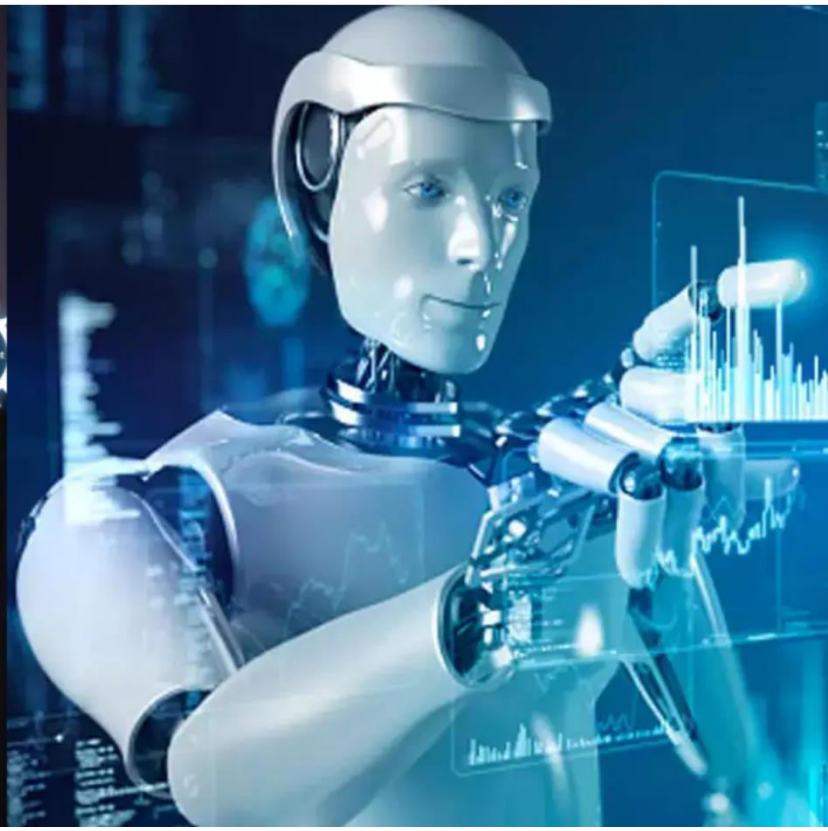
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 6, June 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

Net-Trace, Explainable Ensemble Machine Learning Framework for Network Intrusion Detection

Shreyash Bisne[1], Rehan Shaikh[2], Yaasin Shaikh[3], Ameya Kamble[4], Prof. Shruti Hatwar[5]

U.G. Student, Department of Computer Science, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune (MH), India[1-4]

Professor, Department of Computer Science, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune (MH), India[5]

ABSTRACT: This paper presents the implementation and performance evaluation of Net-Trace, an explainable ensemble machine learning-based Network Intrusion Detection System (NIDS) for real-time cyberattack detection. The proposed framework addresses the limitations of traditional intrusion detection systems by combining Random Forest, Logistic Regression, and a Multi-Class Classifier through an ensemble voting mechanism to improve detection accuracy, robustness, and computational efficiency. The system is trained and evaluated using the NSL-KDD dataset, with preprocessing techniques including data cleaning, feature encoding, normalization, and feature selection to enhance model performance. To improve transparency and trust, SHAP (SHapley Additive exPlanations) is integrated to provide feature-level explanations for each prediction, enabling better understanding of the model's decisions. A web-based dashboard with a MongoDB backend supports real-time monitoring, attack visualization, and alert management. Experimental results demonstrate that the proposed ensemble framework achieves high classification accuracy, strong precision and recall, and low inference latency, making it suitable for deployment in modern cybersecurity environments. The integration of ensemble learning and Explainable AI provides an efficient, scalable, and interpretable solution for network intrusion detection.

KEYWORDS: Network Intrusion Detection System (NIDS), Ensemble Machine Learning, Random Forest, Logistic Regression, Multi-Class Classification, Explainable AI (XAI), SHAP, NSL-KDD, Cybersecurity.

I. INTRODUCTION

The rapid growth of digital technologies and the increasing dependence on interconnected networks have significantly expanded the attack surface for cyber threats. As organizations continue to rely on digital communication and cloud-based services, protecting network infrastructure has become a critical aspect of modern cybersecurity. Network Intrusion Detection Systems (NIDS) play an essential role in monitoring network traffic and identifying malicious activities in real time. However, the emergence of sophisticated attacks, including zero-day exploits, distributed denial-of-service (DDoS) attacks, and advanced persistent threats (APTs), presents significant challenges for conventional intrusion detection techniques.

Traditional signature-based intrusion detection systems are highly effective in detecting previously known attacks but often fail to identify new and evolving threats. To overcome these limitations, machine learning (ML) techniques such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Logistic Regression have been widely adopted for anomaly detection. These models can learn complex patterns from network traffic and adapt to changing attack behaviors, offering improved detection accuracy compared to rule-based approaches. Nevertheless, relying on a single machine learning algorithm may lead to inconsistent performance due to variations in attack types, data imbalance, and feature diversity.

Recent research has shown that ensemble learning techniques can significantly improve intrusion detection by combining the strengths of multiple machine learning models. Ensemble methods reduce overfitting, improve generalization, and enhance classification performance by integrating predictions from different base learners. Models

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

such as Random Forest provide robust feature learning, Logistic Regression offers efficient probabilistic classification, and multiclass classifiers enable accurate identification of diverse attack categories. The combination of these models produces a more reliable and scalable intrusion detection framework suitable for real-time cybersecurity applications.
Another important advancement in modern cybersecurity is the adoption of Explainable Artificial Intelligence (XAI). Although machine learning models can achieve high detection accuracy, their decision-making process is often difficult to interpret. Explainability techniques such as SHAP (SHapley Additive exPlanations) provide feature-level insights into model predictions, allowing security analysts to understand why a particular network flow has been classified as malicious or benign. This transparency improves user trust, simplifies forensic investigations, and supports informed security decision-making.

Motivated by these developments, this paper presents the implementation of Net-Trace, an explainable ensemble machine learning-based Network Intrusion Detection System. The proposed framework combines Random Forest, Logistic Regression, and a Multi-Class Classification model using an ensemble voting mechanism to achieve accurate and efficient intrusion detection. The system is trained and evaluated using the NSL-KDD dataset and incorporates SHAP-based explainability for transparent prediction analysis. Furthermore, a real-time dashboard integrated with MongoDB enables continuous network monitoring, attack visualization, and alert management, making Net-Trace a practical and scalable solution for modern cybersecurity environments.

II. LITERATURE REVIEW

The increasing sophistication of cyberattacks has made Network Intrusion Detection Systems (NIDS) an essential component of modern cybersecurity. Traditional intrusion detection systems primarily relied on signature-based techniques, which identify attacks by matching network traffic against predefined attack signatures. While these methods are highly effective for detecting known threats, they often fail to recognize novel or evolving attacks. To overcome these limitations, machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Logistic Regression have been extensively applied to detect anomalies in network traffic [7], [9]. These techniques provide improved adaptability and learning capability compared to conventional rule-based systems, although the performance of individual models may vary depending on the characteristics of the dataset.

Recent research has increasingly focused on ensemble machine learning techniques for intrusion detection. Ensemble methods combine the predictions of multiple classifiers to improve overall detection accuracy, reduce model bias, and enhance robustness against diverse attack patterns. Random Forest has demonstrated excellent performance in handling high-dimensional network traffic data, while Logistic Regression offers efficient probabilistic classification with low computational complexity. Combining multiple classifiers through ensemble voting enables better generalization and more reliable detection than relying on a single learning algorithm.

Another significant advancement in intrusion detection research is the integration of Explainable Artificial Intelligence (XAI). Although machine learning models provide high predictive performance, understanding the reasoning behind their decisions remains a challenge. Explainability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) identify the contribution of individual network features to model predictions, thereby improving transparency, user trust, and forensic analysis capabilities [6], [15].

Motivated by these developments, this work presents Net-Trace, an explainable ensemble machine learning-based Network Intrusion Detection System. The proposed framework integrates Random Forest, Logistic Regression, and a Multi-Class Classification model using an ensemble voting mechanism to achieve accurate and efficient intrusion detection. Furthermore, SHAP-based explainability enhances model interpretability, making the system more reliable and suitable for real-time cybersecurity applications.

III. PROPOSED METHODOLOGY

The proposed Net-Trace Network Intrusion Detection System (NIDS) is designed to provide accurate, efficient, and interpretable detection of network intrusions using an ensemble machine learning framework. Unlike traditional intrusion detection systems that rely on a single classifier, the proposed approach combines multiple machine learning models to improve detection accuracy, reduce false alarms, and enhance generalization across different attack categories. The system follows a structured workflow consisting of data preprocessing, feature engineering, model

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

training, ensemble prediction, and explainable AI-based interpretation. This integrated methodology enables the framework to efficiently analyze large volumes of network traffic while maintaining low computational overhead and high detection performance.

The proposed system is implemented using the NSL-KDD dataset, where raw network traffic is transformed into structured feature vectors before being supplied to the learning models. During preprocessing, missing values are handled, categorical attributes are encoded, numerical features are normalized, and irrelevant features are removed to improve model efficiency. These preprocessing steps ensure that the generated feature vectors are suitable for effective training and accurate intrusion classification.

3.1 Data Preprocessing and Feature Engineering

The first stage of the proposed methodology focuses on preparing the network traffic data for machine learning. The NSL-KDD dataset contains both numerical and categorical network features representing different communication sessions. To improve data quality and model performance, several preprocessing operations are performed.

Initially, duplicate and inconsistent records are removed from the dataset. Categorical attributes such as Protocol Type, Service, and Flag are transformed into numerical representations using encoding techniques. Numerical features are then normalized using Standard Scaling, ensuring that all attributes contribute equally during model training. Finally, feature selection techniques are applied to eliminate redundant attributes while preserving the most informative network features.

The processed dataset is divided into training and testing subsets, enabling unbiased evaluation of the proposed intrusion detection framework.

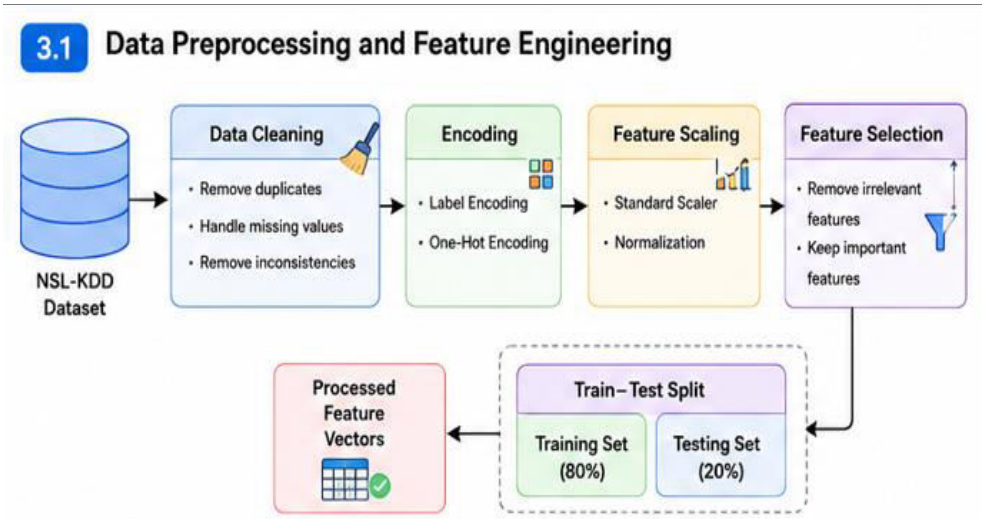


Figure 1. Data Preprocessing and Feature Engineering Pipeline

3.2 Ensemble Learning Framework

The core component of the proposed Net-Trace system is an ensemble machine learning framework that combines the predictions of multiple base classifiers. Three complementary machine learning models are independently trained using the processed dataset:

Random Forest (RF): Learns complex nonlinear relationships through an ensemble of decision trees and provides robust feature learning with high classification accuracy.

Logistic Regression (LR): Performs probabilistic classification with low computational complexity, making it suitable for efficient detection of normal and malicious traffic.

Multi-Class Classifier (MCC): Identifies different categories of network attacks, including Normal, DoS, Probe, R2L, U2R, and other supported attack classes.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

Each classifier independently analyzes the incoming feature vectors and generates a prediction. Since each model captures different characteristics of network traffic, combining their outputs improves the overall robustness and reliability of the intrusion detection system.

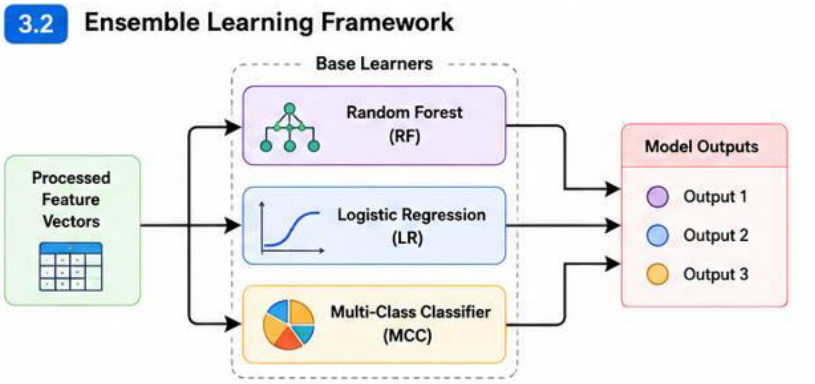


Figure 2. Ensemble Learning Framework with Base Models

3.3 Ensemble Voting Mechanism

Instead of relying on a single prediction, the proposed framework aggregates the outputs of all base learners using an ensemble voting mechanism. During prediction, the outputs generated by the Random Forest, Logistic Regression, and Multi-Class Classifier are combined to determine the final attack label.

For hard voting, the final prediction corresponds to the class receiving the majority of votes from the individual classifiers. Alternatively, probability-based soft voting can be employed by averaging the predicted class probabilities and selecting the class with the highest confidence score. This aggregation strategy minimizes the weaknesses of individual models while improving classification stability, reducing false positives, and increasing detection accuracy across multiple attack categories.

The ensemble framework therefore provides a more reliable intrusion detection mechanism than any individual classifier operating independently.

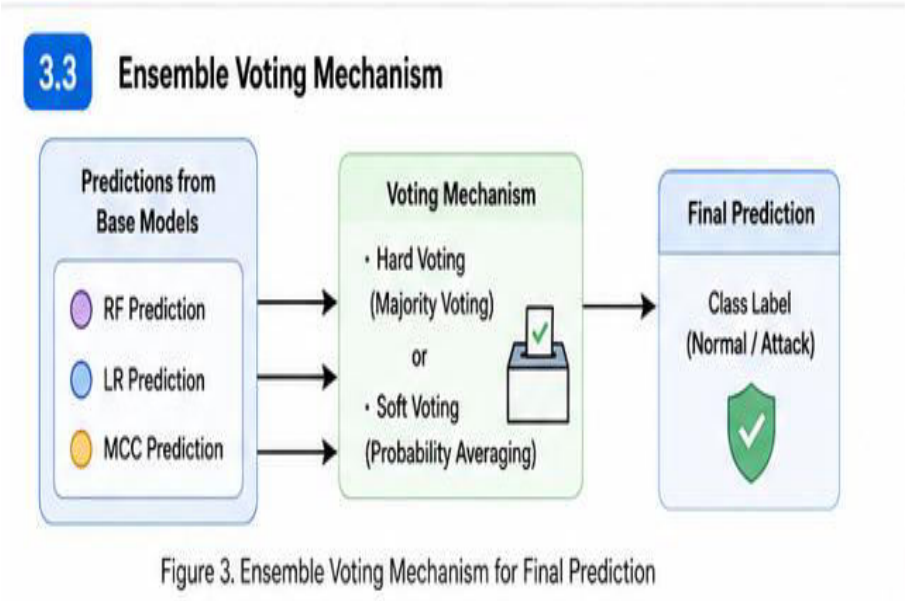


Figure 3. Ensemble Voting Mechanism for Final Prediction

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

3.4 Explainable Artificial Intelligence (XAI)
To improve transparency and trust in automated intrusion detection, the proposed framework incorporates Explainable Artificial Intelligence (XAI) using SHAP (SHapley Additive exPlanations).

For every prediction generated by the ensemble model, SHAP computes the contribution of each network feature toward the final classification outcome. These feature importance values enable cybersecurity analysts to understand why a particular network flow has been classified as either normal or malicious.

The generated explanations are stored together with prediction results and visualized through the monitoring dashboard.

IV. SYSTEM ARCHITECHTURE

The proposed Net-Trace Network Intrusion Detection System (NIDS) is designed as a modular and scalable architecture that enables accurate, efficient, and interpretable intrusion detection using an ensemble machine learning framework. The architecture is divided into three major layers: Data Processing Layer, Ensemble Analysis Layer, and Actionable Intelligence Layer. Together, these components provide an end-to-end pipeline for network traffic analysis, attack detection, prediction explanation, and real-time monitoring.

4.1 Data Processing Layer
The Data Processing Layer is responsible for collecting, transforming, and preparing network traffic for machine learning analysis. The system utilizes the NSL-KDD dataset, where each network connection is represented by 41 network features describing various characteristics such as protocol type, service, flag, connection duration, source bytes, destination bytes, and traffic statistics.

Before model training and prediction, the collected data undergoes several preprocessing operations. Missing or inconsistent records are removed, categorical attributes are converted into numerical representations using encoding techniques, and numerical features are normalized using Standard Scaling. Feature selection is then applied to eliminate redundant attributes while preserving the most informative features. Finally, the processed dataset is divided into training and testing subsets to enable effective model learning and unbiased performance evaluation.

4.2 Ensemble Analysis Layer
The Ensemble Analysis Layer forms the core of the proposed intrusion detection framework. Instead of relying on a single classifier, the system employs three complementary machine learning models: Random Forest, Logistic Regression, and a Multi-Class Classifier.

Each model independently analyzes the processed network feature vectors and generates its own prediction. Random Forest effectively captures complex nonlinear relationships within network traffic, Logistic Regression provides efficient probabilistic classification, while the Multi-Class Classifier accurately distinguishes between multiple attack categories, including Normal, DoS, Probe, R2L, U2R, and other supported attack classes.

The predictions generated by these base learners are combined using an ensemble voting mechanism. Depending on the implementation, either hard voting (majority voting) or soft voting (probability averaging) is employed to determine the final attack label. This collaborative decision-making process improves detection accuracy, reduces false positives, and enhances the robustness of the overall intrusion detection system.

5.3 Actionable Intelligence Layer
The Actionable Intelligence Layer converts prediction results into meaningful insights for cybersecurity analysts. After the ensemble model produces the final classification, the system applies SHAP (SHapley Additive exPlanations) to compute the contribution of individual network features toward the prediction.

The generated explanations identify the most influential features responsible for classifying each network flow as normal or malicious, thereby improving model transparency and supporting forensic investigation. Detection results, SHAP explanations, timestamps, and attack information are stored in a MongoDB database for future analysis.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

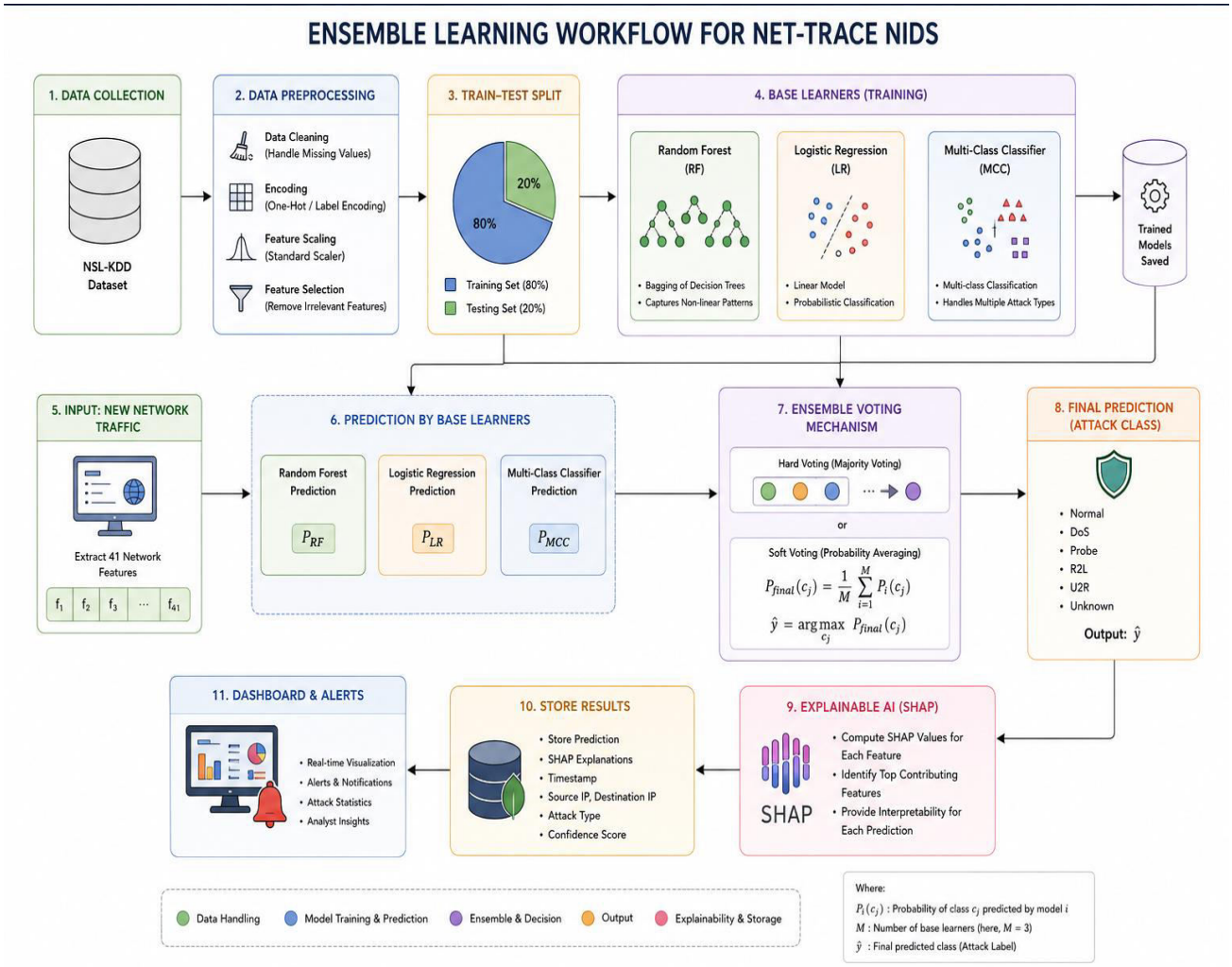
Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

A web-based monitoring dashboard retrieves this information from the database to provide real-time visualization, attack statistics, historical records, and automated alert generation. This layer enables security analysts to quickly interpret detection results, monitor network activity continuously, and respond effectively to potential cyber threats.



V. FUTURE WORK

1. Adaptive Ensemble Learning and Model Optimization

The current Net-Trace framework employs an ensemble of Random Forest, Logistic Regression, and a Multi-Class Classifier using a fixed voting strategy. Future work will focus on developing adaptive ensemble techniques that dynamically adjust the contribution of individual classifiers based on their real-time performance. Advanced hyperparameter optimization methods and automated model selection techniques can also be incorporated to further improve detection accuracy while reducing computational overhead.

2. Enhanced Real-Time Intrusion Detection and Zero-Day Attack Identification

Although the proposed system demonstrates strong performance on the NSL-KDD dataset, its ability to detect emerging and previously unseen cyber threats can be further enhanced. Future research will investigate the integration of online learning, anomaly detection algorithms, and continual learning techniques to improve zero-day attack detection. Additionally, deploying the framework with real-time packet capture tools such as Zeek or Suricata will enable continuous monitoring of live network traffic in enterprise environments.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

3. Explainability, Cloud Deployment, and Automated Response
Future enhancements will focus on extending the explainability capabilities of the framework by integrating additional Explainable Artificial Intelligence (XAI) techniques alongside SHAP to provide richer insights into prediction decisions. The system can also be deployed on cloud and edge computing platforms to improve scalability and support distributed intrusion detection. Furthermore, integrating automated incident response mechanisms and Security Information and Event Management (SIEM) platforms will enable faster threat mitigation and strengthen the practical applicability of the proposed Net-Trace framework in modern cybersecurity infrastructures.

VI. CONCLUSION

The proposed Net-Trace Network Intrusion Detection System (NIDS) presents an efficient and scalable solution for detecting network intrusions using an ensemble machine learning framework. By integrating Random Forest, Logistic Regression, and a Multi-Class Classifier through an ensemble voting mechanism, the system achieves high detection accuracy, improved robustness, and reduced false alarm rates while maintaining low computational overhead suitable for real-time cybersecurity applications. The implementation leverages comprehensive data preprocessing and feature engineering techniques to enhance model performance and ensure reliable classification of both normal and malicious network traffic.

VII. ACKNOWLEDGEMENT

The authors wish to express their sincere gratitude to our guides, Prof. S.S.Sapkal, Prof. Suvarna Bahir for providing invaluable guidance, continuous support, and constructive criticism throughout the conceptualization, design, and analysis phases of the Net Trace NIDS project.

Finally, the completion of this research would not have been possible without the collective effort and intellectual contributions from the wider research community, whose published work on tiered optimization and ensemble learning (cited herein) laid the critical theoretical foundation for this study.

REFERENCES

[1]. Chotima Thana-Aksaneekorn, Somkiat Kosolsombat, and Taweewat Luangwiriya, ``Machine Learning Classification for Intrusion Detection Systems Using the NSL-KDD Dataset," in Proc. IEEE International Conference on Cybernetics and Innovations (ICCI), 2024.
[2]. Martin Husák, Darshan Kumar Manoj, and Priyanka, ``Machine Learning in Intrusion Detection: An Operational Perspective," in Proc. 20th IEEE International Conference on Network and Service Management (CNSM), 2024.
[3]. Pritidhrita Paul Akash, F. M. Shafiullah Fahim, Naila Nahian, Muhammad Nazrul Islam, et al., ``Proactive Threat Detection in Network Security: An Ensemble Machine Learning Approach," in Proc. IEEE International Women in Engineering Conference on Electrical and Computer Engineering (WIECON-ECE), 2024.
[4]. Polasis Sudhakar, Durga Prasanna N, et al., ``Wrapper-Based Feature Selection for Enhanced Intrusion Detection Using Random Forest Classification," in Proc. International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), IEEE, 2024.
[5]. G. Pardeshi Nilesh and V. Patil Dipak, ``Binary and Multiclass Classification Intrusion Detection System Using Benchmark NSL-KDD and Machine Learning Models," in IEEE Conference Proceedings (ICDSNS), 2024.
[6]. Md. Shamimul Islam, Subrata Saha, and Md. Asraf Uddin Alam, ``Intrusion Detection System: An Optimization Based Deep Learning Approach Using NSL-KDD Dataset," in Proc. IEEE 2nd International Conference on Computing, Applications and Systems (COMPAS), 2025.
[7]. ``Machine Learning-Based Intrusion Detection for Network Security Management," in Proc. IEEE International Conference on Network and Service Management (CNSM), 2024.
[8]. ``Ensemble Machine Learning Techniques for Proactive Network Threat Detection," in Proc. IEEE WIE Conference on Electrical and Computer Engineering (WIECON-ECE), 2024.
[9]. ``Feature Selection Methods for Improving Network Intrusion Detection Using Random Forest," in Proc. IEEE ICICNIS, 2024.
[10]. ``Machine Learning-Based Binary and Multi-Class Intrusion Detection Using NSL-KDD Dataset," in IEEE Conference Proceedings, 2024.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 6, June 2026

|DOI: 10.15680/IJIRSET.2026.1506151|

[11]. Goyal.D, Singh.S. K., Kumar.S., Chilkoti.V., Chui.K. T., Hsu.C. H., and Gupta.B. B., ``An Explainable AI Framework for Intrusion Detection Using SHAP and LIME," in Proc. IEEE 14th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 2025.

[12]. Sani.R. I. and Bhakti.M. A. C., ``Intrusion Detection System Using Convolution Neural Network with Feature Selection for Multiclass Classification," in Proc. IEEE International Conference on Intelligent Computing (ICIC), 2024.

[13]. Neha.N. N., Muzahid.A. A. M., Shamsuzzoha.M., Jahid.I., Han.H., Zhang.Y., Hossain.M. M., and Sohel.F., ``Intelligent Intrusion Detection in IoT: Integrating Machine Learning and Feature Automation," in Proc. 17th IEEE International Conference on Computer and Automation Engineering (ICCAE), 2025.

[14]. Valasev.R. S., Priambodo.A. R., and Anggraini.R. N. E., ``Evaluating Contemporary Machine Learning and Deep Learning Strategies for Intrusion Detection," in Proc. IEEE International Conference on Control, Automation, Electronics, Robotics, Internet of Things, and Artificial Intelligence (CERIA), 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details