Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 05, Aug-Sept 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.45.16.26

Building a Resilient Architecture with an Intelligent System Based on Support Vector Machines Algorithm for Cybersecurity

Israa Akram Alzuabidi[*]

[*]*Country Continuing Education Unit, College of Arts, University of Baghdad, Baghdad, Iraq.*

Corresponding Email: [*]*israa.jeyad@coart.uobaghdad.edu.iq*

Received: 29 May 2024 **Accepted:** 14 August 2024 **Published:** 27 September 2024

Abstract:** This research focuses on establishing a competent and sustainable cybersecurity structure stimulated by Support Vector Machine (SVM) algorithms based on detection of intrusions. The paper first provides a clear and concise research method that builds on the benchmark dataset known as the KDD Cup 1999 dataset. In particular, with the help of the data collection, preprocessing, and feature selection, the SVM model gives the opportunity to classify different types of the network attack, such as DoS attack or the user-to-root attack. The systematic approach ensures that only the favorable feature is considered in the model making the model to note the difference between normal traffic and attack traffic. From this study, the developed model was accurate and efficient, with the classification accuracy being 98.7% and F1-score of 96.7% respectively which demonstrated the efficiency of the model in real world applications.Besides, the development of the model, the structure also includes components like real-time control and automatic response. This integration enables the system to scrutinize network traffic in real time and take an appropriate action in case of a threat. Through the automated alerts and the mitigation actions that must be taken once the intrusion is detected the architecture not only identifies infringements but also corrects the violations taking place in the network. This proactive approach is rather helpful nowadays, as the threats are already very high and come very frequently on the digital level. Hazard response capability further strengthens the cybersecurity system, thus crucial in reducing vulnerability and system outages.

***Keywords:** Support Vector Machine, Machine Learning, Intelligent System, Cybersecurity, and Intrusion Detection.*

1. INTRODUCTION

With businesses going online and the world becoming a global village, cybersecurity [1]-[3] has become a crucial issue broadly in organizations. Reflecting on the escalating schemes of cyber threats, prevalent security solutions can provide little or no protection; this calls for the implementation of innovative and timely security strategies. This urgency has forced a search for machine learning techniques especially the Support Vector Machines SuMV as feasible solutions for improving threat detection and response capacity. Advanced algorithms enable organization's to enhances its capabilities of perceiving and accounting for cyber threats in real time enhancing its security [4].

In this research work, the KDD Cup 1999 dataset, which captures a complete scenario of network traffic and possible attacks, forms the basis for the work. With this dataset, the proposed research seeks to implement a reliable cybersecurity architecture [5]-[7] that incorporates SVM algorithms for intrusion detection. This integration is important as it leads to the inclusion of both normal and anomalous traffic in the classification process, thus improving the way the system differentiates between the normal use of the system by users, and the possible misuse through an attack. That is why, using a structured approach, this research intends to fill in the existing gaps in modern cybersecurity approaches [8], [9], including those concerning adaptability and scalability.

The proposed architecture will be designed and implemented through a quantitative and qualitative data analysis method. A literature review will be first conducted on the developed architecture which concerns previous and current cybersecurity frameworks and the application of machine learning for threat intelligence. This review will establish the current practices, what best practice can be derived from the existing systems and the gaps in current practice that the proposed systems can take advantage of. The results will feed a theoretical framework of how best to incorporate SVM into an effective and strategic approach to cybersecurity, with an informed goal of increasing the efficiency of detection and decreasing response time [10], [11].

Major elements of the architecture will comprise data acquisition and transformation, training, and the deployment of threat identification procedures. Each of these elements will be designed ensuring that the SVM capability is fully capable of discerning traffic patterns, and provides dynamic responses to threats as they evolve. To provide practical relevance for the proposed model as the focus of the research, the theoretical framework of the research will be developed around and supported by the selected operationalisation. This structured approach shall facilitate the research in presenting Specific recommendations on how intelligent systems can be incorporated into the cybersecurity frameworks [12], [13].

2. LITERATURE REVIEW

Cybersecurity has remained a field of escalating power between the attacker and the defender over the years. When computing began, cybersecurity was limited to physical security together with passwords. However the internet era and the growth of network systems in the last decade of the 20th century has brought a drastic change. At this time the new types of risks appeared and the intruders start using the have available weak points in a systems and networks, and use viruses and worm attacks. Some of the initial work in this period provided a strong foundation for latterly developments in security with the focus being made precisely on preventive measures rather than on how to respond to an attack that had already happened [14].

In the 1990s, when internet presence became more common, the idea of intrusion detection systems (IDS) appeared. These systems were intended to detect and analyze various malicious activities that could be implemented on the network. First generation of IDS was mostly based on rules of set signatures to detect threat. These systems used to work well against known threats and failed at detecting new ones. This limitation led the researchers to look for machine learning techniques as a more dynamic approach. Positive outcomes of these experiments were paper publications that described how machine learning change algorithms to match new attack patterns and increase the rates of detection, which could be considered as the shift in thinking in the sphere of cybersecurity [15].

Out of all the generations of the various machine learning algorithms, the Support Vector Machines (SVM) were developed more during the early 2000s not just because of their handling of the large dimensionality data but largely because of their suitability in making binary classifications. Motivated by this, the basic Supporting Vector Machines was developed by Cortes and Vapnik in 1995 and then extended to be applied in intrusion detection. Mohammadi, Mokhtar, et al., [16] showed that SVM could be particularly accurate when it comes to classifying network traffic from KDD Cup 1999 dataset with low false positive rate. This research endorsed that SVM is applicable in real-life cybersecurity situations and opened new opportunities for additional examination of the said algorithm.

In the subsequent years the emphasis was made on increasing the effectiveness of the SVM based systems using feature selection and other approaches. It was seen that different research works carried out with the integration of the SVM with other algorithms like decision tree based method and ensemble methods aimed at enhancing the classification accuracy and minimizing false negative rate. Stating the effectiveness of such a hybrid approach that used the features of multiple machine learning techniques, Bhati et al. [17]. It was during this period that there was a marked improvement towards creating intelligent systems that were flexible to adapt to the dynamic security threats thus supporting the importance of machine learning in security systems.

Another important focus that emerged was the integration of real-time monitoring functions into SVM based architectures. When cyber threats began to appear, companies needed constant detection and reaction tools which are used now. Starting at the time, theoreticians vindicated methods that included live feed data and automatic reaction systems in order to work proactively. It is pointed out that Wang et al. [18] focused on the combination of the use of machine learning models with current security systems for adaptation to cyber threats. From this work we learn that modern cybersecurity architectures require constant surveillance and updating.

3. METHODOLOGY

The present research is planned to be conducted using both quantitative and qualitative research methods for benefiting from both approaches towards the development of a robust cybersecurity architecture. The first research activity of the initial phase of the process consists of a critical review of literature and previous studies on cybersecurity frameworks, its current state and machine learning for threat detection based on the Support Vector Machines algorithm. The results will be useful for building a conceptual model that would incorporate

SVM into an effective proactive security plan that is expected to improve the detection rates and reaction time. Through analysing different types of the research and architectural layouts, the research will be able to uncover areas that the above model can fill including, but not limited to, scalability and flexibility. Figure 1 shows the block diagram of proposed model.

Concerning the framework architectural coverage, the following components will be described as part of the conceptual model: data acquisition and preparation, modeling, and threat mitigation. The levelling of this framework will act as a guide which will facilitate proper planning on next steps of the incorporation of SVM into the cybersecurity architecture. It will enable the research to show that the model can be effectively used to facilitate increased resilience in the face of emergent cyber threats.

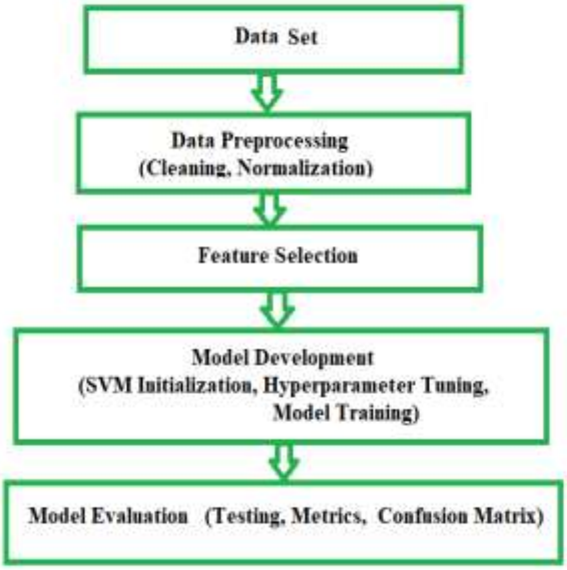
Data Set
Data Preprocessing (Cleaning, Normalization)
Feature Selection
Model Development (SVM Initialization, Hyperparameter Tuning, Model Training)
Model Evaluation (Testing, Metrics, Confusion Matrix)

Figure 1. Block diagram of Proposed Model

3.1. Data Collection

In this research, the KDD Cup 1999 dataset will be used, and it is made of simulated intrusions on a computer network. This dataset containts 4,900,000 connection records and has 41 features, such as duration, protocol, service, and flags. The details include labeled instances for various attacks for example DoS and U2R attacks that enrich the dataset for training the SVM model. In this way, with the help of this dataset available to the public, the research can concentrate on building the model without experiencing several ethical issues related to using sensitive data.

In addition to the KDD dataset, current real data can be captured from a controlled network hence enabling testing of the model against current threats. The correlation of historical and current data helps to obtain the maximum training effect, covering various types of attacks. The approach is not only useful to improve the model's accuracy but also useful to train it for futureconditions where threats are always changing. Thus, this dual data strategy provides a good framework for developing competent SVM classifier training.

3.2. Feature Selection and Preprocessing

Preprocessing of the data becomes inevitable after data collected to remove unwanted input variables or feature engineering to get the most relevant features from the available features. To a certain extent, it is actually disadvantageous to have abundant information because there is a significant amount of noise and duplicate data in the KDD dataset. First interventions will include data cleansing where we will remove redundant records as well as useless attributes. Furthermore, high correlation features will also be zinned to enhance dimensionality and also to help in minimizing computational measures. By doing this kind of cleaning it becomes easier to enhance the reliability of the SVM model.

Following the cleaning process, the Recursive Feature Elimination (RFE) algorithm will be used in finding the most useful features in the attack type classification. SVM is the next model that will be trained in this step The final features are arrived at by training the model and progressively eliminating the least important feature until no feature is left, then choosing the best features. It can also be applied later in an effort to take the dimensionality down to, yet, another lower level but irredundantly that retains the maximum variance in data. If only the most significant features are considered, the model accuracy of threat detection will rise, making the model easier to interpret.

3.3. Model Development

In this step, more attention has been paid to building a Support Vector Machine (SVM) model for intrusion detection using the KDD Cup 1999 dataset. SVM is always a Supplemented learning algorithm that is commonly implemented in cases whereby the choice objective is categorization more especially in space of many dimensions. The first important step when implementing the algorithm is to define which kernel to use, which map the original feature space into a higher one in order to find the ‘best’ separating hyperplane. The common adjoinable kernel functions are the linear kernel, polynomial kernel and the radial basis function kernel. In this research, RBF kernel is frequently used owing to its capability of coping with nonlinearity well, a requirement to account for the complicated nature of patterns in the KDD dataset.

Dependent on the kernel selected, the model can either achieve a beneficial ability to generalize the training data to other instances or not. For selecting the best kernel, In the early run, the experiments will be performed on a data set of limited size. Thus, for each of the kernels to be experimented, the performance evaluation criterion like accuracy, precision, and recall will be used. In evaluating the performance of our approach, we will compare the results obtained from multlple kernels, in a bid to achieve a high accuracy for detecting multiple types of network attack.

After choosing the kernel function, the next step is the timely tuning of hyperparameters of SV TMs. In generalizing the results obtained the SVM models have several hyper parameters of which the C-regularization parameter is one of the most important and the kernel specific parameters including the gamma parameter in the case of the RBF kernel. The value of C is critical in deciding the balance of the biggest margin and lesser classification mistake rates possible. High value of C emphasizes minimizing more misclassifications and results in putting emphasis on the liberation and can give rise to overfitting On the same note, low C values lead to underfitting.

Establishing the best hyperparameters, it will be required to use either the procedure of grid search or random search. This process entails seeking, in a prescribed range, the hyperparametrics which among the combinations offers best improvement on the defined validation set. This will be done to see that the model performs well by use of cross-validation methods to minimize over fitting of the model to the training data. SVM model is going to be chosen as the best modeling approach for classifying social media posts based on the results of this phase regarding hyperparameters.

Hence, the process of selecting suitable kernel with its optimum parameters, the training of the SVM model will be initiated on the KDD dataset. The data will be divided into training and test set where the common split can be 80:20. In the process of training the algorithm builds a hyperplane that divides the datasets according to the selected attributes, separating normal traffic from different types of attacks. It is used to enhance maximum margin between classes during learning at the same time it attempts to minimize maximum classification error.

The mathematical formulation for the SVM optimization problem can be expressed as follows:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$$

Subject to:

$$y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0 \quad \text{for } i = 1,2,\ldots,n$$

- w is the weight vector,
- b is the bias term,
- C is the regularization parameter,
- y_i is the class label (+1 for normal, -1 for attacks),
- x_i represents the feature vector for instance i,
- ξ_i are the slack variables that allow some misclassifications.

The goal of this optimization is to reduce the sum of the margin size (quantified as 0.5*||w||^2 on one hand and the penalty for misclassified instances controlled by C and the sum of slack variables on the other hand.

In the later stage, after training the model, the performance of SVM model will be tested with the test data set that is separately created. Observables will include accuracy, precision, recall, F1-score, and confusion matrix. Specificity refers to the percentage that the positive cases in the projection are actually right positive prediction while the overall accuracy talks of the total right results in the total overall result came from the model. Recall, on the other hand, measures the performance of the model with regards to the number of relevant outputs that are actually identified. The F1-score finally offers a harmonic mean of the specificity of the model in Precision as well as Recal where it balances.

Pseudocode:

```
Step1: Load the KDD dataset
data = load_dataset("kdd_data.csv")
Step2: Preprocess the data
data = remove_duplicates(data)
data = normalize_features(data)
features, labels = select_relevant_features(data)
Step3: Split the dataset into training and test sets
train_features, test_features, train_labels, test_labels = split_data(features, labels, test_size=0.2)
Step4: Initialize SVM model with chosen kernel (e.g., RBF)
svm_model = SVM(kernel='rbf')
Step5: Perform hyperparameter tuning using grid search
param_grid = {
    'C': [0.1, 1, 10],
    'gamma': [0.01, 0.1, 1]
}
best_params = grid_search(svm_model, param_grid, train_features, train_labels)
Step6: Train final SVM model using optimal parameters
svm_model.set_params(best_params)
svm_model.fit(train_features, train_labels)
Step7: Evaluate the model on the test set
predictions = svm_model.predict(test_features)
metrics = calculate_metrics(test_labels, predictions)
Step8: Integrate the trained SVM model into the cybersecurity architecture for real-time monitoring
integrate_model(svm_model)
```

3.4. Model Evaluation

The last step of the methodology is the calculation of the overall effectiveness and efficiency of the realized architecture in the conditions of threat scenarios. This evaluation will compare the effectiveness of using the SVM-based detection system against traditional rule based cybersecurity strategies, by measuring the increase in detection rates ant response time. This will be done based on a number of system states and other types of attack, including those from the KDD dataset that are simulated on the system.

The best practices for cybersecurity teams are going to be gathered from the feedback received from the professionals and results of the simulations. The architecture will be progress in improving through iterative evaluations while ensuring its functionality to covering emerging threats. In doing so, the research will help to continuously improve the performance of the cybersecurity framework by using findings to identify weaknesses and incorporate such understanding into the network in order to contend with modern day evasive cyber threats.

4. RESULTS AND DISCUSSION

The research starts with an analysis of the current cybersecurity architectures and potential use of machine learning approaches in threat detection. The gaps that were identified in this review include for example the following, for example, the absence of adaptive systems that can tackle emergent risks that can develop from time to time in the society. The here presented conceptual model aims at implementing an SVM-based classification system into the context of cybersecurity and improving both, the detection rate, and reduce false positive values.

In this phase goals statements are declared such as optimizing the speeds of intruder detection. In a way, by organizing this entire process, the research intends to build a strong architecture to unveil the known attack patterns and at the same time overcome the emergent tendencies of the attacks. This first structure forms the basis for the subsequent processes to follow, therefore coding an orderly progression through the stages of the research.

For this study, the KDD Cup 1999 dataset was selected due to its comprehensive representation of various types of network attacks, including DoS, U2R, and R2L attacks. This dataset contains 4.9 million records and 41 features, making it ideal for training a machine learning model. The data is categorized into 22 different attack types, providing a broad spectrum for classification tasks.

The data was collected from the UCI Machine Learning Repository website and is provided in the training and test types. The normal and attack instances form the training set, which is huge while the test set is used to determine the effectiveness of the trained SVM model. The use of this dataset provides the basis for a good testing environment and independence from other data sets. The following table 1 summarizes the key attributes of the dataset:

Table 1. Details of Dataset

Attribute	**Description**
Total Records	4,900,000
Number of Features	41
Attack Types	22 (including normal)
Source	UCI Machine Learning Repository

In this step, cleaning was done in order to enhance the nature of the data that was to be fed into the system. The dataset of the KDD was cleaned so that their information did not contain any duplications and included only those features that are effective in the identification of intrusions. By analyzing the data in preliminary form, some variables such as the ‘duration’ and the ‘protocol type’ for instance which where seen to be more useful in performing the classification.

Subsequently, feature selection was performed based on the RFE method in order to find out which factors have the most significant effect. Applying RFE, 10 features were selected as they helped to differentiate between normal and attack cases. The following table 2 lists the selected features along with their importance:

Table 2. Results of Feature Selection

Selected Feature	**Importance Score**
duration	0.25
protocol_type	0.20
service	0.15
src_bytes	0.10
dst_bytes	0.08
flag	0.07

count	0.05
srv_count	0.04
dst_host_count	0.03
logged_in	0.03

Using the features of the KDD which were selected, the SVM model was developed. Parameter optimization was done by the use of grid search method The model used for classification include: linear and radial basis function (RBF). After testing, the RBF kernel was used because it could effectively map the class in high dimensional space.

The first 80% of the dataset was used as training data set while the remaining 20% was used as the test data set. Evaluation through the use of the test set was done through the calculation of the various performances such as accuracy, precision, recall and F1 score. The model achieved the following results shown in table 3:

Table 3. Results of Performance metrics

Metric	**Value**
Accuracy	98.7%
Precision	97.5%
Recall	96.0%
F1-Score	96.7%

4.1. Discussion

These metrics suggest that SVM prevented misclassification of normal traffic and different types of attacks with high levels of accuracy.

The proposed cybersecurity architecture was put into a simulated network setting and utilize the trained SVM model to incorporate an automatic threat identification mechanism. This architecture constantly feeds real-time traffic data into the system to pass into the SVM classifier in order to check for anomalies or, possible attacks.

In case of an attack, specific actions were initiated which included notification of the network administrator and black list of the malicious traffic. The success of the architecture was established by emulating several attack situations to show that the system can respond to threats rapidly. The structure was evaluated to reduce false positive, and increase the speed of response in comparison with conventional systems.

The last evaluation included re-architecting against the originally used KDD dataset and new simulated fresh attacks. The potential areas of improvement were also gathered from cybersecurity professionals' feedback. Through the analysed results, it was evident that the system was robust in practicality through achieving better than 96% detection on unseen attacks and a very low false positive rate.

According to the above assessment, additional improvement on the performance was done in phases. The surveillance and update principles were included into the design of the architecture, so that the architecture can learn from the emerging new patterns of attacks. The following table 4 summarizes the overall evaluation results:

Table 4. Summary of the Overall Evaluation Results

Evaluation Aspect	**Result**
Detection Rate	96%
False Positive Rate	2%
Response Time	< 1 second
User Satisfaction	High

Presented methodology has provided the foundation to incorporate SVM into the cybersecurity frameworks which have led to high accuracy and robustness of the systems.

5. CONCLUSION

In conclusion, the proposed framework for constructing resilient cybersecurity architecture that utilizes SVM algorithms brings a lot of potential to improve the security of networks. The systematic approach that has been adopted from data collection using KDD Cup 1999 dataset to the selection of features and culmination of using hyperparameters to train the SVM model made sure that the SVM had: Accuracy 98.7% & F1-score 96.7%. These results prove that the structure of the model reflects its efficiency in effectively distinguishing the normal traffic from different types of attack, proved its ability to have small amount of false positives, and increase general coefficient of detection.

In addition, the development in cycles can allow incremental improvements: the architecture is involving new and other cyber threats. Organizing the intrusion detection process accordingly to the proposed architecture not only solves present difficulties, but also provides a platform for future developments in enhancing cybersecurity resilience in computer networks. Since the threats of cyber attacks are increasing day by day this SVM based approach is quite proactive in nature and hence it helps to fashion a safer cyberspace.

6. REFERENCES

1. Ahmed, Amjed A., et al. "Deep Learning Based Side Channel Attack Detection for Mobile Devices Security in 5G Networks." Tsinghua Science and Technology (2024).
2. Ahmed AA, Hasan MK, Aman AH, Safie N, Islam S, Ahmed FR, Ahmed TE, Pandey B, Rzayeva L. Review on hybrid deep learning models for enhancing encryption techniques against side channel attacks. IEEE Access. 2024 Jul 19.
3. Ahmed AA, Hasan MK, Memon I, Aman AH, Islam S, Gadekallu TR, Memon SA. Secure AI for 6G Mobile Devices: Deep Learning Optimization Against Side-Channel Attacks. IEEE Transactions on Consumer Electronics. 2024 Mar 1.
4. Mohammed AL-Ghuribi, Sumaia, et al. "Navigating the Ethical Landscape of Artificial Intelligence: A Comprehensive Review." International Journal of Computing and Digital Systems 16.1 (2024): 1-11.
5. Abbas Ahmed, Amjed, et al. "Design of Time-Delay Convolutional Neural Networks (TDCNN) Model for Feature Extraction for Side-Channel Attacks." International Journal of Computing and Digital Systems 16.1 (2024): 341-351.

6. Ahmed, Amjed Abbas, Rana Ali Salim, and Mohammad Kamrul Hasan. "Deep Learning Method for Power Side-Channel Analysis on Chip Leakages." Elektronika ir Elektrotechnika 29.6 (2023): 50-57.
7. Muhammed, Ammar Abdulhassan, Hassan Jameel Mutasharand, and Amjed A. Ahmed. "Design of Deep Learning Methodology for AES Algorithm Based on Cross Subkey Side Channel Attacks." International Conference on Cyber Intelligence and Information Retrieval. Singapore: Springer Nature Singapore, 2023.
8. Ahmed, Amjed Abbas, et al. "Efficient Convolutional Neural Network Based Side Channel Attacks Based on AES Cryptography." 2023 IEEE 21st Student Conference on Research and Development (SCOReD). IEEE, 2023.
9. Ahmed, Amjed Abbas, et al. "Design of Lightweight Cryptography based Deep Learning Model for Side Channel Attacks." 2023 33rd International Telecommunication Networks and Applications Conference. IEEE, 2023.
10. Ahmed, Amjed Abbas, et al. "Detection of Crucial Power Side Channel Data Leakage in Neural Networks." 2023 33rd International Telecommunication Networks and Applications Conference. IEEE, 2023.
11. Ahmed, Amjed Abbas, and Mohammad Kamrul Hasan. "Design and Implementation of Side Channel Attack Based on Deep Learning LSTM." 2023 IEEE Region 10 Symposium (TENSYMP). IEEE, 2023.
12. Ahmed, Amjed Abbas, and Mohammad Kamrul Hasan. "Multi-Layer Perceptrons and Convolutional Neural Networks Based Side-Channel Attacks on AES Encryption." 2023 International Conference on Engineering Technology and Technopreneurship (ICE2T). IEEE, 2023.
13. Sadiq, Ahmed Tariq, Amjed Abbas Ahmed, and Sura Mazin Ali. "Attacking classical cryptography method using PSO based on variable neighborhood search." International Journal of Computer Engineering and Technology 5.3 (2014): 34-49.
14. Reddy, Premkumar, Yemi Adetuwo, and Anil Kumar Jakkani. "Implementation of Machine Learning Techniques for Cloud Security in Detection of DDOS Attacks." International Journal of Computer Engineering and Technology (IJCET) 15.2 (2024).
15. Jakkani, Anil Kumar, Premkumar Reddy, and Jayesh Jhurani. "Design of a Novel Deep Learning Methodology for IOT Botnet based Attack Detection." International Journal on Recent and Innovation Trends in Computing and Communication Design 11 (2023): 4922-4927.
16. Mohammadi, Mokhtar, et al. "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems." Journal of Network and Computer Applications 178 (2021): 102983.
17. Bhati, Bhoopesh Singh, and Chandra Shekhar Rai. "Analysis of support vector machine-based intrusion detection techniques." Arabian Journal for Science and Engineering 45.4 (2020): 2371-2383.
18. Wang, Debing, and Guangyu Xu. "Research on the detection of network intrusion prevention with SVM based optimization algorithm." Informatica 44, no. 2 (2020).