International Volume **14**, Number **3**, September **2023**

Intrusion Detection in Electric Vehicles using Machine Learning with Model Explainability

Md. Abu Sayeed Sardar, Hasi Saha, Md. Nahid Sultan, Md. Fazle Rabbi

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur -5200, Bangladesh
sayeedcse15hstu@gmail.com, hasi@hstu.ac.bd, nahid@hstu.ac.bd, rabbi@hstu.ac.bd

Received May 2023; revised August 2023

Abstract. *In this Fourth Industrial Revolution, the transport system is now at a different level. Electric Vehicle is now gaining importance day by day for their uncommon features. But, intrusion and unwanted attacks are one of the major problems of this kind of vehicle, and detecting the anomalies is an urgent task to run this smoothly. In this study, we proposed an ML-based intrusion detection system that finds the best ML algorithm for detection. We use Explainable AI tools to show how the features impact the performance of the classifiers. The proposed methodology includes missing value handling, outlier removal, normalization, and imbalanced data handling techniques in the preprocessing phase. We use Explainable AI (XAI) tools SHAP to show the feature importance and impact of the features. We use Logistic Regression, Decision Tree, Support Vector Classifier, Gradient Boosting Classifier, Random Forest (RF), Extreme Gradient Boosting, and K Nearest Neighbors Machine Learning (ML) algorithms for analysis purposes. Among all the algorithms, RF performs better than others; it shows 99.64% accuracy with 99% precision, recall, and F1 score. XAI tool SHAP is used on RF, and we get the model explainability.*
Keywords: Intrusion Detection, Electric Vehicles, Machine Learning, Explainable AI

1. **Introduction.** Given the rapid technological evolution, a dramatic change has occurred globally in every industry. Transportation is also developing with the pace of the 4th industrial revolution. In recent times, the positive impact of the advancement in the Internet of Things (IoT) and communication networks are placing on the transportation industry that novel and intelligent applications for vehicles are commencing. To control the surge of carbon footprint, companies are moving towards autonomous electric vehicles (ARVs) [1]. Even, Artificial Intelligence (AI) driven electric vehicles are promoting Intrusion Detection Systems (IDSs) as it alerts drivers about abnormalities [2]. These highly qualified vehicles are controlled through AI-based controlling systems such as Machine Learning (ML) and Deep Learning (DL). ML algorithms are employed in the development of highly secure Industrial Control Systems. ML algorithms help to detect intrusions at the entry level utilizing the data received through network packets. It also identifies anomalies initially in the physical process employing information [3]. Machine learning can be used for electric vehicle intrusion detection to detect and prevent unauthorized access to the vehicle or its systems. This can be achieved by training a machine learning model on data collected from various sensors and systems installed in the vehicle. The model can be trained to detect abnormal behaviour or patterns in the data that could indicate an intrusion attempt. For example, the model can be trained to detect

unusual patterns in the electric motor's behaviour, such as sudden changes in torque or speed, which could indicate an attempt to tamper with the motor. Other types of data that could be used to train the machine-learning model include GPS data, accelerometer data, battery voltage data, and vehicle telemetry data. The model can also be trained to detect anomalies in communication between the vehicle and external systems, such as the charging station. Once the machine learning model is trained, it can be integrated into the vehicle's security system to provide real-time intrusion detection and prevention. In case of an intrusion attempt, the system can trigger an alarm, alert the owner or the authorities, or even take preventive actions, such as shutting down the vehicle or locking the doors.

Overall, machine learning-based intrusion detection systems can improve the security of electric vehicles and protect them from theft or malicious attacks. The contributions of our work are:

- We propose a machine learning-based methodology that can detect network intrusion more accurately than existing works.
- We eliminate features from the dataset that make the dataset well fit the machine learning algorithms.
- Using the XAI tool SHAP, we show the feature importance and the individual features' impact on the output by the partial dependencies.
- A comparative study among the machine learning algorithms is performed to find the best classifier to classify network intrusion detection.

2. **Literature Review.** For the betterment and ensuring the security of mankind, ample initiatives have been taken by different centuries. The primary threat we face on the road is that a minor mistake or concentration break can bring a devastating conclusion. Thereby, scholars are continually contributing to the same world. After stepping on a long path, we are now at the stage where intrusion detection of electric vehicles is the talk of the town.

Complex value neural networks (CVNNs) are the subject of research by Han et al. (2021) in order to safeguard controller area networks (CAN IDs). An automatic secure continuous cloud service availability method was developed by Aloqaily et al. (2019) for use in smart vehicles; this method aids intrusion detection against cyber-attacks and guarantees services that live up to users' expectations in terms of QoE and QoS [5]. Kang and Kang, (2016) introduced a novel Intrusion Detection System (IDS) using a Deep Neural Network (DNN) [6] to speed up the security of in-vehicle networks. Wang et al. (2018) tested the security performance of Controller Area Network (CAN) using actual messages from a single car. As a result, they recommended an innovative IDS to detect intrusions. [7]. Pascale et al. (2021) presented a two-stage Intrusion Detection System (IDS) for automated cars to identify potential cyberattacks [8]. The Intrusion Detection System (IDS) approach was a primary focus of Lokman et al.'s (2021) advocacy effort because of its importance in ensuring the safety of both data and networks [9]. Zhang et al. (2019) used Gradient Descent with Momentum (GDM) and Gradient Descent with Momentum and Adaptive Gain (GDM/AG) [10] to increase the IDS's precision and performance. New deep learning-based Intrusion Detection Systems (IDS) [11] were proposed by Basnet and Ali, (2020) to identify Denial of Service (DoS) attacks in EV charging stations. Inter-vehicle networks, in-vehicle networks, the Internet of Drones (IoD), and ground vehicle power stations are all evaluated in terms of their detection performance by Banafshehvaragh and Rahmani (2022) [12]. Umer et al. (2022) analyzed ML-based ICSs for their performance adaptability and difficulties. To be specific, semi-supervised, supervised, reinforcement learning, and unsupervised [13]. Using spatial-temporal correlation (STC)

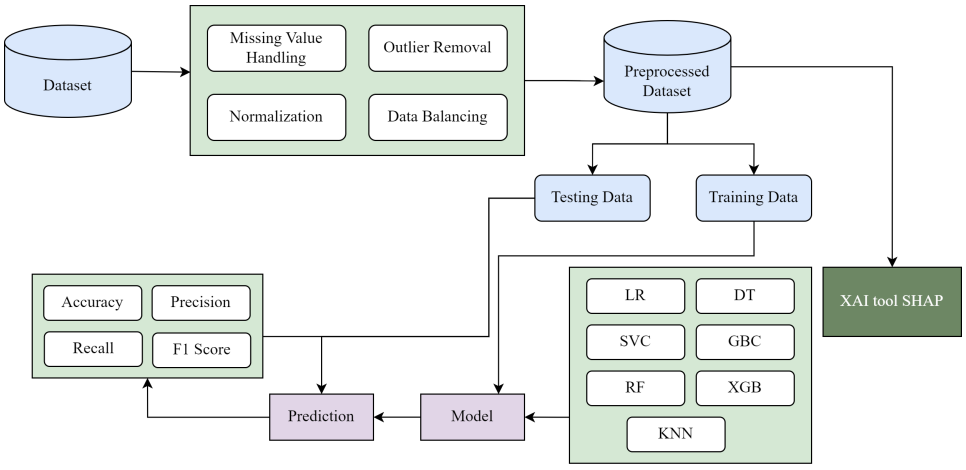

FIGURE 1. Proposed Methodology for Detecting Intrusion on Electric Vehicles

aspects of in-vehicle communication traffic (intrusion detection system (IDS) [14], Cheng et al. (2022) presented a novel model for detecting intrusion.

3. **Proposed Mechanism.** After detecting and handling the missing value of the dataset, we determined the variance threshold using a 0.5 threshold. We identified duplicate features, which were seven in number, and eliminated them to 55 from 62. As the dataset contains seven classes having different values, we employed SMOTE for oversampling and balancing the dataset. Then, we split our dataset into training (80) and testing (20). Then, we trained the dataset using Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Classifier (SVC), k-nearest Neighbour (KNN), Extreme Gradient Boosting (XGB), and Gradian Boosting classifier (GBC). This study used an explainable AI mode SHAP on the classifiers to detect the feature responsible for the intuition detection problem. To determine the impact of a given feature, this study utilizes beswarm plot and prediction probability plot which informs how the change in feature brings a change in the algorithms' output and helps explain which feature is responsible and how.

3.1. **Dataset Description.** Defenses against today's complex and persistent network threats rely heavily on intrusion detection and prevention systems (IDS/IPS). Anomaly-based intrusion detection methods have inconsistent and inaccurate performance evolutions because of the unreliability of test and validation datasets. To better simulate genuine real-world data (PCAPs), the CICIDS2017 dataset includes both benign and the most up-to-date prevalent attacks. Included as well are the time-stamped, source- and destination-IP- and port-labeled, protocol- and attack-labeled CSV files that show the findings of a network traffic analysis performed with CICFlowMeter. The definition of the retrieved features is also made available.

3.2. **Data Preprocessing Techniques.** The process of data preparation entails cleaning and organizing raw data. The authors preprocessed the dataset to identify missing values, outliers, noisy data, and other irregularities before running it. The most well-known data preparation methods employed here are outlined.

Outlier Detection: In this research, Turkey fences were used to identify Q1, Q2, and Q3 quartile outliers and extreme values. For a given data collection, the first quartile (Q1) is the value with which 75% of the values fall below [11]. As a rule of thumb, 25% of all values are above the third quartile or Q3.

$$lower\_limit = Q1–1.5(Q3–Q1) \tag{1}$$

$$upper\_limit = Q3 + 1.5(Q3–Q1) \tag{2}$$

Any value that crosses these upper and lower limit is considered as outlier.

Normalization: The raw data in this study is transformed linearly using min-max normalization (range normalization). Let's pretend that minA and maxA are the lowest and highest possible values for attribute A. In the range [minA, maxA], min-max normalization maps a value of attribute A, d to d'.

$$d' = \frac{d - min_A}{max_A - min_A} \tag{3}$$

Imbalance Data Handling: When the proportions of the different types of information in a dataset are significantly off, we say that the dataset is imbalanced. SMOTE Tomek is used to deal with the inequity issue. To dramatically improve the effectiveness of a classifier model, SMOTETomek combines oversampling and undersampling strategies [15]. First, the SMOTE method is used to oversample the minority group; then, once samples from the majority groups have been found and eliminated from Tomek Links, the distribution has been brought into statistical equilibrium.

3.3. **Description of Algorithms.** LR, DT, RF, SVC, KNN, XGB, and GBC are the seven ML methods used in this investigation. Each algorithm is briefly explained in the following sections:

Logistic Regression (LR): For forecasting the likelihood of a binary result, supervised machine learning approaches like LR work effectively. Regularization is essential in logistic regression to reduce overfitting, especially when there are few training instances or many parameters to learn. Multiclass problems can be classified by employing LR. Since LR has a linear decision surface, it cannot address non-linear issues.

Decision Tree (DT): Machine learning algorithms such as the decision tree classifier employ a tree-like architecture to make decisions according to predefined criteria. The algorithm generates a tree-like structure, with each node representing a feature-based determination and each leaf node representing a categorical label. The training data is subdivided recursively into smaller groups based on the values of the input features in the decision tree classifier. With metrics like information gain and Gini impurity in mind, the algorithm chooses the most appropriate feature for each data split.

Random Forest (RF): RF is a classifier that uses the average to boost the predicting accuracy of a particular dataset by employing a series of decision trees on various subsets of that dataset. Instead of depending on one decision tree, RF forecasts from each tree and projects the ultimate outputs depending on the overwhelming votes. The variables for the classification problem are rated according to their significance [16]. The accuracy improves as the volume of trees in the forest increases, reducing the detrimental effects of overfitting.

Support Vector Classifier (SVC): SVC is a model utilised to fix pattern recognition problems. To precisely separate data into several categories, SVC employs the concept of decision planes which exercise decision boundaries. SVC performs comparatively better on high dimensional spaces and is memory efficient. But, it displays bad performance in datasets having noise.

K-Nearest Neighbors (k-NN): k-Nearest Neighbours (k-NN) is a popular choice for classification algorithms in many contexts. The idea that the example's predicted value is probably comparable to that of neighbours inspired the development of k-NN. The k-NN

method establishes a metric in the vector space of the predictor, plots all candidates for a position, and calculates the probability by calculating the percentage of good risks among the k-nearest points in the training set.

Extreme Gradient Boosting Classifier (XGB): XGB is an advanced implementation of the gradient boosting algorithm for classification and regression tasks. It is an ensemble machine learning algorithm that combines the outputs of several decision trees to make a final prediction. XGBoost improves upon the standard gradient boosting algorithm by adding regularization terms to control overfitting, handling missing values, and parallelizing the algorithm to make it more scalable. The algorithm also uses a novel "gradient-based sampling" technique to select the instances used to train each tree, resulting in faster and more accurate predictions. XGBoost can handle both categorical and numerical features and automatically handles missing values. It can also handle large datasets with high-dimensional features and is known for its high accuracy and speed.

Gradient Boosting Classifier (GBC): GBC is an ensemble machine learning algorithm used for classification tasks. The algorithm works by building a set of decision trees sequentially, with each subsequent tree trying to correct the mistakes of the previous ones. At each stage, the algorithm identifies the instances misclassified by the previous tree and assigns them higher weights so that the next tree can focus more on these instances. The algorithm combines the outputs of all the trees to make the final prediction.

SHAP: SHAP (SHapley Additive exPlanations) is an XAI (Explainable Artificial Intelligence) tool that provides a unified framework to explain the output of any machine learning model. It is based on the Shapley value from cooperative game theory, which assigns a fair distribution of the value of a prediction to each input feature [17]. SHAP values represent the contribution of each feature to the final prediction and enable users to understand how the model arrived at its output. The tool can handle any input data, including structured and unstructured data, images, and text. SHAP is particularly useful for complex models such as deep learning models, where traditional methods of interpretation, such as feature importance or partial dependence plots, may not be sufficient. The tool also provides visualizations that make interpreting and communicating the explanations to stakeholders easy.

3.4. **Performance measure techniques.** Accuracy, Precision, Recall, F-1 Score, and Receiver Operating Characteristic (ROC) were the performance measuring measures used in this work to evaluate ML algorithms.

Accuracy: Accuracy is defined as the percentage of training data that was properly categorised. Accuracy is one of the most fundamental indicators of performance, but it can sometimes lead to erroneous conclusions, especially when dealing with unbalanced data sets. Mathematically,

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4}$$

Precision: For binary classification, precision is calculated as the fraction of true positives (TP) minus the sum of false positives (FP). When the goal is to lower FP, Precision functions flawlessly on skewed data. Precision is a useful statistic to utilise regardless of the FP rate. Mathematically,

$$Precision = \frac{TP}{(TP + FP)} \tag{5}$$

Recall: True Positive Rate (TPR) and Sensitivity are other names for Recall. The recall is often determined by dividing the total number of TP by the sum of the TP and

FN. Using recall is appropriate when trying to eliminate FN from an unbalanced data collection. Mathematically,

$$Recall = \frac{TP}{(TP + FN)} \tag{6}$$

F-1 Score: F-1 score is the mathematical average of the two measures of accuracy, Precision and Recall. It takes more than just precision to determine whether or not a model may be used. Both high Precision and Recall are required for the model to make sense. This is why the F1-Score is used to evaluate the efficacy of different classifiers. The greater the f-1 score, which might be a number between 0 and 1, the better the model. Mathematically,

$$F - 1Score = \frac{2PR}{(P + R)} \tag{7}$$

ROC: The ROC curve is a graphical tool for comparing the accuracy of different classifiers and making a final decision. It is a probability curve showing the true positive rate (TPR) vs the false positive rate (FPR) at various cutoff points along the X-axis.

3.5. **Experimental Environment:** For experimental purpose, we use Python programming language. We use Python sklearn library for the benchmark ML models and use Jupyter Notebook in Anaconda IDE. Besides this, we use Pandas, Numpy and Matplotlib python library for the experiment.

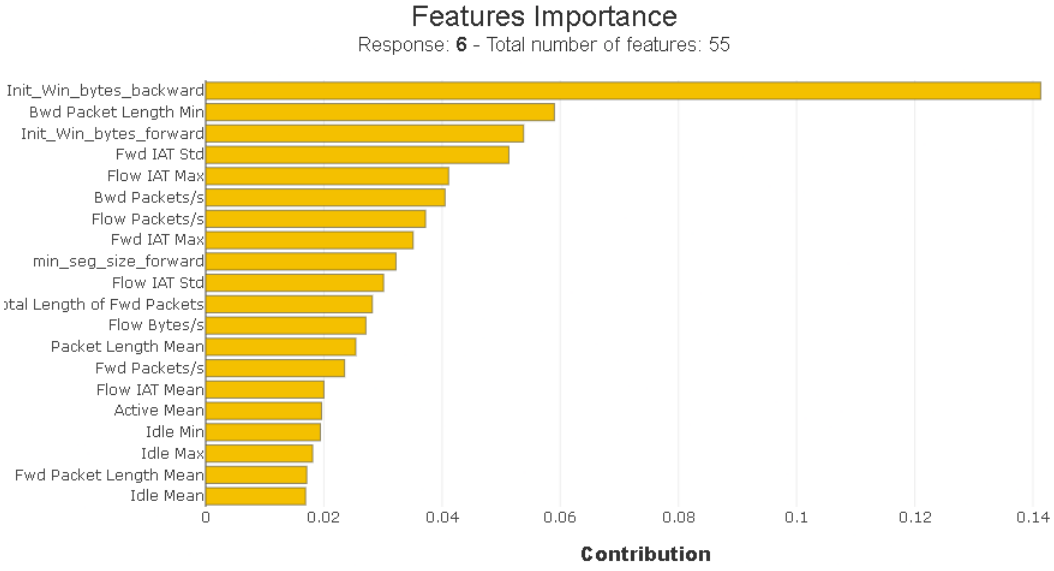

FIGURE 2. Importance of the Features using SHAP

TABLE 1. Performance of the ML algorithms for Intrusion Detection

Algorithms	Precision	Recall	F1	Acc
LR	0.69	0.57	0.53	57.11%
SVC	0.70	0.57	0.55	56.75%
DT	0.99	0.99	0.99	99.25%
RF	0.99	0.99	0.99	99.64%
KNN	0.96	0.96	0.96	96.39%
GB	0.99	0.99	0.99	99.38%
XGB	0.99	0.99	0.99	99.41%

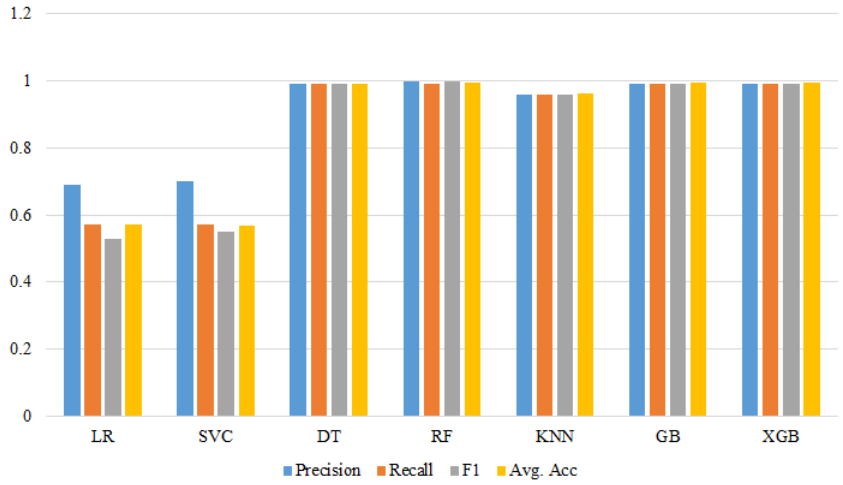

FIGURE 3. Performance of the classifiers to detect intrusion

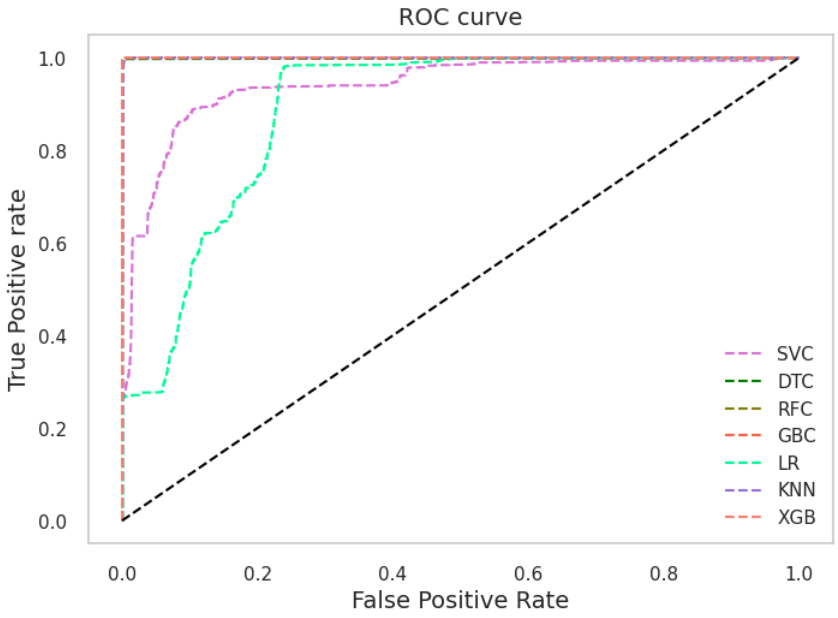

FIGURE 4. ROC curve to show the performance of the classifiers

4. **Results and Analysis.** To find the intrusion on the electric vehicle, we use some benchmark ML models and calculate the accuracy, precision, recall and f1 score for individual algorithms and performances are shown in Table. 1. The performance of the algorithms is high except for LR and SVC. Most algorithms show more than 95% accuracy with good precision and recall. Among all the algorithms, RF outperforms others by showing 99.64% accuracy. RF can be a suitable solution to detect intrusion in electric vehicles. Figure 3 and Figure 4 show the performance of the algorithms. We also find the feature importance of RF to show its explainability. The Figure. 2 shows the importance of the features from top to down. Init_Win_bytes_backward, Bwd Packet Length Min, Init_Win_bytes_forward, Fwd IAT Std, and Flow IAT Max are the top five features indicated by the SHAP on the RF algorithm. Init_Win_bytes_backward carries a significant importance than the other features.

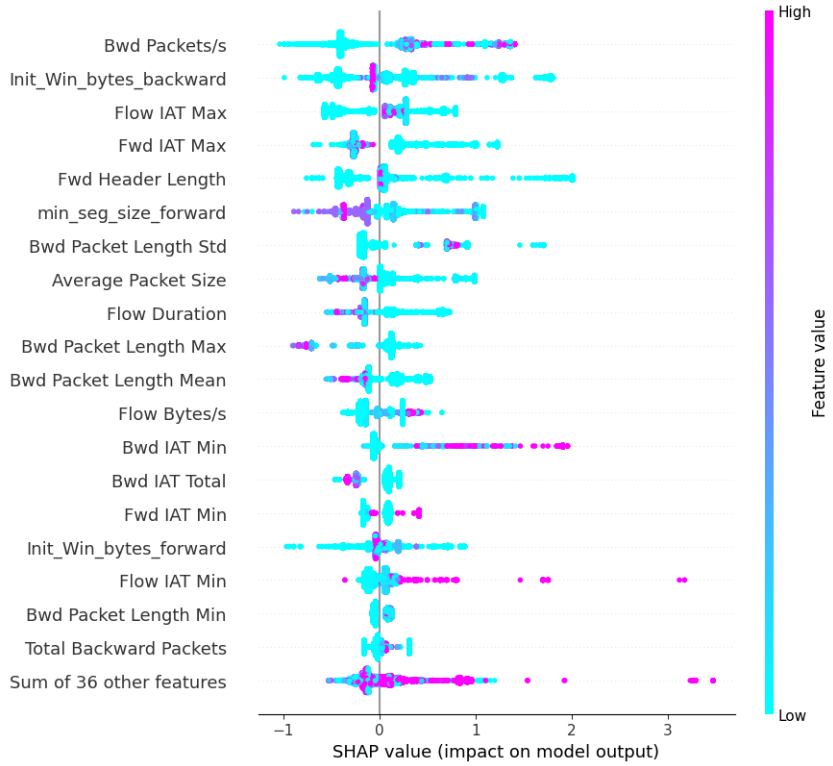

FIGURE 5. Importance and Distribution of the Features using Beswarm Plot

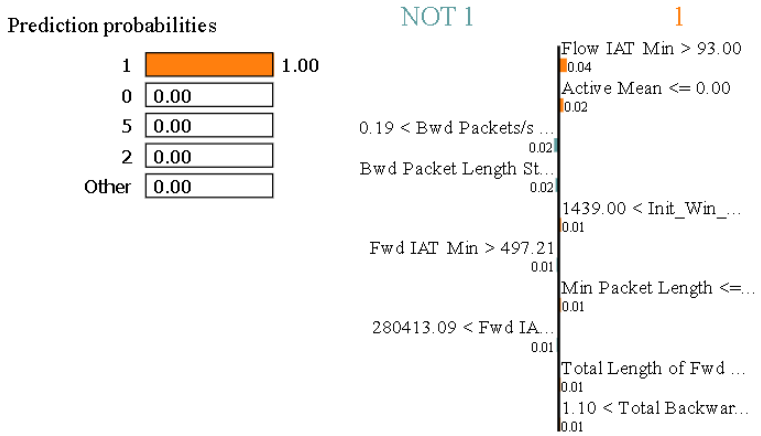

FIGURE 6. Probability of the features for a class

We also show the importance of the features by a beswarm plot in Figure. 5 for all the data points. Most of the datapoint show weak contributions to the model, and few show the strength of the contribution. Also, a local explainability of the model is shown in Figure. 6 that shows 100% probability of prediction for the instances. The top features show a positive impact, and the other features' impact is ignorable.

5. **Conclusion and Future Work.** This research aims to detect intrusion on the electric vehicle using machine learning and explain the top ML model according to the importance of the features. To achieve this aim, our proposed methodology includes data preprocessing techniques (Missing value handling, outlier removal, normalization, and data balancing), analysis using ML algorithms, and the use of eXplainable AI tools SHAP to

find the importance of the features and their impact on the performance of the classifiers. Among all the classifiers, RF outperforms other algorithms regarding accuracy, precision, recall, and f1 score. This research helps to make an ML-based IDS for the domain experts and researchers in this field.

REFERENCES

[1] Alkheir, A. A., Aloqaily, M., & Mouftah, H. T. (2018). Connected and autonomous electric vehicles (caevs). IT professional, 20(6), 54-61.

[2] Tabassum, A., Erbad, A., & Guizani, M. (2019, June). A survey on recent approaches in intrusion detection system in IoTs. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) (pp. 1190-1197). IEEE.

[3] Umer, M. A., Junejo, K. N., Jilani, M. T., & Mathur, A. P. (2022). Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. International Journal of Critical Infrastructure Protection, 100516.

[4] Han, M., Cheng, P., & Ma, S. (2021). PPM-InVIDS: Privacy protection model for in-vehicle intrusion detection system based complex-valued neural network. Vehicular Communications, 31, 100374.

[5] Aloqaily, M., Otoum, S., Al Ridhawi, I., & Jararweh, Y. (2019). An intrusion detection system for connected vehicles in smart cities. Ad Hoc Networks, 90, 101842.

[6] Kang, M. J., & Kang, J. W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. PloS one, 11(6), e0155781.

[7] Wang, Q., Lu, Z., & Qu, G. (2018, September). An entropy analysis based intrusion detection system for controller area network in vehicles. In 2018 31st IEEE International System-on-Chip Conference (SOCC) (pp. 90-95). IEEE.

[8] Lombardi, M., Pascale, F., & Santaniello, D. (2022). Two-Step Algorithm to Detect Cyber-Attack Over the Can-Bus: A Preliminary Case Study in Connected Vehicles. ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg, 8(3).

[9] Lokman, S. F., Othman, A. T., & Abu-Bakar, M. H. (2019). Intrusion detection system for automotive Controller Area Network (CAN) bus system: a review. EURASIP Journal on Wireless Communications and Networking, 2019, 1-17.

[10] Zhang, J., Li, F., Zhang, H., Li, R., & Li, Y. (2019). Intrusion detection system using deep learning for in-vehicle security. Ad Hoc Networks, 95, 101974.

[11] Basnet, M., & Ali, M. H. (2020, September). Deep learning-based intrusion detection system for electric vehicle charging station. In 2020 2nd International Conference on Smart Power & Internet Energy Systems (SPIES) (pp. 408-413). IEEE.

[12] Banafshehvaragh, S. T., & Rahmani, A. M. (2022). Intrusion, Anomaly, and Attack detection in Smart Vehicles. Microprocessors and Microsystems, 104726.

[13] Umer, M. A., Junejo, K. N., Jilani, M. T., & Mathur, A. P. (2022). Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. International Journal of Critical Infrastructure Protection, 100516.

[14] Cheng, P., Han, M., Li, A., & Zhang, F. (2022). STC-IDS: Spatial–temporal correlation feature analyzing based intrusion detection system for intelligent connected vehicles. International Journal of Intelligent Systems, 37(11), 9532-9561.

[15] M. Hasan, M. M. Islam, S. W. Sajid and M. M. Hassan, "The Impact of Data Balancing on the Classifier's Performance in Predicting Cesarean Childbirth," 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 2022, pp. 1-4, doi: 10.1109/ICECTE57896.2022.10114515.

[16] Rabbi, M.F., Moon, M.H., Dhonno, F.T., Sultana, A., Abedin, M.Z. (2022). Foreign Currency Exchange Rate Prediction Using Long Short-Term Memory, Support Vector Regression and Random Forest Regression. In: Derindere Köseoğlu, S. (eds) Financial Data Analytics. Contributions to Finance and Accounting. Springer, Cham. https://doi.org/10.1007/978-3-030-83799-0_8

[17] M. Hasan, P. Roy and A. M. Nitu, "Cervical Cancer Classification using Machine Learning with Feature Importance and Model Explainability," 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 2022, pp. 1-4, doi: 10.1109/ICECTE57896.2022.10114548.