JSL, 8(6), JUNE 2023: 43 - 51

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

OPTIMIZING CYBERSECURITY INTRUSION DETECTION THROUGH INTELLIGENT FEATURE SELECTION AND SUPPORT VECTOR MACHINE-BASED CLASSIFICATION

Adewar Swapnil Prabhakar

Research Scholar, Dept. of Computer Science and Engineering, Sabarmati University, Ahmedabad, Gujarat

Dr. Amit Kishor

Research Supervisor, Dept. of Computer Science and Engineering, Sabarmati University, Ahmedabad, Gujarat

ABSTRACT

The increasing sophistication and frequency of cyberattacks have created a strong need for accurate and efficient intrusion detection systems capable of identifying malicious network activities. Machine learning provides an effective approach for developing intelligent intrusion detection systems by learning patterns associated with normal and abnormal network behaviour. However, high-dimensional network datasets may contain redundant, irrelevant, and correlated features that increase computational complexity and reduce classification efficiency. Intelligent feature selection can address this problem by identifying the most informative characteristics of network traffic before classification. Support Vector Machine (SVM) is particularly suitable for intrusion detection because of its ability to construct effective decision boundaries in complex feature spaces. This paper examines the integration of intelligent feature selection with SVM-based classification for cybersecurity intrusion detection. It discusses feature reduction, preprocessing, classification performance, and practical deployment challenges. The paper argues that combining carefully selected network features with SVM can improve computational efficiency, reduce unnecessary complexity, and support reliable detection while maintaining strong generalization capabilities.

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

Keywords: Cybersecurity, Intrusion Detection System, Machine Learning, Feature Selection, Support Vector Machine, Network Security, Classification, Anomaly Detection

I. INTRODUCTION

The rapid expansion of interconnected systems, cloud computing, Internet of Things devices, mobile networks, and digital services has significantly increased the cybersecurity attack surface. Conventional security mechanisms such as firewalls and signature-based detection systems remain important, but they may have difficulty identifying previously unseen or rapidly changing attack patterns. Intrusion Detection Systems (IDS) therefore play an important role by monitoring network or system activities and identifying behaviour that may indicate security violations. Modern IDS research increasingly incorporates machine learning because algorithms can learn patterns from network traffic and distinguish normal behaviour from malicious activities. Recent surveys indicate that machine learning and deep learning are being widely investigated to improve IDS accuracy, reduce false alarms, and address increasingly complex attacks.

Machine-learning-based intrusion detection, however, depends heavily on the quality and structure of the input data. Contemporary network datasets may contain numerous traffic characteristics, including packet statistics, protocol information, timing characteristics, flow information, and connection-level attributes. Many of these variables can be redundant or weakly related to the target class. High-dimensional input spaces can increase computational requirements and make classification more difficult. Research using UNSW-NB15 has specifically shown the importance of feature reduction because the performance of machine-learning IDS can decrease in high-dimensional spaces and imbalanced datasets. Feature selection is therefore not merely a preprocessing step; it can become a central component of an efficient IDS architecture.

Support Vector Machine is a powerful supervised learning algorithm that seeks an optimal decision boundary between classes. Through the use of kernel functions, SVM can handle nonlinear relationships between network features and attack classes. It has been extensively examined in intrusion detection research because it can perform well in complex classification problems and can generalize effectively when the feature space is appropriately prepared. Nevertheless, SVM performance can be affected by unnecessary features, class imbalance, inappropriate parameter selection, and computational costs associated with large

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

datasets. Combining SVM with intelligent feature selection can address some of these limitations by reducing the dimensionality of the learning problem before classification.

The development of an effective feature-selection-based SVM intrusion detection framework consequently requires several interconnected stages. These include data collection, preprocessing, categorical encoding, normalization, feature ranking or selection, model training, parameter optimization, and performance evaluation. Benchmark datasets such as UNSW-NB15 are useful because they contain modern normal and malicious traffic and include multiple attack categories. The official UNSW dataset contains 2,540,044 records, 49 generated features plus class information, and nine major attack categories. This paper examines the role of intelligent feature selection and SVM classification in optimizing cybersecurity intrusion detection, focusing on the relationship between dimensionality reduction, classification performance, computational efficiency, and practical cybersecurity requirements.

II. INTELLIGENT FEATURE SELECTION FOR NETWORK INTRUSION DETECTION

Feature selection is an essential stage in machine-learning-based intrusion detection because network traffic datasets can contain a large number of variables with different levels of relevance. The objective is to identify a subset of features that provides sufficient information for distinguishing normal traffic from malicious activity. Reducing irrelevant and redundant features can decrease computational requirements, shorten training time, improve model interpretability, and potentially increase classification performance. In intrusion detection, this is particularly important because network monitoring can generate large amounts of information continuously.

Feature-selection approaches are generally divided into filter, wrapper, and embedded methods. Filter methods evaluate features using statistical or information-based criteria without directly depending on a particular classifier. Correlation analysis, mutual information, information gain, and statistical tests are common examples. Wrapper methods evaluate subsets according to classifier performance, while embedded methods perform feature selection during model training. Each approach involves trade-offs between computational cost and predictive performance. For large network datasets, computationally efficient filter and embedded methods can be especially useful.

E-mail: ***editor.jslsci888@gmail.com***

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

Correlation-based selection can remove features that provide information similar to other variables. If two features are strongly correlated, retaining both may increase dimensionality without providing substantial additional information. Mutual information can identify features that have strong dependencies with the target class, including nonlinear relationships. Tree-based importance measures can also provide useful rankings. Research on UNSW-NB15 has demonstrated the use of feature reduction before applying SVM, k-nearest neighbours, logistic regression, artificial neural networks, and decision trees. Such evidence supports the argument that feature reduction can be integrated into a broader machine-learning IDS pipeline.

The quality of feature selection is particularly important in multi-class intrusion detection. A feature that is highly useful for detecting denial-of-service traffic may not be equally useful for identifying reconnaissance or exploitation attacks. Therefore, selection should consider both binary classification and attack-category classification. An effective method should retain features that capture diverse characteristics of malicious behaviour while eliminating variables that contribute little discriminatory information. Feature selection can also reduce the risk of overfitting by limiting the number of dimensions available to the classifier.

Class imbalance is another issue that must be considered. Some attack categories may occur much less frequently than normal traffic or dominant attack classes. If feature selection is performed without considering class distribution, the resulting feature subset may favour majority classes. Techniques such as stratified sampling, class weighting, resampling, or synthetic minority oversampling can be considered alongside feature selection. However, data balancing should be performed carefully to avoid information leakage between training and testing datasets.

Overall, intelligent feature selection should be viewed as a strategic component of IDS optimization rather than simply a method for reducing the number of variables. A well-designed selection process can simplify the learning problem, improve computational efficiency, and support more interpretable models. For SVM-based intrusion detection, this is particularly important because reducing the feature space can make the optimization process more manageable. Feature selection should therefore be performed using training data only, validated through appropriate cross-validation, and evaluated according to its effect on detection accuracy, precision, recall, F1-score, false-positive rate, and computational cost.

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

III. SVM-BASED CLASSIFICATION FOR CYBERSECURITY INTRUSION DETECTION

Support Vector Machine is a supervised classification technique based on finding an optimal separating hyperplane between classes. For intrusion detection, the algorithm can learn the boundary between normal and malicious traffic from labelled network data. The maximum-margin principle allows SVM to seek a boundary that separates classes while maximizing the distance from the closest training observations. This characteristic can contribute to generalization when the data are appropriately scaled and representative of operational conditions.

A major advantage of SVM is its ability to use kernel functions. Linear kernels can be effective when classes are approximately separable in the selected feature space, whereas nonlinear kernels such as the radial basis function can model more complex relationships. In network intrusion detection, attack patterns may not be linearly separable because malicious behaviour can arise from combinations of traffic characteristics. Kernel-based SVM can therefore provide flexibility for modelling nonlinear relationships without requiring extensive manual feature construction.

The effectiveness of SVM depends substantially on preprocessing. Network traffic features may have very different numerical ranges, and variables with large scales can disproportionately influence the optimization process. Standardization or normalization can therefore be applied before model training. Categorical variables such as protocol or service may also need appropriate encoding. Missing values, duplicates, extreme values, and invalid records should be addressed during preprocessing. These steps are essential because classification performance depends not only on the algorithm but also on the quality of the learning data.

Hyperparameter selection is another important consideration. Parameters such as the regularization coefficient and kernel parameters determine how the SVM balances classification errors and model complexity. Poor parameter choices can lead to underfitting or overfitting. Grid search, randomized search, or other optimization methods can be used with cross-validation to identify suitable configurations. The optimization process should be performed only within the training data so that the test set remains an unbiased estimate of final performance.

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

SVM can be applied to both binary and multi-class intrusion detection. Binary classification distinguishes normal traffic from attack traffic and can serve as the first stage of a hierarchical IDS. Multi-class classification attempts to distinguish among different attack categories. A two-stage architecture can first determine whether traffic is suspicious and subsequently classify the attack type. This approach can be useful when the primary objective is rapid detection followed by detailed categorization. The suitability of such architectures depends on computational resources and the operational requirements of the security environment.

Recent empirical research confirms that SVM remains an important algorithm in intrusion detection studies. A 2024 study using UNSW-NB15 compared SVM with logistic regression, decision tree, and random forest after exploratory analysis and feature selection. Other research has also combined feature reduction with SVM and alternative classifiers, demonstrating the continuing relevance of classical machine-learning algorithms for network intrusion detection. Thus, SVM remains a valuable component of cybersecurity research, particularly when combined with appropriate feature engineering and dimensionality reduction.

IV. OPTIMIZATION, EVALUATION, AND PRACTICAL CYBERSECURITY DEPLOYMENT

An optimized SVM intrusion detection framework should follow a structured pipeline beginning with reliable network traffic data. Benchmark datasets such as UNSW-NB15 and CICIDS2017 can provide standardized environments for experimentation. UNSW-NB15 contains realistic normal activity and synthetic contemporary attacks, while CICIDS2017 contains benign traffic and several common attack scenarios. Using more than one dataset can strengthen evaluation by demonstrating whether the model generalizes beyond a single traffic environment.

Preprocessing should precede feature selection and classification. Duplicate records, missing values, inconsistent categorical values, and irrelevant identifiers should be examined carefully. Numerical variables should be scaled when required by SVM, while categorical variables should be appropriately encoded. Data should then be divided into training, validation, and testing sets, preferably using stratification where class distributions are highly uneven. Importantly, feature selection and preprocessing transformations should be fitted

E-mail: ***editor.jslsci888@gmail.com***

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

using training data only to prevent information leakage.

The feature-selection stage can combine multiple criteria to produce a robust subset. For example, highly correlated variables can first be removed, followed by mutual-information ranking and classifier-based selection. An alternative approach can use recursive feature elimination with SVM. The resulting subsets can be compared to identify the smallest feature set that maintains acceptable detection performance. Such optimization is important because the objective should not simply be maximum accuracy; it should also consider model complexity, latency, false alarms, and resource consumption.

Performance evaluation should use multiple metrics. Accuracy provides an overall measure but can be misleading when classes are imbalanced. Precision indicates how many detected attacks are actually malicious, while recall measures the proportion of actual attacks successfully detected. F1-score balances precision and recall. Confusion matrices provide class-level information, while false-positive rate is particularly important in operational IDS because excessive false alarms can overwhelm security analysts. Detection latency, training time, memory consumption, and inference cost should also be considered when the intended application involves real-time monitoring.

A practical IDS should also account for evolving attack patterns. Cyber threats change over time, meaning a model trained on historical traffic may gradually become less effective. Recent research emphasizes concept drift and feature dynamics as important challenges for machine-learning IDS, highlighting the need for adaptive and dynamic detection approaches. An SVM model can therefore be periodically retrained using updated traffic data, while feature-selection procedures can be repeated to identify changes in the relevance of network characteristics. Monitoring model performance after deployment is essential.

Finally, an optimized SVM-based IDS should be integrated with broader cybersecurity operations. The model can generate alerts that are passed to security information and event management platforms or security operations centres for investigation. Automated responses should be carefully controlled because false positives can disrupt legitimate activities. Explainability can also support analyst confidence by showing which selected features contributed most strongly to classification. The strongest practical architecture is therefore not an isolated classifier but a monitored cybersecurity component supported by data governance, model validation, human analysis, continuous retraining, and incident-response

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

procedures.

V. CONCLUSION

Machine-learning-based intrusion detection has become an important approach for strengthening cybersecurity against increasingly sophisticated network threats. However, the effectiveness of an IDS depends heavily on the quality and dimensionality of its input data. Large network datasets may contain redundant, irrelevant, and correlated variables that increase computational requirements and reduce model efficiency. Intelligent feature selection provides a systematic mechanism for addressing these challenges by retaining the most informative characteristics of network behaviour.

Support Vector Machine offers a strong classification framework for intrusion detection because of its maximum-margin learning principle and ability to model nonlinear relationships through kernel functions. When combined with appropriate preprocessing, feature selection, parameter optimization, and class-imbalance strategies, SVM can provide an effective approach for distinguishing normal and malicious network activities. Research using datasets such as UNSW-NB15 demonstrates the continuing relevance of classical machine-learning algorithms, including SVM, for cybersecurity classification.

Nevertheless, high classification accuracy alone is insufficient for real-world cybersecurity deployment. An operational IDS must minimize false positives, detect attacks across different categories, operate with acceptable latency, and remain effective as network behaviour changes. Dataset limitations, concept drift, class imbalance, adversarial behaviour, and changes in attack strategies can all reduce model effectiveness. Therefore, continuous evaluation, retraining, feature monitoring, and validation are essential components of a sustainable intrusion detection architecture.

In conclusion, the integration of intelligent feature selection with SVM-based classification provides a promising framework for optimizing cybersecurity intrusion detection. Feature selection reduces unnecessary complexity, while SVM provides a robust classification mechanism for distinguishing network behaviours. Future systems should combine dynamic feature selection, adaptive learning, explainable AI, multi-dataset evaluation, and real-time monitoring to address evolving cyber threats. The most effective IDS will ultimately be one that balances detection accuracy with efficiency, interpretability, adaptability, and operational usability.

Online ISSN: 2456-1495
CODEN (USA): JSLOBJ

REFERENCES

Aljawarneh, S., Aldwairi, M., &Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152–160.

Chawla, N. V., Bowyer, K. W., Hall, L. O., &Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

Gao, X., Shan, C., Hu, C., Niu, Z., & Liu, Z. (2019). An adaptive ensemble machine learning model for intrusion detection. IEEE Access, 7, 82512–82521.

Khraisat, A., Gondal, I., Vamplew, P., &Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity, 2, 20.

Moustafa, N., & Slay, J. (2015). 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) (pp. 1–6). IEEE.

Moustafa, N., & Slay, J. (2016). The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective, 25(1–3), 18–31.

Niyaz, Q., Sun, W., & Javaid, A. Y. (2016). A deep learning based DDoS detection system. Computers & Security, 66, 129–145.

Pfahringer, B. (2024). Unveiling machine learning strategies and considerations in intrusion detection systems: A comprehensive survey. Frontiers in Computer Science, 6, 1387354.

Ullah, I., & Mahmoud, Q. H. (2020). A hybrid model for intrusion detection in Internet of Things. Journal of Cyber Security Technology, 4(2), 64–93.

Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. Computer Networks, 174, 107247.