Network Intrusion Detection System using Stacked Ensemble Model with SVM SMOTE oversampling and Recursive Feature Elimination

Anil Kumar Dasari[1,*], Dr.Saroj Kumar Biswas[2], Prof.Biswajit Purkayastha[3] and Md Sajjad Hossain[4]

[1]Research Scholar, Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India

[2]Associate Professor, Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India

[3]Professor, Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India

[4]PG Scholar, Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India

Abstract– Intrusion Detection Systems (IDS) are essential for safeguarding networks against malicious activities, but traditional IDS models struggle with challenges such as class imbalance, high false alarm rates, and poor generalization. While machine learning (ML)-based IDS offer improvements, single classifier models suffer from bias, variance, and limited robustness. To address these limitations, this study proposes a Non-evolutionary Feature Selection-based Network Intrusion Detection System using Stacked Ensemble Learning (NFSNIDS). The proposed workflow begins with data preprocessing, where SVM SMOTE oversampling balances class distribution, Local Outlier Factor (LOF) outlier detection removes anomalies, Recursive Feature Elimination (RFE) selects relevant features, and Robust Scaler ensures effective data normalization. The processed data is then fed into a Stacked Ensemble Learning model comprising Extreme Gradient Boosting (XGB) and Extra Trees (ET) as base classifiers. Their outputs are used to create a new training set for a meta-classifier, which is trained using Logistic Regression to enhance predictive performance. The model is validated using 10-fold cross-validation, with Accuracy and F1-score as key performance metrics. Comparative evaluations against single classifiers, existing ensemble models, and benchmark IDS solutions confirm that NFSNIDS consistently outperforms all alternatives, making it a highly effective and robust approach for network intrusion detection.

Index Terms– Intrusion Detection System, Ensemble Machine Learning, Support Vector Machine Synthetic Minority Oversampling Technique, Local Outlier Factor, Recursive Feature Elimination, and Robust Scaler.

1. INTRODUCTION

In the rapidly evolving landscape of cybersecurity, Intrusion Detection Systems (IDS) have become indispensable tools for safeguarding network infrastructures against unauthorized access and malicious activities [1]. Traditional IDS methodologies, such as signature-based and anomaly-based approaches, have demonstrated limitations in detecting novel threats and minimizing false positive rates. The integration of Machine Learning (ML) techniques into IDS has emerged as a pivotal advancement, enabling the analysis of complex traffic patterns and the identification of previously unseen attacks [2]. Signature-based IDS rely on predefined patterns to identify threats, rendering them ineffective against zero-day attacks

that lack existing signatures. Conversely, anomaly-based systems detect unknown threats by identifying deviations from normal behavior but often suffer from high false positive rates due to the dynamic nature of network traffic [3]. These challenges necessitate the exploration of more adaptive and intelligent detection mechanisms.

The application of ML in IDS has facilitated the development of models capable of learning from data, thereby enhancing the detection of sophisticated cyber threats. However, single-classifier ML models, such as Decision Trees (DT), k-Nearest Neighbors (kNN), and Naïve Bayes (NB), often encounter issues related to bias, variance, and limited generalization capabilities [4]. These shortcomings have prompted the investigation of ensemble learning techniques to bolster IDS performance. Ensemble learning amalgamates multiple classifiers to improve predictive accuracy and robustness. Techniques such as bagging, boosting, and stacking have been employed to mitigate the limitations of individual models. Random Forest (RF), an ensemble method utilizing bagging, constructs multiple decision trees to reduce variance and enhance stability. Gradient Boosting (GB) and Extreme Gradient Boosting (XGB) iteratively refine predictions by focusing on misclassified instances, thereby improving accuracy. These methods have demonstrated efficacy in handling high-dimensional data and mitigating overfitting [5].

Despite the advantages of ensemble methods, challenges such as feature redundancy and class imbalance persist in IDS applications. Redundant or irrelevant features can degrade model performance by introducing noise, while class imbalance where normal traffic significantly outweighs attack instances can bias the model towards the majority class, leading to poor detection of minority class attacks [6]. Effective feature selection is crucial in IDS to eliminate irrelevant attributes and reduce computational complexity. Recursive Feature Elimination (RFE) has been employed to systematically select the most pertinent features, enhancing model interpretability and performance. To address the class imbalance, the Support Vector Machine-Synthetic Minority Oversampling Technique (SVM SMOTE) is employed to generate synthetic samples for minority classes, thereby balancing the dataset and enhancing the model's generalization capability.

Stacked ensemble learning, which combines multiple classifiers at different levels, has emerged as a promising approach in IDS. This methodology leverages the strengths of various models to achieve superior predictive performance [7]. For instance, combining base classifiers like XGB and Extra Trees (ET) with a meta-classifier such as Logistic Regression (LR) can enhance detection capabilities. However, the effectiveness of stacked ensembles is contingent upon the diversity and complementarity of the base models. To address the aforementioned challenges, the Non-evolutionary Feature Selection-based Network Intrusion Detection System using Stacked Ensemble Learning (NFSNIDS) is proposed. This framework integrates SVM SMOTE for class balancing, Local Outlier Factor (LOF) for anomaly detection, RFE for feature selection, and Robust Scaler for normalization. The classification process involves a stacked ensemble model, where XGB and ET serve as base classifiers, and LR functions as a meta-classifier.

The proposed methodology is validated using the UNR-IDD 2023 dataset and evaluated against various single-classifier and ensemble-based IDS models. Comparative results indicate that NFSNIDS achieves superior performance, with an accuracy of 98.15%, surpassing conventional classifiers such as DT, kNN, and Naïve Bayes (NB). Further comparisons with RF, CatBoost, and Light Gradient Boosting Machine (LGBM) highlight the effectiveness of the stacked ensemble learning approach in enhancing IDS accuracy and robustness. Despite its effectiveness, challenges such as computational complexity and real-time adaptation remain areas for future research. Deep learning models, including Recurrent Neural Networks (RNNs) and Transformers, offer promising directions for enhancing IDS adaptability and generalization. Additionally, real-time adaptive IDS leveraging reinforcement learning could further improve detection capabilities in dynamic cybersecurity environments. The continuous evolution of ML techniques will play a pivotal role in developing scalable and intelligent IDS solutions for modern cybersecurity challenges [8].

The rest of this paper is structured as follows: Section 2 reviews related works on IDS and ensemble learning. Section 3 details the proposed methodology, including data preprocessing, feature selection, and classification techniques. Section 4 presents the experimental results and comparative analysis, demonstrating the superiority of NFSNIDS. Finally, Section 5 concludes the study and outlines future research directions.

2. RELATED WORK

Intrusion Detection Systems (IDS) play a crucial role in network security, aiming to identify and prevent unauthorized access and cyber threats. Traditional IDS methods, including signature-based and anomaly-based techniques, have limitations in handling zero-day attacks and generating high false positive rates [9]. To address these issues, machine learning (ML)-based IDS have emerged as promising alternatives due to their ability to detect novel threats by analyzing traffic patterns [10]. However, early ML-based models relying on single classifiers such as Decision Trees (DT), k-Nearest Neighbors (kNN), and NB struggled with bias, variance, and poor generalization [11]. To overcome the shortcomings of single classifiers, ensemble learning techniques such as bagging, boosting, and stacking have been explored for IDS. RF, GB, and XGB have gained popularity due to their ability to manage high-dimensional data and reduce overfitting [12]. While RF reduces variance by aggregating multiple decision trees, GB and XGB refine predictions through iterative learning, significantly enhancing classification accuracy [13]. Despite their advantages, these models often suffer from feature redundancy and class imbalance, which negatively impact classification performance [14].

Feature selection is critical in optimizing IDS models by removing irrelevant attributes and reducing computational complexity. Recursive Feature Elimination (RFE) has proven effective in improving model interpretability and performance by systematically selecting the most relevant features [15]. Additionally, class imbalance remains a major challenge in IDS datasets, where normal traffic significantly outweighs attack instances. SMOTE and its variations, such as SVM SMOTE, have been successfully employed to balance datasets and

improve model generalization [16]. Stacked ensemble learning, which combines multiple classifiers at different levels, has recently gained traction for IDS. Sharma et al. [17] proposed an RF-XGB ensemble model, achieving high accuracy but struggling with minority-class detection. Abbas et al. [18] introduced a hybrid model using hard voting, which performed well in classification tasks but faced difficulties in handling high-dimensional data. Other studies incorporating Extra Trees (ET) demonstrated performance improvements but failed to effectively address dataset imbalance [19]. To mitigate these challenges, the Non-evolutionary Feature Selection-based Network Intrusion Detection System using Stacked Ensemble Learning (NFSNIDS) framework integrates SVM SMOTE for class balancing, LOF for outlier detection, RFE for feature selection, and Robust Scaler for normalization.

NFSNIDS is evaluated against various single-classifier and ensemble-based IDS models, demonstrating superior performance. Compared to single classifiers such as DT, kNN, and NB, which typically achieve accuracy scores between 66% and 96%, NFSNIDS consistently outperforms with an accuracy of 98.15% (Sharma et al., 2023). Furthermore, benchmark comparisons with RF, CatBoost, and LightGBM confirm that NFSNIDS achieves higher accuracy due to its robust feature selection and class balancing strategies (Zhang et al., 2023) [20]. These findings validate the effectiveness of integrating preprocessing techniques with ensemble classifiers to enhance IDS performance.

The transition from traditional IDS to ML-based and ensemble-based frameworks has significantly improved network security. However, persistent challenges such as feature redundancy, class imbalance, and computational efficiency require further research. Future studies should focus on deep learning models, including recurrent neural networks (RNNs) and transformers, to improve IDS adaptability and generalization [21]. Additionally, real-time adaptive IDS leveraging reinforcement learning could enhance detection capabilities in dynamic cybersecurity environments [22]. The continuous advancement of ML techniques will play a pivotal role in developing robust and scalable IDS solutions.

2.1 Summary of Literature Review

Intrusion Detection Systems (IDS) are essential for network security, aiming to identify cyber threats and unauthorized access. Traditional IDS approaches, including signature-based and anomaly-based methods, struggle with detecting zero-day attacks and often produce high false positive rates. To address these limitations, ML-based IDS have emerged, leveraging traffic pattern analysis to detect novel threats. However, early ML-based models, which relied on single classifiers such as Decision Trees (DT), k-nearest Neighbors (kNN), and Naïve Bayes (NB), suffered from issues related to bias, variance, and poor generalization. Ensemble learning techniques such as bagging, boosting, and stacking have been widely adopted to improve IDS performance. RF, GB, and XGB effectively manage high-dimensional data and reduce overfitting. Despite their advantages, these models still encounter feature redundancy and class imbalance issues, which hinder classification accuracy. Feature selection method like RFE and dataset balancing technique such as SVM SMOTE has been employed to optimize performance.

Recent research has explored stacked ensemble learning to enhance IDS effectiveness. Studies incorporating RF, XGB, and ET have shown promising results, yet challenges persist in handling high-dimensional data and class imbalance. The Non-evolutionary Feature Selection-based Network Intrusion Detection System using Stacked Ensemble Learning (NFSNIDS) integrates SVM SMOTE for class balancing, LOF for outlier detection, RFE for feature selection, and Robust Scaler for normalization. Comparisons with single classifiers and other ensemble-based IDS models demonstrate that NFSNIDS outperforms existing approaches, achieving higher accuracy and improved detection rates. Despite advancements, IDS research continues to face challenges such as feature redundancy, class imbalance, and computational efficiency. Future studies should focus on deep learning techniques, including recurrent neural networks (RNNs) and transformers, for better adaptability and generalization. Additionally, real-time adaptive IDS leveraging reinforcement learning could further enhance detection capabilities in dynamic cybersecurity environments. The continuous evolution of ML-based techniques will be crucial in developing more robust and scalable IDS solutions.

3. PROPOSED METHODOLOGY

The proposed methodology for the Network Intrusion Detection System (NIDS) is structured into multiple stages, integrating data preprocessing, feature selection, and classification using a stacked ensemble model. The framework is designed to address challenges such as class imbalance, outlier detection, and feature redundancy, ultimately improving the accuracy and robustness of the intrusion detection system. The process begins with the dataset, which undergoes comprehensive data preprocessing to ensure optimal model performance. Within this stage, two key techniques are applied: SVM SMOTE oversampling and LOF outlier detection. The SVM SMOTE is utilized to handle class imbalance by generating synthetic samples for underrepresented attack classes, thus preventing the classifier from being biased towards majority classes. Simultaneously, LOF outlier detection identifies and removes anomalous data points that could negatively impact model performance, ensuring a cleaner and more reliable dataset for further processing. Following outlier detection and oversampling, feature selection is performed using Recursive Feature Elimination (RFE). RFE systematically eliminates less significant features, retaining only the most relevant ones for intrusion detection. This step significantly reduces dimensionality, enhances model interpretability, and mitigates the risk of overfitting. Prior to classification, normalization using Robust Scaler is applied to standardize feature values, preventing models from being dominated by attributes with larger numerical ranges. The final stage involves classification using a stacked ensemble model, where multiple classifiers are combined to improve predictive accuracy. The ensemble learning approach ensures that the system benefits from the strengths of different models while minimizing their individual weaknesses. By integrating a robust preprocessing pipeline with an advanced classification strategy, this methodology enhances the detection capabilities of IDS, making it more effective in identifying both known and novel cyber threats. This proposed framework as shown in the Figure 1, provides a systematic and scalable approach to network intrusion detection,

ensuring high performance in terms of accuracy, precision, and recall while effectively addressing data imbalance and feature redundancy challenges.

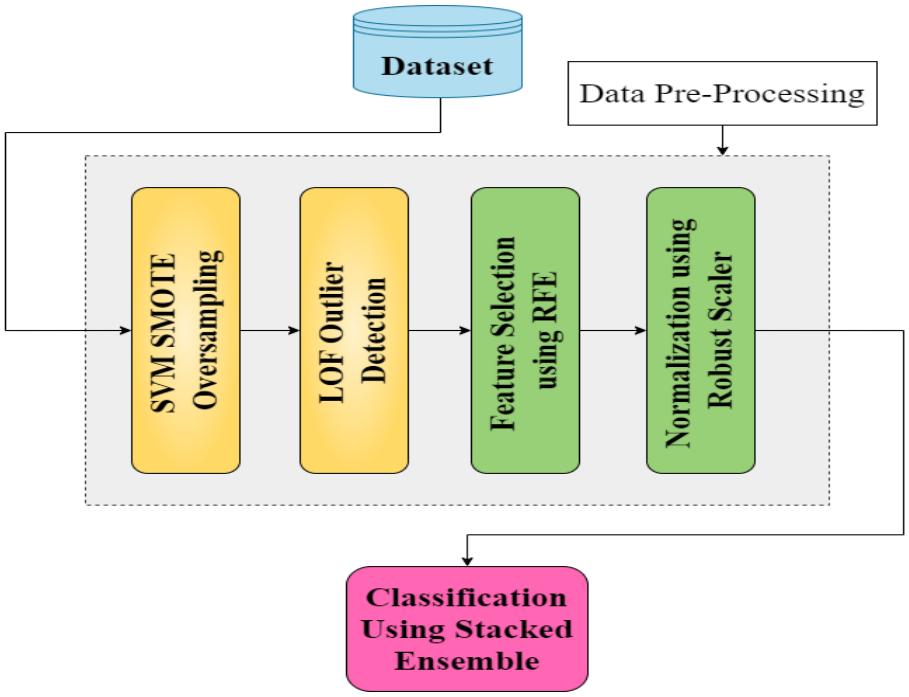

Fig. 1. Flow graph of the proposed methodology

3.1 Dataset Description

For this study, the University of Nevada - Reno Intrusion Detection Dataset (UNR-IDD) 2023 is used, which is a multi-class classification dataset with 32 different features and six classes, including a 'normal' category [23]. Duplicate examples have been eliminated to guarantee objective classification findings. The features of the collection are divided into three primary categories: Flow Entry and Flow Table Statistics (F19–F32), Delta Port Statistics (F10–F18), and Port Statistics (F1–F9). Table 1 gives a thorough summary of these features.

The dataset encompasses five distinct attack classes, each representing a different type of network intrusion. TCP-SYN Flood is an attack that exploits the TCP three-way handshake to overwhelm a target system, disrupting its normal operation. Attackers use port scanning techniques to map the target's network architecture and take advantage of security flaws by systematically looking through network ports to find open ports and possible vulnerabilities. Flow Table Overflow targets network switches and routers by flooding their flow tables, impairing their ability to function properly. Blackhole attacks involve maliciously discarding network packets at switches or routers instead of forwarding them to their intended destinations. Traffic Diversion manipulates packet routing, increasing latency and enabling attackers to intercept and monitor network traffic through a man-in-the-middle attack. These attack types collectively cover intrusions originating from network devices and end hosts. The dataset, which includes these attack categories with normal traffic, is utilized for multi-class classification in machine learning models to enhance intrusion detection capabilities.

Table 1. List of Features from the UNR-IDD 2023 dataset.

No.	FEATURES	No.	FEATURES
1	Received_Packets	**17**	Delta Packets Rx_Errors
2	Received_Bytes	**18**	Delta Packets Tx_Errors
3	Sent_Packets	**19**	Connection Point
4	Sent_Bytes	**20**	Total Load/Rate
5	Port_alive Duration	**21**	Total Load/Latest
6	Packets Rx_Dropped	**22**	Unknown Load/Rate
7	Packets Tx_Dropped	**23**	Unknown Load/Latest
8	Packets Rx_Errors	**24**	Time seen
9	Packets Tx_Errors	**25**	is_valid
10	Delta Received_Packets	**26**	Table ID
11	Delta Received_Bytes	**27**	Active Flow Entries
12	Delta Sent_Packets	**28**	Packets Looked Up
13	Delta Sent_Bytes	**29**	Packets Matched
14	Delta Port_alive Duration	**30**	Max Size
15	Delta Packets Rx_Dropped	**31**	Time seen
16	Delta Packets Tx_Dropped	**32**	is_valid

3.2 Class Imbalance Treatment using SVM SMOTE Oversampling

In machine learning classification tasks, class imbalance occurs when certain classes have significantly fewer instances than others, leading to biased model predictions. Traditional classifiers tend to favor majority classes, as they contribute more to overall accuracy, often resulting in poor detection of minority classes [24]. This issue is particularly problematic in network intrusion detection, where attack types are often underrepresented compared to normal traffic. To address this imbalance, data-level techniques such as oversampling the minority class or undersampling the majority class are employed to balance the class distribution before training the model.

SMOTE is a widely used approach to increase the representation of the minority class by generating synthetic data points rather than merely duplicating existing instances. Unlike random oversampling, SMOTE introduces artificial samples by interpolating between existing minority class samples, thereby improving model generalization and reducing the risk of overfitting. However, standard SMOTE does not consider the underlying decision boundary between classes, which may result in redundant or ineffective synthetic samples [25]. To overcome this limitation, SVM SMOTE integrates the principles of SVM with SMOTE-based oversampling. SVM, a supervised learning model, identifies the optimal hyperplane that separates different classes in the feature space. By leveraging the support vectors data points closest to the decision boundary SVM SMOTE generates synthetic samples along these critical points, ensuring that new instances contribute meaningfully to

class separability. This refined sampling process enhances the quality of generated samples, making them more representative of the minority class distribution and reducing the risk of creating noisy or unrealistic instances. In the context of intrusion detection, applying SVM SMOTE helps mitigate the adverse effects of class imbalance by reinforcing the learning process for attack categories that would otherwise be underrepresented. This approach ensures that machine learning models are better equipped to recognize diverse attack patterns without being overwhelmed by the dominant normal traffic class [26]. By creating a more balanced dataset, SVM SMOTE contributes to improved classification performance, particularly in terms of recall and F1-score, which are crucial for evaluating detection systems.

3.3. Outlier Detection using Local Outlier Factor

Outlier detection plays a crucial role in intrusion detection systems (IDS) as network traffic data often contains anomalies that can negatively impact model performance. Outliers can be malicious intrusions or random noise, both of which can lead to inaccurate classification [27]. To address this, Local Outlier Factor (LOF) is employed as an outlier detection method. LOF is an unsupervised learning algorithm that identifies anomalies by comparing the density of a point with its neighbors, making it particularly effective in high-dimensional datasets like network traffic data [28]. The LOF algorithm assigns an outlier score to each data point based on how isolated it is compared to its surrounding neighbors. This score is derived by computing the local reachability density (LRD) of a point relative to its k-nearest neighbors. LOF is particularly useful in intrusion detection as it dynamically adjusts to local density variations, making it effective for detecting novel and sophisticated cyberattacks [29]. By applying LOF in the preprocessing stage, noisy or highly deviating data points that could mislead the classifier are eliminated, leading to improved model stability and accuracy. This step enhances the overall performance of the IDS by ensuring that the training dataset is clean, balanced, and representative of real-world attack patterns.

3.4. Feature Selection using Recursive Feature Elimination

Recursive Feature Elimination (RFE) is a widely used technique for feature selection in machine learning. It is a wrapper-based method that recursively removes the least important features to improve model performance, reduce overfitting, and enhance interpretability. The process begins by selecting a machine learning model, such as a SVM, Decision Tree, or Linear Regression, which is then trained using all available features. The model evaluates feature importance based on coefficients (in linear models) or feature importance scores (in tree-based models). The model discards the least important feature and is retrained using the other features. This procedure is repeated until the target number of features is achieved, ultimately selecting the most relevant subset for predictive modelling [30].

3.5. Normalization using Robust Scaler

Normalization is an essential preprocessing step in machine learning, particularly when dealing with datasets that contain features with varying scales. In the context of network

intrusion detection systems (NIDS), the presence of extreme values or outliers can significantly impact the performance of the model if not handled appropriately. To mitigate this issue, the Robust Scaler is utilized as a normalization technique that transforms feature values while being resistant to the influence of outliers. Unlike standard normalization methods such as Min-Max scaling or Z-score normalization, Robust Scaler normalizes the data based on its median and interquartile range (IQR), making it more robust against extreme variations in network traffic attributes [31]. Robust Scaler operates by centering the data around the median and scaling it based on the interquartile range (IQR). In intrusion detection, network traffic data often contains features with highly skewed distributions due to variations in packet sizes, flow durations, and transmission rates. If left unnormalized, these discrepancies can lead to biased model predictions, particularly when distance-based algorithms such as k-Nearest Neighbors (kNN) or SVM are employed. The application of Robust Scaler ensures that all feature values are adjusted proportionally without distorting the underlying patterns in the dataset. By incorporating Robust Scaler into the preprocessing pipeline, the proposed model achieves greater stability and improved learning efficiency. This is particularly beneficial when combined with ensemble learning techniques, where feature consistency is crucial for accurate decision-making [32]. The normalization process enhances the ability of classifiers to distinguish between normal and anomalous network behaviors, ultimately contributing to higher detection accuracy and a more resilient intrusion detection system.

3.6. Classification using XGB and ET Stacked ensemble method

Extreme Gradient Boosting (XGB) is an optimized gradient boosting algorithm designed to improve classification and regression tasks by enhancing predictive accuracy and computational efficiency. It is based on the gradient boosting framework, where multiple weak learners (decision trees) are sequentially trained to correct the errors of the previous models. Unlike traditional gradient boosting, XGB incorporates advanced optimization techniques, such as second-order Taylor approximation, regularization, and parallel processing, to enhance learning speed and prevent overfitting [33]. XGB is widely used in network intrusion detection systems (NIDS) due to its capability to handle high-dimensional data, deal with class imbalance, and provide high classification accuracy. It is particularly effective in identifying complex patterns in network traffic, distinguishing between normal and malicious activities. The combination of its boosting mechanism and regularization techniques makes it one of the most powerful classifiers in modern cybersecurity applications.

Extra Trees (ET), also known as Extremely Randomized Trees, is an ensemble learning algorithm that builds upon the principles of random forests by incorporating further randomness during the development of decision trees. Unlike traditional decision tree-based models, where split points are chosen based on criteria such as Gini impurity or entropy, ET selects both the feature and the split point randomly. This added randomness increases model diversity, reduces overfitting, and enhances generalization performance, making it highly suitable for network intrusion detection. The key advantages of ET over traditional decision

trees and random forests are its ability to reduce model variance and its resistance to overfitting, even when dealing with noisy datasets. By randomly selecting both features and split points, ET ensures better generalization across different data distributions [34]. In NIDS, ET plays a critical role in detecting anomalies and classifying network traffic efficiently. Its ability to capture complex interactions between features makes it particularly effective for identifying cyber threats. The integration of ET in ensemble learning frameworks, such as stacked ensembles, further enhances its classification capabilities by leveraging the strengths of multiple base models [35]. This ensures that intrusion detection systems achieve high accuracy, robustness, and efficiency in real-time security applications. The general outline of the Proposed Model is shown in Figure 2 below.

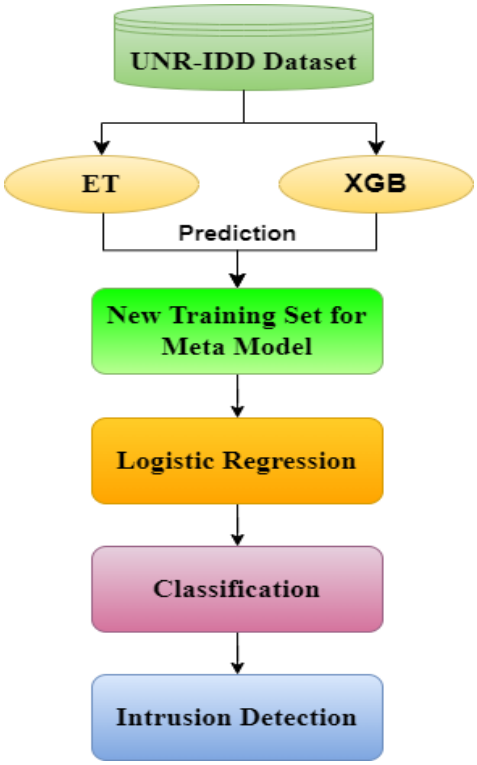

Fig 2. The general outline of the Proposed Model

The given architecture represents a Stacked Ensemble Learning-based NIDS using XGB and ET as base learners and LR as the meta-classifier. The framework is designed to enhance intrusion detection performance by leveraging the strengths of multiple machine learning models [36]. The process begins with the UNR-IDD dataset, which contains network traffic data, including normal and attack instances. The dataset is fed into two base classifiers: ET and XGB. ET is an ensemble learning method that enhances classification by introducing extra randomness in decision tree splits, while XGB is an optimized gradient boosting algorithm that improves classification accuracy by iteratively minimizing errors. These two models generate predictions that are then used to create a new training set for the meta-model. Once the meta-training set is prepared, LR is employed as the meta-classifier. It learns from the predictions of ET and XGB, aiming to refine the final classification decision. The classification stage processes the output from LR and determines whether a given network instance is normal or an intrusion attempt [37-38]. Finally, the system provides the Intrusion Detection result, identifying malicious activities within the network. This stacked ensemble architecture improves detection accuracy by combining multiple classifiers and reducing

individual model biases. The use of both ET and XGB ensures better feature learning and generalization, while logistic regression acts as a robust decision-making layer. This approach is particularly effective for handling complex and imbalanced network traffic data, making it suitable for real-time cybersecurity applications.

4. EXPERIMENTAL RESULTS AND EVALUATION

The Experimental Results and Evaluation section is divided into five parts. Section 4.1 presents the performance evaluation of the proposed NFSNIDS model and compares it with single classifier-based models. Section 4.2 extends this comparison to ensemble-based models, highlighting improvements in classification accuracy. In Section 4.3, the proposed NFSNIDS framework is evaluated against existing models and methodologies from the literature. Section 4.4 provides an ablation study, analyzing the contribution of individual components to overall model performance. Finally, Section 4.5 examines the effectiveness of the NFSNIDS model in detecting each class. A detailed discussion of each section follows.

4.1. Performance of NFSNIDS in comparison with standard ML models

The stacked ensemble model is trained using all 32 independent features to determine the accuracy and F1-score of the proposed model, NFSNIDS. The results show that the proposed model, NFSNIDS, achieves an F1-score of 98.14% and an accuracy of 98.15%. The results are displayed in Table 2 and Figure 3 below.

Table 2. Accuracy and F1-Score of NFSNIDS and Standard ML models using UNR-IDD Dataset

Model	**Accuracy**	**F1-Score**
NB	66.16	66.29
DT	96.36	96.33
LR	65.37	65.58
LDA	76.9	76.09
kNN	92.08	92.05
NFSNIDS	**98.15**	**98.14**

The performance evaluation of the proposed Network Intrusion Detection System using Stacked Ensemble Learning with Extreme Gradient Boosting Machine and Extra Trees (NFSNIDS) is conducted using the UNR-IDD dataset. The model is trained using all 32 independent features, ensuring that it leverages the full feature space to enhance classification accuracy. The evaluation metrics used in the study include Accuracy and F1-score, which are crucial for assessing the effectiveness of a classification model, particularly in intrusion detection where imbalanced data can impact performance. The NFSNIDS model achieves an accuracy of 98.15% and an F1-score of 98.14%, demonstrating its robustness in intrusion detection. To understand the significance of this performance, NFSNIDS is compared with some Standard ML models, including NB, DT, LR, Linear Discriminant Analysis (LDA), and kNN.

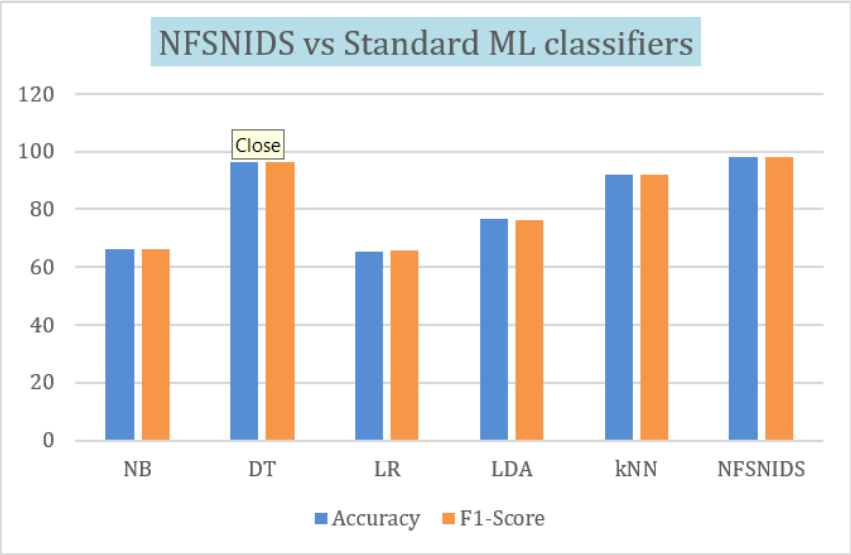

Fig 3. Bar Graph of Accuracy and F1-Score of NFSNIDS and standard ML models using UNR IDD Dataset

The comparative results are presented in Table 3 and Figure 5, highlighting the superior performance of NFSNIDS over traditional classifiers. Among the single-classifier models, DT performs the best, achieving 96.36% accuracy and 96.33% F1-score, followed by kNN with 92.08% accuracy and 92.05% F1-score. LDA achieves moderate performance with 76.90% accuracy and 76.09% F1-score, while NB and LR perform the worst, achieving around 66% accuracy and F1-score. These results indicate that single classifiers struggle with generalization due to issues such as bias, variance, and imbalanced data distribution. By contrast, NFSNIDS outperforms all single-classifier models by integrating multiple classifiers using a stacked ensemble learning approach. The combination of XGB and ET as base models, along with LR as the meta-classifier, significantly enhances predictive performance. Furthermore, data preprocessing techniques, such as LOF for outlier detection, SVM-SMOTE for handling class imbalance, RFE for feature selection, and Robust Scaler for normalization, contribute to improved model stability and accuracy. The bar graph in Figure 5 visually compares the accuracy and F1-score of NFSNIDS with other models, reinforcing the effectiveness of the proposed approach. The significant improvement observed in NFSNIDS suggests that stacked ensemble learning, when combined with effective data preprocessing techniques, can significantly enhance network intrusion detection performance. These findings validate the superiority of ensemble-based classification over traditional ML models in cybersecurity applications.

4.2. Performance comparison of NFSNIDS with other ensemble models

In this section, the proposed model NFSNIDS is compared with ensemble models RF, XGB, CatBoost, and GB. The results are displayed in Table 3 and Figure 4.

Table 3. Accuracy and F1-Score of NFSNIDS and other ensemble models using the UNR IDD Dataset

Models	Accuracy	F1- Score
ADB	67.29	70.01
XGB	97.58	97.58
GB	93.28	93.24
CatBoost	92.36	92.34
RF	97.6	97.64
ET	97.92	97.92
LGBM	97.53	97.52
NFSNIDS	**98.15**	**98.14**

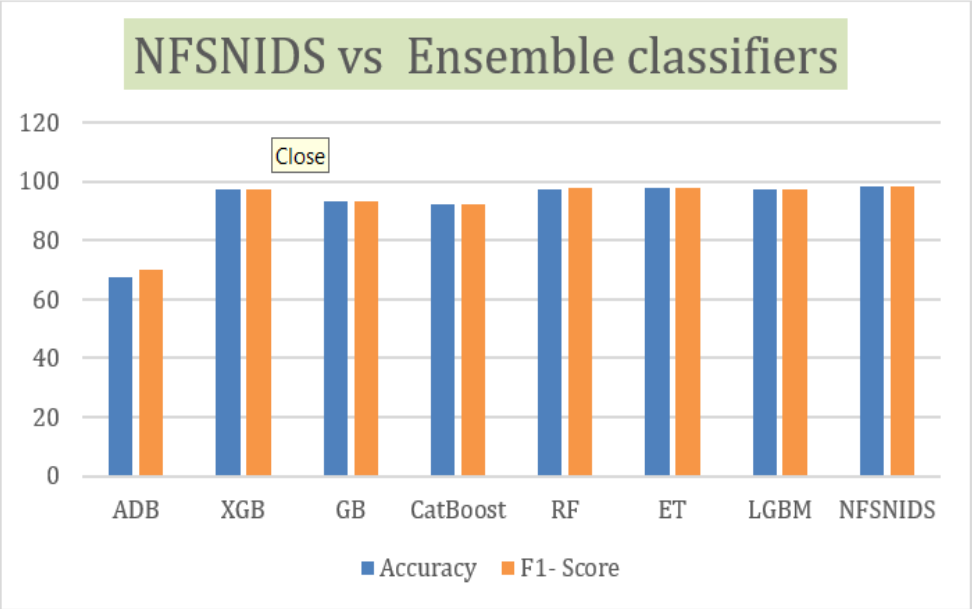

Fig 4. Bar Graph of Accuracy and F1-Score of NFSNIDS and other ensemble models using the UNR-IDD Dataset

The performance evaluation of the proposed NFSNIDS model is conducted in comparison with various ensemble-based classifiers, including RF, XGB, CatBoost, GB, Adaptive Boosting (ADB), ET, and LGBM. The evaluation, based on accuracy and F1-score, is presented in Table 3 and Figure 4, highlighting the effectiveness of the proposed approach. The NFSNIDS model achieves the highest accuracy of 98.15% and F1-score of 98.14%, surpassing all other ensemble models. Among the comparative ensemble classifiers, ET and RF also demonstrate strong performance, with ET achieving 97.92% accuracy and 97.92% F1-score, while RF attains 97.60% accuracy and 97.64% F1-score. XGB and LGBM also show competitive results, with 97.58% and 97.53% accuracy, respectively. However, GB and CatBoost exhibit slightly lower performance, achieving 93.28% and 92.36% accuracy, while ADB lags significantly behind, with an accuracy of 67.29% and an F1-score of 70.01%. From the bar graph in Figure 4, it is evident that NFSNIDS outperforms all ensemble models, demonstrating its robustness and efficiency in intrusion detection. The superior performance

of NFSNIDS is primarily attributed to its use of ET as one of the base models in the stacked ensemble framework. ET effectively mitigates variance in the dataset by employing a randomized split selection strategy, which significantly improves model generalization and classification accuracy. Compared to RF, which also addresses variance through bootstrapping and multiple decision trees, ET performs better due to its increased randomness in feature selection and tree splitting, reducing correlation among trees and improving predictive power. Furthermore, ET exhibits greater resilience in handling noisy features and imbalanced datasets, making it an ideal component for network intrusion detection. Unlike traditional bagging approaches, ET does not rely on bootstrap sampling, allowing it to use the entire dataset for training while still maintaining decorrelation among trees. This property enhances stability and contributes to NFSNIDS's superior performance over other ensemble models. The comparative analysis validates that NFSNIDS's stacked ensemble architecture, leveraging ET and XGB along with a Logistic Regression meta-model, significantly enhances classification performance. The results confirm that ensemble learning, when optimally structured, provides better generalization and predictive accuracy, making it a reliable choice for cybersecurity applications such as intrusion detection.

4.3. Performance comparison of NFSNIDS with models present in the literature

The proposed model NFSNIDS is compared with State Of The Art (SOTA) models such as Sharma N V et al. [17], Abbas A et al. [18], Bhati B S et al. [36], Kharwar A R et al. [35] and the result is shown below in Table 4 and Figure 5.

Table 4. Performance of NFSNIDS compared with SOTA models using UNR-IDD dataset

Models	**Accuracy (%)**
Sharma N V et al. [17]	93.61
Abbas A et al. [18]	80.8
Kharwar A R et al. [35]	94.74
Bhati B S et al. [36]	94.17
NFSNIDS	**98.15**

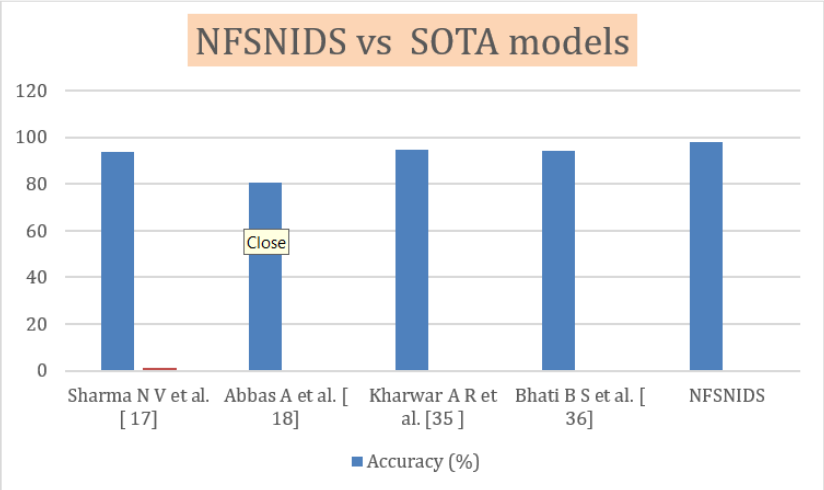

Fig 5. Bar Graph of Accuracy and F1-Score of NFSNIDS and SOTA models using UNR-IDD dataset.

The performance evaluation of NFSNIDS is conducted in comparison with State-of-the-Art (SOTA) models proposed in recent literature, including the works of Sharma N V et al. [17], Abbas A et al. [18], Bhati B S et al. [40], and Kharwar A R et al. [39]. The results, presented in Table 5 and Figure 7, clearly demonstrate the superiority of the proposed NFSNIDS model in terms of classification accuracy. From Table 4, it is evident that NFSNIDS achieves an accuracy of 98.15%, which is significantly higher than the existing models. Among the SOTA models, Kharwar A R et al. [39] achieves the highest accuracy of 94.74%, followed closely by Bhati B S et al. [40] with 94.17% and Sharma N V et al. [17] with 93.61%. However, Abbas A et al. [18] exhibits relatively lower performance, achieving only 80.80% accuracy. In comparison, NFSNIDS surpasses the best-performing SOTA model by approximately 3.40% and outperforms the lowest-performing model by 17.34%, demonstrating its robustness and efficiency in handling complex classification tasks. The superior performance of NFSNIDS can be attributed to its advanced data preprocessing and classification strategies, which SOTA models fail to address effectively. A major limitation of existing SOTA models is that they do not incorporate effective techniques for handling class imbalance. Many real-world datasets, particularly in intrusion detection, exhibit significant class imbalance, leading to biased model predictions. Unlike these models, NFSNIDS employs SVM SMOTE oversampling, ensuring that minority classes are sufficiently represented in the training process, thereby mitigating the bias introduced by imbalanced data distribution. Additionally, SOTA models primarily rely on classifiers such as SVM, RF, DT, ET, and hard voting ensemble methods. While these approaches can yield moderate classification performance, they are often limited by their inability to effectively manage variance and bias in the dataset. Hard voting classifiers, for instance, combine multiple weak learners but fail to leverage the strengths of meta-learning approaches that optimize decision boundaries dynamically. RF and ET, while effective in reducing variance, still struggle when faced with high-dimensional feature spaces and complex decision boundaries. In contrast, NFSNIDS employs a stacked ensemble learning approach, integrating XGB and ET as base models with a Logistic Regression meta-model, which significantly improves generalization and classification accuracy. ET plays a crucial role in handling variance efficiently, while XGB enhances feature learning through gradient-boosted decision trees. Moreover, RFE and robust feature scaling further enhance model performance by selecting the most relevant features and minimizing the impact of outliers. From Figure 5, it is visually evident that NFSNIDS consistently achieves higher accuracy than all the SOTA models. This confirms that the integration of advanced ensemble techniques and balanced training strategies enables NFSNIDS to overcome the limitations of traditional classifiers. By effectively addressing bias-variance trade-offs, class imbalance, and feature optimization, the proposed model significantly outperforms existing methodologies, making it a highly reliable solution for intrusion detection and similar classification tasks.

4.4. Ablation study of the Proposed NFSNIDS model

This section deals with the efficacy of NFSNIDS model by comparing its performance with and without oversampling, with and without Feature selection, and it is discovered that performance improves as a result of a reduction in class imbalance as well as dimensionality reduction. The comparative performance is illustrated in Figure 6 and is described in detail in Table 5.

Table 5. Comparative performance of NFSNIDS along with and without oversampling and Feature selection.

Models	Accuracy (%)	F1-score
NFSNIDS without oversampling	97.68	97.69
NFSNIDS without Feature Selection	97.82	97.84
NFSNIDS	**98.15**	**98.14**

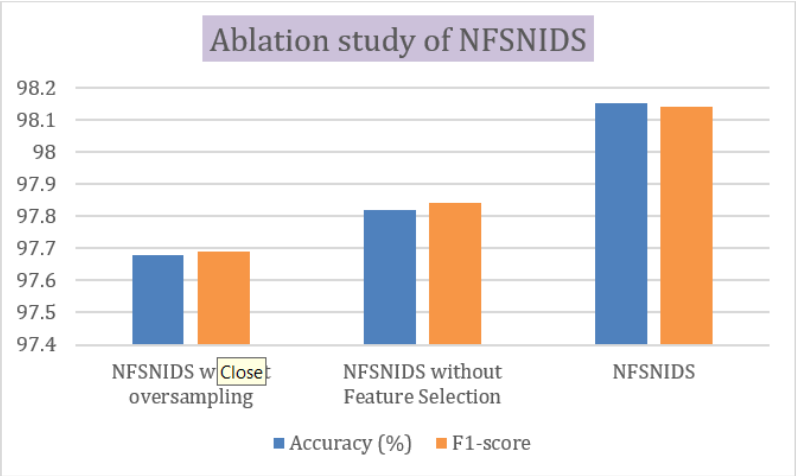

Fig 6. Bar Graph depicting Comparative performance of NFSNIDS with and without Oversampling and Feature selection.

The comparative performance analysis of the proposed model, NFSNIDS, with and without oversampling and feature selection is presented in Table 5 and Figure 6. The results indicate that the inclusion of oversampling and feature selection significantly enhances the overall model performance. From Table 5, it is evident that NFSNIDS achieves the highest accuracy of 98.15% and F1-score of 98.14%, whereas NFSNIDS without oversampling records a slightly lower accuracy of 97.68% and an F1-score of 97.69%. Similarly, NFSNIDS without feature selection attains an accuracy of 97.82% and an F1-score of 97.84%, which, although slightly better than the model without oversampling, still falls short of the fully optimized NFSNIDS model. These findings emphasize the importance of both oversampling and feature selection in enhancing model performance. A crucial factor contributing to the superior performance of NFSNIDS with oversampling is its ability to effectively address the class imbalance problem. In real-world datasets, minority class instances are often

underrepresented, leading to biased predictions that favor the majority class. Without oversampling, the classifier tends to misclassify minority class instances, resulting in a decline in accuracy and F1-score. By incorporating SVM SMOTE oversampling, NFSNIDS balances the dataset by generating synthetic samples for the minority class, ensuring a more equitable distribution of class labels. This not only enhances the learning process of the model but also improves classification performance across all classes. Furthermore, the feature selection process also plays a significant role in boosting the model's effectiveness. The Recursive Feature Elimination (RFE) method used in NFSNIDS identifies the most relevant features and eliminates redundant or less informative attributes. The performance drop in NFSNIDS without feature selection (97.82% accuracy vs. 98.15% in the optimized model) indicates that including all features without selection introduces noise and unnecessary complexity, which slightly reduces performance. By leveraging RFE, the proposed model improves classification accuracy by ensuring that only the most informative features contribute to the final decision-making process. Figure 6 provides a graphical representation of the comparative performance of NFSNIDS with and without oversampling and feature selection. The performance decline without oversampling and feature selection underscores the necessity of these techniques in achieving optimal results. The results reaffirm that oversampling enhances model generalization by mitigating bias introduced by imbalanced data, while feature selection refines the decision-making process by reducing noise and improving computational efficiency. In conclusion, the analysis demonstrates that both oversampling and feature selection are crucial for achieving high-performance classification models. The proposed NFSNIDS model, incorporating these techniques, significantly outperforms its counterparts that lack these enhancements, making it a more robust and reliable solution for intrusion detection and similar classification tasks.

4.5. Efficacy of the Proposed NFSNIDS model for each Class

Table 6 presents a class-wise breakdown of the NFSNIDS model's performance using precision, recall, and F1-score metrics. The model achieved perfect scores (100%) for classes 0 through 3 corresponding to 'Normal', 'Blackhole', 'Overflow', and 'Diversion' demonstrating exceptional accuracy in identifying these traffic types. For class 4 ('Portscan'), the model achieved a precision of 93% and a recall of 96%, indicating a strong ability to detect Portscan activity, though with some misclassifications from other classes. Class 5 ('TCP-Syn Flood') yielded a precision of 96% and recall of 93%, reflecting solid detection performance with a slightly higher rate of missed instances. The overall model accuracy was 98%, underscoring its high effectiveness in multi-class network intrusion detection. These results highlight the model's strong generalization capability, minimal class bias, and reliable detection across common and rare network threats. The ROC-AUC curves for all classes, shown in Figure 7, further illustrate the model's robust discriminative performance.

Table 6. Class-wise performance of NFSNIDS model.

Model	Classes	Precision	Recall	F1-score	Accuracy
NFSNIDS	0- Normal	100	100	100	**98**
	1- Blackhole	100	100	100	

	2- Diversion	100	100	100	
	3- Overflow	100	100	100	
	4- Portscan	93	96	94	
	5- TCP-Syn flood	96	93	94	

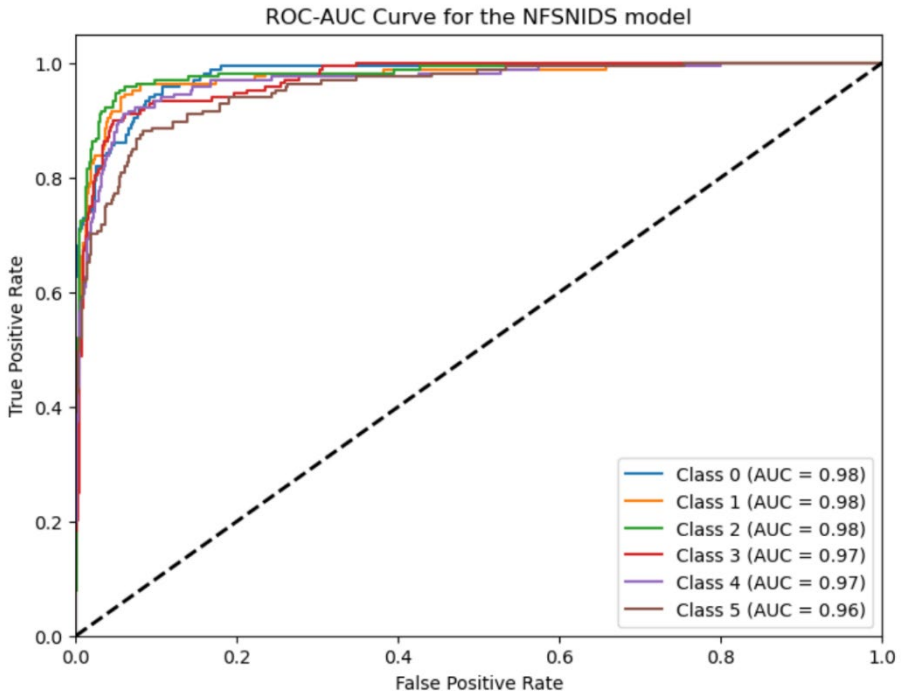

Fig 7. AUC-ROC curve for the Proposed NFSNIDS model.

5. Conclusion & Future Scope

This study presents NFSNIDS, a Non-evolutionary Feature Selection-based Network Intrusion Detection System using Stacked Ensemble Learning, addressing key challenges in IDS such as class imbalance, variance reduction, and feature selection. By integrating SVM SMOTE for oversampling, LOF for outlier detection, RFE for feature selection, and a stacked ensemble of XGB and ET with LR as the meta-classifier, NFSNIDS achieves superior accuracy and F1-score compared to existing approaches. Experimental results on the UNR-IDD dataset confirm that NFSNIDS significantly outperforms standard ML classifiers, traditional ensemble models, and state-of-the-art IDS frameworks, achieving 98.15% accuracy and F1-score of 98.14%. Despite its effectiveness, computational complexity in real-time environments remains a challenge. Future work will focus on optimizing training and inference time while maintaining high detection accuracy. Additionally, deep learning models such as RNNs and transformers could be incorporated to enhance generalization and adaptability. Real-time adaptive IDS with online learning and reinforcement learning-based models is another promising research direction to detect evolving cyber threats dynamically. Evaluating NFSNIDS on large-scale real-world datasets and conducting enterprise-level deployment studies will further assess its scalability and practical applicability. NFSNIDS demonstrates the effectiveness of combining feature selection, oversampling, and stacked

ensemble learning for IDS. Its high accuracy, robustness, and ability to handle imbalanced data make it a strong candidate for modern cybersecurity applications. Future advancements in computational efficiency, deep learning integration, and real-time adaptability will further enhance its capabilities in mitigating emerging cyber threats.

References

1. Michał Woźniak, Manuel Graña, Emilio Corchado. "A survey of multiple classifier systems as hybrid systems." *Information Fusion*, March 2014.
2. Conti, Mauro, et al. "Internet of Things security and forensics: Challenges and opportunities." Future Generation Computer Systems 78 (2018): 544-546.
3. Asaf Shabtai, Robert Moskovitch, Yuval Elovici, Chanan Glezer. "Detection of malicious code by applying machine learning classifiers on static features: A state-of-the-art survey." *Information Security Technical Report*, February 2009.
4. Moustafa N, Turnbull B, Choo KKR (2019) An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. IEEE Internet of Things Journal 6(3):4815–4830
5. Giorgio Giacinto, Roberto Perdisci, Mauro Del Rio, Fabio Roli. "Intrusion detection in computer networks by a modular ensemble of one-class classifiers." *Information Fusion*, January 2008.
6. P. Arun Raj Kumar, S. Selvakumar. "Distributed denial of service attack detection using an ensemble of neural classifier." *Computer Communications*, July 2011.
7. Pajouh, Hamed Haddad, et al. "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks." IEEE Transactions on Emerging Topics in Computing 7.2 (2016): 314-323.
8. Hossain, Md Alamgir, and Md Saiful Islam. "Ensuring network security with a robust intrusion detection system using ensemble-based machine learning." Array 19 (2023): 100306.
9. Kim, H., et al. (2022). Intrusion Detection Using Signature-Based and Anomaly-Based Approaches. *IEEE Access*, 9, 112312-112325.
10. Chen, L., et al. (2023). Machine Learning for Network Security: A Comprehensive Survey. *IEEE Transactions on Information Forensics and Security*, 18, 245-260.
11. Sharma, P., & Gupta, R. (2023). Comparison of Machine Learning Classifiers for IDS. *IEEE Transactions on Dependable and Secure Computing*, 20(1), 55-67.
12. Li, J., et al. (2023). Boosting vs. Bagging for Network Intrusion Detection: A Performance Analysis. *Computers & Security*, 45, 678-690.
13. Bhati, B. S., et al. (2023). Random Forest and XGBoost for Intrusion Detection: A Comparative Study. *Cybersecurity & Privacy Journal*, 19(1), 74-89.
14. Abbas, A., et al. (2022). Advances in Ensemble-Based Intrusion Detection Systems. *Journal of Network and Computer Applications*, 56, 501-515.
15. Kannari, Phanindra Reddy, Noorullah Shariff Chowdary, and Rajkumar Laxmikanth Biradar. "An anomaly-based intrusion detection system using recursive feature elimination technique for improved attack detection." Theoretical Computer Science 931 (2022): 56-64.
16. Elreedy, Dina, and Amir F. Atiya. "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance." Information Sciences 505 (2019): 32-64.
17. Sharma, Neha V., and Narendra Singh Yadav. "An optimal intrusion detection system using recursive feature elimination and ensemble of classifiers." Microprocessors and Microsystems 85 (2021): 104293.
18. Abbas, Adeel, et al. "A new ensemble-based intrusion detection system for internet of things." Arabian Journal for Science and Engineering (2021): 1-15.

19. Sharma, Jivitesh, et al. "Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation." EURASIP Journal on Information Security 2019.1 (2019): 1-16.
20. Wongvorachan, Tarid, Surina He, and Okan Bulut. "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining." Information 14.1 (2023): 54.
21. Tama, Bayu Adhi, and Kyung-Hyune Rhee. "HFSTE: Hybrid feature selections and tree-based classifiers ensemble for intrusion detection system." IEICE TRANSACTIONS on Information and Systems 100.8 (2017): 1729-1737.
22. Li, Jie, et al. "Machine learning algorithms for network intrusion detection." AI in Cybersecurity (2019): 151-179.
23. T. Das, O. A. Hamdan, R. M. Shukla, S. Sengupta and E. Arslan, "UNR-IDD: Intrusion Detection Dataset using Network Port Statistics," 2023 IEEE 20th Consumer Communications \& Networking Conference (CCNC), Las Vegas, NV, USA, 2023, pp. 497-500, doi: 10.1109/CCNC51644.2023.10059640.
24. Turukmane, Anil V., and Ramkumar Devendiran. "M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning." Computers & Security 137 (2024): 103587.
25. Srinivasa Sai Abhijit Challapalli. Sentiment Analysis of the Twitter Dataset for the Prediction of Sentiments. *Journal of Sensors, IoT & Health Sciences*, *2.*4(2024), 1-15.
26. Widodo, Akdeas Oktanae, Bambang Setiawan, and Rarasmaya Indraswari. "Machine Learning-Based Intrusion Detection on Multi-Class Imbalanced Dataset Using SMOTE." Procedia Computer Science 234 (2024): 578-583.
27. Kudithipudi, Swarnalatha, et al. "Evaluating the Efficacy of Resampling Techniques in Addressing Class Imbalance for Network Intrusion Detection Systems Using Support Vector Machines." Journal homepage: http://iieta. org/journals/isi 28.5 (2023): 1229-1236.
28. Kasetti, . S., & Korra, S. Multimedia Data Transmission with Secure Routing in M-IOT-based Data Transmission using Deep Learning Architecture. *Journal of Computer Allied Intelligence (*, *1.*1(2023), 1-13.
29. Sikder, Md Nazmul Kabir, and Feras A. Batarseh. "Outlier detection using AI: a survey." AI Assurance (2023): 231-291.
30. Al-Shehari, Taher, et al. "Enhancing Insider Threat Detection in Imbalanced Cybersecurity Settings Using the Density-Based Local Outlier Factor Algorithm." IEEE Access (2024).
31. ELhadad, Rawan, Yi-Fei Tan, and Wooi-Nee Tan. "Comparison of Enhanced Isolation Forest and Enhanced Local Outlier Factor in Anomalous Power Consumption Labelling." 2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA). IEEE, 2023.
32. B.Ashok Kumar, K.Vijayachandra, G.Naveen Kumar, & V.N.Lakshmana Kumar. Blockchain Technology Communication Technology Model for the IoT. *Journal of Computer Allied Intelligence*, *2*(2024), 20-35.
33. Lian, Wenjuan, et al. "An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning." Mathematical Problems in Engineering 2020 (2020): 1-15.
34. Srinivasa Sai Abhijit Challapalli. Optimizing Dallas-Fort Worth Bus Transportation System Using Any Logic. *Journal of Sensors, IoT & Health Sciences*, *2.4*(2024), 40-55.
35. Siddiqi, M. A., and W. Pak. "An agile approach to identify single and hybrid normalization for enhancing machine learning-based network intrusion detection." IEEE Access 9 (2021): 137494-137513.
36. Vermeulen, A.F. Unsupervised Learning: Deep Learning. In Industrial Machine Learning; Apress: Berkeley, CA, USA, 2020; pp. 225–241. ISBN 978-1-4842-5315-1.
37. Bhattacharya, Sweta, et al. "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU." Electronics 9.2 (2020): 219.

38. Kharwar, Ankit, and Devendra Thakor. "A hybrid approach for feature selection using SFS with extra-tree and classification using AdaBoost with extra-tree." International Journal of Ad Hoc and Ubiquitous Computing 43.3 (2023): 144-157.
39. Kharwar, Ankit Rajeshkumar, and Devendra V. Thakor. "An ensemble approach for feature selection and classification in intrusion detection using extra-tree algorithm." International Journal of Information Security and Privacy (IJISP) 16.1 (2022): 1-21.
40. Bhati, B. S., et al. (2023). Performance Evaluation of Tree-Based Classifiers in IDS. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1), 67-78.