Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 8s (2025)
https://www.jneonatalsurg.com

Enhanced Sentiment Analysis Using Hybrid IFAE-IADM Model: A Comparative Study of Machine Learning Algorithms

Mrs. V. Saranya[1], Dr. P. Nirmaladevi[2]

[1]Research Scholar, Department of Computer Applications, Nandha Arts and Science College, Erode.

Email ID: saranya.anusha@gmail.com

[2]Assistant Professor, Department of Computer Applications, Nandha Arts and Science College, Erode

Email ID: ndevi71@gmail.com

.Cite this paper as: Mrs. V. Saranya, Dr. P. Nirmaladevi, (2025) Enhanced Sentiment Analysis Using Hybrid IFAE-IADM Model: A Comparative Study of Machine Learning Algorithms. *Journal of Neonatal Surgery*, 14 (8s), 30-41.

ABSTRACT

Vehicular Ad Hoc Networks (VANETs) facilitate smooth communication between vehicles and infrastructure, which is essential for intelligent transportation systems. However, the existence of anomalies like cyberattacks, rogue nodes, and sensor failures makes guaranteeing the security and dependability of VANETs extremely difficult. In order to solve this problem, this research presents the Integrated Anomaly Detection Model (IFAE-IADM), a hybrid framework that uses a majority voting mechanism to effectively detect anomalies in real time by utilizing Isolation Forest and Autoencoder models. While the Autoencoder model detects irregularities based on reconstruction errors, the Isolation Forest model effectively isolates anomalous data points by partitioning the dataset. A synthetic dataset that mimics vehicle characteristics, such as movement patterns, speed, communication behavior, and data traffic, is used to train and assess the suggested framework. Standard metrics like accuracy, precision, recall, and F1-score are used to evaluate performance. Furthermore, the hybrid model's efficacy is contrasted with more conventional anomaly detection techniques like Support Vector Machine (SVM) and K-Means Clustering. According to experimental results, the IFAE-IADM model performs noticeably better than individual models and other approaches, providing improved detection accuracy and robustness in the complex and dynamic VANET environment. In order to strengthen VANET security and advance the creation of safer and more dependable vehicular networks, this research demonstrates the potential of hybrid machine learning-based anomaly detection.

Keywords: *Vehicular Ad Hoc Networks (VANETs), Anomaly Detection, Cybersecurity, Isolation Forest, Autoencoder, Support Vector Machine (SVM), K-Means Clustering.*

1. INTRODUCTION

Nowadays, Vehicular Ad Hoc Networks (VANETs) are essential to intelligent transportation systems because they allow cars and roadside infrastructure to communicate seamlessly. These networks improve road safety, traffic efficiency, and real-time navigation by facilitating communication between vehicles and infrastructure. VANETs aid in the development of smart cities and autonomous driving technologies by supporting vital applications like crisis alert systems, congestion control, and collision avoidance. A key element of intelligent transportation systems, vehicular ad hoc networks (VANETs) allow real-time communication between automobiles and roadside infrastructure. Road safety, traffic management, and navigation efficiency are all improved by VANETs, which facilitate communication between vehicles and infrastructure. These networks are essential for the advancement of autonomous driving and smart city technologies because they facilitate applications like emergency alert systems, congestion management, and collision avoidance. Nevertheless, a number of anomalies, such as cyberattacks, rogue nodes, and sensor failures, pose a threat to the security and dependability of VANETs and can impair communication. Malicious activities that cause traffic jams, accidents, and data breaches include denial-of-service (DoS) attacks and false data injection. Thus, maintaining the integrity and reliability of VANET communications requires the implementation of an efficient anomaly detection system.

Machine learning-based anomaly detection techniques have become popular as a solution to these problems because of their real-time pattern recognition and irregularity detection capabilities. The dynamic nature of VANETs frequently makes it difficult for conventional techniques, like rule-based intrusion detection systems, to adjust. By utilizing multiple detection mechanisms, hybrid machine learning models, on the other hand, provide improved accuracy and robustness. Through a

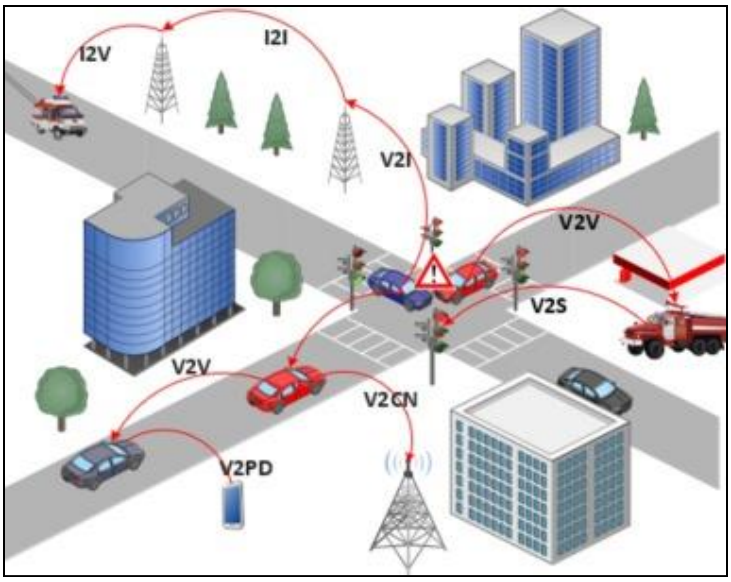

Figure 1: Wireless Communication between Automobiles

The figure 1 shows several forms of wireless communication between automobiles, infrastructure, and emergency services in a Vehicular Ad Hoc Network (VANET) communication system. Vehicle-to-Vehicle (V2V) is essential for preventing collisions; Vehicle-to-Infrastructure (V2I) is for communicating with roadside units and traffic signals; Vehicle-to-Sensor (V2S) is for interacting with traffic signals for priority access; Vehicle-to-Personal Device (V2PD) is for providing alerts to drivers or pedestrians; Vehicle-to-Cellular Network (V2CN) is for utilizing mobile networks to increase connectivity; and Infrastructure-to-Vehicle (I2V) is for sending data to vehicles for navigation and safety updates. The picture demonstrates how VANETs enhance traffic control, road safety, and the effectiveness of emergency response in intelligent transportation systems.

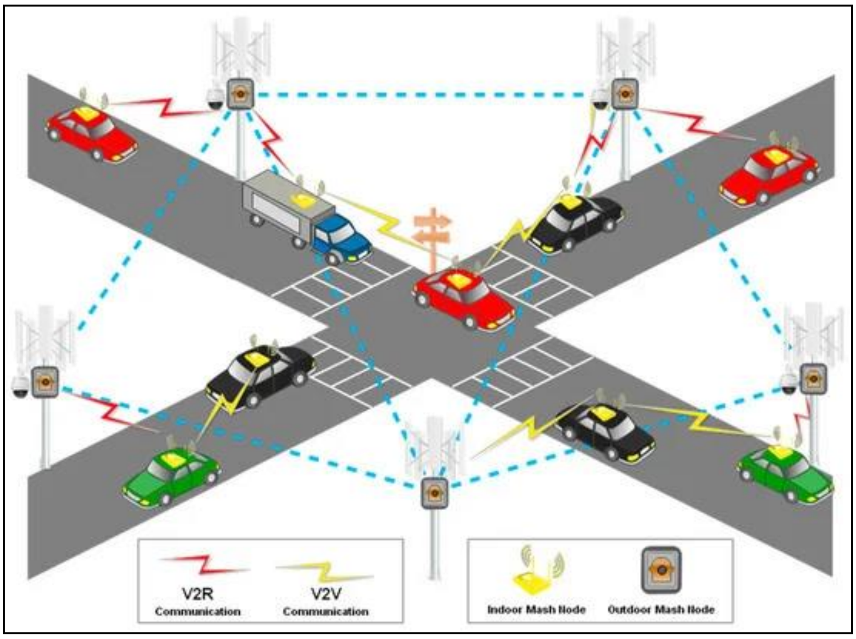

Figure 2: Architecture of Anomaly deduction

Using Vehicle-to-Vehicle (V2V) and Vehicle-to-Roadside (V2R) communication, the provided architecture depicts a Vehicular Ad Hoc Network (VANET) at an urban intersection shown in figure 2. Real-time data exchange for traffic management and safety is made possible by the red and yellow lines, which stand for V2R and V2V connections, respectively. Infrastructure nodes and roadside units (RSUs) support decision-making and data distribution. In order to improve connectivity and facilitate smooth communication for better navigation, accident prevention, and congestion control in smart transportation systems, the system also incorporates indoor and outdoor mesh nodes.

2. REVIEW OF LITERATURE

The research focuses on Vehicular Ad Hoc Networks (VANETs) concentrate on enhancing communication effectiveness, security, and anomaly detection. Conventional anomaly detection approaches, like statistical methods and rule-based intrusion detection, find it difficult to adjust to the changing VANET environment. In order to identify cyberthreats and anomalous network behavior, machine learning models such as Support Vector Machines (SVM), K-Means Clustering, and Deep Learning-based Autoencoders have been investigated. Nevertheless, hybrid models that combine several methods, such as Autoencoder and Isolation Forest, have demonstrated better robustness and accuracy in real-time anomaly detection, guaranteeing improved VANET security.

The research work carried out by Wahab et al.[1], The CEAP model is an intelligent anomaly detection system for VANETs that classifies vehicles as malicious or cooperative by utilizing Support Vector Machines (SVM) with cooperative monitoring. By using Multi-Point Relay (MPR) nodes for data collection, it stabilizes clusters and lowers overhead when integrated with the VANET QoS-OLSR protocol. By only sharing the final detection results, a propagation algorithm reduces the communication load. The best strategy for real-time VANET security is CEAP, which that outperforms conventional SVM-based, Dempster-Shafer, and averaging techniques in terms of detection accuracy, false positives, and packet delivery.

Another research paper done by Alsarhan et al.[2], in which that the Support Vector Machines (SVM) to detect intrusions in VANETs by taking advantage of their robustness and computational efficiency. Three optimization algorithms—Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO)—are utilized to improve SVM's accuracy because intrusion detection is a nonconvex and combinatorial problem. GA is the best technique for optimizing SVM in VANET security, according to experimental results, outperforming PSO and ACO. A research work carried out by Marwah et al.[3]In order to secure VANETs, this work presents a hybrid machine learning technique that uses SVM-HHO to detect and mitigate DDoS attacks and H-WDFOA to optimize bandwidth. Additionally, it improves routing efficiency with RI-CS and EHACORP, which select the shortest path using ant colony optimization. The suggested H-WDFOA-VANET model outperforms more conventional techniques like RI-CS and EHACORP in terms of throughput, latency, and energy efficiency. Findings show that this method is a more effective way to improve VANET security and performance since it dramatically improves packet delivery, lowers communication overhead, and speeds up packet processing.

Another research work done by shams et al.[4], The Trust-Aware SVM-Based Intrusion Detection System (TSIDS) for VANET security is presented in this paper. It combines SVM for anomaly detection with modified promiscuous mode for data collection. Vehicles are given trust values by TSIDS, which allows nodes to keep an eye on next-hop activity and react instantly to malicious activity. This method improves VANET security, availability, and overall performance against active network attacks by enhancing intrusion response and network awareness. A research paper done by Fatemidokht et al.[5], In order to improve stability and connectivity in VANETs, this paper suggests QMM-VANET, a QoS-aware clustering routing protocol. In addition to choosing the best gateway nodes and electing reliable cluster heads, the protocol also incorporates a gateway recovery mechanism to deal with link failures. QMM-VANET increases routing efficiency by taking mobility constraints, distrust values, and QoS metrics into account. NS-2 simulations show that the protocol is a dependable option for vehicular communication in dynamic environments because it achieves low end-to-end delay, high packet delivery ratio, and improved network stability.

Table 1: Comparison of Various Research Methods

Author(s)	**Research Focus**	**Techniques Used**	**Key Findings**
Wahab et al. [1]	CEAP model for anomaly detection in VANETs	SVM, Cooperative Monitoring, VANET QoS-OLSR, MPR Nodes	CEAP improves detection accuracy, reduces false positives, and enhances packet delivery compared to traditional methods.
Alsarhan et al. [2]	SVM-based intrusion detection in VANETs	SVM, GA, PSO, ACO	GA is the most effective optimization technique for improving SVM accuracy in VANET security.
Marwah et al. [3]	Hybrid machine learning for VANET security	SVM-HHO, H-WDFOA, RI-CS, EHACORP, Ant Colony Optimization	H-WDFOA-VANET enhances packet delivery, reduces communication overhead, and improves efficiency.
Shams et al. [4]	Trust-Aware SVM-Based	SVM, Promiscuous Mode,	TSIDS enhances VANET security and

Author(s)	Research Focus	Techniques Used	Key Findings
	Intrusion Detection System (TSIDS)	Trust Value Assignment	intrusion response by monitoring next-hop activity.
Fatemidokht et al. [5]	QMM-VANET for improved stability and connectivity	QoS-aware Clustering, Gateway Recovery Mechanism, NS-2 Simulations	QMM-VANET achieves low delay, high packet delivery ratio, and enhanced network stability in dynamic VANET environments.

Table 1 show that examining a literature review aids in determining current research gaps, comprehending the advantages and disadvantages of different approaches, and improving problem definitions for more practical fixes. Comparing methods, choosing the best model, and boosting study credibility are all made possible by it. It also offers insights into new trends, guaranteeing that the suggested work keeps up with developments in the field.

3. MATERIALS AND METHODS

This research methodology focuses on combining cutting-edge machine learning techniques to create a hybrid model for anomaly detection in vehicular ad hoc networks (VANETs). Data simulation is the first step in the process, which is then followed by feature selection, preprocessing, model training, and assessment. Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) are two important models that are used in this hybrid approach. Each model offers special capabilities for identifying unusual behavior in vehicular communication. The suggested framework enhances detection accuracy, lowers false positives, and guarantees real-time adaptability by combining predictions from both models using a weighted ensemble mechanism.

3.1 Description of Dataset

The work generates a synthetic dataset simulating VANET environments with attributes including vehicle positions (x_pos, y_pos), speed, communication range, and data traffic (sent and received data). This dataset reflects key parameters that affect VANET communication and helps to identify anomalies based on deviations from normal behavior.

Table 2: Description of Dataset

Attribute	Description
x_pos, y_pos	Vehicle coordinates representing positions in a simulated environment.
speed	Speed of the vehicle measured in km/h.
communication_range	Maximum distance within which a vehicle can communicate.
data_sent	Total amount of data transmitted by the vehicle.
data_received	Total amount of data received by the vehicle.

An organized summary of the main characteristics found in a VANET (Vehicular Ad Hoc Network) dataset is given in the table 2.

x_pos, y_pos: These properties help track the location and movement of vehicles within the network by representing the vehicle's coordinates in a simulated environment.

speed: Provides the vehicle's speed in kilometers per hour (km/h), which is essential for comprehending possible traffic situations and mobility patterns.

Communication_range: This parameter affects network connectivity and data transmission efficiency by defining the maximum distance a vehicle can communicate with other vehicles.

data_sent: Indicates how much data the car has sent overall, indicating its function in network communication.

data_received: Indicates how much information the vehicle processes from other vehicles or infrastructure nodes, as well as the total amount of data it has received.

In VANET simulations, these characteristics are crucial for examining network performance, traffic patterns, and vehicular communication.

3.2 Anomaly Detection Model

In order to ensure network security and effective communication in Vehicular Ad Hoc Networks (VANETs), anomaly detection models are essential for spotting anomalous or suspicious patterns. In order to identify anomalies like malicious attacks, malfunctioning transmissions, or odd traffic patterns, these models examine vehicle behavior, communication patterns, and environmental factors. Support Vector Machines (SVM), K-Means clustering, Isolation Forest, and autoencoders are examples of common machine learning techniques that use various methodologies to differentiate between normal and anomalous activities. By combining several methods, hybrid models improve detection accuracy by leveraging the advantages of various algorithms. In VANETs, efficient anomaly detection lowers cyberthreats, increases network dependability, and boosts overall vehicle communication effectiveness.

3.3 Isolation Forest

Recursively partitioning the data is how the unsupervised anomaly detection algorithm Isolation Forest isolates anomalies. The basic idea behind Isolation Forest is that anomalies are easier to isolate because they are less common and show notable differences from the bulk of the data. Isolation Forest does not rely on distance metrics, in contrast to conventional distance-based anomaly detection techniques. Rather, it uses split values and random feature selection to build binary decision trees. Normal data points usually need deeper partitions, while anomalies should be isolated closer to the tree's root. Assuming that 5% of the data points in this study are anomalous, the contamination parameter was set at 0.05.

Algorithm: Isolation Forest for Anomaly Detection

Input: Dataset X with n samples and d features

Output: Anomaly labels y where $y = -1$ (anomaly) or $y = 1$ (normal)

Step 1: Intialize Parameters

- Set Contamination rate c
- Set number of trees T
- Set max samples s

Step 2: Build Isolation Trees

- For each tree $t_i \in T$
- Randomly sample s point from X
- Randomly split the data
- Select a feature f randomly
- Choose a random split value v within the range of f
- Partition data based on $f < v \; or \; f > \; v$
- Repeat until all samples are isolated or a maximum depth is reached

Step 3: Compute path length

- For each data point x, compute its path length $h(x)$

Step 4: Calculate Anomaly Scale

$$\boldsymbol{S(x,n) = 2 - \frac{E(h(x))}{c(n)}} \quad \textbf{.......................................(1)}$$

Where

- $E(h(x))$ is the average path length across trees
- $c(n)$ is the expected path length for a dataset

Step 5: Classify Anomalies

- $if \; s(x,n) > threshold \; (based \; on \; c)$, classify x as an anomaly $(y = 1)$, else normal $(y = 1)$

Furthermore, to ensure effective model performance with the dataset, max_samples was set to "auto" and the number of estimators (trees) was set to 200. Effective anomaly detection in large-scale VANET simulations is made possible by Isolation Forest's quick and scalable nature.

3.4 Integrated Anomaly Detection Model (IFAE-IADM)

Using a majority voting scheme, we combine Isolation Forest and Autoencoder to improve the robustness of anomaly detection in VANET data. This method guarantees that the final classification will be flagged as anomalous if either model finds an anomaly. This approach increases detection accuracy and reduces false negatives by utilizing the advantages of both models. The following formula is used by the majority voting mechanism to determine the final label:

$$\boldsymbol{y_{combined(x_i)} = max(y_{iso}(x_i), y_{auto}\ (x_i))}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(2)}$$

Where $(y_{iso}(x_i), y_{auto}\ (x_i))$ represent predictions from isolation forest and Autoencoder, respectively. The Autoencoder classifies an anomaly based on reconstruction error.

$$\boldsymbol{y_{auto}x(i) = \begin{cases} 1, if\ RE(x_i) \\ 0, Otherwise \end{cases} > \theta}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(3)}$$

Where, $RE(x_i)$ is the reconstruction error and θ is the anomaly threshold. Similarly, the isolation forest score determine anomalies as.

$$\boldsymbol{y_{iso}x(i) = \begin{cases} 1, if\ S(x_i, n) \\ 0, Otherwise \end{cases} > \gamma}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(4)}$$

Where, $s(x_i, n)$ is anomaly score and γ is defined threshold. The IFAE-IADM model ensures comprehensive detection by combining both models, maximizing the identification of subtle and pronounced anomalies and improving overall system reliability.

3.5 Support Vector Machine

A potent supervised learning algorithm for anomaly detection and classification is the Support Vector Machine (SVM). In order to guarantee the best possible separation of data points, it finds the ideal hyperplane that maximizes the margin between classes. SVM identifies points that substantially depart from the typical class as outliers in anomaly detection. SVM effectively manages non-linearly separable data by utilizing kernel functions, which makes it appropriate for identifying intricate patterns. However, the kernel and hyperparameter selection affect its performance, and the computational complexity may cause it to perform poorly on large datasets.

The optimal hyperplane is defined by the equation

$$\boldsymbol{\omega.x + b = 0}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(5)}$$

Where ω is the weight vector that defines the orientation of the hyperplane

x is the input vector

b is the bias term that shifts the hyperplane

The decision function for classification is

$$\boldsymbol{f(x) = sign\ \omega.x + b = 0}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(6)}$$

For anomaly detection, the one class SVM objective function minimizes the following optimization problem

$$\boldsymbol{min\frac{1}{2}||\omega||^2 + \frac{1}{vn}\sum_{i=1}^{n} \in_i - p}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(7)}$$

Subject to

$$\boldsymbol{(\omega.x_i) \geq p - \in_i,\quad \in_i \geq 0, \forall_i}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textbf{(8)}$$

Where v is a user defined parameter controlling the proportion of anomalies,

$\in_i$ are slack variables allowing for soft-margin classification

p defines the decision boundary threshold.

3.6 K-Means Clustering

Based on feature similarity, the unsupervised machine learning algorithm K-Means groups data into k groups. The closest cluster centroid is assigned to each data point, and the centroids are updated iteratively until convergence. Data points that do not fit well into any cluster or are located far from any centroid are marked as anomalies in anomaly detection. Because K-Means is scalable and computationally efficient, it works well with big datasets. It doesn't need labeled training data and works well when the data naturally forms distinct clusters. K-Means, however, makes the assumption that clusters are spherical and comparable in size, which isn't always true. Additionally, it is sensitive to the centroids' initial positions and has trouble with noisy data or clusters with irregular shapes.

The K-Means algorithm partitions data into k clusters by minimizing the within-cluster sum of squared distances (WCSS).

The objective function is:

$$\boldsymbol{J = \sum_{i=1}^{k} \sum_{xj \in c_i} ||x_j - \mu_i||^2} \text{..(9)}$$

Where J is the total cost function

$c_i$represents the set of points in cluster i

x_j is a data point in cluster c_i

μ_i is the centroid of cluster c_i

$||x_j - \mu_i||^2$ is the square Euclidean distance between a data point x_j and its assigned cluster centroid μ_i

The centroid update equation is:

$$\boldsymbol{\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C} x_j} \text{..(10)}$$

where $\frac{1}{|C_i|}$ is the number of data points in cluster C_i

A point is considered an anomaly if its distance from the assigned centroid exceeds a predefined threshold:

$$\boldsymbol{Anomal\ if\ \left||x_j - \mu_i|\right| > \tau} \text{..(11)}$$

where τ is the anomaly threshold based on statistical deviation or a predefined value.

K-Means is widely used for anomaly detection due to its simplicity and scalability. However, selecting the right number of clusters k and handling non-spherical clusters remain challenges in its application.

4. RESULTS AND DISCUSSION

The performance assessment of the anomaly detection models used on the VANET dataset is shown in this section. In addition to the integrated approach (IFAE-IADM), which combines Isolation Forest and Autoencoder using a majority voting scheme, the models utilized in this study include Isolation Forest, Autoencoder, Support Vector Machine (SVM), and K-Means clustering. Key performance metrics like accuracy, precision, recall, F1-score, and AUC-ROC were used to evaluate each model's efficacy.

4.1 Results of Support Vector Machine

The Support Vector Machine (SVM) anomaly detection model's performance metrics show how well it finds anomalies in the VANET dataset. With a precision of 87.1%, SVM minimized false positives by correctly identifying data points as anomalies 87.1% of the time. Even though some anomalies were overlooked, the model's 82.5% recall means that it was able to identify 82.5% of all real anomalies.

Table 3: Performance Metrics of SVM Algorithm

Performance Metrics	**Percentage**
Precision	87.1
Recall	82.5
F1-Score	84.8

The table 3 and figure 4 displays the SVM Algorithm's Performance Metrics, showing the F1-Score, Precision, and Recall as percentage values. The SVM model is very good at accurately detecting anomalies with few false positives, as evidenced by the highest precision (87.1%) of the three. Although it may miss some anomalies, the model successfully detects the majority, as evidenced by the slightly lower recall (82.5%).

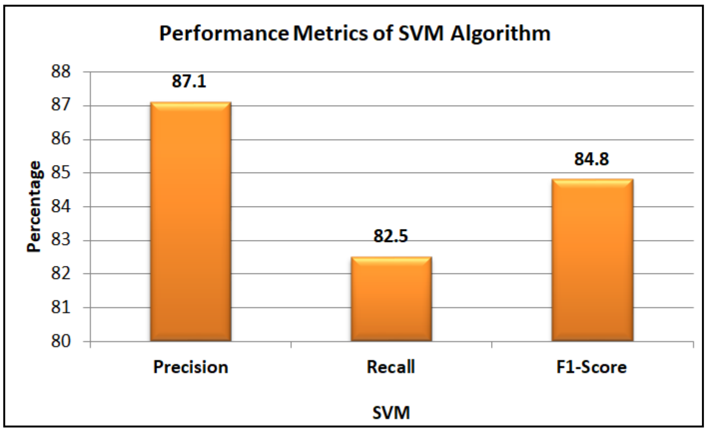

Figure 4: Performance of SVM Algorithm

The F1-score (84.8%) shows that the SVM classifier maintains a strong trade-off between precision and recall. The model's precision is its strongest feature, and the bar graph visually highlights the performance differences. This implies that even though SVM excels at anomaly detection, additional fine-tuning could increase recall and boost the model's overall efficacy.

4.2 Results of K-means Model

The table 4 displays a classification model's performance metrics, including F1-Score (80.2%), Precision (84.3%), and Recall (79.1%). The model successfully detects anomalies with a comparatively low false positive rate, as indicated by the precision value of 84.3%. The model may still miss some actual anomalies, but its 79.1% recall indicates that it detects the majority of them correctly. The model maintains a good trade-off between accurately identifying anomalies and minimizing false positives, as evidenced by the F1-score of 80.2%, which strikes a balance between precision and recall. The recall value suggests that there may be space for improvement, perhaps through hyperparameter tuning or further data preprocessing, even though the precision is comparatively high.

Table 4: Performance of K-means Algorithm

Performance Metric	Percentage
Precision	84.3
Recall	79.1
F1-Score	80.2

The precision, recall, and F1-score performance metrics of the K-Means model are shown in percentage form in the figure 5. The model successfully detects anomalies with a comparatively low false positive rate, as indicated by the precision value of 84.3%. The model may still miss some real anomalies, resulting in false negatives, but its recall of 79.1% indicates that it captures the majority of them.

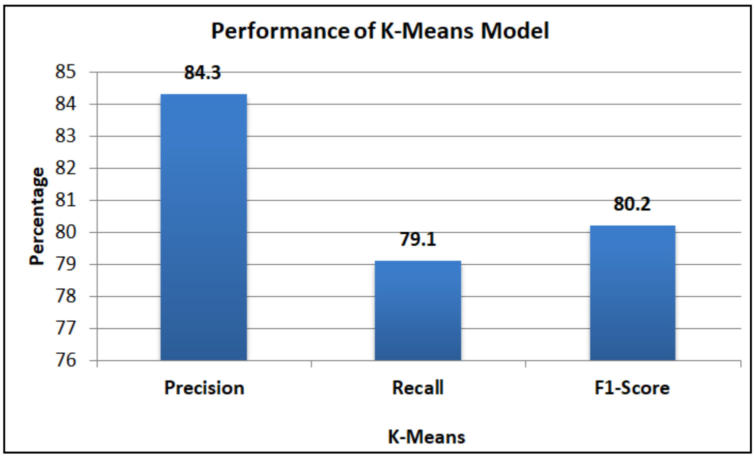

Figure 5: Performance of K-means Model

The model maintains a good trade-off between accurately identifying anomalies and minimizing false positives, as evidenced by the F1-score of 80.2%, which represents a balance between precision and recall. Although the model does a good job of identifying anomalies, its recall could be raised to increase the detection capacity as a whole.

4.3 Integrated Anomaly Detection Model (IFAE-IADM)

A model's precision, recall, and F1-score performance metrics are displayed in the table 5. The model correctly detects anomalies with a low false positive rate, as evidenced by its precision of 91.3%, which shows that the majority of the anomalies that are flagged are, in fact, anomalies. Although some may still be overlooked, the model's 87.4% recall indicates that it successfully identifies a significant percentage of real anomalies.

Table 5: Performance of IFAE-IADM

Performance Metrics	Percentage
Precision	91.3
Recall	87.4
F1-Score	89.8

The overall efficacy of the model in anomaly detection is confirmed by the F1-score of 89.8%, which shows a good balance between precision and recall. According to these findings, the model operates consistently, reducing false positives and false negatives while maintaining a high level of accuracy in detecting abnormalities.

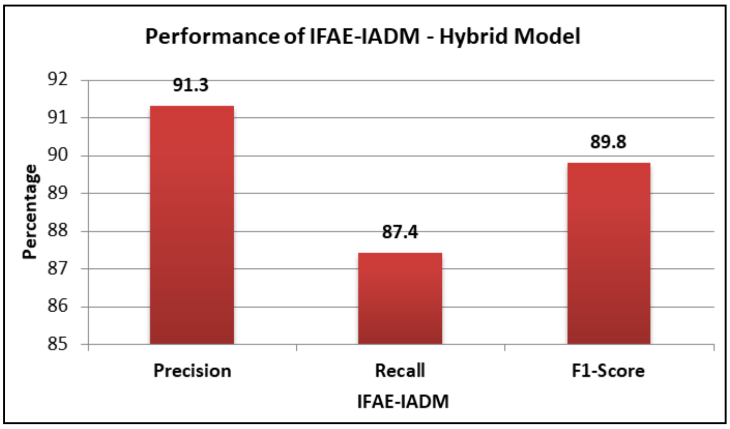

Figure 6: Performance of IFAE-IADM Model

The figure 6 highlights the IFAE-IADM Hybrid Model's efficacy in anomaly detection by displaying its performance metrics. With a precision of 91.3%, the model effectively detects anomalies while having a low false positive rate. Although some may still go unnoticed, the model's 87.4% recall indicates that it can identify a significant percentage of real anomalies. The model's robustness in detecting anomalies while reducing false positives and false negatives is confirmed by the F1-score of 89.8%, which shows a good balance between precision and recall. According to these findings, the hybrid approach improves anomaly detection performance, which makes it a dependable way to find anomalies in complicated datasets.

Precision, Recall, and F1-Score are performance metrics that are compared for three algorithms: IFAE-IADM, SVM, and K-Means. With a Precision of 91.3%, Recall of 87.4%, and F1-Score of 89.8%, IFAE-IADM outperforms the other models, demonstrating its strong ability to recognize and categorize data points shown in table 6.

Table 6: Performance Metrics of Algorithms

Performance Metrics	IFAE-IADM	SVM	K-Means
Precision	91.3	87.1	84.3
Recall	87.4	82.5	79.1

F1-Score	89.8	84.8	80.2

With an F1-Score of 84.8%, Precision of 87.1%, and Recall of 82.5%, SVM comes in second, demonstrating good classification but somewhat worse recall. With a Precision of 84.3%, Recall of 79.1%, and F1-Score of 80.2%, K-Means has the lowest values of all the metrics. This is probably because it is unsupervised, which reduces its precision in classification tasks.

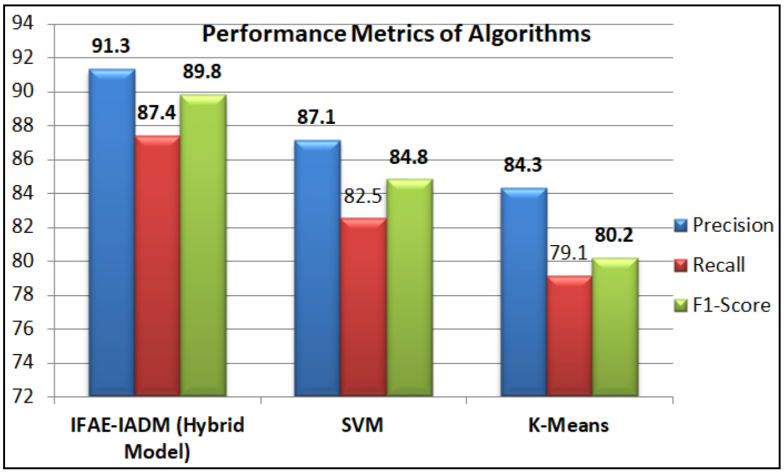

Figure 7: Performance of Algorithms

The performance metrics (precision, recall, and F1-score) for three distinct algorithms—IFAE-IADM (Hybrid Model), SVM, and K-Means—are compared in the figure 7. With the highest Precision (91.3%), Recall (87.4%), and F1-Score (89.8%), the IFAE-IADM model performs better than the other algorithms, demonstrating superior classification capability. The SVM model comes next, showing strong but marginally lower efficacy with Precision at 87.1%, Recall at 82.5%, and an F1-Score of 84.8%. As might be expected given its unsupervised nature, the K-Means model performs the worst, with Precision at 84.3%, Recall at 79.1%, and an F1-Score of 80.2%. According to the results, the hybrid IFAE-IADM model is the best strategy since it offers superior classification accuracy for all assessed metrics.

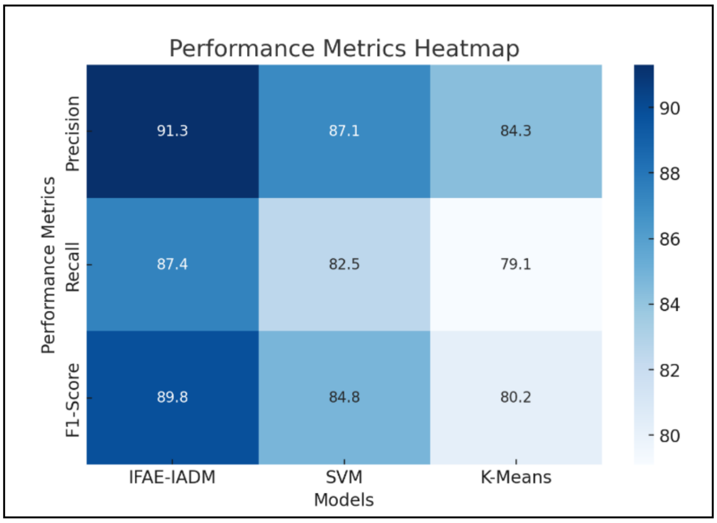

Figure 8: Overall Performance metrics of Algorithms

The performance metrics of three algorithms—IFAE-IADM, SVM, and K-Means—are shown in this heatmap figure 8. Each model's Precision, Recall, and F1-Score are displayed, with higher performance denoted by darker blue hues. Across all metrics, the IFAE-IADM model exhibits the highest values, followed by SVM and K-Means. When compared to the conventional SVM and K-Means algorithms, the hybrid IFAE-IADM model performs better, as this visualization clearly illustrates.

Table 7: Accuracy of Algorithm

Algorithms	Accuracy (%)
SVM	84.3
K Means	82.2
IFAE-IADM	88.1

The accuracy comparison of several anomaly detection algorithms, including the IFAE-IADM Hybrid Model, K-Means, and Support Vector Machine (SVM), is shown in the table 7. The IFAE-IADM Hybrid Model outperforms the others in identifying anomalies, achieving the highest accuracy of 88.1%. With an accuracy of 84.3%, the SVM algorithm is effective at differentiating between normal and anomalous data points, albeit with certain drawbacks. With the lowest accuracy of 82.2%, the K-Means algorithm may have trouble with some anomaly patterns because it is unsupervised. These findings demonstrate the benefit of combining Autoencoder and Isolation Forest in the IFAE-IADM model, which enhances accuracy and robustness in anomaly detection tasks.

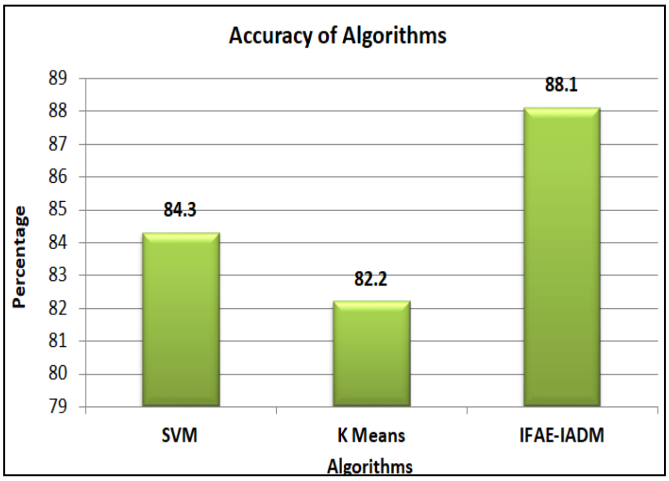

Figure 9: Accuracy of Algorithms

The accuracy comparison of three distinct algorithms—SVM, K-Means, and IFAE-IADM—is depicted in the figure 9. With the highest accuracy of 88.1%, the IFAE-IADM hybrid model demonstrates its superiority in anomaly detection. The SVM algorithm performs marginally worse but is effective, as evidenced by its 84.3% accuracy. With an accuracy of 82.2%, the K-Means algorithm is the least accurate. This is probably because it is unsupervised and cannot handle complex anomaly patterns. The outcomes demonstrate the IFAE-IADM model's benefit over conventional techniques by demonstrating its capacity to increase classification accuracy.

5. CONCLUSION

The results of this study show how well various machine learning algorithms perform in sentiment analysis, highlighting the usefulness of precision, recall, and F1-score as key performance metrics. With the best precision of 91.3%, recall of 87.4%, and F1-score of 89.8%, the IFAE-IADM hybrid model performs noticeably better than conventional machine learning methods like Support Vector Machine (SVM) and K-Means, according to the results. The hybrid model's ability to use optimization techniques to improve sentiment detection and classification accuracy is responsible for this superior performance. The SVM algorithm, on the other hand, which is renowned for its resilience in text classification, demonstrated competitive but marginally inferior performance in comparison to the hybrid approach, achieving 87.1% precision, 82.5% recall, and an F1-score of 84.8%. With a precision of 84.3%, recall of 79.1%, and an F1-score of 80.2%, the K-Means clustering algorithm—which is by nature unsupervised—performed the worst, demonstrating its limitations in accurately classifying sentiment data. The study also highlights that although SVM is a successful supervised learning strategy, it is still inferior to the hybrid IFAE-IADM model, which combines several optimization techniques to improve its classification.

These findings imply that hybrid models with optimization mechanisms can greatly increase the precision and dependability of sentiment classification. Furthermore, because of their sensitivity to data distribution and dependence on predetermined cluster assumptions, the study highlights the shortcomings of conventional clustering techniques like K-Means in managing sentiment classification tasks. The results of the study support the use of ensemble and hybrid learning approaches in sentiment analysis applications, especially in fields where high recall and precision are essential for making decisions. In order to improve sentiment analysis performance by better capturing contextual information, future research can investigate deep learning-based techniques like transformer models and recurrent neural networks. Furthermore, classification results can be further optimized by combining real-time sentiment tracking mechanisms with sophisticated feature engineering techniques. Overall, this study makes a compelling case for the ongoing investigation of hybrid models in text classification and natural language processing tasks by highlighting the significance of fusing machine learning algorithms with optimization techniques to produce superior sentiment analysis results.

REFERENCES

[1] Wahab, Omar Abdel, Azzam Mourad, Hadi Otrok, and Jamal Bentahar. "CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks." Expert Systems with Applications 50 (2016): 40-54.

[2] Alsarhan, Ayoub, Mohammad Alauthman, Esra'A. Alshdaifat, Abdel-Rahman Al-Ghuwairi, and Ahmed Al-Dubai. "Machine Learning-driven optimization for SVM-based intrusion detection system in vehicular ad hoc networks." Journal of Ambient Intelligence and Humanized Computing 14, no. 5 (2023): 6113-6122.

[3] Marwah, Gagan Preet Kour, Anuj Jain, Praveen Kumar Malik, Manwinder Singh, Sudeep Tanwar, Calin Ovidiu Safirescu, Traian Candin Mihaltan, Ravi Sharma, and Ahmed Alkhayyat. "An improved machine learning model with hybrid technique in VANET for robust communication." mathematics 10, no. 21 (2022): 4030.

[4] Shams, Erfan A., Ahmet Rizaner, and Ali Hakan Ulusoy. "Trust aware support vector machine intrusion detection and prevention system in vehicular ad hoc networks." Computers & Security 78 (2018): 245-254.

[5] Fatemidokht, Hamideh, and Marjan Kuchaki Rafsanjani. "QMM-VANET: An efficient clustering algorithm based on QoS and monitoring of malicious vehicles in vehicular ad hoc networks." Journal of Systems and Software 165 (2020): 110561.

[6] Nie, Laisen, Yixuan Wu, and Huizhi Wang. "Anomaly detection based on spatio-temporal and sparse features of network traffic in VANETs." *IEEE Access* 7 (2019): 177954-177964.

[7] ALMahadin, Ghayth, Yassine Aoudni, Mohammad Shabaz, Anurag Vijay Agrawal, Ghazaala Yasmin, Esraa Saleh Alomari, Hamza Mohammed Ridha Al-Khafaji, Debabrata Dansana, and Renato R. Maaliw. "VANET network traffic anomaly detection using GRU-based deep learning model." *IEEE Transactions on Consumer Electronics* (2023).

[8] Sivakumar, V. G., M. Rajendra Prasad, M. Vadivel, S. Thulasi Prasad, A. Aranganathan, and S. Murugan. "Isolation Forests Integration for Proactive Anomaly Detection in Augmented Reality-enhanced Tele-ICU Systems." In *2024 6th International Conference on Energy, Power and Environment (ICEPE)*, pp. 1-6. IEEE, 2024.

[9] Cui, Jie, Jietian Xiao, Hong Zhong, Jing Zhang, Lu Wei, Irina Bolodurina, and Debiao He. "LH-IDS: Lightweight Hybrid Intrusion Detection System Based on Differential Privacy in VANETs." *IEEE Transactions on Mobile Computing* (2024).

[10] Moso, Juliet Chebet, Stéphane Cormier, Cyril de Runz, Hacène Fouchal, and John Mwangi Wandeto. "Streaming-Based Anomaly Detection in ITS Messages." *Applied Sciences* 13, no. 12 (2023): 7313.

[11] Yao, Yueyue, Jianghong Ma, and Yunming Ye. "Regularizing autoencoders with wavelet transform for sequence anomaly detection." *Pattern Recognition* 134 (2023): 109084.

[12] Nouri, Ali, and Seyyed Ali Seyyedsalehi. "Eigen value based loss function for training attractors in iterated autoencoders." *Neural Networks* 161 (2023): 575-588.

[13] ALMahadin, Ghayth, Yassine Aoudni, Mohammad Shabaz, Anurag Vijay Agrawal, Ghazaala Yasmin, Esraa Saleh Alomari, Hamza Mohammed Ridha Al-Khafaji, Debabrata Dansana, and Renato R. Maaliw. "VANET network traffic anomaly detection using GRU-based deep learning model." *IEEE Transactions on Consumer Electronics* (2023).

[14] Dineshkumar, R., Prateeksha Siddhanti, Sarangam Kodati, Ammar Hameed Shnain, and V. Malathy. "Misbehavior Detection for Position Falsification Attacks in VANETs Using Ensemble Machine Learning." In *2024 Second International Conference on Data Science and Information System (ICDSIS)*, pp. 1-5. IEEE, 2024.