Analyzing Firewall Log Classification Performance with Supervised Machine Learning and Deep Learning Algorithms

Sooji Kim

Abstract: Firewalls are the first line of defense for any organization's internet-facing services, and are thus a critical part of its security posture. By analyzing firewall logs, security practitioners can monitor and evaluate network traffic to make informed decisions about what firewall rules should be put in place. This study attempts to determine which machine learning model (SVM, DNN, and RF) is best for classifying internet traffic into one of four classes which correspond to firewall actions - 'Allow', 'Drop', 'Deny', and 'Reset-both', and what impact, if any, feature selection has on classification performance. The random forest algorithm was found to be the best performing supervised learning algorithm for firewall log data classification.

1. Introduction

Firewalls are fundamental components of modern network security architectures, acting as gatekeepers that monitor and control incoming and outgoing traffic based on predefined security rules. They serve as the first line of defense against unauthorized access, malicious activity, and other potential cyberattacks. As network environments grow increasingly complex, the volume of generated firewall logs becomes overwhelming, making manual inspection and rule-based approaches infeasible.

Machine learning techniques offer a promising solution for automating firewall log analysis. By framing the problem as a classification task, ML models can learn patterns in network traffic to accurately classify benign and malicious activities, or categorize them into actions such as 'Allow' or 'Deny'. This enables real-time decision support, anomaly detection, and adaptive security responses.

This study explores the application of the following supervised classification algorithms: Support Vector Machines (SVM), Random Forests (RF), and Deep Neural Networks (DNN), on the UCI Internet Firewall Data dataset, evaluating their performance in terms of accuracy, precision, recall, and F1 score. Additionally, the impact of feature selection on model efficiency and effectiveness is investigated to identify the most informative attributes for automated firewall decision-making.

2. Related Work

Automated classification of internet firewall data has emerged as a valuable approach in strengthening network security and detecting potential threats in real time. As the volume of log data continues to grow, the use of machine learning techniques has become increasingly common for parsing, interpreting, and classifying network events. Two major factors that influence the success of these techniques are the choice of classification algorithm and the quality of the feature selection process.

Ahmad et al. [1] explored this in the context of intrusion detection by comparing the performance of three popular machine learning models: Support Vector Machine (SVM), Random Forest (RF), and Extreme Learning Machine (ELM). Their experiments, conducted on the NSL-KDD benchmark dataset, revealed that ELM consistently achieved superior results across evaluation metrics, including accuracy, precision, and recall. This underscores the importance of selecting a suitable model when working with network security data.

In parallel, Ghaddar and Naoum-Sawaya [2] addressed the challenge of high-dimensional datasets by developing a feature selection technique integrated with an SVM classifier. Their method introduces an adjustable sparsity constraint during model training, allowing the classifier to focus only on the most relevant features. This approach not only reduced the number of input features but also preserved or improved classification performance across various domains, demonstrating the potential of embedded feature selection techniques for improving model efficiency.

Together, these studies highlight the value of combining thoughtful model selection with robust feature reduction strategies to optimize classification outcomes in cybersecurity contexts. This paper extends these ideas by applying similar principles to firewall data classification, using multiple algorithms and evaluating the effects of different feature selection methods on model performance.

3. Data

The firewall data was downloaded from the UCI Machine Learning Repository, and the data was collected from a university's internet firewall. The dataset contains 12 features and 65532 samples. The 'Action' feature is the class or label for each entry. This label represents the action that the firewall took for each connection request. The remaining features include the source port, destination port, number of packets sent, and other information about the network connection.

4. Methods

4.1. Data Preprocessing

Initial preprocessing included removing missing values and duplicate records to ensure data integrity. All features were scaled using the Min-Max normalization method to map each feature to a [0, 1] range. One-hot encoding was applied to transform the categorical class labels into integer values suitable for machine learning classifiers.

4.2. Feature Selection

Feature selection was used to reduce model complexity and improve classification performance. The chi-square method was used to select the most significant features to use for SVM and DNN since it is appropriate for large datasets with categorical features. For the Random Forest classifier, feature importance scores were derived from the trained model and ranked, and the top five features were retained for additional experiments.

4.3. Model Implementation

Three supervised learning models for were evaluated for multi-class classification of the firewall data: Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Network (DNN). Each model was trained and tested using two distinct feature sets: all available features and a reduced set of the top five most important features determined through feature selection. Training and evaluation were performed using an 80:20 split.

Support Vector Machine (SVM): The SVM classifier employed a radial basis function (RBF) kernel and the default hyperparameter of C=1.

Deep Neural Network (DNN): A feedforward neural network was constructed using Keras, with two hidden layers of 20 ReLU-activated neurons each and a softmax output layer. The model was trained using the Adam optimizer for 50 epochs with a batch size of 32.

Random Forest (RF): A 100-tree ensemble was used with default impurity splitting.

5. Results

This section presents the performance of the following classifications models: Support Vector Machine (SVM), Deep Neural Network (DNN), and Random Forest (RF), on both the full feature set and the reduced feature set obtained through feature selection. Evaluation metrics include accuracy, precision, recall, and F1-score, all computed using macro-averaging to accommodate the multi-class nature of the classification task.

5.1. Model Performance Using All Features

All three models demonstrated strong classification capabilities when trained on the full feature set. As shown in Table 1, the DNN and RF classifiers both achieved an accuracy of 0.9977, while the SVM model followed closely with an accuracy of 0.9783. Despite similar accuracy scores, the Random Forest classifier achieved the highest precision (0.8717), recall (0.7789), and F1-scores (0.7972), suggesting superior performance across all classes. In contrast, the SVM model yielded the lowest scores for these metrics, possibly due to the data distribution being relatively imbalanced across classes.

Table 1. Model performance using all features.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
SVM	0.9783	0.7216	0.7309	0.7259
DNN	0.9977	0.7475	0.7475	0.7475
RF	0.9977	0.8717	0.7789	0.7972

5.2. *Model Performance Using Selected Features*

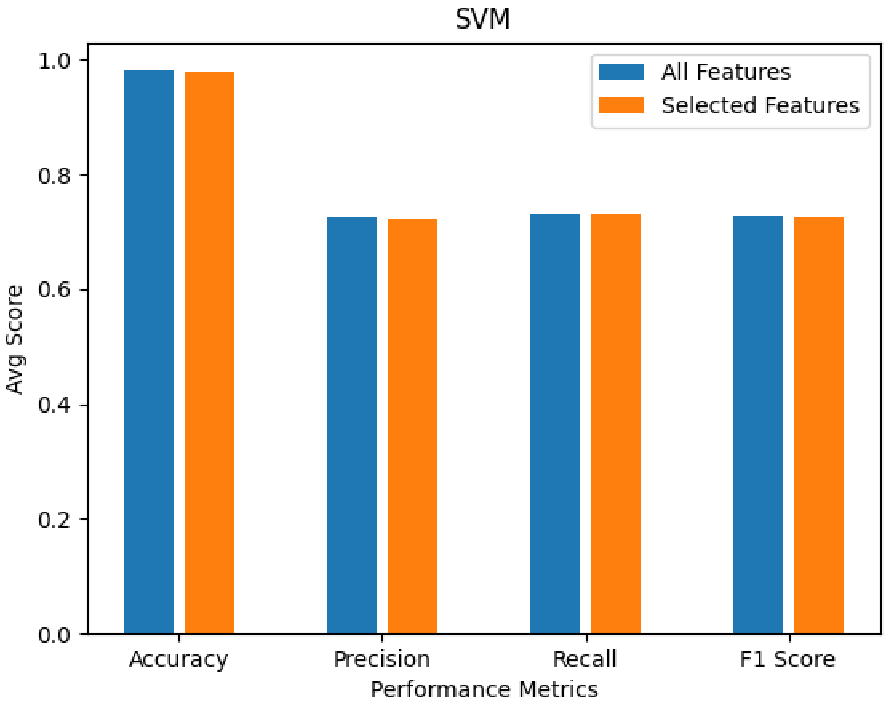
Training models on a reduced feature set resulted in performance improvements for both the Random Forest and SVM classifiers. Notably, the RF model achieved an accuracy of 0.9990, precision of 0.9435, recall of 0.9672, and an F1-score of 0.9546, the highest values observed in this study. These results demonstrate that dimensionality reduction can lead to performance gains by eliminating irrelevant or redundant features.

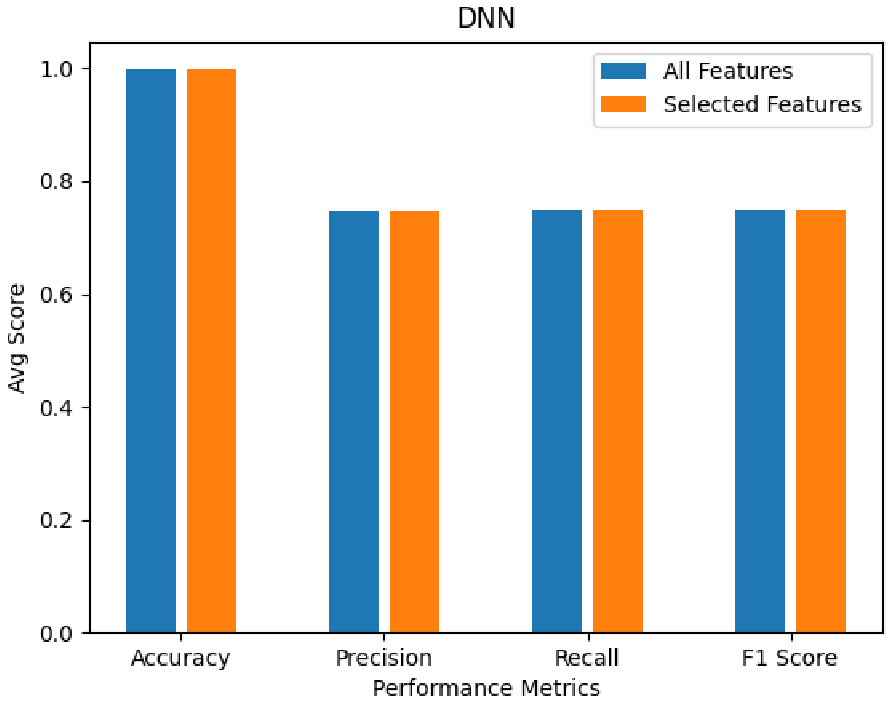
The SVM model also showed slight improvements across all metrics, with an increase in F1-score from 0.7259 to 0.7287. In contrast, the DNN model's performance remained unchanged when using the reduced feature set, suggesting its robustness to irrelevant features or a limited benefit from reduced dimensionality.

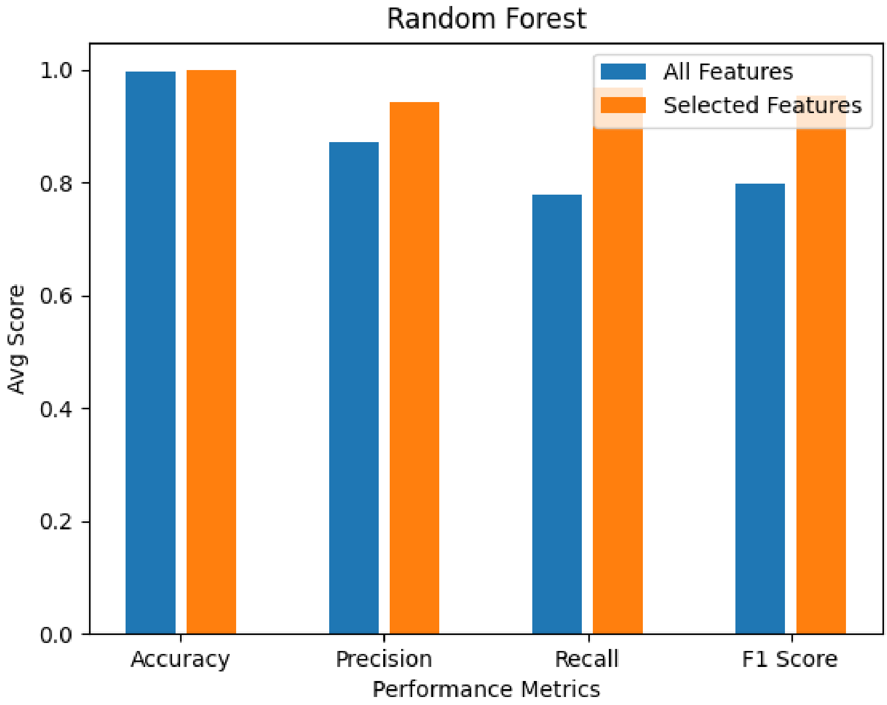
Table 2. Model performance using selected features.

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.9811	0.7262	0.7320	0.7287
DNN	0.9977	0.7475	0.7475	0.7475
RF	0.9990	0.9435	0.9672	0.9546

5.3. *Overall Comparison*







The above figures provide a view of all models on both full and reduced feature sets. Among the three models, Random Forest benefited the most from feature selection, significantly outperforming the others in all evaluation metrics. While the DNN model maintained high accuracy, its moderate F1-score suggests it may be less effective on underrepresented classes. The SVM classifier showed balanced yet moderate performance, with slight gains from dimensionality reduction.

6. Discussion

This study explored the use of supervised machine learning models to classify firewall action data, with a specific focus on the impact of feature selection on classification performance. Three popular models: Support Vector Machine (SVM), Deep Neural Network (DNN), and Random Forest (RF), were trained and evaluated using both the full set of features and a reduced set of the top five selected features.

The Random Forest classifier demonstrated the most significant improvement when trained on the reduced feature set, achieving the highest overall performance in terms of accuracy (0.9990), precision (0.9435), recall (0.9672), and F1-score (0.9546). This result highlights the importance of effective feature selection, particularly for tree-based models that can benefit from the removal of noisy or irrelevant features.

The SVM model also showed measurable improvement in all metrics after feature selection, though its overall performance remained lower than that of the RF and DNN models. This finding suggests that while SVM can benefit from dimensionality reduction, its performance may be more sensitive to the nature of the selected features.

Interestingly, the DNN model did not show improved performance when using the reduced feature sets. This could indicate that the neural network is inherently more robust to irrelevant features due to its deep architecture and capacity to learn hierarchical feature representations. However, it may also suggest that the reduced feature space lacked sufficient complexity to further enhance performance.

6.1. Conclusion

The results of this study demonstrate that feature selection can have a significant positive impact on the performance of classification models, particularly for Random Forest classifiers. While all models achieved high accuracy, only RF and SVM benefited from reduced input dimensionality. These findings suggest that feature selection is a valuable step in the machine learning pipeline for firewall data classification and can lead to more efficient and interpretable models without sacrificing accuracy.

6.2. Future Work

Future research could expand on this work in several directions. First, additional feature selection techniques such as mutual information, L1-based selection, or recursive feature elimination (RFE) could be explored and compared. Second, model performance could be further improved by employing more advanced hyperparameter optimization strategies.

Another area for future exploration is the application of ensemble techniques that combine multiple classifiers to improve robustness and accuracy. Additionally, the models could be evaluated using other real-world datasets or tested in a real-time deployment scenario to assess their practical effectiveness and scalability in dynamic network environments.

References

1. I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, “Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection,” *IEEE Access*, vol. 6, pp. 33789–33795, 2018. doi: 10.1109/ACCESS.2018.2841987.
2. B. Ghaddar and J. Naoum-Sawaya, “High Dimensional Data Classification and Feature Selection using Support Vector Machines,” *European Journal of Operational Research*, vol. 265, 2017. doi: 10.1016/j.ejor.2017.08.040.
3. UCI Machine Learning Repository, “Internet Firewall Data Set,” Available at: https://archive.ics.uci.edu/ml/datasets/Internet+Firewall+Data. Accessed April 2025.