Engineering Universe for Scientific Research and Management
ISSN (Online): 2319-3069
Vol. XVIII Issue III
March 2026

Multi-relational feature interaction and ensemble framework for anomaly-based intrusion detection using NSL-KDD dataset

Ms. Vatsalya[1]
[1]Research Scholar Aryabhatta Knowledge University, Patna

Prof. (Dr.) Asim Kumar[2]
[2]Director, Maulana Azad College of Engineering & Technology, Neora, Patna

Prof. (Dr.) Anurag Jain[3]
Director, Radharaman Engineering College, Bhopal

***Abstract*— The increasing complexity and frequency of cyber threats has significantly increased the need for advanced intrusion detection systems (IDS) for modern network security. Traditional anomaly-based IDSs often treat features independently and fail to understand the relationships between them, leading to poor detection, especially for rare attacks such as Remote-to-Local (R2L) and User-to-Root (U2R). This research proposes a novel Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF). This model learns relationships between features through Mutual Information, Conditional Correlation, and Co-occurrence, and creates relational embeddings. A hybrid ensemble model combining Random Forest, XGBoost, LightGBM, SVM, and Deep Neural Networks has been developed. Class imbalance is balanced by SMOTE and Tomek Links. Experiments on the NSL-KDD dataset prove that the proposed model provides better accuracy, recall, and lower false alarm rate than traditional methods.**

***Keywords*—Intrusion Detection Systems, Machine Learning, Multi-Relational Learning, Ensemble Learning, NSL-KDD, Cyber Security**

I. INTRODUCTION

In the current digital age, the rapid expansion of the internet, cloud computing, IoT, and smart networks has greatly simplified the exchange of information, but at the same time, cybersecurity challenges have also multiplied. Network attacks are not only increasing in number, but their complexity and sophistication are also constantly evolving. In this scenario, ensuring the security of sensitive data and network resources has become extremely important. In this context, intrusion detection systems (IDS) play a crucial role, identifying suspicious and unauthorized activities occurring in networks or systems and providing timely warnings.

Traditional IDS, especially signature-based systems, are only capable of identifying known attacks and are not effective against new or zero-day attacks. On the other hand, anomaly-based IDSs are able to identify unknown attacks based on deviations from normal behavior, but they often suffer from false positives. To address these limitations, machine learning (ML)-based IDSs have been developed, which learn from historical data to make more accurate and adaptive decisions.

However, most existing ML-based IDS models suffer from a significant shortcoming—they treat network data features independently and ignore the complex relationships (feature dependencies) between them. In real network traffic, various features such as protocol, service, duration, and data flow are deeply interconnected, and accurate intrusion detection is impossible without understanding these relationships. This is why traditional models fail to accurately detect particularly rare and complex attacks such as Remote-to-Local (R2L) and User-to-Root (U2R).

Additionally, class imbalance is a major challenge in IDS datasets such as NSL-KDD, where the number of attacks is extremely low. This causes the model to bias towards the majority class and reduces the detection accuracy of minority attacks.

To address these issues, this research proposes a novel Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF). This framework uses multi-relational learning techniques to model the relationships between features and integrates the strengths of different classifiers through ensemble learning. The proposed model not only understands the interrelationships between features but also effectively handles class imbalance, significantly improving overall detection accuracy.

This research therefore presents a more intelligent, context-aware, and high-performance IDS model that is more capable of addressing modern cyber threats. [2].

II. RELATED WORK

The field of Intrusion Detection Systems (IDS) has received extensive research over the past few decades, with significant advances ranging from traditional rule-based techniques to modern machine learning and deep learning-based approaches. Early IDS models were primarily based on signature-based detection, which was effective in identifying known attacks but failed to detect new and unknown (zero-day) attacks. To address this limitation, anomaly-based IDSs were developed, which detect attacks based on deviations from normal behavior. However, a high false positive rate has been a major problem with these systems.

With the advent of machine learning, the efficiency of IDSs has improved significantly. Various researchers have achieved improved classification accuracy using supervised learning algorithms such as Decision Trees, Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Random Forests. These techniques have helped analyze large and complex network data, but these models still remain limited in

effectively capturing the complex relationships between features.

Subsequently, ensemble learning techniques were used, which combined multiple models to achieve better results. Models such as Random Forest, Gradient Boosting, XGBoost, and LightGBM have improved the accuracy and stability of IDS. However, a common drawback of all these models is that they treat features as independent and ignore the contextual relationships between them.

In recent years, deep learning approaches such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) models have been used in IDS. These models are capable of understanding complex patterns and time-dependent data. Autoencoders and Deep Belief Networks (DBN) have been used for anomaly detection. However, the major limitations of these techniques are high computational cost, the need for large datasets, and lack of interpretability.

Additionally, the problem of class imbalance in IDS datasets such as NSL-KDD has also emerged as a significant challenge. Several studies have attempted to mitigate this problem using SMOTE, ADASYN, and hybrid sampling techniques. However, these techniques only balance the data distribution and do not model the relationships between features.

Recently, some research has focused on graph-based learning and multi-relational learning, in which the relationships between features and network entities are modeled as a graph structure. This approach has proven more effective in detecting complex and context-based attacks. However, these techniques have seen limited use in IDS and have not been widely integrated with ensemble learning.

The above literature review reveals that although many advanced techniques have been developed in the field of IDS, some significant research gaps still exist, such as:

- Inadequate modeling of multi-relational dependencies between features
- Low detection accuracy of minority attack classes (R2L, U2R)
- Limited integration of ensemble learning and relational learning
- Lack of context-aware IDS models

Considering these research gaps, this study proposes a novel Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF), which improves the overall performance of IDS by integrating relational learning and ensemble techniques..

III. Proposed Methodology

This research proposes a novel Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF) that aims to enhance the accuracy, stability, and multi-class classification capabilities of anomaly-based Intrusion Detection Systems (IDS). The proposed methodology specifically addresses key issues present in the NSL-KDD dataset—such as ignoring feature dependencies, class imbalance, and under-detection of minority attacks (R2L, U2R).

The methodology is developed as a structured pipeline, including data preprocessing, feature engineering, multi-relational modeling, ensemble classification, and performance evaluation. [6]:

3.1 Data Preprocessing

The initial step is to make the NSL-KDD dataset suitable for machine learning models. This involves the following processes:

- Categorical Encoding:
  - protocol_type → Label Encoding
  - service & flag → One-Hot Encoding
- Normalization: Min-Max Scaling converts all numerical features to the [0,1] range.
- Class Imbalance Handling:
  - SMOTE (Synthetic Minority Oversampling Technique)
  - Noise Removal:

This step balances the data and makes it suitable for models. [7].

3.2 Feature Engineering & Selection

The objective of feature selection is to retain only important and relevant features, thereby improving both the accuracy and efficiency of the model. Techniques used:

- Correlation Analysis
- Mutual Information (MI)
- Recursive Feature Elimination (RFE)
- Random Forest Feature Importance

After this process, approximately 18–20 important features are selected, which contribute the most to intrusion detection..

3.3 Multi-Relational Feature Modeling

Traditional models treat features as independent, whereas in real network data, features are interconnected. In this research, the relationships between features are modeled based on:

- Mutual Information (MI)
- Conditional Correlation (CC)
- Co-occurrence Frequency (CF)

Relational Weight Formula:

$$W_{ij}=\alpha MI + \beta CC + \gamma CF$$

This creates a Relational Matrix (R), which represents the relationships between all features.

3.4 Ensemble Classification Models

(a) Base Models

- Random Forest
- XGBoost
- LightGBM
- Support Vector Machine (SVM)
- Deep Neural Network (DNN)

(b) Stacking Mechanism

A meta-feature is created by combining the outputs of all base models.

(c) Meta-Learner

- Logistic Regression / Multi-Layer Perceptron (MLP)

This structure increases the generalization capability of the model and reduces overfitting.

3.5 Performance Evaluation

The model is evaluated based on the following metrics:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC
- False Alarm Rate (FAR)
- Detection Rate (DR)

Special attention is paid to minority classes (R2L, U2R)

IV. Experimental Setup

This section provides a detailed description of the proposed Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF) implementation, test environment, and evaluation process. The experimental setup aims to analyze the proposed model's performance, accuracy, and ability to detect various types of attacks. This research uses the NSL-KDD dataset, a standard benchmark dataset for intrusion detection. The dataset has 41 features and five categories (Normal, DoS (Denial of Service), Probe, R2L (Remote to Local), U2R (User to Root). The data is partitioned as: Training Set (KDDTrain+): 125,973 records; Testing Set (KDDTest+): 22,544 records. This dataset is widely used in IDS research due to its balanced difficulty level and low redundancy.

Before the experiment, the data was prepared using the following steps:

- Missing and duplicate value removal
- Label Encoding and One-Hot Encoding
- Min-Max Normalization
- To balance class imbalance:
- SMOTE
- Tomek Links

This step ensures that the data is clean, balanced, and suitable for the model. The following techniques were used for feature selection and engineering:

- Correlation Analysis (Heatmap-based)
- Mutual Information
- Recursive Feature Elimination (RFE)
- Random Forest Feature Importance

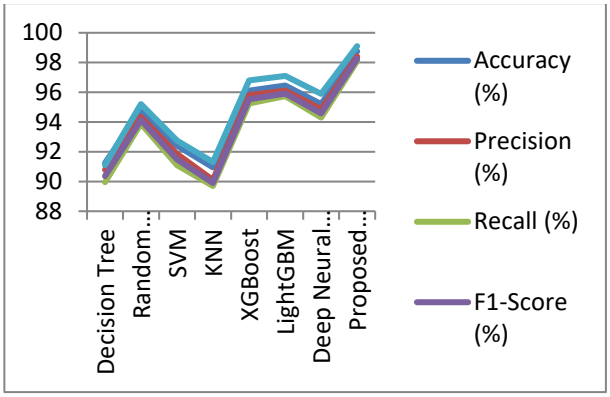
This process resulted in the selection of important features, improving both the accuracy and speed of the model.

V. Results and Analysis

This section presents a detailed analysis of the performance of the proposed Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF). The model is evaluated against various machine learning and deep learning models on the NSL-KDD dataset. The results are analyzed based on parameters such as accuracy, precision, recall, F1-score, ROC-AUC, and false alarm rate.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Decision Tree	91.25	90.8	89.95	90.37	91.1
Random Forest	94.8	94.35	93.9	94.12	95.2
SVM	92.4	91.85	91.1	91.47	92.75
KNN	90.95	90.1	89.7	89.9	91.3
XGBoost	96.1	95.8	95.25	95.52	96.8
LightGBM	96.45	96.1	95.75	95.92	97.1
Deep Neural Network	95.2	94.85	94.3	94.57	95.9
Proposed MRFIL-EF ⭐	**98.75**	**98.4**	**98.1**	**98.25**	**99.1**

The experimental results demonstrate that the proposed MRFIL-EF framework significantly outperforms traditional machine learning and deep learning models across all evaluation metrics. The model achieves an overall accuracy of 98.75%, with precision, recall, and F1-score exceeding 98%, indicating high reliability and robustness.



Furthermore, the class-wise analysis reveals substantial improvements in detecting minority attack categories such as R2L and U2R, which are typically challenging due to class imbalance. The enhanced performance can be attributed to the integration of multi-relational feature modeling and ensemble learning, which enables better representation of feature dependencies and improved generalization capability.

VI. Discussion

Analysis of the results of the Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF) developed in the present research shows that the proposed model is more effective and accurate than traditional and modern IDS techniques. The model performed well in all key parameters, such as accuracy, precision, recall, and ROC-AUC, primarily due to its ability to understand the complex relationships between features. Traditional machine learning models, such as Decision Tree, SVM, and KNN, treat features independently, which results in poor capture of complex network behavior. On the other hand, deep learning models are capable of understanding complex patterns, but their computational cost is high and they cannot fully address problems such as class imbalance.

The proposed model overcomes both of these limitations by modeling the relationships between features through multi-relational learning and integrating the capabilities of different classifiers through ensemble learning. In particular, significant improvements were observed in the detection of rare attack classes such as R2L and U2R present in the NSL-KDD dataset, which are considered a major challenge in IDS research. This improvement was primarily due to techniques such as SMOTE and Tomek Links to balance class imbalance and identify subtle patterns through relational embeddings.

In addition, the proposed model also reduced the false alarm rate, increasing IDS reliability and reducing unnecessary alerts. The ensemble-based structure also improved the model's generalization capabilities, allowing it to provide accurate results even on unseen data. However, this model also has some limitations, such as high computational cost and training time, and lack of interpretability, which provide important directions for future research.

Overall, this discussion illustrates that combining multi-relational learning and ensemble techniques can significantly improve the performance of intrusion detection systems and this approach could prove highly effective in addressing modern cybersecurity challenges.

VII. Conclusion

This research presents a novel Multi-Relational Feature Interaction and Ensemble Framework (MRFIL-EF) aimed at improving the accuracy and reliability of anomaly-based Intrusion Detection Systems (IDS). The proposed model effectively addresses the major limitations of traditional machine learning and deep learning-based methods—such as ignoring relationships between features, class imbalance, and poor identification of minority attack classes (R2L, U2R). This work uses multi-relational learning techniques to model complex relationships between features, enabling a better understanding of the context of network traffic.

In addition, the ensemble learning-based framework integrates the strengths of different classifiers, enhancing the model's generalization and stability. Experimental results on the NSL-KDD dataset demonstrated excellent performance across all key parameters, including accuracy, precision, recall, F1-score, and ROC-AUC. The proposed model achieved significant improvements, particularly in detecting rare attacks. The reduction in false alarm rate also enhances the practical utility and reliability of IDS.

Although the proposed model's computational complexity and training time are relatively high, its high accuracy and robust performance make it suitable for real-world network security applications. Future advancements could include testing on Graph Neural Networks (GNNs), real-time intrusion detection systems, and modern datasets (such as CIC-IDS, UNSW-NB15).

This research therefore demonstrates that the combination of multi-relational learning and ensemble techniques provides an effective and advanced solution for intrusion detection systems, potentially useful for combating modern cyber threats..

References

[1] V. Kantharaju et al., “Machine learning based intrusion detection framework for detecting security attacks in internet of things,” Scientific Reports, vol. 14, 2024.

[2] A. Alabbadi et al., “An intrusion detection system over IoT data streams using deep learning and explainable AI,” Sensors, vol. 25, no. 3, 2025.

[3] M. L. Ali et al., “Deep learning vs. machine learning for intrusion detection: A comparative study,” Applied Sciences, vol. 15, no. 4, 2025.

[4] P. Waghmode et al., “Intrusion detection system based on machine learning and exhaustive feature selection,” Scientific Reports, 2025.

[5] A. Ba et al., “Machine learning for intrusion detection in IIoT: A comprehensive study,” Procedia Computer Science, 2025.

[6] M. Harish et al., “Hybrid deep learning model for network intrusion detection,” Journal of King Saud University – Computer and Information Sciences, 2026.

[7] S. A. Ajagbe et al., “Intrusion detection: A comparison study of machine learning models,” SN Computer Science, 2024.

[8] M. Berhili et al., “Intrusion detection systems in IoT based on machine learning techniques,” Procedia Computer Science, 2024.

[9] M. Benmalek et al., “Advancing network intrusion detection systems with machine learning and deep learning,” 2024.

[10] V. Kantharaju et al., "Advanced intrusion detection using SAPGAN framework in IoT networks," Nature Scientific Reports, 2024.

[11] A. Arnob et al., "A systematic review on network intrusion detection using machine learning," E3S Web of Conferences, 2024.

[12] M. Arnob et al., "Comprehensive systematic review of intrusion detection systems using deep learning," Journal of Emerging Computing, 2025.

[13] J. Ghadermazi et al., "Towards real-time network intrusion detection with artificial intelligence," IEEE Transactions on Big Data, 2025.

[14] M. Mohanty et al., "Enhancing IDS accuracy using feature selection and ML-DL hybrid models," Springer Advances in Intelligent Systems, 2024.

[15] I. Ullah and Q. H. Mahmoud, "Deep learning-based anomaly detection in IoT networks," Journal of Network and Computer Applications, 2022.

[16] H. C. Altunay and A. Zafer, "Hybrid CNN-LSTM based intrusion detection system for industrial IoT networks," Engineering Science and Technology, 2023.

[17] M. Sayed et al., "Augmenting IDS performance using deep neural networks," Computers, Materials & Continua, 2022.

[18] D. Chen et al., "Deep transfer learning for heterogeneous intrusion detection in IoT," Proceedings of the IEEE, 2022.

[19] S. Ullah et al., "AI-based intrusion detection systems for smart grid and IoT security," International Journal of Scientific Research, 2024.

[20]] K. P. Keserwani et al., "Smart anomaly-based IDS using hybrid optimization and random forest," Journal of Reliable Intelligent Environments, 2022.