iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80392

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Cyber-Attack Detection using Machine Learning

Prof. Afrin Sheikh[1], Pramod Pandurang Jadhav[2], Kanchan Jayvant Pande[3], Abdulsamad Shaikh[4], Yasin Salim Shaikh[5]
Computer Dept., K.J College of Engineering and Management, Research, Pune, India

Abstract: ***The growing use of digital communication and cloud computing has significantly increased the number of potential entry points for cyber attacks. Conventional intrusion detection systems based on static rules are ineffective against novel and dynamic threats. This paper proposes a machine learning-based system for cyber attack detection. It has the ability to distinguish between normal and malicious network traffic by discovering underlying patterns in data. The system comprises tasks of data preparation and analysis, feature extraction, and the use of a Support Vector Machine (SVM) classifier to quickly and accurately identify attacks. The experimental results demonstrate the effectiveness of this method in improving accuracy and minimizing false positives, and its ability to learn about novel attack behaviors. This paper demonstrates the potential of intelligent detection techniques to improve contemporary cybersecurity systems.***

Keywords: Cyber-Attack Detection, Machine Learning, Intrusion Detection System, SVM, Anomaly Detection, Network Security

I. INTRODUCTION

A. Background

Cyber-attacks have become a major threat to people, companies, and governments around the world. Types of attacks like malware, denial-of-service, phishing, and network intrusions can damage the security, reliability, and accessibility of important systems. Old methods that rely on known attack signatures are not enough anymore because hackers keep changing their tactics to avoid being caught by fixed security rules.

Machine learning offers a flexible and responsive way to detect cyber-attacks.

It uses past data to learn how systems normally behave and spots unusual activity as it happens. By looking at a lot of network data, smart systems can tell the difference between regular and harmful actions more accurately than traditional methods.

The system has a modular design that includes steps for preparing data, finding important features, classifying threats, and showing results. It is built to work well in real-world settings, be scalable, and provide accurate protection.

B. Problem Statement

The fast growth of digital networks has led to more and more cyberattacks happening often and getting more complicated. Old security systems that rely on fixed rules and known threat signatures can't keep up with new and unknown dangers. Modern networks create huge amounts of data that can't be checked by humans in real time. Attackers keep changing their methods to get past regular security tools. This causes threats to go unnoticed for longer, leading to data leaks and money loss. There's a big need for a smart system that can learn about attack patterns on its own.

Machine learning can spot unusual activities in large data sets. .But choosing the right features and models is still a big challenge.

Too many false alarms and problems with scaling the system also cut down how reliable the detection is.

So, there's a need for a smart, accurate cyberattack detection system that uses machine learning.

C. Objective

More and more people are using cloud services, i nternet-connected devices, and tools that let them access systems from far away. This has caused a big increase in cyberattacks. Because there's so much data flowing through networks, it's impossible to check everything by hand. Machine learning provides an automatic an d flexible way to handle large amounts of complex d ata and spot unusual behavior quickly and effectively.:

1) To learn about various types of cyber attacks like malware, phishing, denial of service, and intrusions.
2) To look into the weaknesses of old detection methods that rely on rules and known attack signatures.
3) To create a smart system for detecting cyber attacks using machine learning.
4) To clean up network traffic data to get rid of noisy and inconsistent information.
5) To find and choose important features that help in accurately identifying attacks.

II. METHODOLOGY

This includes scaling values, data normalization, and the elimination of unnecessary features. Categorical data is transformed into numerical data to be processed by learning algorithms. After data cleaning, feature extraction is performed to identify the most important network features. Statistical and correlation-based methods are employed to decrease the number of features. This increases the processing speed and improves the accuracy of attack detection. The cleaned dataset is divided into two sets: one for training the model and another for testing the model. The training dataset is used to build the machine learning model. A Support Vector Machine (SVM) is selected because of its efficiency in classification problems and its ability to effectively process complex data. The SVM model is trained using labeled network traffic data. Important parameters, such as the type of kernel and regularization parameters, are tuned to improve the performance of the model. After the model has been trained, it is tested using the testing dataset. Important parameters, such as accuracy, precision, recall, and F1-score, are employed to assess the performance of the model. The trained model is then incorporated into a desktop- based system developed using Python. A friendly interface is developed using Tkinter, which enables users to upload their datasets easily. The uploaded data is then subjected to the same preprocessing and feature extraction processes as the training data. The trained SVM model then evaluates the incoming data in real-time. The output of the detection process is then presented to the user through the interface. An SQLite database is used to store user login information, activity logs, and detection results. An admin interface is also incorporated to aid in the management of users and datasets. The system enables the retraining of the model using new data to keep abreast of new attack types. Data integrity and access control mechanisms are in place to ensure the security of the system. Validation of the system is done to determine its accuracy andreliability. The modular design makes it easy to maintain, expand, and improve in the future.

D. Implementation Details

The proposed system of cyber attack detection would be constructed with Python and would employ machine learning techniques. It would be modular to aid in scaling, securing, and maintaining it. It would consist of four main components: frontend, backend, machine learning, and database.

1) Technology Stack

a) Programming Language: Python 3
b) Machine Learning Library: Scikit-learn
c) GUI Framework: Tkinter
d) Database: SQLite3
e) Data Processing: NumPy, Pandas
f) Image Handling: Pillow (PIL) 2

2) Module-wise Implementation

a) User Interface (GUI): The frontend is created using Python's Tkinter. The main window, named main.py, is considered the entry point. It displays login and sign-up details. The login and sign-up operations are facilitated by separate modules, namely, login.py and registration.py. The upload feature is provided on the dashboard, and buttons are utilized to initiate the operations in the backend by using commands in subprocess.
b) Data Preprocessing Module The data uploaded is read using the Pandas library. The missing values will be replaced with the average or median. The categorical variables will be converted into numerical values using label encoding. The numerical variables will be scaled using Min-Max scaling. In addition, noisy values will be removed.
c) Feature Extraction Module: Correlation analysis helps to eliminate features that are not essential. Statistical methods are used to select the most important features. The set of features thus cleaned is then input into the machine learning model.
d) Machine Learning Module: A Support Vector Machine classifier is constructed utilizing the Scikit-learn library. The dataset is split into training and testing parts. The SVM model is trained using labeled data. Hyperparameters are tuned with the grid search. The model has been saved using joblib for future usage.
e) Detection Module: When the user adds a new data set, the data is processed first and then passed on to the trained model. It decides whether the activity is Normal or Attack. The results are displayed on the graphical user interface.
f) Admin Module: Admins enter the system with secure login credentials. They are able to see all users, approve accounts, and upload new training data. They can retune the model when new data is available.

g) Database Module For storing user details and logs, SQLite3 is employed. User authentication and data handling are secured through SQL query mechanisms. This system ensures that it is precise, adaptable, protected, and can identify threats in real time.

III. SYSTEM ARCHITECTURE

The cyberteam attack detection system is designed in a modular structure, such that each part carries out a specific responsibility along the workflow chain of discovering threats. The system mainly has two major users: Admin and User.

The Admin is responsible for user management, the upload of training data, and also the retraining of the machine learning model.

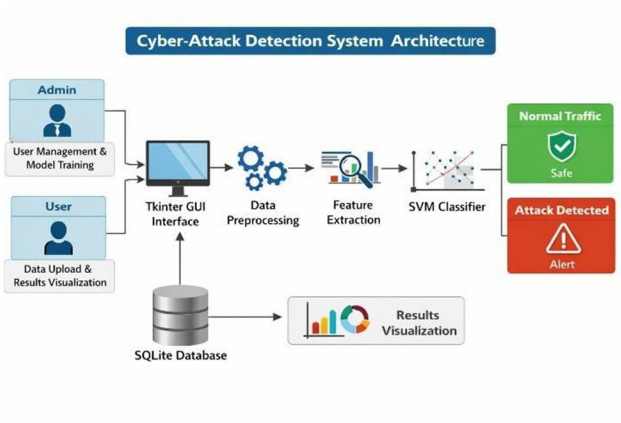
The User can upload network traffic data and see the results of the detection. The front part of the application is the Tkinter GUI interface, used both by the Admin and the User. This interface transmits the data uploaded to the Data Preprocessing Module for cleaning, adjustment, and preparation against further processing.

The cleaned data is then sent to the Feature Extraction Module for selecting the most important parts of the data which will be analyzed.

These selected features become the inputs of the SVM Classifier searching for patterns to decide whether the data is:

Normal Traffic [Safe] Attack Detected(Alert)

User information, logs, and results are all stored in an SQLite Database. The Results Visualization Module is used to represent the output of detection in an understandable manner.



IV. RESULT AND DISCUSSION

A dataset that included both normal and harmful network traffic was used to test the proposed Cyber- Attack Detection System. The dataset required some steps of cleaning, handling missing parts, and making all data consistent before using it to train and test the machine learning model.

A. Results

The classifier Support Vector Machine was trained on 80% of the data and validated on the remaining 20% to determine the effectiveness of the model.

The model worked well and showed its ability to differentiate between normal and malicious network activity quite well.

The system was good at detecting various types of attacks, like denial-of-service, malware, and unauthorized access attempts.

That means it had a much lower false alarm rate compared to older, rules-based systems; thus fewer normal activities were incorrectly tagged as threats. It can also be observed that the model's precision, recall, and F1-score were dependable in finding real attacks while keeping mistakes to a minimum. The confusion matrix also presented the correctness of most attack samples.

B. Discussion

The test results indicate that it is appropriate to use machine learning compared to traditional methods, which use known signatures in identifying cyber attacks. Although the SVM model was able to learn from the network data and detect unseen attack patterns. The design of the system, with its parts for cleaning the data, identifying important features, and making predictions, enabled the system to operate quickly and accurately.

Selection of the best features helps improve performance by reducing the complexity and computationally intensive nature of the data. However, the effectiveness of this system depends on the quality and type of training data.

The accuracy performance of the system may also decrease if the given dataset is not quite diverse in types of attacks. Moreover, sometimes running big datasets may require higher system computers. The results of the experiment have indicated that the proposed system is an efficient, flexible, and deployable system for real- world application.

V. LITERATURE SURVEY

Due to the increasing occurrences and severity of such cyber-attacks on modern connected systems, detection has become one of the major focuses in the realm of cybersecurity research. Academic studies have developed various methods, ranging from older techniques relying on already known attack patterns to newer models making use of machine learning techniques for better detection.

A. Traditional Intrusion Detection Approaches

In the past, most research on intrusion detection systems has used systems that rely on established attack models, like the Snort package.

These systems worked by checking the activity on a network and comparing it with a set of signatures that define various attacking behaviors. They were very effective at detecting known threats, but they were ineffective at detecting new threats because they did not have the ability to learn.

Denning, in 1987, pointed out a big problem with the intrusion detection systems in dealing with strange patterns of attacks.

Often, they missed the threats, hence the prevalence of false negatives. This indicated the need for better systems that could identify threats without exclusively relying on known patterns.

B. Statistical and Rule-Based Methods

To solve the problems associated with signature-based systems, the use of statistical anomaly detection methods has been proposed. Such systems had learned what normal network activity looked like and would send alerts if anything unusual was detected. However, these approaches had a major limitation: they would send a lot of false alarms, particularly if normal behavior was changing frequently over time.

C. Machine Learning for Attack Detection

The advancement in the technology of machine learning has greatly improved the capability of detecting attacks:

Support Vector Machines (SVM): SVM is typically used for applications where a decision between two choices is to be made.

It works particularly well with data with a lot of features. Research has indicated that this algorithm can be even more accurate than other technologies, like neural networks and decision trees, in identifying intruding activities in computer systems with the help of an appropriate mathematical tool, known as kernel function.

Random Forest, K-Nearest Neighbors (KNN): The combination prediction method, like the Random Forest model, and the prediction method using the comparison between data points, like the KNN model, are also discussed.

Research has revealed that Random Forest gives impressive results due to its aggregation of multiple models; however, this technique requires more computational power. Neural Networks and Deep Learning: In recent years, different deep learning techniques like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been used for analyzing data sequences. These methods can find complex patterns on their own, but they require a great deal of data and powerful computer technology to be successful.

VI. CONCLUSION AND FUTURE SCOPE

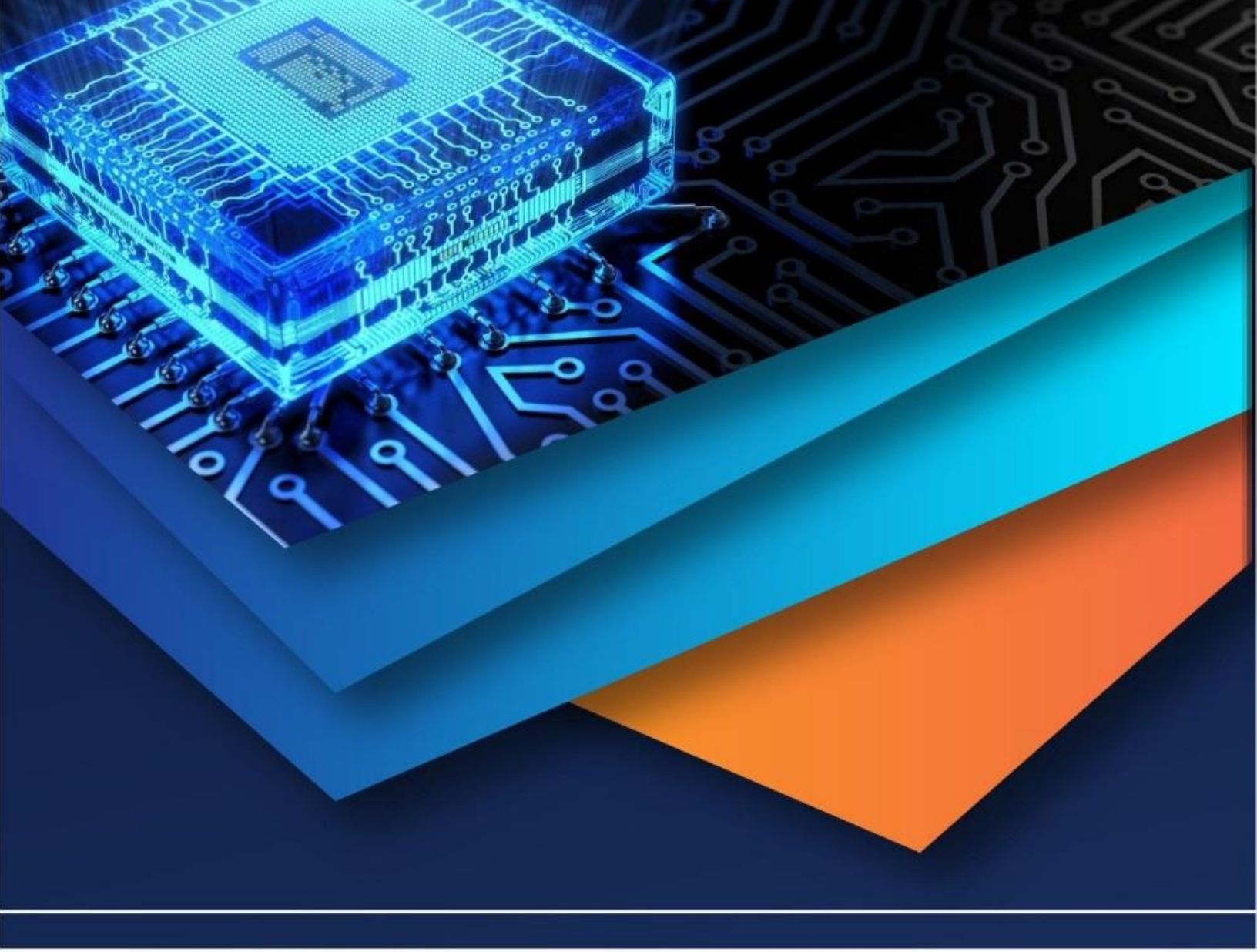
In this project, a system was developed by employing the concept of machine learning that can find cyber attacks and attempt to improve the conventional methods of security. There are various steps adopted by the system, including data processing, selection of vital features, and usage of the SVM classifier to identify the malicious activity in the network. From the results of the test procedure, it can be seen that the system is good at distinguishing between normal and malicious activities. It has a well-defined structure, simple interface, and the capability to track malicious activities in real time, making the system ready for use in the field. The above project indicates the importance of the usage of intelligent systems to improve the security of the systems from cyber threats in today's electronic world.

A. Future Scope

The present system does quite well for detection; however, here are a few ways in which it can be improved in the future:
More complicated forms of attack may be recognized using advanced models like CNNs and LSTMs.
It would be capable of sniffing packets in real time, allowing us to look at live network traffic instead of just looking at CSV files.
Putting it in the cloud will make it easier to scale up and also access from anywhere.
Automating the process of retraining the models would help the system keep up with new and evolving threats.
The support for devices and networks deployed in IoT and mobile environments can also be included.
This might connect the system with firewalls and SIEM tools to enable automated responses in cases of threats.
Improving the dashboards on data visualization would make monitoring and understanding of the system performance easy.
Therefore, other techniques, such as ensemble learning, might provide more appropriate results.
It would be much better to implement the functionality of classifying different kinds of attacks for more detailed analysis.
Meanwhile, utilizing blockchain for logging could provide records of security that are immune from unauthorized modification.

REFERENCES

[1] B. H. Aishwarya, B. S. Akki, H. M. Harshitha, N. R. Navyashree, and D. E. Vedananda, “A Survey on Intrusion Detection System using Machine Learning Techniques”, International Journal for Research in Applied Science & Engineering Technology, 2023.
[2] M. Alkasassbeh, “An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods”, arXiv, 2017. T.-H. Chua and I. Salam, “Evaluation of Machine Learning Algorithms in Network-Based Intrusion Detection System”, arXiv, 2022.
[3] J. Vitorino, R. Andrade, I. Praça, O. Sousa, and E. Maia, “A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection”, arXiv, 2021. A comprehensive survey and taxonomy of SVM- based intrusion detection systems, Journal of Network and Computer Applications, 2021.
[4] S. Dalal, U. K. Lilhore, N. Faujdar et al., “Next- generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree”, Journal of Cloud Computing, 2023.
[5] A. Pinto, L.-C. Herrera, Y. Donoso and J. A. Gutierrez, “Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure”, Sensors, 2023. Khadija M. Abuali, Liyth Nissirat and Aida Al- Samawi, “Advancing Network Security with AI: SVM-Based Deep Learning for Intrusion Detection”, Sensors, 2023.
[6] “Machine learning methods for cyber security intrusion detection: Datasets and comparative study”, Computer Networks, Elsevier, 2021.
[7] Suad M. Othman, F. M. Ba-Alwi and N. T. Alsohybe, “Intrusion detection model using machine learning algorithm on Big Data environment”, Journal of Big Data, 2018.
[8] Haripriya A. P., Meenu V., Malavika M. Hari et al., “A Survey on Machine and Deep Learning Based Intrusion Detection Systems for IoT”, 2023. (Optional) A Literature Review on Machine Learning Methods Used in Intrusion Detection System to Detect Cyber Attack, IEEE Xplore Conference PaperA. Pinto, L.-C. Herrera, Y. Donoso and J. A. Gutierrez, “Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure”, Sensors, 2023. Khadija M. Abuali, Liyth Nissirat and Aida Al- Samawi, “Advancing Network Security with AI: SVM-Based Deep Learning for Intrusion Detection”, Sensors, 2023.
[9] “Machine learning methods for cyber security intrusion detection: Datasets and comparative study”, Computer Networks, Elsevier, 2021. Suad M. Othman, F. M. Ba-Alwi and N. T. Alsohybe, “Intrusion detection model using machine learning algorithm on Big Data environment”, Journal of Big Data, 2018.
[10] Haripriya A. P., Meenu V., Malavika M. Hari et al., “A Survey on Machine and Deep Learning Based Intrusion Detection Systems for IoT”, 2023. (Optional) A Literature Review on Machine Learning Methods Used in Intrusion Detection System to Detect Cyber Attack, IEEE Xplore Conference Paper



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY