AN INTELLIGENT DATA DRIVEN MODEL TO SECURE INTRA VECHILE COMMUNICATIONS BASED ON MACHINE LEARNING

1. Sandhya belidha, Assistant professor, Department of CSE, Teegala Krishna Reddy Engineering college, Telangana, India. sandhyabelidha@gmail.com.

2. A. Suchitha, Department of CSE, Teegala Krishna Reddy Engineering college, Telangana, India. suchithareddy2232@gmail.com.

3. B. Shiva Sathvik, Department of CSE, Teegala Krishna Reddy Engineering college, Telangana, India. shivasathvik27@gmail.com.

4. D. Srinath, Department of CSE, Teegala Krishna Reddy Engineering college, Telangana, India. dharavathsrinathhh@gmail.com.

Abstract - The high relying of electric vehicles on either in vehicle or between-vehicle communications can cause big issues in the system. This paper is going to mainly address the cyber-attack in electric vehicles and propose a secured and reliable intelligent framework to avoid hackers from penetration into the vehicles. The proposed model is constructed based on an improved support vector machine model for anomaly detection based on the controller area network (CAN) bus protocol. In order to improve the capabilities of the model for fast malicious attack detection and avoidance, a new optimization algorithm based on social spider (SSO) algorithm is developed which will reinforce the training process offline. Also, a two-stage modification method is proposed to increase the search ability of the algorithm and avoid premature convergence. Last but not least, the simulation results on the real data sets reveal the high performance, reliability and security of the proposed model against denialof-service (DoS) hacking in the electric vehicle

Keywords: *- Electric Vehicles, Cyber-Attack, Anomaly Detection, Support Vector Machine (SVM), Controller Area Network (CAN) Protocol, Social Spider Optimization (SSO) Algorithm, Malicious Attack Detection, Denial-of-Service (DoS) Hacking, Security Framework, Simulation Results.*

I. INTRODUCTION

The automotive industry is witnessing a profound transformation marked by the growing reliance on electric vehicles (EVs) and the integration of advanced communication systems. As societies increasingly prioritize sustainability, EVs have emerged as a pivotal solution, minimizing environmental impact. This shift, however, brings forth a new frontier of challenges, particularly in the realm of intra vehicle communication. The symbiotic relationship between electric vehicles and communication systems demands a proactive approach to cybersecurity. The risks associated with cyber attacks necessitate the establishment of a secure and reliable intelligent framework, embodying a commitment to innovation, safety, and the seamless integration of EVs into the fabric of modern transportation. In, an intrusion detection system based on the traffic entropy of in-vehicle network communication system of the CAN bus is suggested.

In, an anomaly detection approach is developed which is capable of detecting faults of known and unknown type without requiring the setting of expert parameters. This paper aims to propose an intelligent and highly secure method to equip the electric vehicles with a powerful anomaly detection and avoidance mechanism. The proposed method is constructed based on support vector machine and the concept of oneclass detection system to avoid any malicious behavior in the vehicle.

Here the experimental CAN bus data are used to let the support vector machine learn the normal frequency of the different message frames at different commands. In order to get into the maximum capability of the model, a new optimization algorithm based on social spider optimization (SSO) algorithm is proposed to adjust the SVR setting parameters, properly.

Due to the high complexity and nonlinearity of the electric vehicle CAN bus dataset, a new two-stage modification method based on crossover and mutation operators of genetic algorithm is developed which can increase the algorithm population diversity and at the same time avoid premature convergence. The feasibility and satisfying performance of the proposed model are examined using the real datasets gathered from an electric vehicle.

As societies increasingly prioritize sustainability, EVs have emerged as a pivotal solution, minimizing environmental impact. This shift, however, brings forth a new frontier of challenges, particularly in the realm of intravehicle communication. The symbiotic relationship between electric vehicles and communication systems demands a proactive approach to cybersecurity. The risks associated with cyber-attacks necessitate the establishment of a secure and reliable intelligent framework, embodying a commitment to innovation, safety, and the seamless integration of EVs into the fabric of modern transportation.

II. LITERATURE SURVEY

Technically, vehicles are composed of many hardware modules namely called electronic control units (ECUs) being controlled by different software tools. All sensors installed in a vehicle will send their data to the ECU, where this data are processed and the required orders are sent to the relevant actuators [1]. Such a highly complex hardware software data transfer process may happen through the use of different network protocols such as CAN, LIN, FlexRay or MOST [2]. Among these protocols, CAN bus is the most popular one not only in vehicles, but also in medical apparatuses, agriculture, etc due to its high capability and promising characteristics. Some of the main advantages of the CAN bus standard may be briefly named as allowing up to 1Mbps data rate transfer, reducing the wiring in the device saving cost and time due to the simple wiring, auto retransmission of lost messages and error detection capability [3].

Unfortunately, since CAN bus protocol was devised at a time where vehicles were almost isolated, this standard suffers from some security issues in the new dynamic environment of smart grids. This will motivate the hackers to attack the electric vehicles through the ECU and inject malicious messages into their systems. In [4], some cyber intrusion scenarios are modeled and applied on electric vehicles to assess their vulnerabilities and possible side effects getting finally into the power grid. In [5], a new classification method is developed for cyber intrusion detection in vehicles. In [6], a data intrusion detection

system is developed which can detect cyber-attack based on the CAN bus message frequency increase or CAN message ID misuse. This will help the driver to detect that an attack has happened so to stop the vehicle immediately.

In [7], authors suggest that all CAN messages should pass a data management system to avoid any cyber intrusion. In [8], an algorithmic solution is used to stop attacks of types of denial-of-service or error flag in the vehicle. In [9], it is suggested to assign an ECU as the master ECU in the manufacturing stage of the vehicle so to run an attestation 8 process in the system. In, a firewall is introduced for the vehicle to sit between the CAN bus and the communicating system and stop the cyber-attack commands to the CAN bus.

In the existing system, traditional machine learning algorithms play a pivotal role in detecting anomalies within electric vehicle communication networks, particularly those based on the Controller Area Network (CAN) bus protocol[1]. These algorithms, including K-Nearest Neighbors (KNN), Decision Tree, and Conventional Support Vector Machine (SVM), have been extensively utilized due to their established effectiveness in anomaly detection tasks.

K-Nearest Neighbors (KNN) is a simple yet powerful algorithm that classifies data points based on the majority class of their k-nearest neighbors in the feature space. In the context of electric vehicle communication, KNN can effectively identify anomalies by comparing the communication patterns of individual data points with those of their neighboring data points[4]. This approach enables the detection of abnormal communication behavior, such as unusual message frequencies or unexpected data packet sizes.

Decision Tree algorithms, on the other hand, employ a hierarchical tree-like structure to recursively partition the feature space based on attribute values. By evaluating different decision paths, Decision Trees can discern patterns within electric vehicle communication data that may indicate anomalous behavior. For example, Decision Trees can identify specific combinations of CAN bus message attributes that deviate from normal communication patterns, signaling potential security threats or system malfunctions.

Support Vector Machine (SVM) algorithms are widely recognized for their ability to separate data points in high-dimensional feature spaces using hyperplanes. In the context of anomaly detection in electric vehicle communication, SVMs excel at identifying complex patterns and nonlinear relationships that may indicate anomalous behavior. By leveraging the inherent flexibility and discriminative power of SVMs, 9 the existing system can effectively differentiate between normal and abnormal communication patterns within the CAN bus network.

Overall, the existing work on anomaly detection in electric vehicle communication networks demonstrates the efficacy of traditional machine learning algorithms, such as KNN, Decision Tree, and SVM, in identifying anomalous behavior and safeguarding the integrity and security of vehicle communication systems. These algorithms serve as foundational components in the development of robust cybersecurity solutions tailored to the unique challenges posed by electric vehicles and their communication protocols.

III. METHODOLOGY

Modules:

- Importing required Packages
- Exploring the dataset - Phishing URL Feature Data
- Data Processing - Using Pandas Data frame
- Visualization using seaborn & matplotlib
- Label Encoding using Label Encoder
- Feature Selection
- Train & Test Split
- Training and Building the model
- Trained model is used for prediction
- Final outcome is displayed through front-end

A) System Architecture

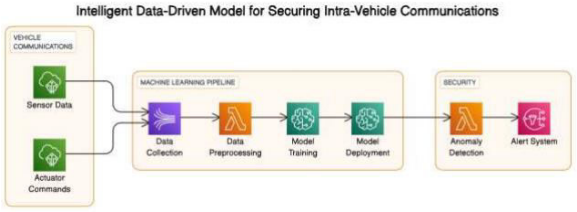


Fig 1: System Architecture

Proposed work

The system architecture of our project is designed to seamlessly integrate both client-side and server-side components, facilitating smooth interaction to deliver the desired functionalities. On the client side, the user interface elements form the backbone, encompassing web pages tailored for registration, login, data upload, and notification settings. These elements enable users to interact with the application intuitively. Complementing these interface elements, client-side logic, typically implemented in JavaScript, handles tasks such as form validation and processing user inputs, ensuring data accuracy and enhancing user experience. The architecture of our project is meticulously designed to foster seamless communication between client-side and server-side components.

The proposed system integrates the Social Spider Optimization (SSO) algorithm into the framework for securing electric vehicles against cyber-attacks, specifically focusing on anomaly detection within the Controller Area Network (CAN) bus protocol. In this system, the SSO algorithm optimizes the parameters of a Support Vector Machine (SVM) model, enhancing its performance in identifying and classifying anomalous communication patterns indicative of potential cyber threats.

The system operates by defining an objective function that evaluates the fitness of potential anomaly detection models based on metrics such as message frequency, inter-arrival times, and statistical properties of communication data. Spiders, representing potential solutions, collaborate to explore the solution space, exchanging information and conducting local search operations to refine their models. Periodic updates to the population guide optimization towards regions with higher anomaly detection capability.

The advantages of SSO over traditional algorithms, including superior global exploration, convergence speed, robustness, and scalability, ensure efficient and effective optimization in tackling the complexities of anomaly detection in intra-vehicle communication networks. Overall, the proposed system offers a robust and adaptive approach to enhancing the security, safety, and reliability of electric vehicles against cyber threats.

B) Dataset Collection

The dataset collection for the proposed intelligent, data-driven model to secure intra-vehicle communications in electric vehicles encompasses several key components. Firstly, it involves gathering comprehensive datasets representative of normal vehicular communication behavior, encompassing various operational scenarios and environmental conditions. Additionally, to accurately train the support vector machine (SVM) model for anomaly detection, datasets containing instances of cyber-attacks and malicious intrusions are crucial. These datasets should cover a spectrum of attack types, including denial-of-service (DoS) attacks, unauthorized access attempts, and data manipulation. Moreover, datasets reflecting the intricacies of the Controller Area Network (CAN) bus protocol are essential. These datasets should capture the nuances of message transmission, network topology, and communication protocols within the vehicle's internal systems. Furthermore, given the emphasis on enhancing model performance through the Social Spider Optimization (SSO) algorithm, datasets that facilitate algorithm optimization and parameter tuning are necessary. These datasets might include simulated attack scenarios, training data for algorithm refinement, and validation sets to assess model robustness. Overall, the dataset collection process should be comprehensive, covering both normal vehicular operation and potential cyber-threat scenarios, to effectively train and validate the proposed intelligent security framework for electric vehicles.

C) Pre-processing

Preprocessing for securing intra-vehicle communications using an intelligent data-driven model involves several crucial steps to ensure the effectiveness and reliability of the proposed framework. Firstly, data collected from the Controller Area Network (CAN) bus protocol needs to be preprocessed to remove noise and irrelevant information, enhancing the model's ability to detect anomalies accurately. This includes filtering out non-essential messages and normalizing data to a consistent format for analysis. Secondly, feature extraction techniques can be applied to extract relevant features from the CAN bus data, such as message frequency, message length, and timing information. These features provide valuable insights into the communication patterns within the vehicle network and aid in identifying potential cyber-attacks. Furthermore, data augmentation methods may be employed to enhance the diversity and robustness of the training dataset, thereby improving the model's generalization capability. Techniques such as oversampling rare attack instances and introducing synthetic data points can help address class imbalance and prevent the model from being biased towards normal behavior. Lastly, data scaling and dimensionality reduction techniques, such as standardization and principal component analysis (PCA), can be utilized to normalize feature values and reduce the computational complexity of the model without sacrificing performance. By implementing these preprocessing steps, the proposed intelligent framework can effectively mitigate cyber-attacks and ensure the security and reliability of intra-vehicle communications in electric vehicles.

D) **Training & Testing**

The training process of the proposed intelligent data-driven model involves several key steps. Initially, historical data comprising normal and anomalous communication patterns within electric vehicles is collected and preprocessed. Subsequently, the improved support vector machine (SVM) model is trained using this data to distinguish between normal and potentially malicious traffic on the controller area network (CAN) bus.

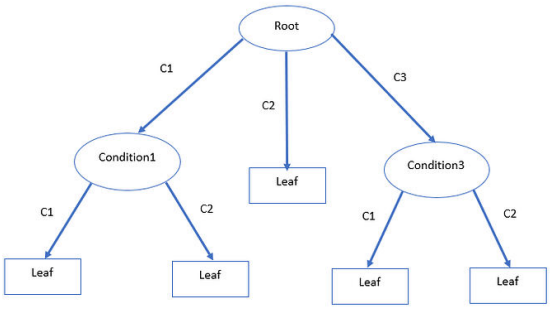
To enhance the model's ability to swiftly detect and mitigate cyber-attacks, the training is reinforced offline with a novel optimization algorithm derived from the social spider optimization (SSO) algorithm. This algorithm iteratively refines the SVM parameters, improving its accuracy in identifying anomalies while minimizing false positives.

Once trained, the model undergoes rigorous testing to evaluate its efficacy in real-world scenarios. Testing involves exposing the model to simulated cyber-attacks, including denial-of-service (DoS) attacks, and assessing its ability to promptly detect and mitigate them. The performance metrics, such as detection rate, false alarm rate, and response time, are carefully analyzed to validate the model's effectiveness in securing intra-vehicle communications in electric vehicles.

E) Algorithms.

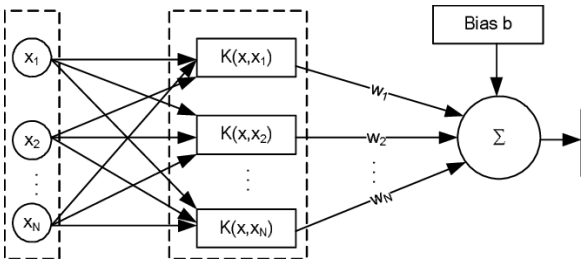
Decision Trees:

Decision trees are a popular supervised learning method used for classification and regression tasks. They partition the data into subsets based on feature values, with each partition representing a node in the tree. At each node, the algorithm selects the feature that best splits the data, aiming to maximize information gain or another criterion. This process continues recursively until a stopping criterion is met, such as reaching a maximum depth or having nodes with homogeneous class labels. Decision trees are interpretable, as they represent simple if-then-else decision rules. However, they are prone to overfitting, especially with complex datasets, and may require pruning or ensemble methods like random forests to improve generalization.



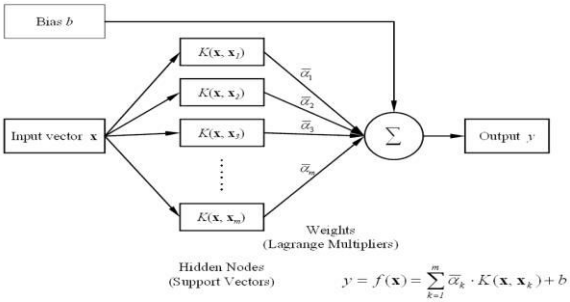
Conventional Support Vector Machines (SVM):

Support Vector Machines (SVM) are a powerful supervised learning algorithm used for classification and regression tasks. SVM aims to find the hyperplane that best separates the classes in the feature space, maximizing the margin between the closest data points from different classes. In conventional SVM, the algorithm seeks to find the optimal hyperplane without explicitly considering the relationships between data points. This approach works well for linearly separable data but may struggle with complex, nonlinear datasets. Additionally, SVM requires solving a convex optimization problem, which can be computationally expensive for large datasets. However, SVM is effective in high-dimensional spaces and is robust against overfitting when appropriate regularization is applied.



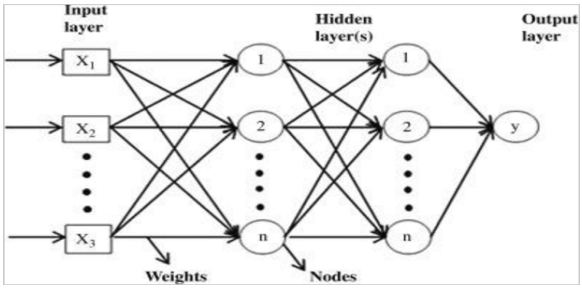
SSO with SVM:

Single-Source Outlier (SSO) detection with Support Vector Machines (SVM) is a technique used to identify outliers or anomalies in a dataset when only one class of outliers is present. Instead of using a conventional binary SVM to classify data points into two classes, SSO-SVM focuses on detecting outliers from the majority class. This is achieved by training an SVM on the majority class data points and identifying instances that lie farthest from the decision boundary as outliers. SSO-SVM is particularly useful in scenarios where anomalous instances are rare and only represented by one class. It can effectively detect outliers in high-dimensional spaces and is robust against noise.



K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple yet powerful supervised learning algorithm used for classification and regression tasks. In KNN, the prediction of a data point is determined by the majority class (for classification) or the average value (for regression) of its nearest neighbors in the feature space. The number of neighbors, represented by the parameter 'k,' is a crucial hyperparameter in KNN, influencing the model's performance and generalization ability. KNN is non-parametric and lazy, meaning it does not make explicit assumptions about the underlying data distribution and defers computation until inference time. While KNN is easy to implement and interpret, it can be computationally expensive for large datasets, especially in high-dimensional spaces. Additionally, it is sensitive to irrelevant features and requires proper preprocessing and tuning for optimal results.



IV. EXPERIMENTAL RESULTS

A) Tkinter GUI.

To run project double click on ‘run.bat’ file to get below screen

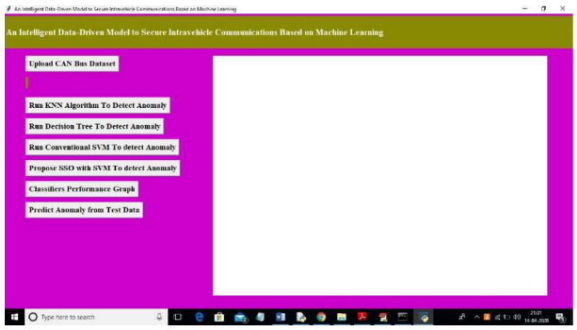


Fig 2

In above screen click on ‘Upload CAN Bus Dataset’ button and upload dataset

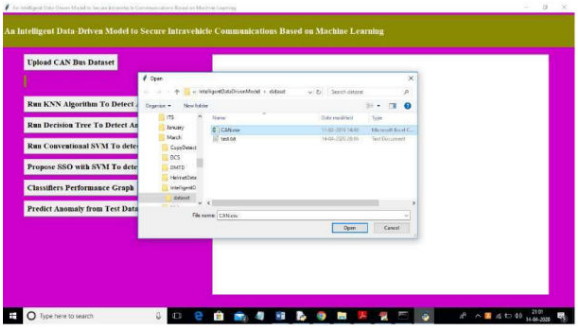


Fig 3

In above screen I am uploading 'CAN.csv' dataset and after uploading dataset will get below screen

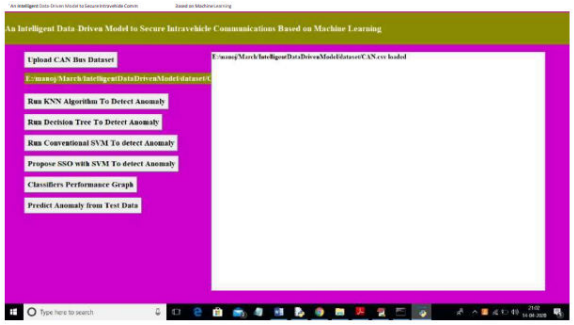


Fig 4

Now click on 'Run KNN Algorithm To Detect Anomaly' button to build KNN classifier train model to detect anomaly and evaluate its performance based on 4 indices

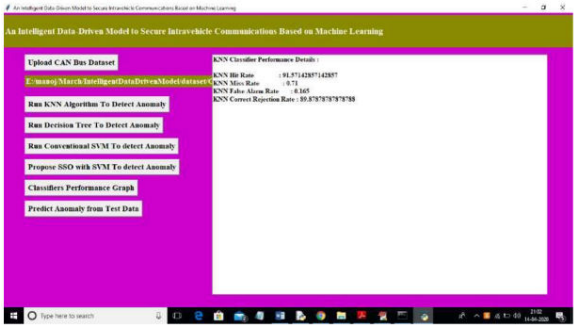


Fig 5

In above screen we got 4 indices values for KNN algorithm and now click on 'Run Decision Tree To Detect Anomaly' button to evaluate decision tree performance.

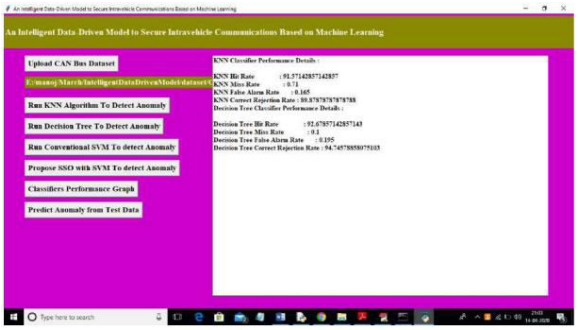


Fig 6

In above screen we got decision tree data and now click on 'Run Conventional SVM To detect Anomaly' button to evaluate conventional SVM performance.

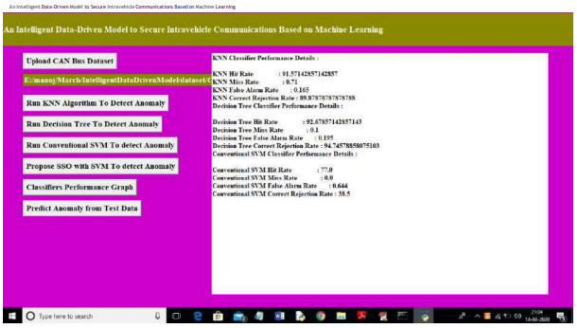


Fig 7

In above screen we got SVM performance data and now click on 'Propose SSO with SVM To detect Anomaly' button to run propose SSO with SVM classifier and evaluate its performance. (Note: when u run SSO then application will open 4 empty windows and you just close newly open empty window and keep working from first window only).

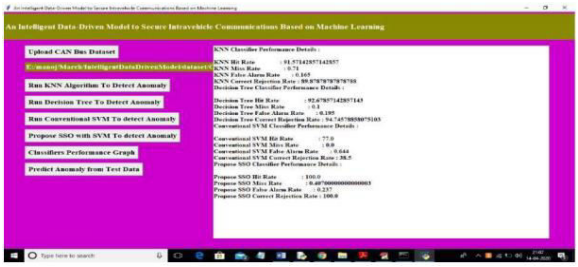


Fig 8

In above screen for SSO we got performance metric as 100% and MR and FR is not mandatory so we can ignore as said in paper. Now click on 'Classifiers Performance Graph' button to get performance graph between all classifiers.

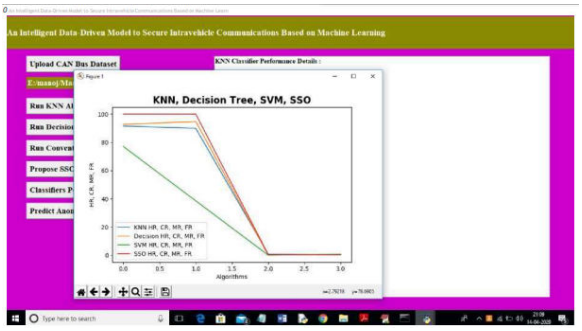


Fig 9

In above graph propose SSO has given high performance compare to other algorithms. In above graph y-axis represents HR, MR, FR and CR values. Now click on 'Predict Anomaly from Test Data' button to upload test data and predict it label.

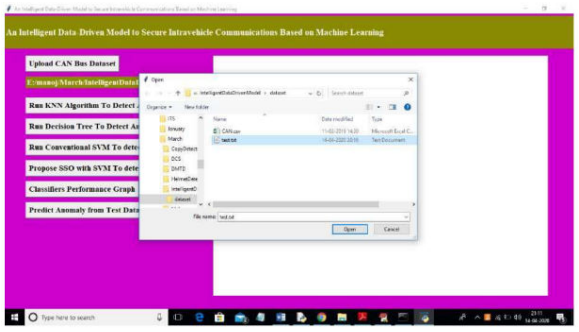


Fig 10

In above screen I am uploading 'test.txt' file and now click on 'Open' button to predict uploaded test file class label.

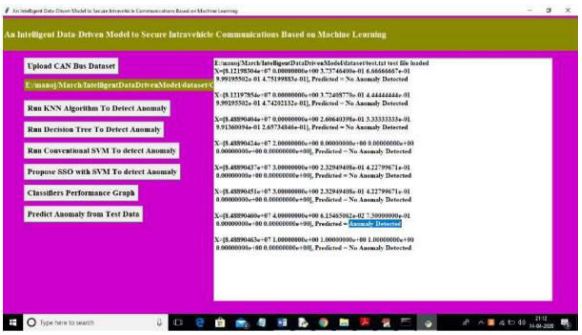


Fig 11

In above screen in text area we can see uploaded test data and its predicted class label.

All records contains normal packet data accept one record. So by using machine learning algorithms we can analyse packets and if packet contains attack then we ignore processing such packets.

B) Frontend

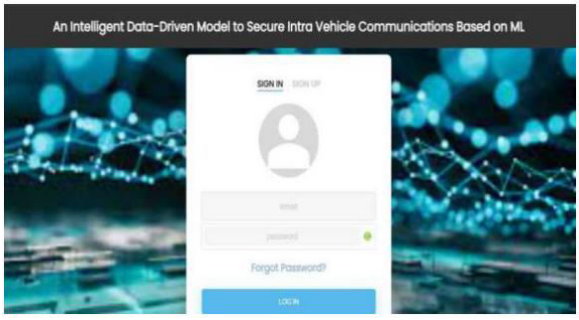


Fig 12: Landing Page

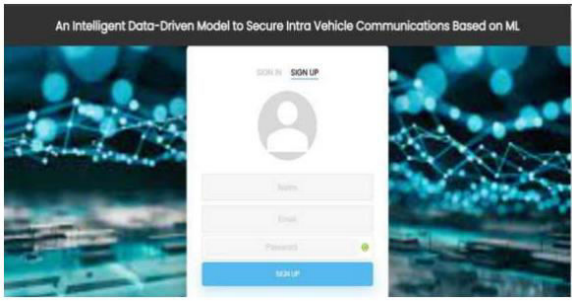


Fig 13: Registration Page

Fig 14: Email Verification Page

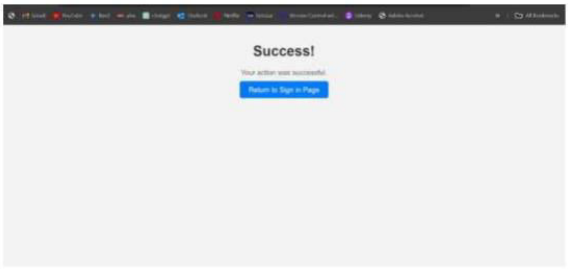


Fig 15 Registration Successful Page

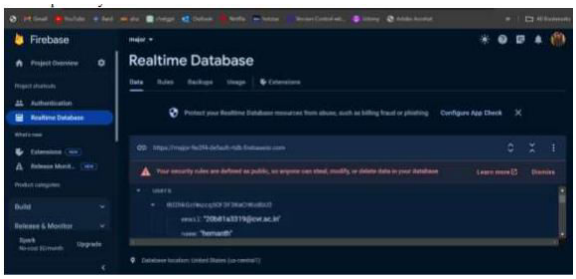


Fig 16 Firebase Real-time Database

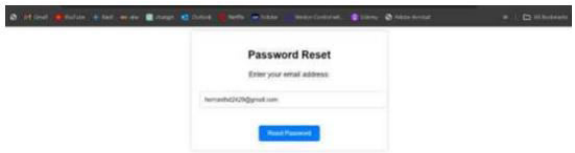


Fig 17 Reset Password Page

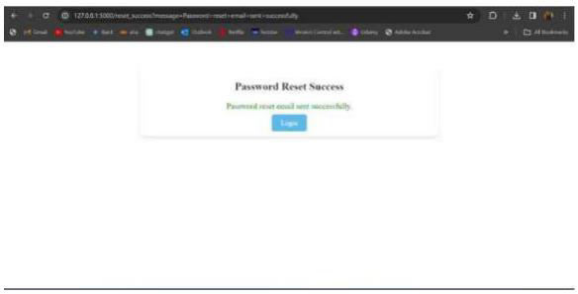


Fig 18: Password Updated

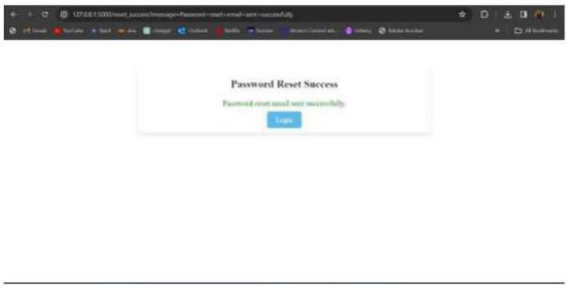


Fig 19: Reset Password Successfully Sent

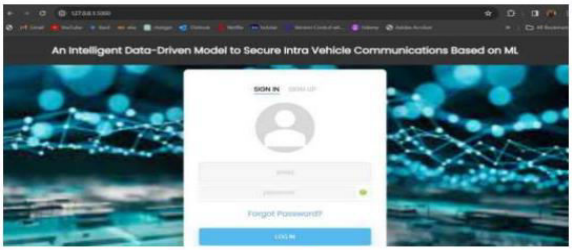


Fig 20: Login Page

Fig 21: Home Page

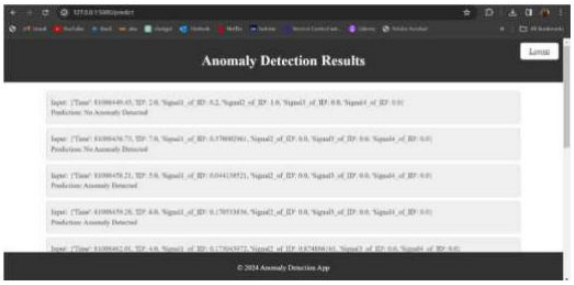


Fig 22 Prediction Results Page

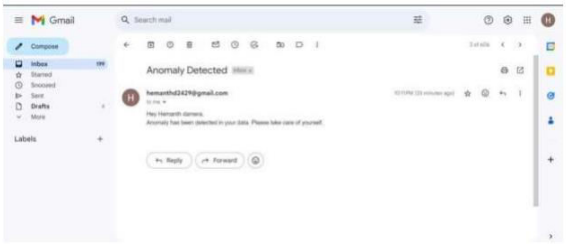


Fig 23 Email Alert Page

V. CONCLUSION

The proposed model leverages an improved support vector machine (SVM) model enhanced by the Modified Support Vector Optimization (MSSO) algorithm. This combination demonstrates a robust approach to anomaly detection, particularly in the context of cybersecurity. By utilizing this model, the system can effectively differentiate between malicious behaviors and trusted message frames in the Controller Area Network (CAN) protocol. The model's performance is evaluated using various metrics, including HR% (True Positive Rate), FR% (True Negative Rate), MR% (False Positive Rate), and CR% (False Negative Rate). The high values of HR% and FR% indicate the model's ability to accurately identify both malicious and trusted behaviors. Conversely, the low values of MR% and CR% suggest minimal misclassification of message frames, underscoring the model's reliability. The success of this proposed model signifies a significant advancement in cybersecurity measures, particularly in the context of CAN protocol security. Its ability to effectively detect anomalies while allowing legitimate message frames to transmit demonstrates its practical applicability and relevance in real-world scenarios.

VI. FUTURE SCOPE

In future works, the authors plan to further explore the performance of different anomaly detection models in the face of various cyberattacks. This ongoing assessment will contribute to continuous improvement and refinement of cybersecurity measures, ensuring robust protection against emerging threats.

REFERENCES

[1] A. Monot; N. Navet ; B. Bavoux ; F. Simonot-Lion, “Multisource Software on Multicore Automotive ECUs—Combining Runnable Sequencing With Task Scheduling”, IEEE Trans. Industrial Electronics, vol. 59, no. 10. Pp. 3934-3942, 2012.

[2] T.Y. Moon; S.H. Seo; J.H. Kim; S.H. Hwang; J. Wook Jeon, “Gateway system with diagnostic function for LIN, CAN and FlexRay”, 2007 International Conference on Control, Automation and Systems, pp. 2844 - 2849, 2007.

[3] B. Groza; S. Murvay, “Efficient Protocols for Secure Broadcast in Controller Area Networks”, IEEE Trans. Industrial Informatics, vol. 9, no. 4, pp. 2034-2042, 2013.

[4] B. Mohandes, R. Al Hammadi, W. Sanusi, T. Mezher, S. El Khatib, “Advancing cyberphysical sustainability through integrated analysis of smart power systems: A case study on electric vehicles”, International Journal of Critical Infrastructure Protection, vol. 23, pp. 33-48, 2018.

[5] G. Loukas, E. Karapistoli, E. Panaousis, P. Sarigiannidis, T. Vuong, A taxonomy and survey of cyber-physical intrusion detection approaches for

vehicles, Ad Hoc Networks, vol. 84, pp. 124-147, 2019.

[6] Hoppe T, Kiltz S, Dittmann J. Security threats to automotive can networks. practical examples and selected short-term countermeasures. Reliab Eng Syst Saf vol. 96, no. 1, pp. 11-25, 2011.

[7] Schulze S, Pukall M, Saake G, Hoppe T, Dittmann J. On the need of data management in automotive systems. In: BTW, vol. 144; pp. 217-26, 2009.

[8] Ling C, Feng D. An algorithm for detection of malicious messages on can buses. 2012 national conference on information technology and computer science. Atlantis Press; 2012.

[9] Oguma H, Yoshioka X, Nishikawa M, Shigetomi R, Otsuka A, Imai H. New attestation based security architecture for in-vehicle communication. In: Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE. IEEE; pp. 16, 2008.