

ME-IDS: An Ensemble Transfer Learning Framework Based on Misclassified Samples for Intrusion Detection Systems

Arka Ghosh

Supervisor: Dr. Qiang Ye

Faculty of Computer Science

November 28, 2023



Thesis Outline

- Overview
- Motivation
- Thesis Contributions
- Related Work
- System Overview of ME-IDS Framework
- Performance Evaluation
- Conclusion

Overview – Intrusion Detection System

What is Intrusion Detection System (IDS)?

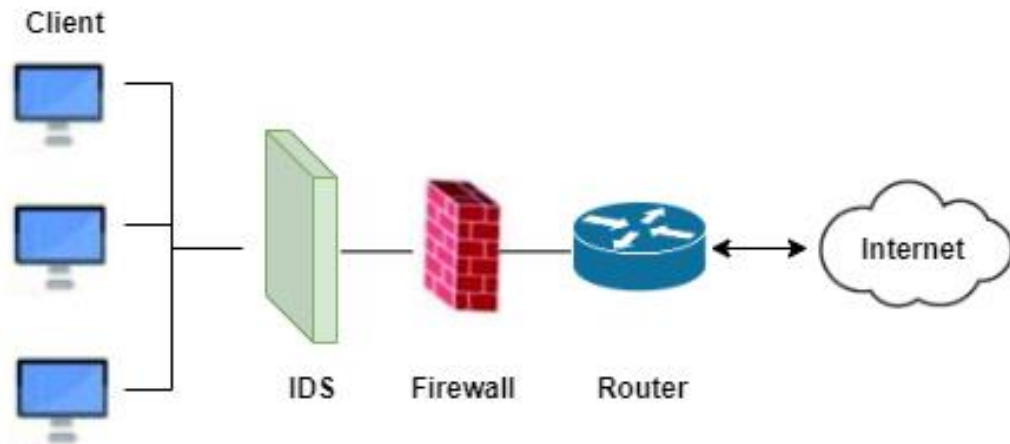


Figure: IDS Positioned between Clients and Network Perimeters

- A security solution designed to monitor network traffic for suspicious or malicious activity.
- Examines packets, protocols, and other network characteristics to identify malicious activity on a network.
- Generates alerts or notifications when suspicious activity is detected.

Overview – Intrusion Detection System

- **Signature-based IDS**

- Relies on pre-defined signatures or patterns to identify known threats.
- Compares network traffic against a database of known attack signatures.
- Raises an alert when the IDS identifies an attack with a matching signature from its directory.

- **Anomaly-based (ML/DL) IDS**

- Establishes a baseline model of network traffic behavior using ML or statistical analysis.
- Raises an alert whenever there is deviation of this behavior and this model.
- Effective against new and previously unseen threats.

- **Hybrid IDS**

- Integrates both signature-based and anomaly based detection methods.
- Combines the strengths of signature-based system's in identifying known threats and anomaly-based system's in detecting novel or evolving attacks.

Motivation

- **ML-based IDS:**

- First generation IDS are ML-based.
- Faces challenges in identifying unforeseen network traffic due to considerable diversity in network flows.
- Requires manual feature engineering and extraction.

- **DL-based IDS:**

- Addresses the challenges of ML-based IDS.
- Demonstrate superior performance in the presence of vast training dataset.
- Poses challenges for DL-based IDS due to scarcity of significant amount of data within IoT environment.
- Demands substantial computational resources and time in terms of training a DL-based IDS from scratch.

Motivation

- Transfer Learning-based solutions:
 - Addressed the challenges in training DL models by leveraging knowledge from pre-trained models, offering a powerful solution and improving performance on diverse tasks.
 - Delivered impressive results in the field of computer vision over the years.
 - Untapped potential for improvement, specifically in merging the principles of transfer learning and ensemble learning, within the IDS domain.

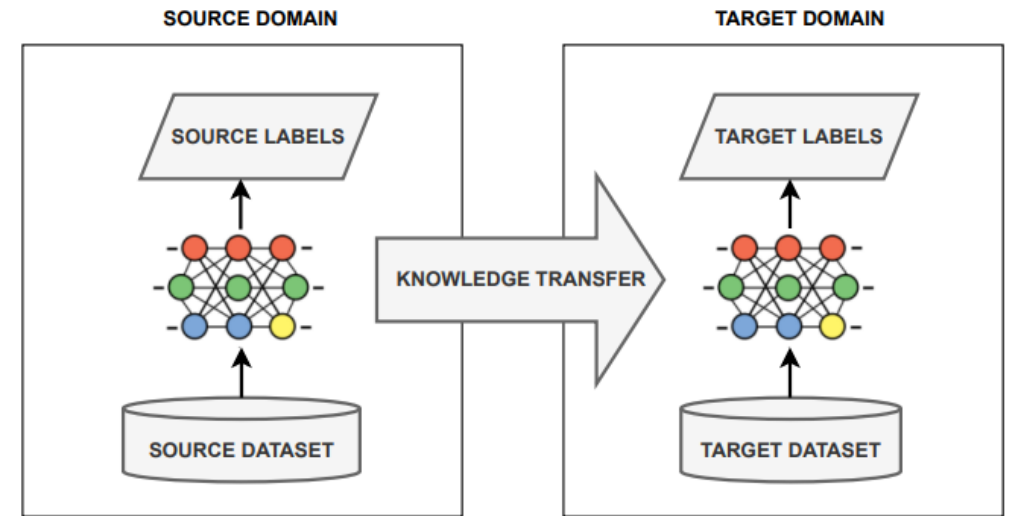


Figure: Process of Transfer Learning

Thesis Contribution

In this thesis, a novel IDS framework, named ME-IDS, is proposed that:

- Employed frequency encoding to encode categorical features and incorporated random noise with a minimal mean and standard deviation to mitigate the tied frequency counts among categories.
- Utilized a filter-based feature selection method, using a p-value threshold, to select important features.
- Adopted a novel weight optimized ensemble scheme that is based on three hyper-tuned variants of a lightweight pre-trained TL model.
- Assessed the performance of the ME-IDS on a publicly available IDS dataset and compared the results with other state-of-the-art IDS schemes.

Related Work

Stacked Ensemble Learning (Thockchom et al., 2023)	<ul style="list-style-type: none">▪ Implemented a stacked ensemble learning based IDS using DT, GNB, and LR as base classifiers and SGD as meta-classifier.▪ Evaluated the proposed IDS scheme on KDD Cup 1999, UNSW-NB15, and CICIDS2017 datasets.
Multistage spectrogram- based IDS (MS-ADS) (Ahmed et al., 2023)	<ul style="list-style-type: none">▪ Generated 2D spectrogram images using Short-Fourier Transformation.▪ Implemented a CNN model utilizing multiple stacks of CNN and Pooling layers.▪ Assessed the performance of the proposed scheme using BoT-IoT dataset.
IoT attack detection using ResNet (Hussain et al., 2020)	<ul style="list-style-type: none">▪ Adopted a chunk-based method to transform tabular data into RGB images after manually eliminating irrelevant features.▪ Trained and tested the generated RGB images using a pre-trained TL model, ResNet18 using CICDDoS2019 dataset.
Confidence Averaging and Concatenation Ensemble (Yang et al., 2022)	<ul style="list-style-type: none">▪ Proposed two novel ensemble learning approach, using the top 3 best performing pre-trained TL models out of 5 models.▪ Evaluated the performance of both the approaches on Car-Hacking and CICIDS2017 Dataset.

System Overview of ME-IDS Framework

The following stages are involved in the ME-IDS framework:

- Real-world IDS dataset
- Data pre-processing
- Tabular data to image conversion
- Transfer learning
- Hyper-Parameter Optimization
- Weighted Ensemble Learning

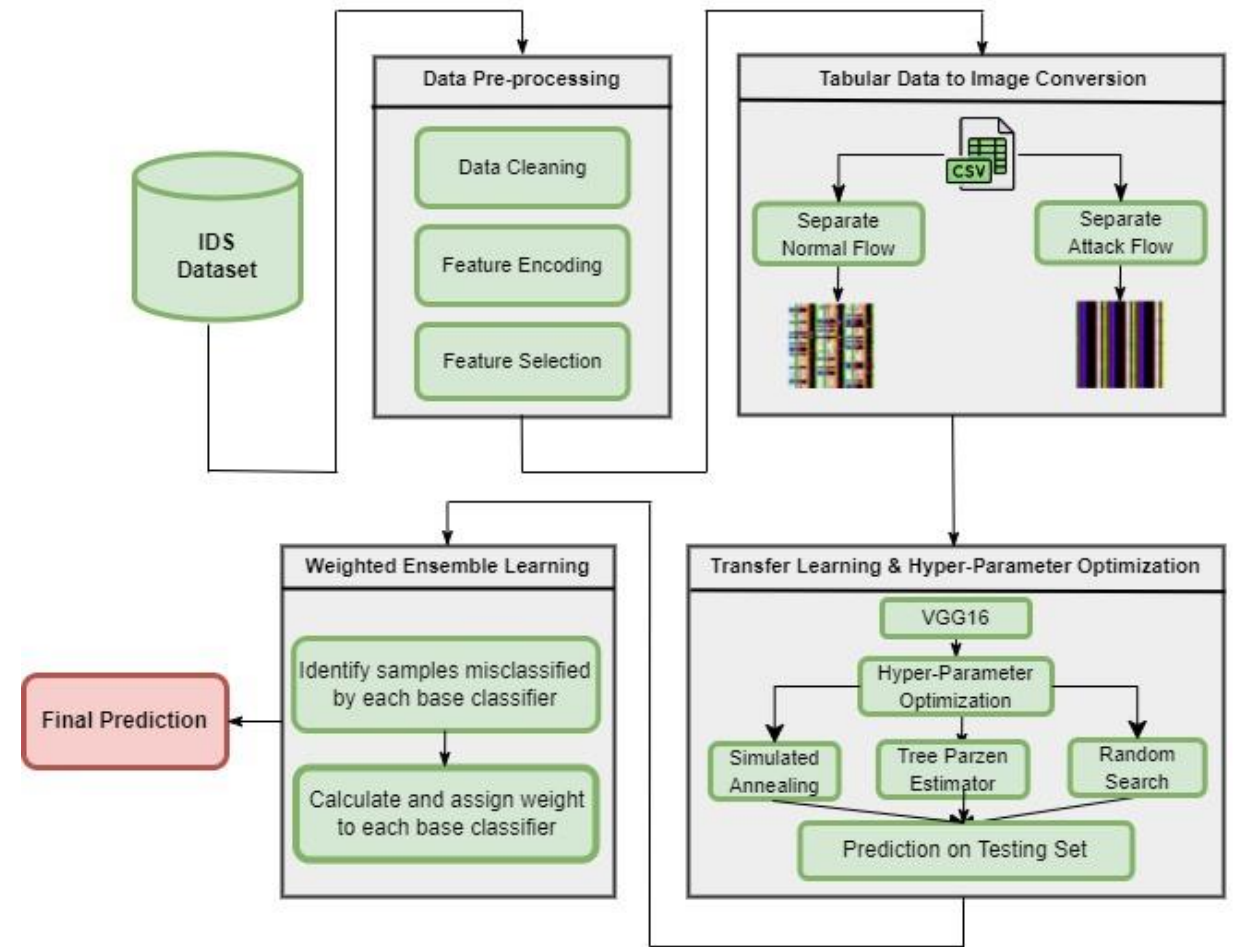


Figure: Overview of ME-IDS Framework

Dataset Overview: UNSW-NB15

- Curated by Cyber Range Lab of Australian Centre for Cyber Security.
- Comprised of 2,57,673 network records, portioned into separate training and testing subset.
- Each record is characterized by 43 features, including 40 numerical and 3 categorical features.
- Includes two labels, one for binary classification and another one for multi-class classification.

Dataset Overview: UNSW-NB15

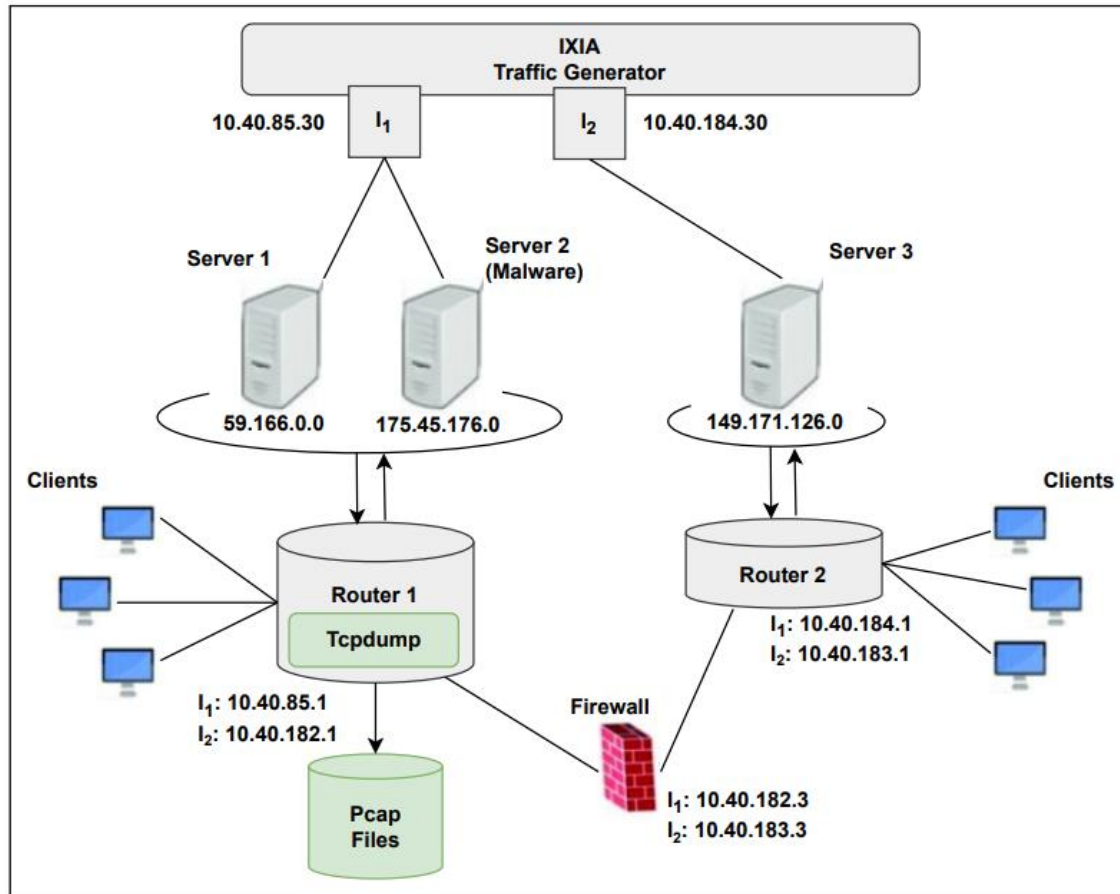


Figure: Testbed Architecture for Generating UNSW-NB15 [3]

Network Type	Training Set		Testing Set	
	Number of Records	Percentage (%)	Number of Records	Percentage (%)
Normal Flow	56,000	31.94	37,000	44.94
Attack Flow	119,341	68.06	45,332	55.06

Figure: Data Distribution of UNSW-NB15 Dataset

Data Pre-Processing: Data Cleaning

- Systematically checked for missing or corrupted (i.e., NaN) values within the dataset.
- No presence of any missing or corrupted values within the training and testing subset.
- Eliminated the “id” column as it presents the unique identifier for each network traffic.
- Retained the remaining 42 features for further analysis.

Data Pre-Processing: Feature Encoding

- **Label Encoding:**

- Represents the categories by integers sequentially, starting from 0.
- Assumes an ordinal relationship between the assigned integers, implying a hierarchy (i.e., BSc: 0 > MSc: 1 > PhD: 2).
- Such hierarchy is not applicable in terms of IDS Dataset.

- **One-hot Encoding:**

- Represents the categorical features by creating a new binary column for each unique category.
- Indicates the presence and absence of each category by 1 or 0, respectively.
- Leads to curse of dimensionality due to large number of categories.

Data Pre-Processing: Feature Encoding

id	color
1	red
2	blue
3	green
4	blue

One Hot Encoding

id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

Figure: Example of One-hot Encoding

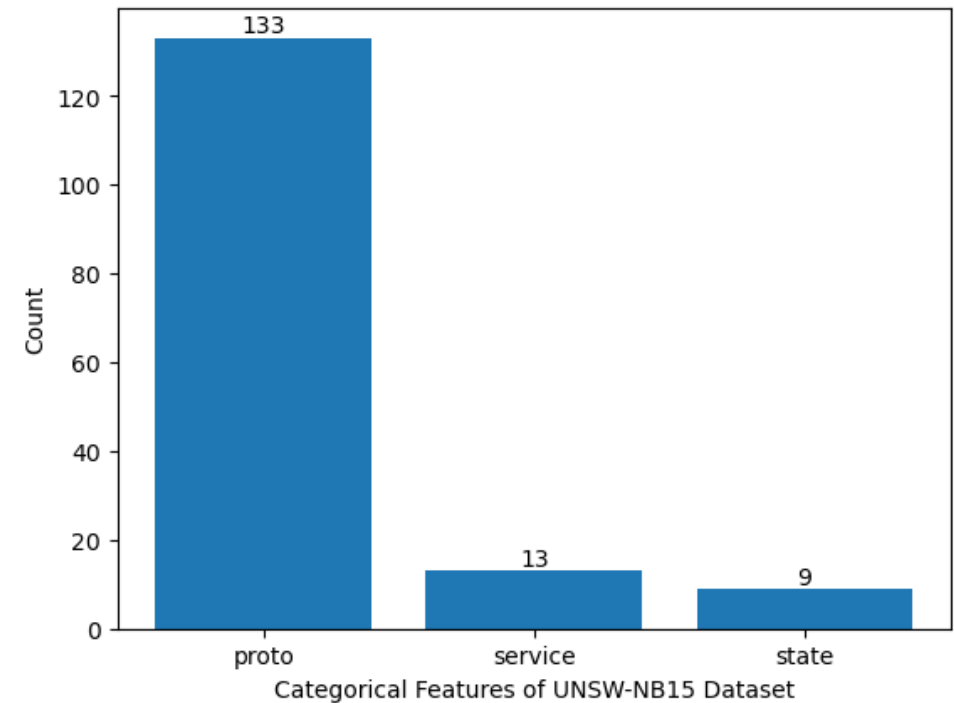
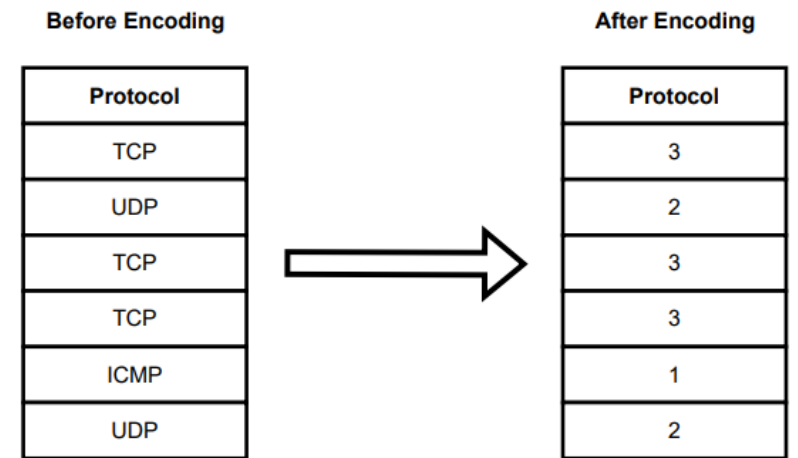


Figure: Frequency of Unique Categories of Categorical Features of UNSW-NB15 Dataset

Data Pre-Processing: Feature Encoding

- **Frequency Encoding:**

- Assigns numerical value to each unique categories based on their total frequency count.
- Provides a numerical representation to the categories of a categorical features that reflects the popularity of each category.
- Poses challenges when there is tied frequency count within two or more categories.
- A minimal noise using a normal distribution of with mean 0 and standard deviation of 0.01 is introduce to mitigate the problem of tied frequency count.



Protocol
TCP
UDP
TCP
TCP
ICMP
UDP

Protocol
3
2
3
3
1
2

Figure: Example of Frequency Encoding

Data Pre-Processing: Feature Selection

- Utilized a filter-based method, Chi-Square, for feature selection.
- Unlike wrapper and embedded based feature selection methods, filter-based methods are not classifier dependent and comparatively faster.
- Assesses the independence between a categorical feature and a target variable by calculating a p-value.
- Features with p-values below a chosen significance level (p-value ≤ 0.05) are considered significant.
- Retained the remaining 39 features for further analysis.

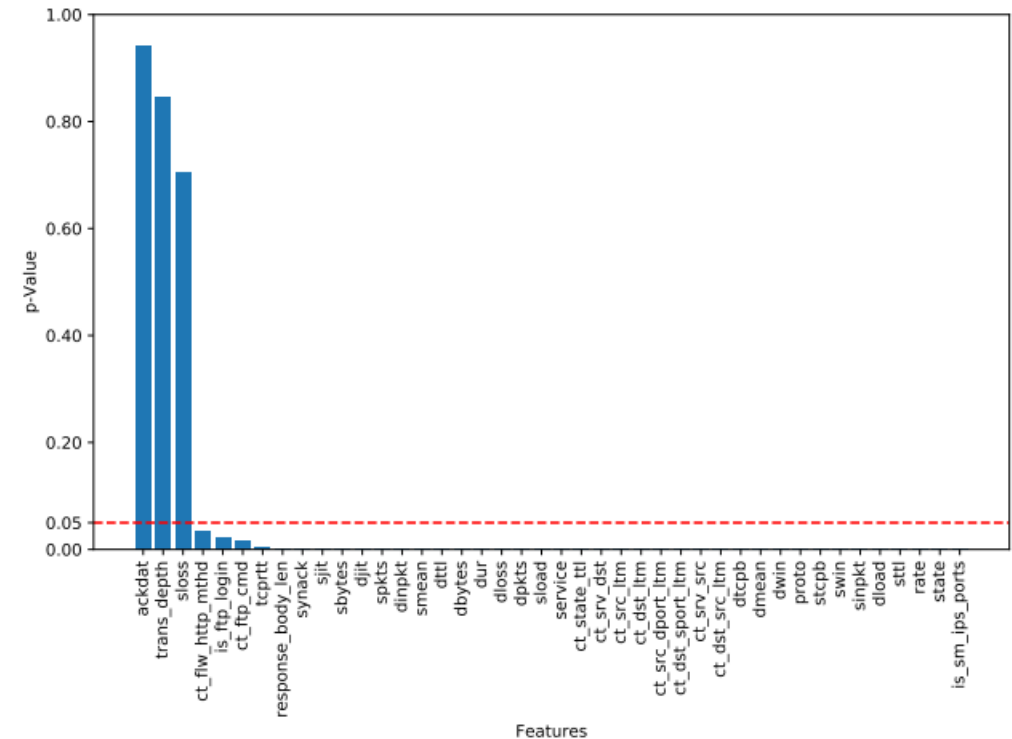


Figure: Feature Significance Analysis Using Chi-Square p-Value.

Tabular Data to RGB Image Conversion

- Adopted a chunk-based method to form square-shaped RGB images.
- Separated the normal and attack flows into two distinct dataframes.
- Selected a total of $39 \times 3 = 117$ rows iteratively in each chunk.
- Each chunk's first 39 feature rows arranged into an image matrix for channel 1, next 39 feature rows for channel 2, and the final 39 samples for channel 3 of RGB image using OpenCV.
- Applied the process to both training and testing set until all samples are turned into RGB images.

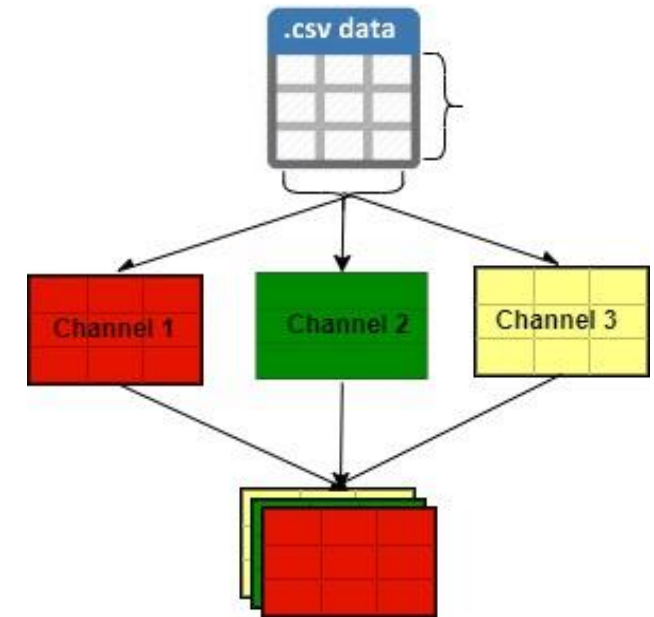
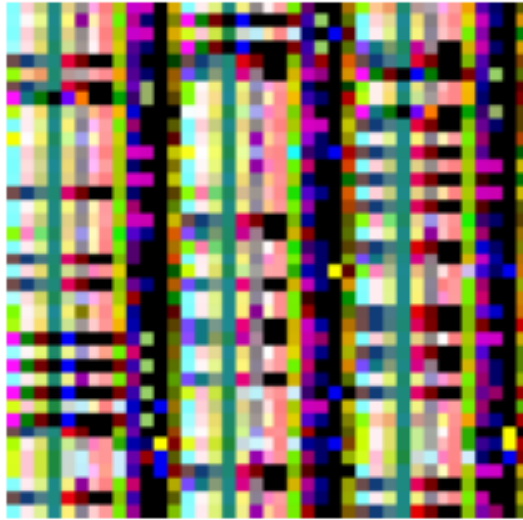
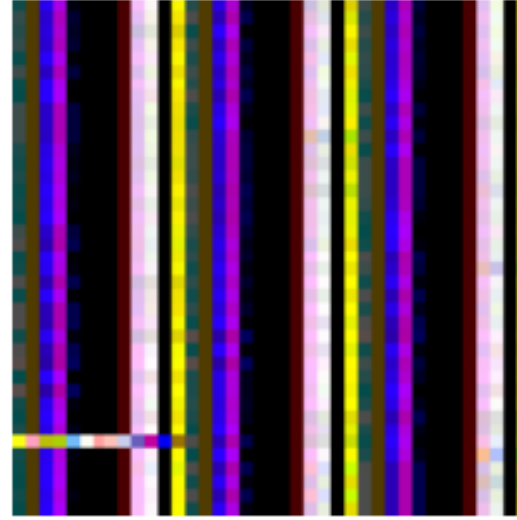


Figure: Process of Transforming Tabular Data to RGB Images

Tabular Data to Image Conversion



(a)



(b)

Figure: Example of Generated RGB Images from UNSW-NB15 Tabular Data:
(a) Normal flow; (b) Malicious flow

Transfer Learning

- Used VGG16 as the base architecture within the ME-IDS framework.
- Initial architecture of VGG16 comprises of a total of 16 layers, including 13 CNN layers followed by 3 fully connected layers.
- Utilized the CNN blocks of VGG16 with partial freezing, followed by a Flatten Layer, a Fully Connected layer with Dropout and final Sigmoid activation layer.

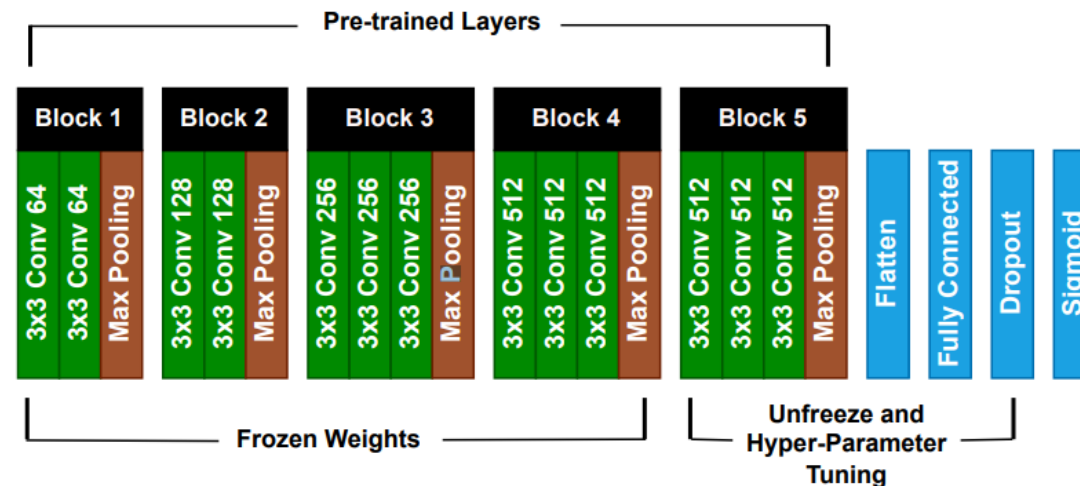


Figure: Proposed VGG16 Architecture adapting to the UNSW-NB15 dataset.

Hyper-Parameter Optimization

- Simulated Annealing (SA):
 - Inspired by the annealing process in metallurgy.
 - Starts with a high probability of accepting worse solutions and gradually decreases it over time.
- Tree Parzen Estimator (TPE):
 - Employs a tree-structured technique that uses Bayesian optimization handle hyper-parameters.
 - Maintains two probability distributions for each hyper-parameter: one for successful configurations and one for unsuccessful ones.
- Random Search (RS):
 - Proposed to address the limitation of Grid Search.
 - Focuses on exploring the hyper-parameter space without considering the performance of previously sampled configurations.

Hyper-Parameter Optimization

- Parameters related to model's architecture:
 - Number of Frozen Layers
 - Filter numbers in fully connected layers
 - Dropout Percentage
- Parameters involved into model's training:
 - Learning Rate
 - Number of Epochs
 - Patience

Hyper-Parameter	Search Space	Step Size	Optimal Value		
			SA	TPE	RS
Frozen Layer	15 - 18	1	17	16	16
Filter	128 - 512	128	256	384	256
Dropout Rate	0.1 - 0.5	0.1	0.2	0.5	0.2
Learning Rate	0.001 - 0.005	0.001	0.004	0.003	0.002
Epoch	10 - 30	5	15	10	20
Patience	2 - 10	1	7	8	8

Figure: Hyper-parameter Configuration Space for VGG16 on RGB images generated using 39 features

Weighted Ensemble Learning

- Proposed a weighted ensemble learning approach with optimized weights in ME-IDS framework.
- Three fine-tuned variants of VGG16 serve as base classifiers.
- The weighted ensemble learning can be denoted as:

$$y_{ensemble} = \sum_{i=1}^n W_i F_i(x)$$

Here,

- $F_i(x)$ => Prediction made by i^{th} base classifier for input x ,
- W_i => Weight assigned to i^{th} classifier,
- $y_{ensemble}$ => Final ensemble prediction for input x

Weighted Ensemble Learning

- The weights are inversely proportioned to the samples misclassified by each base classifiers:

$$w_i = \frac{1}{m_i + \varepsilon}$$

Here,

- m_i => Total number of misclassification made by i^{th} classifier
 - ε => A small constant value of 0.001 in case of rare cases when $m_i = 0$
 - w_i => Initial weight for each base classifier.
- Initial calculated weights are normalized between 0 to 1 using:

$$W_i = \frac{w_i}{\sum_{i=1}^n w_i}$$

Performance Evaluation: ML-Based IDS

- Six classifiers has been used to build ML-based IDS:
 - Decision Tree and Random Forest (Tree-based)
 - Gaussian Naïve Bayes (Probabilistic model)
 - Logistic Regression (Linear model)
 - K-Nearest Neighbors (Instance-based learning)
 - Stochastic Gradient Descent (Optimization-based)
- Selected due to their versatility in wide range of classification tasks and algorithmic approach.

Performance Evaluation: ML-Based IDS

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	85.96	87.57	84.87	85.41
RF	87.03	89.64	85.74	86.41
GNB	75.20	76.79	73.64	73.83
LR	80.40	84.91	78.47	78.86
kNN	84.78	87.12	83.46	84.04
SGD	80.93	87.05	78.80	79.17

Figure: Performance of ML-Based IDS with All Features

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	86.01	87.67	84.91	85.46
RF	87.18	89.79	85.89	86.56
GNB	75.99	78.81	74.15	74.29
LR	80.40	84.92	78.47	78.86
kNN	84.87	87.20	83.56	84.14
SGD	80.93	87.05	78.80	79.17

Figure: Performance of ML-Based IDS with Selected Features

Performance Evaluation: ML-Based IDS

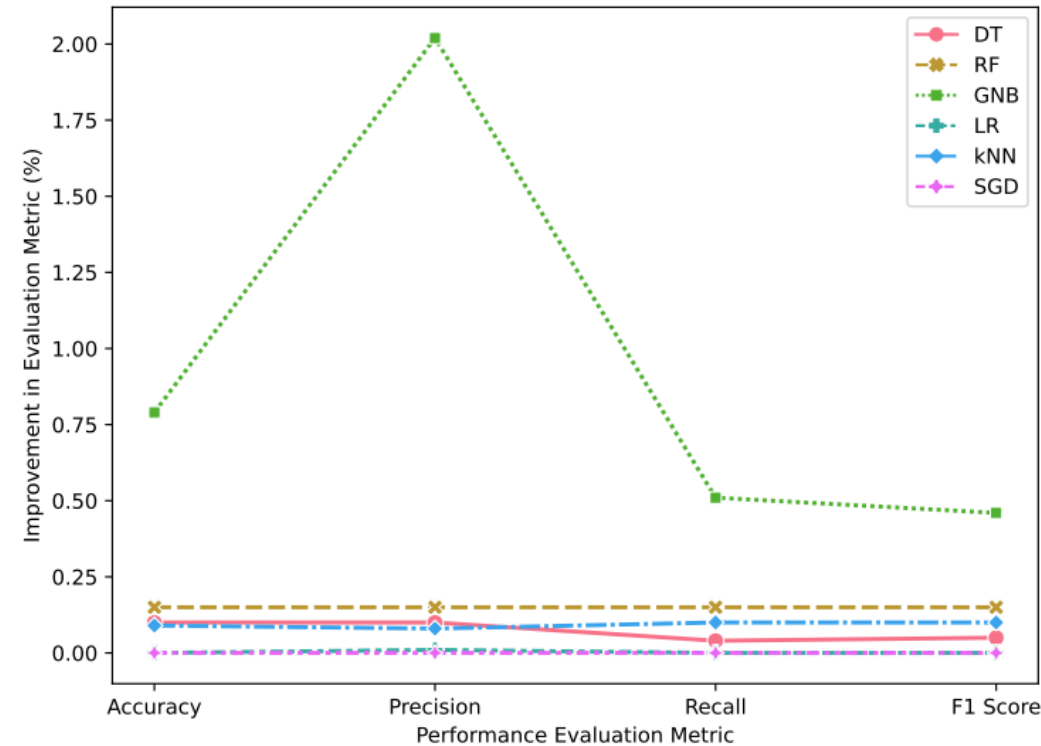


Figure: Performance Improvement of Traditional ML-based IDS with Selected Feature over All Features

Performance Evaluation: ME-IDS & Benchmark IDS Frameworks

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Stacked Ensemble [1]	85.89	87.45	84.81	85.34
Concatenation Ensemble [2]	95.71	96.38	95.22	95.61
Confidence Averaging [2]	97.39	97.73	97.09	97.34
VGG16-SA	98.16	98.33	97.98	98.13
VGG16-TPE	98.31	98.34	98.24	98.29
VGG16-RS	97.69	97.99	97.44	97.66
ME-IDS	98.77	98.87	98.67	98.76

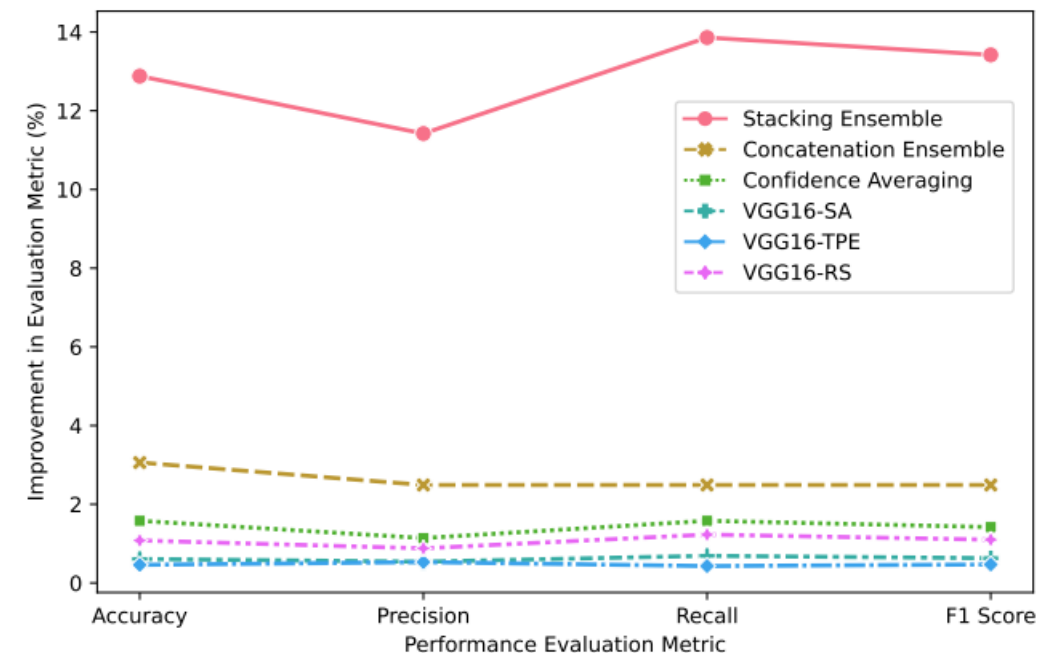


Figure: Performance of ME-IDS and other Benchmark IDS Frameworks with RGB Images Generated from All Features

Performance Evaluation: ME-IDS & Benchmark IDS Frameworks

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Stacked Ensemble [1]	85.85	87.47	84.75	85.29
Concatenation Ensemble [2]	95.31	96.07	94.78	95.20
Confidence Averaging [2]	97.72	97.96	97.50	97.69
VGG16-SA	98.72	98.79	98.63	98.70
VGG16-TPE	99.15	99.17	99.11	99.14
VGG16-RS	99.43	99.43	99.43	99.43
ME-IDS	99.72	99.74	99.68	99.71

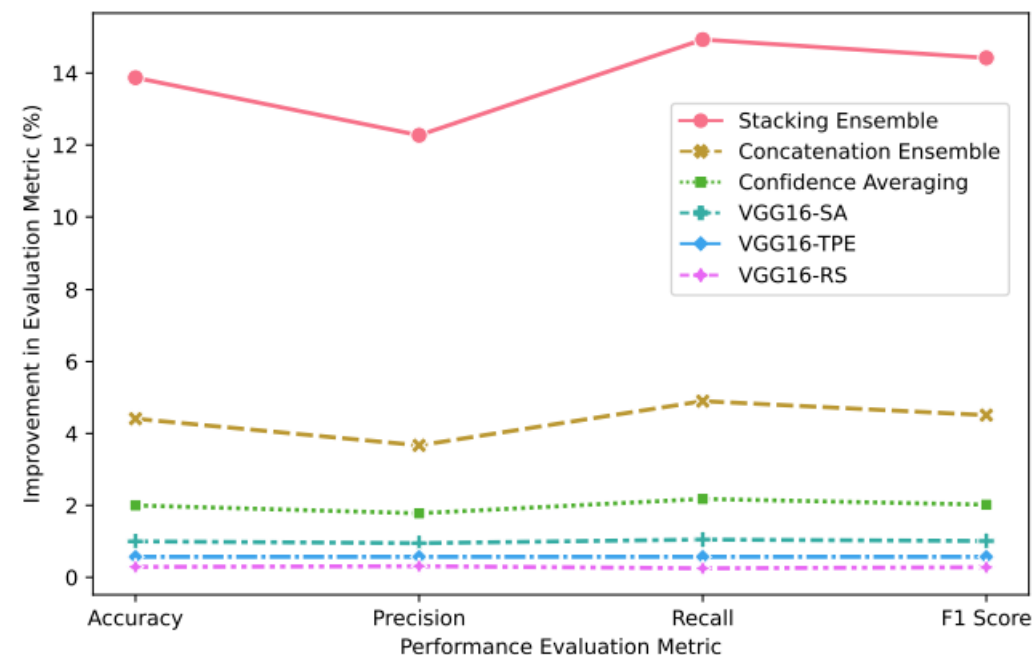


Figure: Performance of ME-IDS and other Benchmark IDS Frameworks with RGB Images Generated from Selected Features

Conclusion

- We proposed a novel Misclassified sample based Ensemble transfer learning framework for IDS (ME-IDS) which:
 - Encodes the categorical features using frequency encoding with minimal noise to mitigate the issue of tied frequency count.
 - Utilizes a filter-based feature selection method to choose the most relevant feature for classification.
 - Leverages the strength of a lightweight pre-trained TL model to create a robust IDS framework.
 - Employs a weighted ensemble learning stage with optimized weights based on the performance of the base classifiers.
- **Publication:**
 - A. Ghosh, and Q. Ye, “ME-IDS: An Ensemble Transfer Learning Framework Based on Misclassified Samples for Intrusion Detection Systems”, Globecom 2024 - IEEE Global Communications Conference (Accepted).

References

1. Ngamba Thockchom, Moirangthem Marjit Singh, and Utpal Nandi. A novel ensemble learning-based model for network intrusion detection. *Complex & Intelligent Systems*, pages 1–22, 2023.
2. Li Yang and Abdallah Shami. A transfer learning and optimized cnn based intrusion detection system for internet of vehicles. In *ICC 2022-IEEE International Conference on Communications*, pages 2774–2779. IEEE, 2022.
3. Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In *2015 military communications and information systems conference (MilCIS)*, pages 1–6. IEEE, 2015.
4. Zeeshan Ahmad, Adnan Shahid Khan, Kartinah Zen, and Farhan Ahmad. Ms-ads: Multistage spectrogram image-based anomaly detection system for iot security. *Transactions on Emerging Telecommunications Technologies*, page e4810, 2023.
5. Faisal Hussain, Syed Ghazanfar Abbas, Muhammad Husnain, Ubaid U Fayyaz, Farrukh Shahzad, and Ghalib A Shah. Iot dos and ddos attack detection using resnet. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6. IEEE, 2020.

Thank You!