ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

# IOT Network Intrusion Detection System Using Machine Learning Techniques

Duddeda Manikanta

21N81A62A1

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

Manikantaduddeda07@gmail.com

Palvatla Harish reddy

22N85A6204

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

Palvatlaharishreddy@gmail.com

Bommideni Vamshi

21N81A62A0

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

bvamshi9030@gmail.com

Mrs. P.Sandhya Rani

Assistant HOD

Computer Science and Engineering (Cyber Security)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

**ABSTRACT:** This study focuses on evaluating the effectiveness of two supervised machine learning techniques—Support Vector Machine (SVM) and Convolutional Neural Networks (CNN)—for use in network intrusion detection systems (IDS). Given the increasing reliance on internet-based services, systems are increasingly exposed to various forms of cyberattacks. An IDS is designed to inspect incoming network traffic and identify whether the request is legitimate or potentially harmful. The IDS is first trained using labeled datasets that include examples of both normal and malicious traffic. Once trained, the system can analyze new incoming requests and determine whether they should be allowed or blocked based on the learned patterns. To enhance the accuracy of detection and reduce processing overhead, feature selection techniques such as Correlation-Based Selection and Chi-Square Selection are applied. These methods help eliminate redundant or irrelevant data, leading to a more efficient and accurate classification process. In this work, the performance of SVM and CNN is compared based on their ability to detect network intrusions. Experimental results demonstrate that CNN provides superior accuracy compared to SVM, particularly when effective feature selection methods are used to refine the input data.

***Keywords*** – SVM, CNN, IDS.

## INTRODUCTION

As internet usage continues to grow globally, so does the frequency and complexity of cyber threats. One of

the first lines of defense against such threats is **intrusion detection**, which has become a vital component in the design of modern cybersecurity systems. Tools such as **Firewalls**, **Intrusion Detection Systems (IDS)**, **Unified Threat Management (UTM)** platforms, and **Intrusion Prevention Systems (IPS)** are actively being researched and implemented to enhance security measures. An IDS functions by collecting and analyzing data from networks and systems to identify unauthorized activities. In particular, **network-based IDS** focuses on monitoring packet flows within a network to detect suspicious behavior. These detection techniques fall into two categories: **signature-based** and **anomaly-based**. While signature-based detection—relying on known patterns of attacks—has proven effective in commercial environments, **anomaly detection**, which aims to identify deviations from normal behavior, remains an area with significant research potential due to its ability to detect new and unknown threats.

However, anomaly-based detection presents challenges. It requires the system to intelligently distinguish between legitimate and harmful traffic without relying on predefined patterns. This need for adaptive intelligence has led to the exploration of **machine learning algorithms**, which offer powerful tools for learning and predicting complex behaviors. It's important to recognize that IDS technologies are not a complete solution to all security issues. For example, they cannot fix weak password systems or insecure network protocols. Though the concept of intrusion detection dates back to the early 1980s—with foundational models introduced in the late 1980s—progress in this domain has been slower than anticipated, despite heavy investment and research. Traditional **signature-based IDS** has seen successful deployment across many organizations, but **anomaly-based IDS** still faces hurdles such as high false alarm rates. These often stem from the difficulty in accurately modeling what constitutes "normal" behavior in diverse and dynamic network environments. **Vector Machines (SVM)**, **k-Nearest Neighbors (k-NN)**, **Naive Bayes**, **Decision Trees**, **Genetic Algorithms**, and **Gaussian Mixture Models**. Among these, **SVM** has been widely used due to its strong performance across various classification tasks. Nevertheless, most of these models suffer from high false positive rates, primarily because of the complexity involved in learning accurate representations of normal network activity. More recently, **Convolutional Neural Networks (CNNs)** have been applied to the field of anomaly detection. These deep learning models, commonly trained using the **backpropagation algorithm**, have demonstrated promising results in learning intricate patterns from large datasets. A significant challenge in evaluating IDS performance lies in the scarcity of robust, real-world datasets. While many studies have used the outdated **KDD Cup 1999** dataset, our work uses the improved **NSL-KDD** dataset to compare the effectiveness of SVM and CNN models in detecting network intrusions.

## 1. LITERATURE REVIEW

### ''A macro-social exploratory analysis of the rate of interstate cyber-victimization,''

This study examines whether macro-level opportunity indicators affect cyber-theft victimization. Based on the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to determine if variation in social structure impacted cyber-victimization across states. The current study found that structural conditions such as unemployment

and non-urban population are associated with where users access the internet. Also, this study found that the proportion of users who access the internet only at home was positively associated with state-level counts of cyber-theft victimization. The theoretical implications of these findings are discussed.

.

**‘‘Incremental anomaly-based intrusion detection system using limited labeled data,’’**

With the proliferation of the internet and increased global access to online media, cybercrime is also occurring at an increasing rate. Currently, both personal users and companies are vulnerable to cybercrime. A number of tools including firewalls and Intrusion Detection Systems (IDS) can be used as defense mechanisms. A firewall acts as a checkpoint which allows packets to pass through according to predetermined conditions. In extreme cases, it may even disconnect all network traffic. An IDS, on the other hand, automates the monitoring process in computer networks. The streaming nature of data in computer networks poses a significant challenge in building IDS. In this paper, a method is proposed to overcome this problem by performing online classification on datasets. In doing so, an incremental naive Bayesian classifier is employed. Furthermore, active learning enables solving the problem using a small set of labeled data points which are often very expensive to acquire. The proposed method includes two groups of actions i.e. offline and online. The former involves data preprocessing while the latter introduces the NADAL online method. The proposed method is compared to the incremental naive Bayesian classifier using the NSL-KDD standard dataset. There are three advantages with the proposed method: (1) overcoming the streaming data challenge; (2) reducing the high cost associated with instance labeling; and (3) improved accuracy and Kappa compared to the incremental naive Bayesian approach. Thus, the method is well-suited to IDS applications.

**‘‘Modeling and implementation approach to evaluate the intrusion detection system,’’**

This study examines whether macro-level opportunity indicators affect cyber-theft victimization. Based on the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to determine if variation in social structure impacted cyber-victimization across states. The current study found that structural conditions such as unemployment and non-urban population are associated with where users access the internet. Also, this study found that the proportion of users who access the internet only at home was positively associated with state-level counts of cyber-theft victimization. The theoretical implications of these findings are discussed.

## 2. SYSTEM MODEL

The architecture of the proposed system includes two core modules: a feature selection process and a machine learning model, as depicted in Figure 1. The feature selection stage plays a crucial role in identifying the most significant attributes from the dataset that contribute to accurate classification. By focusing on only the most relevant features, the system becomes more efficient and potentially more accurate. Next, the machine learning component uses the refined dataset to develop a classification model. This is done by training the model on labeled data, allowing it to learn patterns and make informed predictions. Once training is complete, the model is tested on new, unseen data to assess how accurately it can classify instances, which helps evaluate its overall performance.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

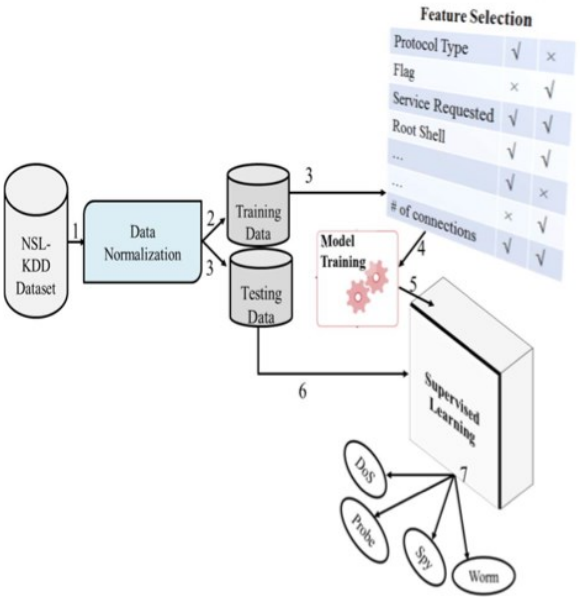


Fig 1: Proposed supervised machine learning classifier System

### A. Feature Selection

Feature selection is a vital process in machine learning, helping to reduce the number of input variables and improve model efficiency. Two commonly used techniques for feature selection are filter methods and wrapper methods. Filter methods work by evaluating each feature independently based on its statistical relationship with the target output. They rank features using statistical measures such as correlation or significance tests, and the least relevant features are removed before training the model. This approach does not rely on any specific machine learning algorithm, making it fast and scalable—especially useful when dealing with very large datasets. On the other hand, wrapper methods assess subsets of features by training and testing a model on different combinations of inputs. The performance of each subset is evaluated, and the one that gives the best accuracy is selected. This method is model-dependent and more computationally intensive, as it involves repeated training cycles. To find the best feature subset, wrapper methods use various search techniques like random search, breadth-first search, depth-first search, or hybrid approaches. In summary, filter methods are better suited for high-dimensional datasets common in data mining, whereas wrapper methods tend to provide better performance for machine learning tasks when computational cost is not a major concern.

### B. Building Machine Intelligence

After selecting the most important features, a machine learning model is created using those features. The model is trained using a **supervised learning approach**, where each entry in the training data includes both the input features and the correct output label. The algorithm uses this labeled data to recognize patterns and learn how different feature combinations relate to specific classes. The structure and behavior of the resulting model depend on the type of machine learning algorithm applied during training.

### C. Support Vector Machine (SVM)

In Support Vector Machines (SVM), the classifier is defined by a separating hyperplane, which varies based on the nature of the problem and the dimensionality of the dataset. When dealing with one-dimensional data, this hyperplane reduces to a single point. For two-dimensional data, it becomes a line, as illustrated in Fig. 2. In three dimensions, it takes the form of a plane, and in even higher dimensions, it is referred to as a hyperplane. For datasets that are linearly separable, the classifier or decision boundary can be represented using a specific mathematical function.

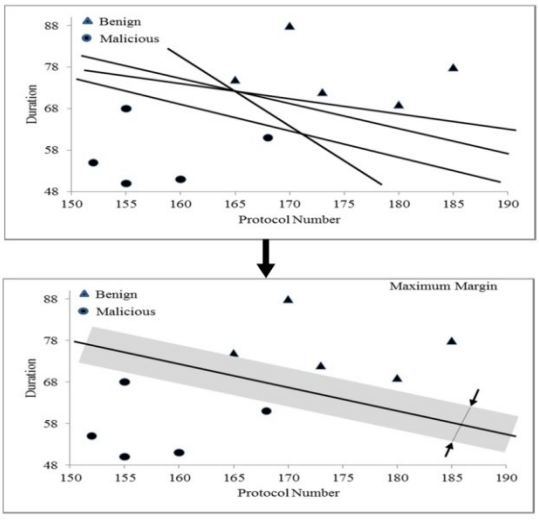


Fig 2: SVM classifier in two dimensional problem spaces

$$ax\ by\ c + + = 0 \quad (1)$$

For a given data points (x,y), the above decision function will classify the point in one class if ax + by ≥ c or it will categorize if ax + by < c. The equation of a line y=ax+b can be rewritten as y−ax−b=0 that can be represent using two vectors as below-

$$\mathbf{w}\begin{pmatrix}-b\\-a\\1\end{pmatrix} \text{ and } \mathbf{x}\begin{pmatrix}1\\x\\y\end{pmatrix} \quad (2)$$

which says we can write the linear equation of a line using two vectors as below-

$$\mathbf{w}^T\mathbf{x} = (-b)\times(1)+(-a)\times x+1\times y, \text{ or}$$
$$\mathbf{w}^T\mathbf{x} = y-ax-b \quad (3)$$

The reason of using the hyper plane equation wTx instead of y=ax+b is because it is easier to work in more than two dimensions with this notation and the vector w will always be normal to the hyper plane. Once the hyper plan with maximum margin has been found, this hyper plane can be used to make predictions [11]. The hypothesis function h will be-

$$h(x_i) = \begin{cases}+1; \ if \ \ \mathbf{w.x}+b \geq 0\\ -1; \ if \ \ \mathbf{w.x}+b < 0\end{cases} \quad (4)$$

### D. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of machine learning model particularly effective for recognizing complex patterns within large datasets. In the context of an Intrusion Detection System (IDS), CNNs are used to analyze network traffic and determine whether the behavior is normal or indicative of a cyberattack. They do this by detecting patterns and irregularities that may signal malicious activity. Typically, a CNN is composed of an input layer, one or more hidden layers, and an output layer, as illustrated inFig.3.

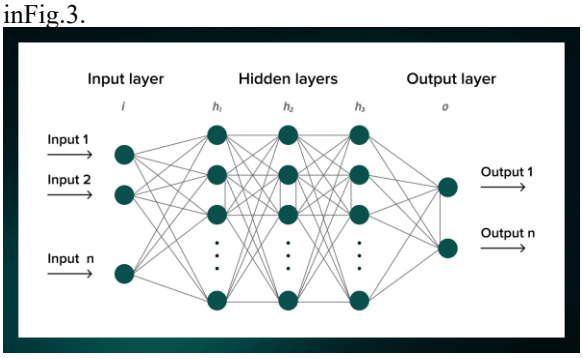


Fig 3: Convolutional neural network showing the input, output and hidden layers

## 3. METHODOLOGIES

❖ Graphical user interface
- Upload dataset
- Preprocess dataset
- Augmentation
- Generate training model
- Run cnn algorithm
- Run Naïve Bayas algorithm
- Run svm algorithm
- Predict attack from test data
- View accuracy graph

❖ **Graphical user interface**

A Graphical User Interface (GUI) is a visual interface that allows users to interact with computer software and hardware systems using graphical elements like icons, buttons, and menus. In this we have buttons like upload data , preprocess data and above as mentioned they will perform different tasks as they are assigned.

- **Upload data**
  It allows user to select and load a CSV dataset and displays its initial attack distribution.

- **Preprocess dataset**
  Transforms categorical features into numerical ones and prepares the data for meachine learning.

- **Augmentation**
  Addresses class imbalance in the dataset using SMOTE (Synthetic Minority Over-sampling Technique)

- **Generate training model**
  It divides the preprocessed and augmented dataset into training and testing sets.

- **Run CNN Algorithm**
  Implements , trains, and evaluats the proposed CNN model. (Convolutional Neural Networks are being used in Intrusion Detection Systems (IDS) to improve their accuracy and efficiency in identifying and classifying network intrusions.)

- **Run SVM Alogrithm**
  Trains and evaluates the support vector meachine algorithm. (Support Vector Machines (SVM) are a widely used machine learning algorithm in Intrusion Detection Systems (IDS) due to their ability to effectively classify network traffic and detect malicious activity.)

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

- **Run Navie Bayes Algorithm**
  Trains and evaluates the Gaussian Naïve Bayas Algorithm.

- **Predict attack from test data**
  Allows the user to upload new data and predict attack types using the trained cnn model.

- **View accuracy graph**
  Generates and displays comparison graphs for all trained algorithm.

## 4. EXPERIMENTAL RESULTS

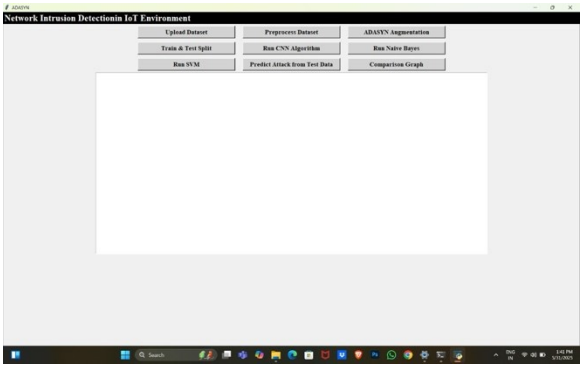


Fig.1: Graphical User Interface.

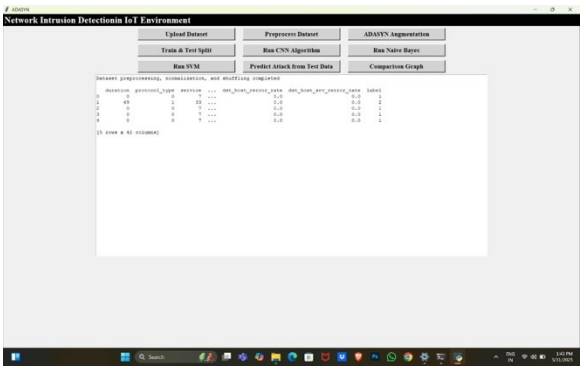


Fig 2: upload data

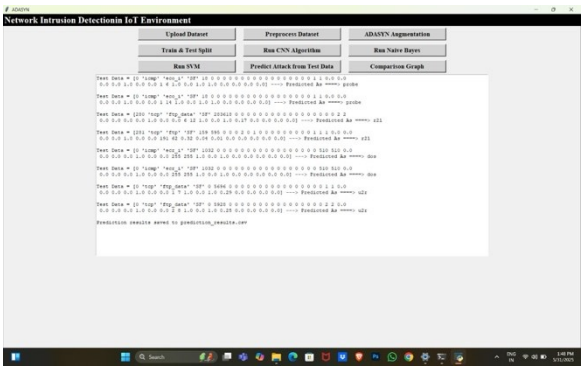


Fig 3: predicted attack from test data

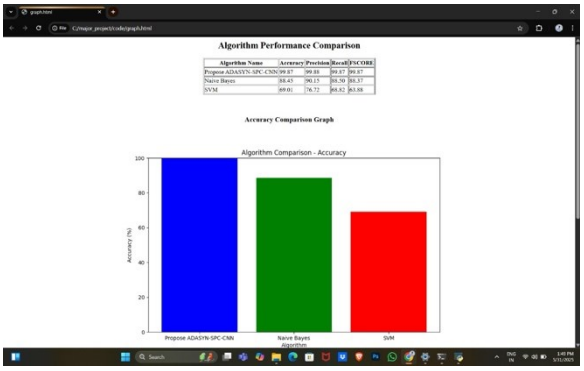


Fig.4: Accuracy graph

## 5. CONCLUSION

In this study, we explored various machine learning models developed using different algorithms and feature selection techniques to identify the most effective approach. The results indicate that the model incorporating a Convolutional Neural Network (CNN) combined with wrapper-based feature selection delivered the highest performance, achieving a detection accuracy of 99.87% in classifying network traffic. These findings are expected to support ongoing research in developing more efficient intrusion detection systems. While current IDS technologies are capable of identifying known threats, detecting previously unseen or zero-day attacks remains a significant challenge, primarily due to the high false positive rates associated with existing solutions.

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

**REFERENCES**

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of inter-state cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXivpreprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016

. [10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniqu