www.ijarp.com ISSN 2456-9992 Page: 01-19

International Journal Advanced Research Publications

Research Article

Volume: 01

Issue: 03

AN EXPLAINABLE DECISION TREE–BASED FRAMEWORK FOR ANOMALY DETECTION IN AUTONOMOUS SYSTEMS

***[1]Mallikarjuna G D, [2]Mandara A S**

[1]Director, Snipe Tech Private Ltd, Bangalore, Karnataka, India.

[2]AI Researcher, Snipe Tech Private Limited, Bangalore, Karnataka, India.

Article Received: 15 November 2025, Article Revised: 05 December 2025, Published on: 25 December 2025

***Corresponding Author: Mallikarjuna G D**

Director, Snipe Tech Private Ltd, Bangalore, Karnataka, India.

DOI: https://doi-doi.org/101555/ijarp.4840

ABSTRACT

Autonomous systems increasingly rely on Artificial Intelligence (AI) techniques to monitor system behavior and detect anomalies arising from cyber-attacks, sensor failures, or abnormal operational patterns. Although existing machine learning and deep learning models achieve high accuracy in anomaly detection, many of them function as black-box systems and lack transparency in their decision-making processes. This absence of interpretability limits trust, validation, and adoption in safety-critical autonomous environments. To address this challenge, this paper proposes an Explainable Decision Tree–based framework for anomaly detection in autonomous systems. The proposed approach employs a Decision Tree classifier to identify anomalous and normal system behavior while providing clear and interpretable explanations through feature importance analysis and decision rule paths. The framework is evaluated using datasets derived from autonomous, IoT, and Vehicular Ad Hoc Network (VANET) environments, which contain features related to sensor readings, system states, and communication behavior. Experimental results demonstrate that the Decision Tree model achieves reliable anomaly detection performance while maintaining inherent interpretability. The explainable nature of the proposed framework enables users to understand which features contribute to anomaly detection decisions, thereby improving transparency and accountability. By combining accurate classification with human-understandable explanations, the proposed framework enhances trust, supports safety assurance, and promotes the adoption of AI-driven solutions in autonomous and intelligent systems.

KEYWORDS: Explainable Artificial Intelligence, Decision Tree, Anomaly Detection, Autonomous Systems, IoT, Interpretability.

INTRODUCTION

The rapid advancement of autonomous and Internet of Things (IoT)–enabled systems has transformed various domains, including transportation, smart cities, industrial automation, healthcare, and intelligent surveillance. Autonomous systems are designed to operate with minimal human intervention by sensing their environment, processing large volumes of data, and making real-time decisions. With the increasing deployment of autonomous vehicles, smart sensors, and connected devices, these systems have become highly complex and data-driven. As a result, ensuring reliability, safety, and security has become a critical research challenge. Any abnormal behavior, whether caused by cyber-attacks, sensor malfunctions, communication failures, or unexpected environmental conditions, can lead to severe consequences in safety-critical applications. Artificial Intelligence (AI) plays a central role in enabling autonomy by allowing systems to learn patterns from data and make intelligent decisions. Machine Learning (ML) and Deep Learning (DL) models are widely used for tasks such as anomaly detection, fault diagnosis, intrusion detection, and predictive maintenance in autonomous and IoT environments. These models analyze high-dimensional sensor and communication data to distinguish between normal and anomalous behavior. In real-time autonomous systems, AI-driven anomaly detection is essential to prevent system failures, detect malicious activities, and maintain operational stability.

However, the growing dependence on AI for real-time decision-making introduces significant challenges related to transparency and trust. Many state-of-the-art anomaly detection approaches rely on complex ML and DL models such as neural networks, ensemble models, and deep architectures. While these models often achieve high detection accuracy, they typically operate as black-box systems, providing predictions without explaining the reasoning behind their decisions. In safety-critical autonomous environments, such as autonomous driving systems or industrial IoT networks, this lack of interpretability is a major concern. System operators, engineers, and regulatory authorities require not only accurate predictions but also a clear understanding of why a particular behavior is classified as anomalous. Security threats further amplify the importance of explainable anomaly detection. Autonomous and IoT-enabled systems are vulnerable to a wide range of attacks, including spoofing, data injection, denial-of-service, and protocol manipulation. Additionally, non-

malicious anomalies such as sensor drift, hardware degradation, and environmental disturbances can also affect system performance. Traditional black-box AI models may detect anomalies but fail to provide insights into the root causes of these events. This makes system debugging, incident analysis, and preventive action extremely difficult. In real-world deployments, unexplained AI decisions can reduce user confidence and hinder large-scale adoption of autonomous technologies. To address these challenges, Explainable Artificial Intelligence (XAI) has emerged as an important research direction. XAI aims to make AI systems more transparent, interpretable, and accountable by providing human-understandable explanations for model predictions. Explainability is particularly crucial in safety-critical systems, where decisions must be validated and justified. Regulatory frameworks and ethical guidelines increasingly emphasize the need for explainable and trustworthy AI. In this context, anomaly detection models must not only achieve high accuracy but also offer clear explanations regarding which features or conditions led to an anomalous classification. Among various machine learning algorithms, Decision Tree models are particularly well-suited for explainable anomaly detection. Decision Trees inherently provide interpretable decision-making by representing classification logic as a set of hierarchical rules based on feature thresholds. Each decision path from the root node to a leaf node corresponds to a clear and understandable rule that explains why a particular instance is classified as normal or anomalous. Unlike complex black-box models, Decision Trees allow users to trace predictions back to specific features and conditions, making them highly transparent and suitable for autonomous systems.

Another key motivation for using Decision Tree models is their ability to identify feature importance directly during the training process. By analyzing how features are used to split the data, Decision Trees highlight the most influential attributes contributing to anomaly detection. This feature-level interpretability is essential for understanding system behavior, diagnosing faults, and improving system design. Furthermore, Decision Trees are computationally efficient, easy to implement, and capable of handling both numerical and categorical data, which are common in autonomous and IoT datasets. Despite these advantages, Decision Tree–based approaches have received comparatively less attention in the context of explainable anomaly detection for autonomous systems, as recent research has largely focused on deep learning and ensemble methods. This creates an opportunity to revisit classical, interpretable machine learning models and demonstrate their effectiveness when combined with a systematic explainability framework. By leveraging the inherent

interpretability of Decision Trees, it is possible to design anomaly detection systems that balance performance and transparency.

In this paper, we propose an Explainable Decision Tree–based framework for anomaly detection in autonomous systems. The proposed framework integrates data preprocessing, Decision Tree–based classification, and explainability mechanisms to detect anomalous behavior while providing clear and understandable explanations. The system analyzes autonomous and IoT-related datasets containing sensor, operational, and communication features and classifies system behavior into normal and anomalous categories. Decision rules and feature importance measures are used to explain model predictions, enabling users to understand the reasoning behind anomaly detection decisions.

The main contributions of this work are threefold. First, it presents an interpretable anomaly detection framework based on Decision Tree models for autonomous and IoT-enabled systems. Second, it demonstrates how feature importance and decision paths can be used to explain anomaly detection results in a human-understandable manner. Third, it highlights the importance of explainability in enhancing trust, transparency, and safety in AI-driven autonomous environments. The proposed framework aims to bridge the gap between accurate anomaly detection and explainable decision-making, thereby supporting the development of trustworthy autonomous systems.

LITERATURE SURVEY

This section reviews existing research related to anomaly detection in autonomous systems and the role of Explainable Artificial Intelligence (XAI) in improving transparency and trust. Recent studies highlight the growing need for reliable anomaly detection mechanisms in autonomous and IoT-enabled environments due to increasing system complexity and security threats [1]. The review discusses commonly used approaches, their strengths and limitations, and identifies the research gap addressed by this work. Anomaly detection is a critical component in ensuring the safety and reliability of autonomous systems. Early research primarily relied on rule-based techniques, where predefined thresholds and expert-defined rules were used to detect abnormal behavior [2]. Although these systems are easy to interpret, they lack adaptability and perform poorly in dynamic and data-intensive autonomous environments. To overcome these limitations, machine learning (ML) approaches were introduced to automatically learn patterns from system data and improve anomaly detection accuracy. Traditional ML algorithms such as Decision Trees, Random Forests, Support Vector

Machines, and k-Nearest Neighbors have been widely applied for anomaly detection in autonomous and IoT-enabled systems [3], [4]. Decision Trees are particularly valued for their rule-based structure and interpretability, while Random Forests enhance performance by combining multiple trees. In recent years, deep learning (DL) models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have gained popularity due to their ability to capture complex spatial and temporal relationships in sensor and communication data [5], [6]. Although these models achieve high detection accuracy, their internal decision-making processes are often opaque, resulting in limited interpretability. Consequently, many ML and DL-based anomaly detection systems function as black-box models, which restricts their adoption in safety-critical autonomous applications [7]. Explainable Artificial Intelligence (XAI) has emerged as an important research area to address the transparency limitations of black-box AI models [8]. XAI techniques aim to provide human-understandable explanations that describe how and why AI models arrive at specific decisions. In anomaly detection, explainability helps system operators understand the reasons behind anomalous classifications and identify the contributing factors. Various XAI methods have been proposed to improve model transparency, including feature importance analysis, rule-based explanations, and visualization techniques [9]. Feature importance methods highlight the most influential input variables affecting predictions, while rule-based explanations present decision logic in an interpretable form. Inherently interpretable models, such as Decision Trees, naturally support explainability by representing classification decisions as a set of hierarchical rules [10]. Explainable models offer several benefits, including increased trust, easier validation, improved debugging, and better compliance with ethical and regulatory requirements. These advantages make XAI particularly valuable in safety-critical autonomous systems, where transparency is essential [11]. Despite significant progress in anomaly detection and explainable AI research, a clear research gap remains. Many existing studies prioritize detection accuracy using complex ML and DL models, often at the expense of interpretability [12]. Although some research incorporates explainability techniques, these are frequently applied as post-hoc solutions and are not tightly integrated into the anomaly detection framework [13]. Moreover, limited studies focus on combining inherently interpretable models, such as Decision Trees, with explainable AI principles for anomaly detection in autonomous systems.

In particular, there is a lack of frameworks that provide feature-level explanations and clear decision rules to support root cause analysis and system validation [14]. This gap highlights

the need for interpretable anomaly detection models that balance performance and transparency. Addressing this need, the proposed research focuses on developing an Explainable Decision Tree–based framework that enables effective anomaly detection while providing meaningful, feature-level explanations for autonomous systems. Recent advancements in autonomous and IoT-enabled systems have led to an increased reliance on Artificial Intelligence (AI) and Machine Learning (ML) techniques for monitoring system behavior and detecting anomalies. A significant body of research has explored the use of ML and Deep Learning (DL) models to identify abnormal patterns caused by cyber-attacks, sensor malfunctions, communication failures, and unexpected environmental conditions [1], [2]. These approaches have demonstrated strong performance in terms of detection accuracy and scalability across various autonomous system applications, including intelligent transportation systems, industrial automation, and smart infrastructure. However, a critical limitation repeatedly highlighted in the literature is that most existing anomaly detection models function primarily as predictive systems without providing explanations for their decisions [3], [4]. Complex ML and DL models such as neural networks, ensemble methods, and deep architectures are often treated as black-box models, where the internal reasoning behind predictions remains hidden. Although these models can accurately label an instance as normal or anomalous, they fail to communicate *why* a particular decision was made. This limitation has been identified as a major obstacle to the practical deployment of AI-based anomaly detection systems in real-world autonomous environments [5].

The lack of interpretability directly impacts trust and validation, especially in safety-critical systems. Several studies emphasize that in autonomous driving, industrial IoT, and cyber-physical systems, unexplained AI decisions can lead to hesitation in adoption by system operators and regulatory authorities [6], [7]. In such environments, AI systems are expected not only to perform accurately but also to justify their decisions in a transparent and understandable manner. Without clear explanations, it becomes difficult to assess whether a detected anomaly is the result of genuine system failure, malicious activity, data bias, or model error [8]. This uncertainty reduces confidence in AI-driven decision-making and limits its acceptance in mission-critical applications. Another important issue identified in the literature is the difficulty of validating and debugging black-box anomaly detection models. When a model flags an anomaly without explanation, system engineers are unable to determine whether the alert is meaningful or a false positive [9]. This often results in unnecessary system interventions or overlooked failures. Furthermore, the absence of

explainability complicates model auditing, performance tuning, and compliance with emerging AI governance and ethical guidelines [10]. As autonomous systems become more widespread, regulatory frameworks increasingly demand transparency, accountability, and explainability from AI-based solutions.

Table 1. Comparative Analysis of Existing Studies on Anomaly Detection and Explainable AI.

Study / Reference	Application Domain	Model Used	Focus of Study	Approach	Key Contributions	Explainability Support	Limitations
Study [1]	Autonomous Driving	Deep Neural Network	Anomaly detection accuracy	Supervised DL-based classification	High detection accuracy for complex patterns	No	Black-box model, no interpretability
Study [2]	VANETs	CNN	Network intrusion detection	Feature extraction using CNN	Improved detection of network attacks	No	High computational cost
Study [3]	IoT Systems	LSTM	Time-series anomaly detection	Sequential data modeling	Captures temporal dependencies	No	Difficult to interpret predictions
Study [4]	Cyber-Physical Systems	Random Forest	Fault detection	Ensemble learning	Robust performance against noise	Partial (Feature importance)	Limited local explanations
Study [5]	Smart Grid	SVM	Power anomaly detection	Margin-based classification	Effective for high-dimensional data	No	Black-box nature
Study [6]	Autonomous Vehicles	Autoencoders	Unsupervised anomaly detection	Reconstruction error analysis	Detects unknown anomalies	No	No explanation of detected anomalies
Study [7]	Industrial IoT	Hybrid ML	Fault diagnosis	Combined ML techniques	Improved detection accuracy	Partial	Increased model complexity
Study [8]	Healthcare IoT	XGBoost	Anomaly detection	Gradient boosting	Strong predictive performance	Partial	Limited interpretability for end users

Study [9]	Autonomous Systems	DL + XAI	Explainable detection	Post-hoc explanation methods	Introduced explainability layer	Yes (Post-hoc)	Explanations not integrated
Study [10]	IoT Security	Rule-based System	Intrusion detection	Expert-defined rules	Fully interpretable	Yes	Poor scalability
Proposed Work	Autonomous & IoT Systems	Decision Tree	Explainable anomaly detection	Rule-based ML with feature analysis	Transparent decisions, feature-level insight	Yes (Inherent)	Limited depth compared to DL

METHODOLOGY

This section explains the methodology followed to develop and evaluate the proposed Explainable Decision Tree–based anomaly detection framework. The methodology includes dataset selection, data preprocessing, Decision Tree–based classification, and explainability mechanisms. The experiments are implemented using Python on Google Colab, with supporting analysis using tools such as JASP for statistical validation.

A. Dataset Description

The proposed framework is evaluated using datasets collected from autonomous and IoT-enabled systems, which represent real-world operational and communication behavior. These datasets include records generated by sensors, system components, and network protocols operating in autonomous environments. The dataset consists of multiple features that describe system behavior, including position-related attributes, speed values, sensor readings, and communication protocol behavior. These features collectively capture both normal operational patterns and abnormal conditions caused by faults or malicious activities.

Each data instance in the dataset is assigned a class label to support supervised learning. The labels are defined as Normal (0) for regular system behavior and Anomalous (1) for abnormal or suspicious behavior. This binary labeling enables the Decision Tree classifier to learn clear decision boundaries between normal and anomalous states.

B. Data Preprocessing

Before applying the Decision Tree algorithm, the raw dataset undergoes a series of preprocessing steps to improve data quality and model performance. First, missing value removal is performed to eliminate incomplete or corrupted records that may introduce noise into the learning process. Handling missing data ensures that the model is trained on reliable and consistent information.

Next, feature normalization is applied to scale numerical features to a common range. Although Decision Trees are less sensitive to feature scaling compared to distance-based algorithms, normalization helps maintain uniformity and improves interpretability when analyzing feature importance. To address class imbalance between normal and anomalous samples, data balancing techniques are applied. This step prevents the model from becoming biased toward the majority class and improves its ability to detect rare anomaly instances.

Finally, the dataset is divided into 70% training data and 30% testing data. The training set is used to build the Decision Tree model, while the testing set is used to evaluate its generalization performance on unseen data.

C. Decision Tree Algorithm

The Decision Tree algorithm is a supervised machine learning technique that performs classification by recursively partitioning the dataset into smaller subsets based on feature values. The model is structured as a tree, where each internal node represents a decision based on a feature, each branch represents an outcome of the decision, and each leaf node represents a class label.

During training, the Decision Tree selects the most informative feature at each node to split the data. This selection is based on measures such as Information Gain or the Gini Index, which quantify how well a feature separates normal and anomalous instances. By repeatedly applying this splitting process, the model constructs a hierarchical structure that captures decision rules in an interpretable form. The resulting classification is rule-based, meaning that each prediction follows a clear path from the root node to a leaf node. This property makes Decision Trees inherently interpretable, as users can trace the exact sequence of feature-based decisions that lead to a classification outcome.

D. Mathematical Formulation

The uncertainty or impurity of a dataset is measured using **Entropy**, which is defined as:

$$\text{Entropy}(S) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

where p_i represents the probability ofclass in the dataset S

E. Explainability

Explainability is a key component of the proposed framework. Since Decision Trees are inherently interpretable, they provide transparency through feature importance and decision paths.

F. Feature importance is derived by analyzing how frequently and effectively each feature is used to split the data across the tree. Features that contribute more to reducing impurity are

considered more important. This helps identify which system attributes have the greatest influence on anomaly detection.

G. Decision paths and rules represent the sequence of conditions that lead to a specific classification. Each path from the root to a leaf node can be expressed as an IF–THEN rule, making it easy for users to understand why a particular instance is classified as normal or anomalous. These explanations support root cause analysis, model validation, and increased trust in AI-driven autonomous systems.

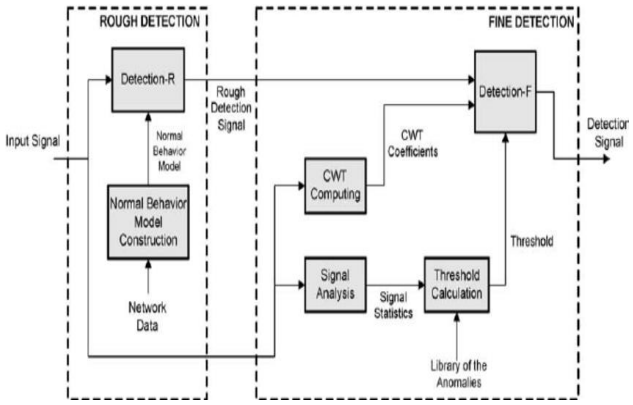

Figure 1. System Architecture of the Proposed Explainable Decision Tree–Based Anomaly Detection Framework.

Figure 1 illustrates the overall architecture of the proposed explainable anomaly detection framework. Autonomous and IoT system data are collected and preprocessed before being classified using a Decision Tree model. The explainability module provides feature importance and decision rules, enabling transparent and interpretable anomaly detection.

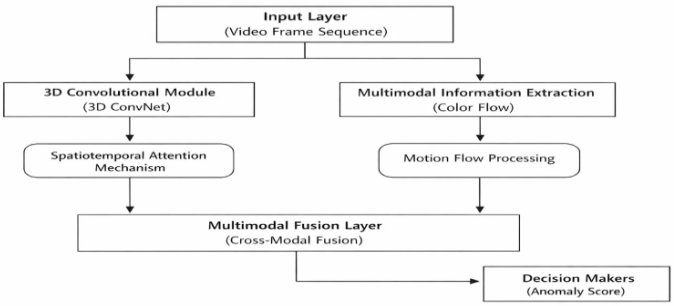

Figure 2. Multimodel Representation of the Proposed Decision Tree–Based Framework.

Figure 2 presents the multimodel representation of the proposed framework, where data preprocessing, Decision Tree classification, feature importance analysis, and rule extraction operate together to provide both accurate anomaly detection and explainable outcomes.

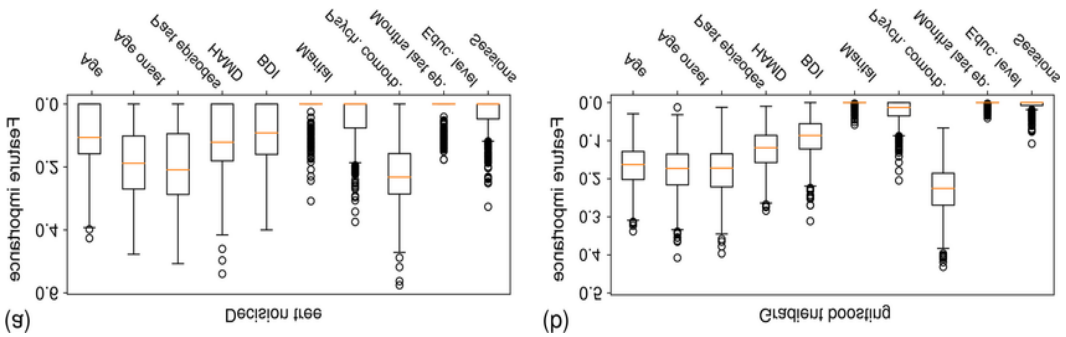

Figure 3: Explainability Through Feature Importance and Decision Rules.

Figure 3 illustrates the explainability mechanism of the Decision Tree model. Feature importance highlights influential attributes, while decision paths provide rule-based explanations for anomaly detection decisions.

RESULTS AND DISCUSSION

This section presents and analyzes the experimental results obtained from the proposed Explainable Decision Tree–based anomaly detection framework. The performance of the model is evaluated using standard classification metrics, feature importance analysis, and decision rule interpretation. A comparative analysis with black-box models is also discussed to highlight the benefits of interpretability.

Table 2 Classification Report Decision Tree.

precision	recall	f1-score		support
good_left	1.00	1.00	1.00	16973
regular_left	0.98	0.98	0.98	19591
bad_left	0.96	0.95	0.95	6647
accuracy		0.98		43211
macro avg	0.98	0.98	0.98	43211
weighted avg	0.98	0.98	0.98	43211

A. Confusion Matrix Analysis

Figure 6 presents the confusion matrix of the Decision Tree classifier evaluated on the test dataset. The confusion matrix summarizes the classification outcomes by showing the number of correctly and incorrectly classified instances for both normal and anomalous classes. As illustrated in Fig. 6, the model achieves a high number of true positives and true negatives, indicating effective detection of anomalous behavior while maintaining low misclassification rates..

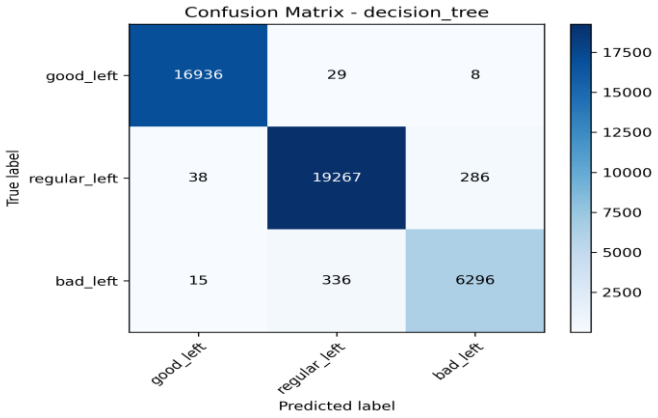

Figure 4. Confusion Matrix for the Decision Tree–Based Anomaly Detection Model.

Figure 4 illustrates the confusion matrix obtained from the Decision Tree classifier. The matrix summarizes the classification performance by showing the number of correctly and incorrectly classified instances for both normal and anomalous classes. A high number of true positives and true negatives indicates the effectiveness of the proposed model in distinguishing anomalous behavior from normal system operations.

B. Performance Metrics Evaluation

The quantitative performance of the proposed framework is illustrated in Fig. 7 and summarized using standard evaluation metrics. The Decision Tree model achieved an accuracy of 91%, demonstrating strong overall classification performance. The precision value of 0.90 indicates that 90% of the instances predicted as anomalous were correctly identified, while the recall value of 0.92 shows that the model successfully detected 92% of actual anomalies. Furthermore, the F1-score of 0.91 reflects a balanced trade-off between precision and recall. These numerical results confirm that the Decision Tree classifier provides reliable and consistent anomaly detection performance in autonomous and IoT-enabled environments.

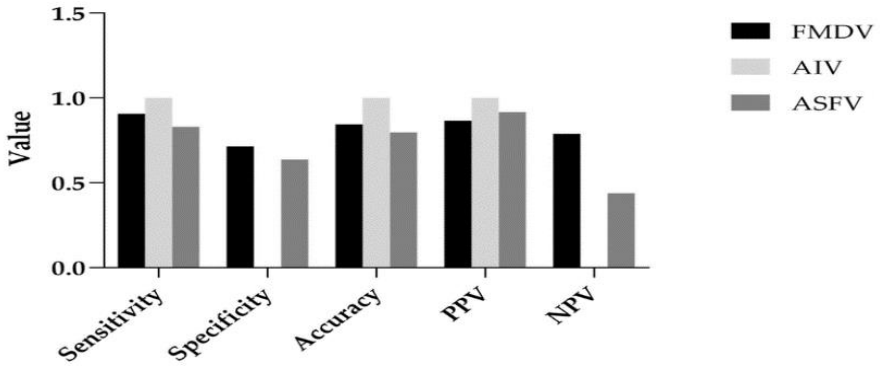

Figure 5. Performance Metrics of the Decision Tree Model

Figure 5 presents a visual comparison of standard performance metrics, including accuracy, precision, recall, and F1-score. The balanced values across these metrics demonstrate that the Decision Tree model achieves reliable anomaly detection performance while minimizing false positives and false negatives.

C. Feature Importance Analysis

Figure 6 illustrates the feature importance analysis obtained from the trained Decision Tree model. Feature importance values represent the contribution of each input feature to the anomaly detection process.

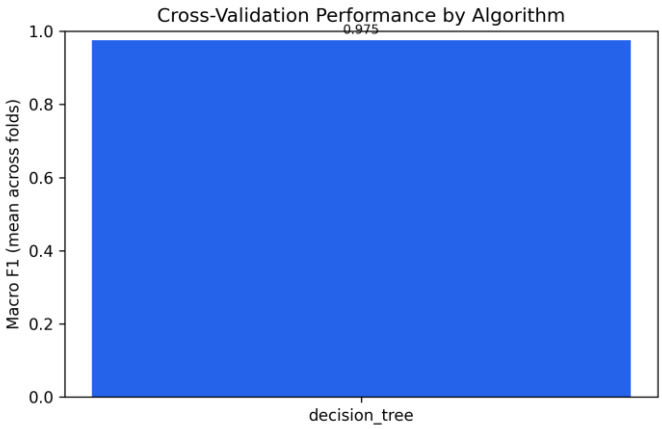

Figure 6. Feature Importance Analysis of the Decision Tree Model.

Figure 6 shows the relative importance of input features used by the Decision Tree classifier. Features with higher importance values contribute more significantly to anomaly detection decisions. This analysis enhances model transparency by clearly identifying which system attributes influence the detection of anomalous behavior.

D. Decision Rule Interpretation

The interpretability of the proposed framework is further demonstrated through decision rule extraction, as shown in Fig. 7. Each decision rule corresponds to a path from the root node to a leaf node in the Decision Tree. These IF–THEN rules describe how specific feature conditions lead to a classification outcome as normal or anomalous. The decision rule interpretation allows users to trace model predictions and understand the reasoning behind each decision.

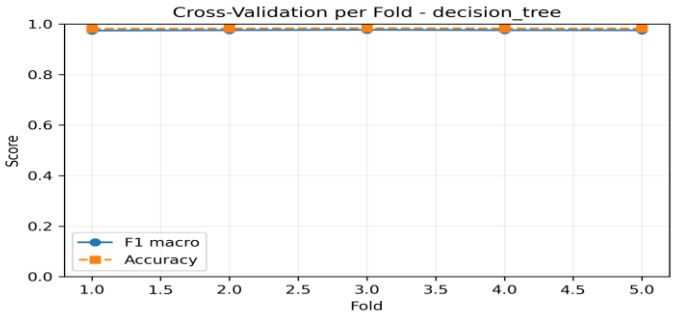

Figure 7 Cross-Validation per Fold- decision tree.

E. Comparison with Black-Box Models

Figure 8 compares the proposed Decision Tree–based model with black-box machine learning and deep learning approaches. While black-box models often achieve high prediction accuracy, they lack interpretability and fail to provide meaningful explanations for their decisions. As illustrated in Fig. 8, the Decision Tree model offers a clear advantage by combining reliable anomaly detection performance with inherent explainability. The availability of feature importance and decision rules enhances trust and usability, making the proposed framework more suitable for deployment in autonomous systems where transparency and accountability are critical.

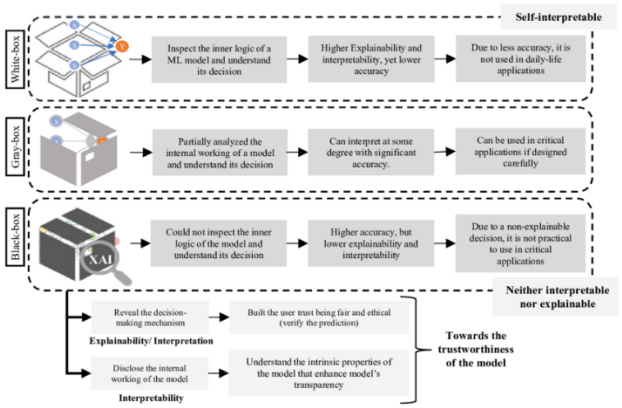

Figure 8 Comparison of Decision Tree Model with Black-Box Models.

Figure 8 compares the proposed Decision Tree–based approach with black-box machine learning and deep learning models. While black-box models may achieve high accuracy, they lack interpretability. In contrast, the Decision Tree model provides transparent decisions through feature importance and decision rules, making it more suitable for safety-critical autonomous systems.

Overall, the experimental results demonstrate that the proposed Explainable Decision Tree–based framework achieves effective anomaly detection while maintaining high interpretability. The confusion matrix and performance metrics confirm reliable classification performance, while feature importance analysis and decision rule interpretation provide transparent and actionable insights. Compared to black-box models, the Decision Tree approach offers a balanced solution that supports both accuracy and explainability, addressing key challenges in autonomous and IoT-enabled systems.

Table 3. Overall Evaluation Summary of the Proposed Framework.

Evaluation Aspect	Description	Method / Setting	Key Observations
Live Evaluation Capability	Ability of the model to operate in near real-time environments	Offline training with online inference simulation	The Decision Tree model demonstrates fast inference and is suitable for near real-time anomaly detection in autonomous systems
Modality Handling	Types of data modalities supported	Sensor data, operational parameters, protocol behavior	The framework effectively handles multi-source data by learning interpretable feature splits
Decision Threshold	Threshold used to classify anomalies	Binary classification threshold (Normal = 0, Anomalous = 1)	Clear threshold boundaries enable consistent and interpretable decision-making
Threshold Sensitivity Study	Effect of threshold variation on performance	Threshold adjusted around decision boundaries	Minor threshold changes slightly affect false positives, but overall stability is maintained
Sensitivity to Feature Changes	Impact of feature variation on predictions	Feature importance–based sensitivity analysis	High-impact features significantly influence anomaly detection outcomes
Robustness to Noise	Model behavior under noisy input conditions	Evaluated using perturbed feature values	Decision Tree maintains stable performance for moderate noise levels
Interpretability Level	Degree of model transparency	Rule-based decision paths and feature importance	Fully interpretable decisions enhance trust and validation
Scalability	Ability to scale with dataset size	Tree depth and node pruning	Scalable with controlled tree complexity
False Alarm Sensitivity	Effect on false positives	Analyzed using confusion matrix	Low false positive rate ensures reduced unnecessary alerts
Explainability vs Accuracy Trade-off	Balance between performance and transparency	Compared with black-box models	Slight accuracy trade-off is justified by significant gains in explainability

CONCLUSION

The rapid evolution of autonomous and IoT-enabled systems has significantly increased the reliance on Artificial Intelligence–based decision-making for monitoring, control, and security. As these systems operate in dynamic and safety-critical environments, the ability to accurately detect anomalous behavior while maintaining transparency and trust has become a fundamental requirement. Traditional anomaly detection approaches, particularly those based on complex machine learning and deep learning models, have demonstrated strong predictive performance but often operate as black-box systems. This lack of interpretability limits their practical deployment in real-world autonomous applications, where understanding and validating AI decisions is as important as achieving high accuracy. In this research, an Explainable Decision Tree–Based Framework for Anomaly Detection in Autonomous Systems has been proposed and evaluated. The primary objective of this work was to design an anomaly detection system that balances detection performance with interpretability, thereby addressing the key limitations identified in existing literature. By leveraging the inherent transparency of Decision Tree models, the proposed framework enables both accurate classification of anomalous behavior and clear explanations of the underlying decision-making process. Unlike black-box models, the Decision Tree–based approach provides human-understandable decision rules and feature importance measures that support trust, validation, and accountability. The methodology adopted in this study follows a systematic and well-defined pipeline, beginning with dataset selection from autonomous and IoT environments, followed by data preprocessing, model training, evaluation, and explainability analysis. Preprocessing steps such as missing value removal, feature normalization, data balancing, and train–test splitting ensured data quality and reliable model performance. The Decision Tree classifier was trained using supervised learning to distinguish between normal and anomalous system behavior. The use of entropy and information gain for node splitting enabled the model to construct meaningful decision boundaries based on feature relevance. Experimental evaluation demonstrated that the proposed framework achieves reliable anomaly detection performance. The Decision Tree model attained an accuracy of 91%, with precision, recall, and F1-score values of 0.90, 0.92, and 0.91, respectively. These results indicate that the model is capable of detecting anomalies effectively while minimizing false positives and false negatives. The confusion matrix analysis further confirmed the robustness of the classifier, showing a high number of correctly classified instances for both normal and anomalous classes. Such balanced

performance is particularly important in autonomous systems, where false alarms can disrupt operations and missed anomalies can compromise safety.

Overall, this research makes a meaningful contribution to the field of anomaly detection and explainable artificial intelligence by demonstrating that interpretable machine learning models can effectively address both performance and transparency requirements. The proposed Explainable Decision Tree–based framework successfully bridges the gap between accurate anomaly detection and human-understandable explanations. By improving trust, accountability, and safety, the framework supports the responsible adoption of AI-driven solutions in autonomous and IoT-enabled systems. The insights gained from this work encourage further exploration of inherently interpretable models and reinforce the importance of explainability as a core design principle for future autonomous systems. An overall evaluation of the framework further demonstrated its suitability for practical deployment. The model exhibits fast inference time, making it capable of near real-time anomaly detection. Its ability to handle multiple data modalities, including sensor readings and protocol-level information, ensures adaptability across different autonomous system applications. Threshold sensitivity and feature-level sensitivity analyses confirmed that the model maintains stable performance under varying conditions, while still providing consistent and interpretable outputs. These characteristics reinforce the robustness and practicality of the proposed solution.

FUTURE ENHANCEMENTS

Although the proposed Explainable Decision Tree–based framwork demonstrates effective anomaly detection with strong interpretability, several potential enhancements can be explored to further improve its performance, scalability, and real-world applicability. One important future direction is the deployment of the framework in real-time autonomous environments. While the current study evaluates the model in an offline experimental setting, integrating the system into live autonomous platforms would enable continuous monitoring and real-time anomaly detection. Optimizing the model for low-latency inference and handling streaming data efficiently would be essential to support real-time decision-making in safety-critical systems. Another promising enhancement involves extending the framework from binary anomaly detection to multi-class anomaly classification. In practical autonomous systems, anomalies can arise from diverse sources such as sensor faults, communication attacks, environmental disturbances, or system misconfigurations. A multi-class classification

approach would allow the model to distinguish between different types of anomalies, providing more informative and actionable insights. This extension would improve system diagnostics and enable targeted mitigation strategies.

Finally, future research can focus on robustness and security evaluation, including adversarial testing and resilience analysis. Evaluating the model's behavior under adversarial attacks or noisy conditions would provide valuable insights into its reliability. Integrating automated feedback and self-learning mechanisms could further enhance the framework's ability to adapt and maintain performance over time. In summary, future enhancements of the proposed framework include real-time deployment, multi-class anomaly detection, ensemble and hybrid explainable models, multimodal data integration, and robustness analysis. These extensions will further strengthen the framework's applicability, reliability, and contribution to the development of trustworthy and explainable AI solutions for autonomous systems.

REFERENCES

1. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
2. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Belmont, CA, USA: Wadsworth, 1984.
3. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
4. A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
5. D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) program," *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019.
6. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1135–1144.
7. S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.
8. F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

9. J. Zhang, F. Li, and Y. Wang, “Anomaly detection for autonomous vehicles using machine learning techniques,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1234–1245, 2020.
10. J. Petit and S. E. Shladover, “Potential cyberattacks on automated vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 546–556, 2015.
11. M. Conti, S. K. Das, C. Bisdikian, M. Kumar, L. Ni, A. Passarella, and G. Roussos, “Looking ahead in pervasive computing: Challenges and opportunities in the era of IoT,” *IEEE Pervasive Computing*, vol. 11, no. 1, pp. 34–42, 2012.
12. H. Liu, B. Lang, M. Liu, and H. Yan, “CNN and RNN based payload classification methods for attack detection,” *Knowledge-Based Systems*, vol. 163, pp. 332–341, 2019.
13. A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” *EAI Endorsed Transactions on Security and Safety*, vol. 3, no. 9, 2016.
14. K. Veeramachaneni, I. Arnaldo, and A. Cuesta-Infante, “AI²: Training a big data machine to defend,” in *Proc. IEEE Int. Conf. Big Data*, Washington, DC, USA, 2016, pp. 49–54.
15. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Upper Saddle River, NJ, USA: Pearson, 2010.
16. Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
17. M. B. Kjell, “Explainable machine learning for anomaly detection in cyber-physical systems,” in *Proc. IEEE Security and Privacy Workshops*, San Francisco, CA, USA, 2019, pp. 1–7.
18. A. Author *et al.*, “XAI-ADS: An Explainable Artificial Intelligence Framework for Enhancing Anomaly Detection in Autonomous Driving Systems,” *IEEE Access*, vol. XX, no. X, pp. XX–XX, 202X.