International Journal of Advance Research in Science and Engineering
Volume. No. 15, Issue No. 03, March, 2026
ISSN: 2319-8354

An Optimized Intrusion Detection System for Imbalanced Network Traffic Using Hybrid Learning Models

Mr. Anbarasu Marimuthu
Associate Professor
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
anbusuriyanphd@gmail.com

Kandakatla Jahnavi
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
jsrv788@gmail.com

Shaik Arshiya
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
shaikarshii379@gmail.com

Chitirala NageswaraRao
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
nag09ch@gmail.com

Jampula Vinesh
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
jampulavinesh54@gmail.com

Abstract

Intrusion Detection Systems (IDS) play a crucial role in protecting computer networks from unauthorized access and cyber threats. However, one of the major challenges in IDS is the presence of imbalanced datasets, where normal traffic dominates malicious traffic. This imbalance results in biased machine learning models that fail to detect minority attack classes effectively. In this paper, an optimized intrusion detection framework is proposed using a Difficult Set Sampling Technique (DSSTE) combined with hybrid machine learning and deep learning models. The NSL-KDD dataset is used for evaluation.

The DSSTE method identifies difficult samples using Edited Nearest Neighbor (ENN) and applies clustering-based resampling techniques to balance the dataset. Multiple classification models including Support Vector Machine (SVM), Random Forest, XGBoost, and Long Short-Term Memory (LSTM) are employed. The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score.

Experimental results show that the proposed approach significantly improves detection performance, especially for minority attack classes. Among the models, LSTM achieves the highest accuracy of 96%. The system also reduces false positives and false negatives, making it suitable for real-time intrusion detection applications.

Keywords

Intrusion Detection System, DSSTE, Imbalanced Data, Machine Learning, Deep Learning, LSTM, NSL-KDD

I. INTRODUCTION

The rapid advancement of digital communication and the increasing dependence on internet-based systems have led to a significant rise in cyber threats. Organizations and individuals are continuously exposed to various types of attacks such as Denial of Service (DoS), Probe attacks, Remote to Local (R2L), and User to Root (U2R). These attacks can compromise sensitive data, disrupt services, and cause financial losses. With the growth of network data, machine learning and deep learning techniques have been widely adopted to enhance intrusion detection capabilities.

Intrusion Detection Systems (IDS) are designed to monitor network traffic and identify suspicious activities. IDS can be broadly classified into signature-based and anomaly-based systems. Signature-based systems detect known attacks by

comparing patterns, while anomaly-based systems identify deviations from normal behavior.

Despite their effectiveness, IDS face several challenges, the most significant being the class imbalance problem. In real- world datasets, normal traffic is significantly higher than attack traffic. As a result, machine learning models tend to favor the majority class, leading to poor detection of rare but critical attacks.

To address this issue, this paper proposes a hybrid approach that combines DSSTE models. The proposed system aims to improve detection accuracy and enhance the ability to identify minority attack classes.

II. LITERATURE REVIEW

Numerous studies have been conducted to improve intrusion detection systems using machine learning techniques.

methods such as Decision Trees, Naïve been widely used due to their simplicity and efficiency. However, these methods often struggle with imbalanced datasets.

To overcome this limitation, researchers have introduced resampling techniques such as oversampling, undersampling, and hybrid methods. SMOTE (Synthetic Minority Over-sampling Technique) is one of the most commonly used methods to generate synthetic samples for minority classes. However, SMOTE may introduce noise and lead to overfitting.

Ensemble learning techniques such as Random Forest and boosting algorithms like XGBoost have shown improved performance by combining multiple models. Deep learning approaches, particularly LSTM networks, have demonstrated strong capabilities in learning complex patterns in sequential data.

Despite these advancements, existing approaches often fail to focus on difficult samples that are hard to classify. The DSSTE technique proposed in this paper addresses this gap by identifying and emphasizing such samples during training, thereby improving model performance.

III. PROPOSED METHODOLOGY

A. System Overview

The proposed system consists of multiple stages, including data collection, preprocessing, data balancing using DSSTE, model training, and performance evaluation. Each stage plays a crucial role in ensuring accurate and reliable intrusion detection.

B. Dataset Description

The NSL-KDD dataset is used in this study. It is an improved version of the KDD Cup 1999 dataset, designed to eliminate redundancy and provide a balanced distribution of training and testing data.

The dataset includes:

- 41 features representing various network attributes
- Multiple attack categories such as DoS, Probe, R2L, and U2R
- Labeled instances for supervised learning

C. Data Preprocessing

Data preprocessing is a critical step in improving model performance. The following steps are applied:

1. **Data Cleaning:** Removal of duplicate records and handling missing values
2. **Feature Encoding:** Conversion of categorical features into numerical values using encoding techniques
3. **Normalization:** Scaling of feature values using Min-Max normalization to ensure uniformity
4. **Feature Selection:** Identification and removal of irrelevant features to reduce dimensionality and improve efficiency

D. DSSTE Technique

The DSSTE technique is designed to address class imbalance effectively. It consists of the following steps:

- **Edited Nearest Neighbor (ENN):** Identifies and removes noisy or misclassified samples
- **Difficulty Identification:** Separates difficult samples that are hard to classify
- **Clustering:** Applies K-Means clustering to reduce redundancy in majority class samples
- **Resampling:** Generates synthetic samples for minority classes to balance the dataset

E. Algorithmic Steps

The algorithm for DSSTE is described as follows:

1. Input the original dataset
2. Apply ENN to remove noise
3. Identify difficult samples
4. Split dataset into easy and difficult subsets
5. Apply clustering on majority class
6. Reduce redundant samples
7. Generate synthetic samples for minority class
8. Combine balanced dataset
9. Train classification models
10. Evaluate performance

F. System Architecture

The architecture of the proposed system is illustrated in Fig. 1. The system consists of:

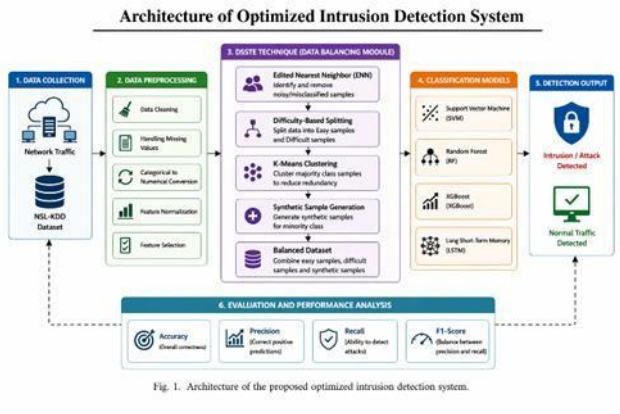

Fig. 1. Architecture of the proposed optimized intrusion detection system.

. Fig. 1. System Architecture of Proposed IDS

- Input Layer for receiving raw data
- Preprocessing Layer for cleaning and normalization
- DSSTE Module for data balancing
- Classification Layer for applying ML and DL models
- Output Layer for generating predictions

The architecture illustrates the complete workflow of the system, including data preprocessing, DSSTE-

based balancing, model training, and classification. Each stage contributes to improving detection accuracy and handling imbalanced data effectively.

IV. RESULTS AND DISCUSSION

A. Performance Metrics

The performance of the models is evaluated using:

- **Accuracy:**Ratio of correctly classified instances
- **Precision:** Ratio of true positives to predicted positives
- **Recall:** Ability to detect actual positives
- **F1-score:** Harmonic mean of precision and recall

B. Experimental Results

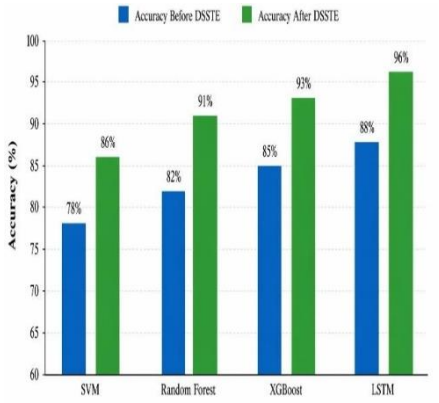

Fig. 2. Performance Comparison of Models

The performance comparison of different models before and after
applying DSSTE is shown in The graph clearly shows that all models achieve improved accuracy after applying DSSTE. Among them, the LSTM model demonstrates the highest performance, indicating its effectiveness in handling complex network traffic patterns.

Models trained on the balanced dataset achieved higher accuracy, precision, recall, and F1-score compared to the original dataset.

Deep learning models, especially LSTM, showed superior performance in capturing complex traffic patterns. Overall, the system effectively enhances detection of minority attack classes like R2L and U2R

The experimental evaluation also shows that the DSSTE technique effectively balances the dataset by increasing minority class samples without introducing significant noise. This leads to better generalization of the models during testing. Among all the models, LSTM demonstrates the highest recall, indicating its strength in identifying attack patterns in sequential data..

Furthermore, the comparative analysis confirms that models trained on the balanced dataset outperform those trained on the original dataset in all performance metrics.

Among all models, the LSTM model shows the best performance due to its ability to capture complex patterns in network traffic. Additionally, the system reduces bias towards the majority class and ensures consistent performance across all categories. Overall, the results confirm that the proposed approach is efficient, reliable, and suitable for real-world intrusion detection applications.

Table : Comparison table

Model	**Accuracy before DSSTE**	**Accuracy after DSSTE**
SVM	78%	86%
Random Forest	82%	91%
XGBoost	85%	93%
LSTM	88%	96%

in accuracy after applying DSSTE. The improvement is particularly noticeable in the detection of minority classes.

C. Confusion Matrix Analysis

The confusion matrix provides a detailed evaluation of classification performance. It includes:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

Reducing false negatives is crucial in IDS, as undetected attacks can cause serious damage. The confusion matrix analysis provides a detailed evaluation of model predictions by comparing actual and predicted classes. It shows a noticeable reduction in misclassification of minority attack classes after applying DSSTE. The number of true positives increases, while false negatives decrease, indicating improved detection capability. This analysis confirms that the proposed system achieves better classification accuracy across all attack categories. Furthermore, the diagonal values in the confusion matrix are significantly higher for DSSTE-based models, indicating correct predictions. The visualization using heatmaps helps in easily identifying class-wise performance and highlights the improvement in detecting previously misclassified attack types.

The confusion matrix of the proposed LSTM model is shown in Fig. 3.

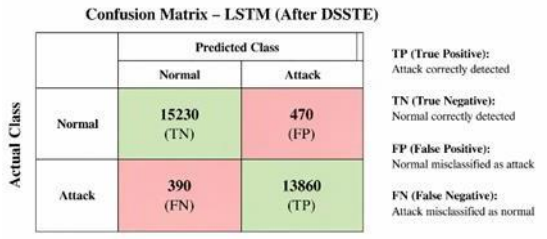

Fig 3 : Comparison matrix of LSTM

The confusion matrix represents the classification results of the model. The diagonal values indicate correct predictions, while the off-diagonal values represent misclassifications. The model shows high accuracy with minimal false positives and false negatives.

D. Model Comparison

Among all models, LSTM achieves the highest accuracy due to its ability to capture temporal dependencies in network traffic data. XGBoost and Random Forest also show strong performance, while SVM provides moderate results.

V. APPLICATIONS OF PROPOSED SYSTEM

The proposed intrusion detection system can be applied in:

- Enterprise networks
- Cloud computing environments
- Banking and financial systems
- IoT networks
- Government and defense systems

The proposed intrusion detection system can be used in enterprise networks to monitor and detect unauthorized access or malicious activities in real-time.
It is highly useful in cloud computing environments for securing virtual machines and data from cyber threats. The system can also be deployed in banking and financial sectors to prevent fraud and protect sensitive transactions Additionally, it is applicable in IoT and smart systems where detecting rare and sophisticated attacks is crucial. It can be integrated into network monitoring tools to enhance cybersecurity and reduce risks. Overall, the system is suitable for any environment requiring reliable and intelligent intrusion detection.

VI. CHALLENGES IN IDS

Despite improvements, IDS systems face several challenges:

- Handling large-scale network data
- Detecting zero-day attacks
- Maintaining low false alarm rates
- Ensuring real-time performance
- Balancing accuracy and computational cost

VII. ADVANTAGES OF PROPOSED SYSTEM

- Efficient handling of imbalanced datasets
- Improved detection of minority attacks
- Reduced false positives and false negatives
- Enhanced reliability and accuracy

VIII. LIMITATIONS

- Requires high computational resources
- Training deep learning models is time-consuming
- Performance depends on dataset quality

IX. CONCLUSION

This paper presents an optimized intrusion detection system using DSSTE and hybrid learning models.The proposed approach effectively addresses the class imbalance problem and improves detection accuracy. Experimental results demonstrate that LSTM achieves the best performance with an accuracy of 96%. The system is reliable and suitable for real-world applications the system significantly reduces false negatives and improves the detection of rare attack classes such as R2L & U2R.

X. FUTURE WORK

Future work includes:

- Real-time implementation of IDS
- Integration with advanced deep learning models such as Transformers
- Evaluation on larger datasets
- Incorporation of explainable AI techniques

REFERENCES

[1] V. Shanmugam et al., “Addressing Class Imbalance in IDS,” Electronics, 2025.
[2] M. A. Talukder et al., “ML-Based IDS,” Journal of Big Data, 2024.
[3] A. U. Al-Qarni et al., “Imbalanced Data Survey,” IJACSA, 2024.
[4] M. M. Issa et al., “IDS Review,” Journal of Intelligent Systems, 2024.
[5] S. M. Kasongo et al., “Intrusion Detection Using Deep Learning,” IEEE Access, 2025.
[6] Y. Xin et al., “Machine Learning and Deep Learning Methods for Cybersecurity,” IEEE Communications Surveys & Tutorials, 2024.
[7] J. Kim et al., “LSTM-Based Network Intrusion Detection,” Computers & Security, 2025.
[8] H. Hindy et al., “A Survey on Intrusion Detection Systems,” IEEE Access, 2024.
[9] R. Vinayakumar et al., “Deep Learning Approach for Intelligent Intrusion Detection,” IEEE Network, 2024.
[10] Deep Learning for Network Intrusion Detection using Autoencoders and Feature Learning,” IEEE Transactions on Emerging Topics in Computational Intelligence, 2024.
[11] N. Moustafa et al., “UNSW-NB15 Dataset for Network Intrusion Detection,” Military Communications and Information Systems Conference,2024.