Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 10, pp. 3535–3544
DOI : 10.47974/JIOS-2481

An adaptive hybrid feature representation framework for modelling of complex network attack and classifications

Arikatla Annapurna *
M. Thillaikarasi †
Department of Computer Science & Engineering
Annamalai University
Annamalainagar 608002
Tamil Nadu
India

K. Subba Reddy §
Department of Computer Science and Engineering
Rajeev Gandhi Memorial College of Engineering and Technology
Nandyal 518501
Andhra Pradesh
India

Abstract

Network intrusion detection systems are playing a significant role in protecting communication infrastructures against sophisticated cyber-attacks. Existing machine learning classification methods are needed to achieve accurate classifications of network data. Network cyber-attacks are diverse and dynamic in nature. Thus, existing methods sometimes fail due to a lack of multi-view learning for cyber-attacks, which captures the complex, high-dimensional, and non-linear relationships inherent in network traffic data. These limitations lead to reduced detection accuracy and poor generalisation for diverse and evolving attack patterns. This paper proposes a Hybrid Feature Representation Framework (HFRF) that integrates multi-stage feature refinement, multi-view transformation, and adaptive feature selection for effective network data classification. The proposed framework initially performs variance- and mutual information-based filtering to retain relevant features, followed by dual transformations using Principal Component Analysis (PCA) and Non-Negative Matrix Factorization (NMF) to capture both global and localised data structures. These complementary representations are then fused and optimised using an ensemble-

* *E-mail:* arikatlaannapurna4@gmail.com (Corresponding Author)
† *E-mail:* selvathillai69@gmail.com
§ *E-mail:* mrsubbareddy@yahoo.com

© Taru Publications

based feature-importance mechanism to select the most discriminative features dynamically. Extensive experiments are conducted on benchmark datasets, including CICIDS2018 and UNSW-NB15, to demonstrate the efficacy of the proposed method compared to other methods.

Subject Classification: *68T05, 68T09, 68T10.*

Keywords: *Feature representation, Adaptive feature selection, Machine learning, Feature fusion, Classification.*

1. Introduction

Rapid developments in communication technology and the integration of networks have increased the risk of cyber-attacks on network infrastructure. With the ever-growing sophistication of modern networks and the complexity of their design, ensuring secure data transmission remains a serious challenge. Network Intrusion Detection Systems (NIDS), which are used to detect potential attacks and prevent breaches of data integrity and confidentiality, provide important solutions in this context. While traditional signature-based detection strategies are effective at identifying well-known attacks, their effectiveness is limited when applied to new and evolving threats. For that reason, machine-learning methods, which can help identify attack patterns by analyzing datasets, have drawn growing attention. Although offering unique capabilities for intrusion detection, current machine learning approaches in Sambangi and Gondi [18] have limitations that restrict their applicability in certain situations. In particular, Hossain and Islams [17] proposed network security problems are associated with the specific nature of network traffic data, which is often high-dimensional, heterogeneous, and dynamic. That means it includes many redundant, irrelevant, and noisy features, which can decrease classification accuracy and negatively affect overall algorithm performance. Most of the currently developed classification algorithms, such as Naive Bayes, Random Forests, Logistic Regression, and Decision Trees, require careful selection and manipulation of features to improve the quality of the output results. Furthermore, current models cannot leverage multi-view features for detecting cyber-attacks, resulting in low detection efficiency.

Traditional IDS approaches struggle to handle the complexity of modern network data, underscoring the importance of feature engineering and representation learning. Techniques such as variance thresholding, PCA, and NMF help reduce dimensionality and improve data representation. Variance thresholding and PCA capture global data patterns, while NMF focuses on localized feature representations. However, these methods do not fully utilize the complementary information present in network traffic. Feature selection is equally important, with relevance and redundancy commonly used as criteria, though neither is sufficient alone. To address these limitations, this paper

proposes a Hybrid Feature Representation Framework (HFRF) that enhances the discriminative capability of traffic features through multi-perspective data analysis.

This paper's main contributions can be summed up as follows: (i) the development of a new hybrid feature representation model incorporating multi-stage feature enhancement and multi-view transformation, (ii) an efficient data fusion technique using both PCA and NMF, (iii) adaptive feature selection to improve classification results, and (iv) rigorous experiments showing better results compared to traditional approaches. This approach can prove useful for building scalable, adaptive network IDS and can be further extended to other cybersecurity-related problems involving high-dimensional, complex datasets.

2. Background Work

Recent research on improving network intrusion detection systems has emphasised increasing the efficiency of classification models through feature engineering, ensemble learning, and deep learning. The work on developing a hybrid feature selection and stacking ensemble for intrusion detection shows that a mutual information-based wrapper method is effective for detecting attacks in networks. These works show that redundant and irrelevant features degrade classification performance, and hence Alsaffar, Nouri-Baygi, and Zolbanin [1] demonstrated that hybrid feature selection plays an important role in improving classification efficiency. In these works, ensemble learning is mostly used, but not multi-view feature representation modelling. Another study on feature selection in DDoS detection used statistical methods to reduce feature dimensionality. These techniques help reduce training complexity but fail to capture non-linear relationships, Araujo et al. [2] noted that this may limit their effectiveness against complex attacks. Some ensemble learning techniques, including correlation analysis, mutual information, and dimensionality reduction, are found to be more efficient at detecting attacks and reducing false positives.

As mentioned by Hossain [3] most of the transformation techniques studied consider the network traffic from one point of view which is not sufficient to analyse the complexity of network traffic. Ahmed et al. [4], Cheng et al. [5] and Basha et al. [24] demonstrated deep learning-based IDSs' ability to recognize complex traffic patterns and enhance detection by hierarchical feature extraction. Hybrid approaches have been shown to be effective in detecting complex attacks by lopes et al. [6], but computational complexity and interpretability issues have been raised by Jadhav and Pellakuri [7] and Samaan and Jeiad [8] respectively. While Bakar et al. [9] introduced adaptive frameworks to enhance the scalability and efficiency in the varying environments, the approaches do not use multi-view learning and are still dependent on the predetermined techniques.

Akter et al. [10] highlighted that adversarial attacks make robust feature representations essential for IDSs, since detection performance depends heavily on feature quality. Sarker [11], Dasari and Devarakonda [12], and Rajesh et al.

[13] noted that no single classifier can detect every attack type, making adaptive hybrid IDSs important for modern cloud and IoT environments, as emphasized by Pasha et al. [14] and Aljuhani [15]. Building on these findings and the real-time detection work of Musumeci et al. [16], the proposed multi-view hybrid framework combines feature selection, PCA, and NMF to capture global and local data characteristics.

3. Proposed Hybrid Feature Representation Framework (HFRF) Method

The proposed Hybrid Feature Representation Framework (HFRF) is designed to enhance network intrusion detection by transforming raw high-dimensional network traffic data into an optimized and discriminative feature space through a sequence of mathematically defined steps. For the data, $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^d$represents the input feature vector and $y_i$denotes the corresponding class label. Initially, the data is preprocessed by removing duplicates, handling missing and infinite values, encoding categorical labels, and normalizing features using standard scaling such that $X' = \frac{X-\mu}{\sigma}$, ensuring uniform feature distribution. To reduce noise and redundancy, a variance threshold is applied to eliminate features with low variability, producing a reduced feature set X_{var}, followed by mutual information-based selection where the dependency between each feature and the target variable is quantified using $MI(x_j, y) = \sum p(x_j, y) log \frac{p(x_j,y)}{p(x_j)p(y)}$, and the top informative features are retained to form X_{sel}.

Algorithm: Hybrid Feature Representation Framework (HFRF)

Input: Dataset D with features and labels

Output: Optimized feature set and predicted class labels

1. Load the dataset and extract features and labels.
2. Preprocess the data by removing duplicates, handling missing and infinite values, encoding labels, selecting numerical features, and normalizing the data.
3. Apply variance-based filtering to remove low-variance features and reduce noise.
4. Perform mutual information-based feature selection to retain the most relevant features.
5. Transform the selected features using PCA to capture global structure.
6. Apply NMF to extract localized feature representations.
7. Fuse the selected features with PCA and NMF features to form a hybrid feature set.
8. Apply adaptive feature selection using a tree-based model to select the most important features.
9. Split the dataset into training and testing sets.
10. Train classification models using the optimized feature set.

To capture complex relationships in the data, a multi-view transformation strategy is employed. Principal Component Analysis (PCA) projects the selected features onto a lower-dimensional orthogonal subspace by maximizing variance, computed from the covariance matrix $\Sigma = \frac{1}{N} X_{sel}^T X_{sel}$, resulting in transformed features X_{pca} that represent global data structure. In parallel, the feature matrix is divided into two non-negative matrices using Non-Negative Matrix Factorization (NMF). $X_{sel} \approx WH$, where W captures localized, parts-based representations, producing X_{nmf}. These complementary representations are then fused by concatenating the original selected features with the transformed features as $X_{fused} = [X_{sel} \;||\; X_{pca} \;||\; X_{nmf}]$, thereby integrating global and local information into a unified feature space.

To further refine this representation, an adaptive feature selection mechanism is applied using a tree-based ensemble model, where feature importance scores I_j are computed and the most significant features are selected to form the optimized feature subset X_{opt}. This step ensures that only the most discriminative and relevant features contribute to the classification process, reducing dimensionality while preserving predictive power. The optimized feature set is then used to train classification models $\hat{y} = f(X_{opt})$, where different classifiers such as Random Forest, Logistic Regression, Decision Tree, and Naive Bayes can be employed. The learning objective minimizes the empirical loss $L = \frac{1}{N} \sum l(y_i, \hat{y}_i)$, For classification problems, cross-entropy is usually used. Standard measures including accuracy, precision, recall, F1-score, and AUC are used to assess the model's performance. which quantify classification effectiveness and generalisation.

4. Experimental Analysis and Results Discussion

Experimental validation is conducted on two benchmark datasets: the CICIDS2018 dataset reported by Sharafaldin, Lashkari, and Ghorbani [19] and the UNSW-NB15 dataset introduced by Moustafa and Slay [20]. Using the performance metrics discussed by Powers [23], the proposed HFRF framework will be compared with established machine-learning classifiers.

Table 1

Performance Comparative Analysis for Attacks Classification Using CICIDS2018 Dataset

Method	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	0.9998	0.9998	0.9998	0.9998	1.0000
Logistic Regression	0.9568	0.9612	0.9568	0.9576	0.9785
Decision Tree	0.9998	0.9998	0.9998	0.9998	0.9999
NaiveBayes	0.7359	0.8658	0.7359	0.7498	0.8483
HFRF_Random Forest	1.0000	1.0000	1.0000	1.0000	1.0000
HFRF_LogisticRegression	0.9983	0.9983	0.9983	0.9983	0.9999
HFRF_Decision Tree	0.9996	0.9996	0.9996	0.9996	0.9997
HFRF_NaiveBayes	0.8335	0.8921	0.8335	0.8421	0.9625

Table 2
Performance Comparative Analysis for Attacks Classification Using UNSW-NB15 Dataset

Method	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	0.9601	0.9600	0.9601	0.9600	0.9940
Logistic Regression	0.9701	0.9701	0.9701	0.9701	0.9959
Decision Tree	0.8637	0.8658	0.8637	0.8600	0.9178
NaiveBayes	0.9673	0.9673	0.9673	0.9673	0.9647
HFRF_Random Forest	0.8477	0.8463	0.8477	0.8465	0.9014
HFRF_Logistic Regression	0.9695	0.9695	0.9695	0.9695	0.9959
HFRF_Decision Tree	0.9758	0.9758	0.9758	0.9758	0.9969
HFRF_NaiveBayes	0.9093	0.9099	0.9093	0.9081	0.9704

The results in Table 1 and Table 2 show that the proposed HFRF framework consistently improves intrusion detection performance on both datasets. While classifiers already perform well on CICIDS2018, HFRF further enhances their results, with HFRF_Random Forest achieving the best performance. On the more challenging UNSW-NB15 dataset, HFRF notably improves accuracy and AUC, demonstrating its ability to capture complex patterns and strengthen feature representation.

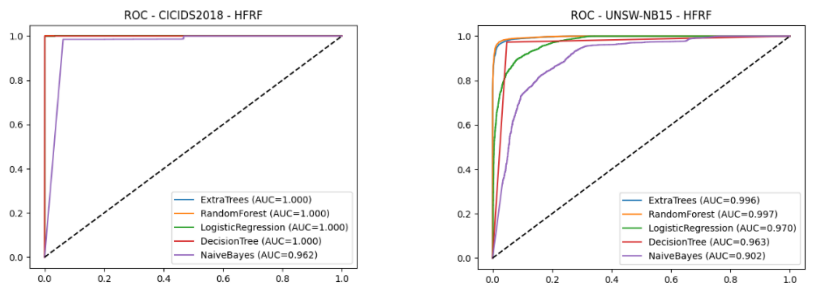


Figure 1
Performance Analysis with ROC Curves for the Proposed HFRF methods using CICIDS2018 and UNSW-NB15 Datasets

Figure 1 presents the ROC curves discussed by Bradley [21], showing strong classification performance of the proposed HFRF models on both the CICIDS2018 and UNSW-NB15 datasets. HFRF–Random Forest and HFRF–Logistic Regression achieve near-perfect AUC values. Similarly, the DET curves reported by Martin et al. [22]. In Figure 2 indicate that the models produce low error rates and a better balance between false positives and false negatives, ensuring reliable intrusion detection even on the more challenging UNSW-NB15 dataset.

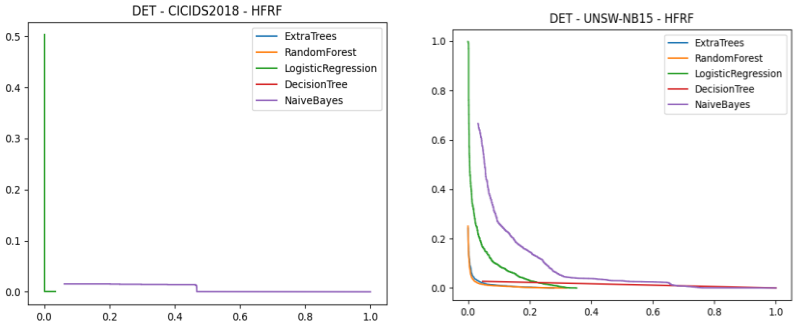


Figure 2
Performance Analysis with DET Curves for the Proposed HFRF methods using CICIDS2018 and UNSW-NB15 Datasets

5. Conclusion and Scope of the Work

It is evident that HFRF is highly effective at improving the performance of network intrusion detection systems. This technique has achieved perfect classification (100% across all evaluation metrics) using Random Forest and has increased Logistic Regression's effectiveness from 95.68% to 99.83%. Moreover, this technique has increased Decision Tree efficiency from 86.37% to 97.58%, with an AUC of 0.9969. Therefore, this shows that the proposed technique can enhance feature discriminability and generalizability. In the future, some other techniques include integrating with Using real-time intrusion detection systems, deep learning models, implementing explainable AI to increase the model's interpretability, implementing large-scale intrusion detection systems, and applying them to the IoT and cloud domains.

References

[1] A. M. Alsaffar, M. Nouri-Baygi, and H. M. Zolbanin, "Shielding networks: Enhancing intrusion detection with hybrid feature selection and stack ensemble learning," *Journal of Big Data*, vol. 11, Art. no. 133 (2024), doi: 10.1186/s40537-024-00994-7.

[2] P. H. H. N. de Araujo, A. Silva, N. Ferraz Junior, F. Cabrini, A. Santiago, A. Guelfi, and S. Kofuji, "Impact of feature selection methods on the classification of DDoS attacks using XGBoost," *Journal of Communication and Information Systems*, vol. 36, no. 1, pp. 200–214 (2021).

[3] M. A. Hossain, "Enhanced ensemble-based distributed denial-of-service (DDoS) attack detection with novel feature selection: A robust cybersecurity approach," *Artificial Intelligence Evolution*, vol. 4, no. 2, pp. 165–186 (2023), doi: 10.37256/aie.4220233337.

[4] S. Ahmed, Z. A. Khan, S. M. Mohsin, S. Latif, S. Aslam, H. Mujlid, M. Adil, and Z. Najam, "Effective and efficient DDoS attack detection using deep learning algorithm, multi-layer perceptron," *Future Internet*, vol. 15, no. 2, Art. no. 76 (2023), doi: 10.3390/fi15020076.

[5] J. Cheng, Y. Liu, X. Tang, V. S. Sheng, M. Li, and J. Li, "DDoS attack detection via multi-scale convolutional neural network," *Computers, Materials & Continua*, vol. 62, no. 3, pp. 1317–1333 (2020).

[6] I. O. Lopes, D. Zou, F. A. Ruambo, S. Akbar, and B. Yuan, "Towards effective detection of recent DDoS attacks: A deep learning approach," *Security and Communication Networks*, vol. 2021, Art. no. 5710028 (2021), doi: 10.1155/2021/5710028.

[7] A. D. Jadhav and V. Pellakuri, "Highly accurate and efficient two phase-intrusion detection system (TP-IDS) using distributed processing of HADOOP and machine learning techniques," *Journal of Big Data*, vol. 8, Art. no. 131 (2021), doi: 10.1186/s40537-021-00521-y.

[8] S. S. Samaan and H. A. Jeiad, "Feature-based real-time distributed denial of service detection in SDN using machine learning and Spark," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2302–2312 (2023).

[9] R. Abu Bakar, X. Huang, M. S. Javed, S. Hussain, and M. F. Majeed, "An intelligent agent-based detection system for DDoS attacks using automatic feature extraction and selection," *Sensors*, vol. 23, no. 6, Art. no. 3333 (2023), doi: 10.3390/s23063333.

[10] M. S. Akter, H. Shahriar, I. Iqbal, M. D. Hossain, M. A. Karim, V. Clincy, and R. Voicu, "Exploring the vulnerabilities of machine learning and quantum machine learning to adversarial attacks using a malware dataset: A comparative analysis," in *Proc. 2023 IEEE Int. Conf. Software Services Engineering (SSE)*, pp. 222–231 (2023).

[11] I. H. Sarker, "CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks," *Internet of Things*, vol. 14, Art. no. 100393 (2021), doi: 10.1016/j.iot.2021.100393.

[12] K. B. Dasari and N. Devarakonda, "Detection of different DDoS attacks using machine learning classification algorithms," *Ingénierie des Systèmes d'Information*, vol. 26, no. 5, pp. 461–468 (2021), doi: 10.18280/isi.260505.

[13] S. Rajesh, M. Clement, S. S. B. Sooraj, S. H. Al Shifan, and J. Johnson, "Real-time DDoS attack detection based on machine learning algorithms," pp. 1–5 (2021), doi: 10.2139/ssrn.3974241.

[14] M. J. Pasha, K. P. Rao, A. MallaReddy, and V. Bande, "LRDADF: An AI-enabled framework for detecting low-rate DDoS attacks in cloud computing environments," *Measurement: Sensors*, vol. 28, Art. no. 100828 (2023), doi: 10.1016/j.measen.2023.100828.

[15] A. Aljuhani, "Machine learning approaches for combating distributed denial of service attacks in modern networking environments," *IEEE Access*, vol. 9, pp. 42236–42264 (2021).

[16] F. Musumeci, A. C. Fidanci, F. Paolucci, F. Cugini, and M. Tornatore, "Machine-learning-enabled DDoS attacks detection in P4 programmable networks," *Journal of Network and Systems Management*, vol. 30, no. 1, Art. no. 21 (2022), doi: 10.1007/s10922-021-09633-5.

[17] M. A. Hossain and M. S. Islam, "Ensuring network security with a robust intrusion detection system using ensemble-based machine learning," *Array*, vol. 19, Art. no. 100306 (2023), doi: 10.1016/j.array.2023.100306.

[18] S. Sambangi and L. Gondi, "A machine learning approach for DDoS (distributed denial of service) attack detection using multiple linear regression," *Proceedings*, vol. 63, no. 1, Art. no. 51 (2020).

[19] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th International Conference on Information Systems Security and Privacy (ICISSP)*, Funchal, Madeira, Portugal, pp. 108–116 (2018).

[20] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," in *Proc. 2015 Military Communications and Information Systems Conference (MilCIS)*, Canberra, Australia, pp. 1–6 (2015), doi: 10.1109/MilCIS.2015.7348942.

[21] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159 (1997), doi: 10.1016/S0031-3203(96)00142-2.

[22] A. F. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. A. Przybocki, "The DET Curve in Assessment of Detection Task Performance," in *Proc. Eurospeech '97*, Rhodes, Greece, pp. 1895–1898 (1997).

[23] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies,* vol. 2, no. 1, pp. 37–63 (2011).

[24] M. S. Basha, K. R. Prasad, S. K. Mouleeswaran, R. C. Poonia, and S. Sebastian, "Multi-disease detection system with X-ray images using deep learning techniques," *J. Inf. Optim. Sci.,* vol. 45, no. 5, pp. 1379–1388 (2024).

Received March, 2026