Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.4, January : 2025

CLOUD-BASED INTRUSION DETECTION APPROACH USING HYBRID LEARNING TECHNIQUE

Prashanth Kumar, M. Tech Scholar, Department of Computer Science and Engineering, Technocrats Institute of Technology, Bhopal, MP. prashanthmandal143@gmail.com
Dr. Bhupendra Verma, Professor, Department of Computer Science and Engineering, Technocrats Institute of Technology, Bhopal, MP

ABSTRACT
Cloud computing has revolutionized data storage and processing by offering scalability, flexibility, and cost efficiency. However, the increasing adoption of cloud platforms has exposed them to a wide range of security threats, including intrusions, cyberattacks, and data breaches. Intrusion Detection Systems (IDS) play a critical role in safeguarding cloud environments by identifying and mitigating unauthorized activities. This paper presents a novel cloud-based intrusion detection approach using a hybrid learning technique that combines Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The proposed model leverages the strengths of ANN in learning complex patterns and SVM in binary classification to enhance the detection accuracy of malicious activities. Extensive experimentation is conducted on benchmark datasets, showcasing the model's robustness in detecting diverse attack patterns while minimizing false positives. The study also highlights the scalability of the hybrid approach in handling large-scale cloud environments. The results demonstrate that the proposed ANN-SVM hybrid model outperforms traditional IDS solutions in terms of precision, recall, and overall efficiency, making it a promising tool for securing cloud-based infrastructures.
Keywords: IOT, Cyber, ANN, SVM, Deep learning, Security.

I. Introduction
Cloud computing has emerged as a transformative paradigm in the modern digital ecosystem, enabling businesses and individuals to store, process, and access data and applications over the internet. The flexibility, scalability, and cost-effectiveness offered by cloud platforms have driven their widespread adoption across various domains, including healthcare, finance, education, and e-commerce. However, as cloud environments become increasingly integrated into critical infrastructures, they also become prime targets for cyberattacks and intrusions. This growing reliance on cloud systems necessitates the development of robust security mechanisms to ensure data integrity, confidentiality, and availability.

Intrusion Detection Systems (IDS) have long been recognized as an essential component of cybersecurity frameworks. By monitoring network traffic and system activities, IDS can detect potential threats and unauthorized access attempts in real time. Traditional IDS methods, while effective in certain scenarios, often struggle to adapt to the dynamic and distributed nature of cloud environments. The sheer volume of data generated in cloud systems, coupled with the evolving complexity of cyber threats, demands advanced and intelligent approaches to intrusion detection.

In this context, machine learning (ML) techniques have emerged as a powerful tool for enhancing the capabilities of IDS. Among the various ML methods, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) stand out due to their distinct strengths. ANN excels in identifying complex, non-linear patterns in data, making it well-suited for analyzing network traffic and identifying subtle anomalies. On the other hand, SVM is renowned for its effectiveness in binary classification tasks and its ability to handle high-dimensional data efficiently. While both techniques have shown promise individually, their combination offers the potential to create a hybrid system that leverages the advantages of both approaches.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.4, January : 2025

The primary objective of this study is to develop a cloud-based intrusion detection approach using a hybrid learning model that integrates ANN and SVM. By combining the deep learning capabilities of ANN with the classification precision of SVM, the proposed model aims to achieve higher detection accuracy and reduce false positive rates. The hybrid approach is designed to address the unique challenges of cloud environments, such as scalability, high data throughput, and the need for real-time threat detection.

This paper is structured as follows: an overview of related work highlights the limitations of existing IDS solutions and the potential of hybrid learning techniques. The methodology section details the architecture and implementation of the proposed ANN+SVM hybrid model, including data preprocessing, feature selection, and model training. The experimental results demonstrate the effectiveness of the model on benchmark datasets, comparing its performance with traditional and standalone ML techniques. Finally, the paper concludes with a discussion of the implications, limitations, and future research directions for cloud-based IDS.

By addressing the critical need for advanced intrusion detection in cloud environments, this research aims to contribute to the ongoing efforts to secure cloud infrastructures and protect sensitive data from cyber threats. The proposed hybrid learning approach represents a significant step forward in the development of intelligent and scalable IDS solutions, paving the way for more secure and resilient cloud ecosystems.

II. Proposed Methodology

The proposed work is discuss using the following flow chart-

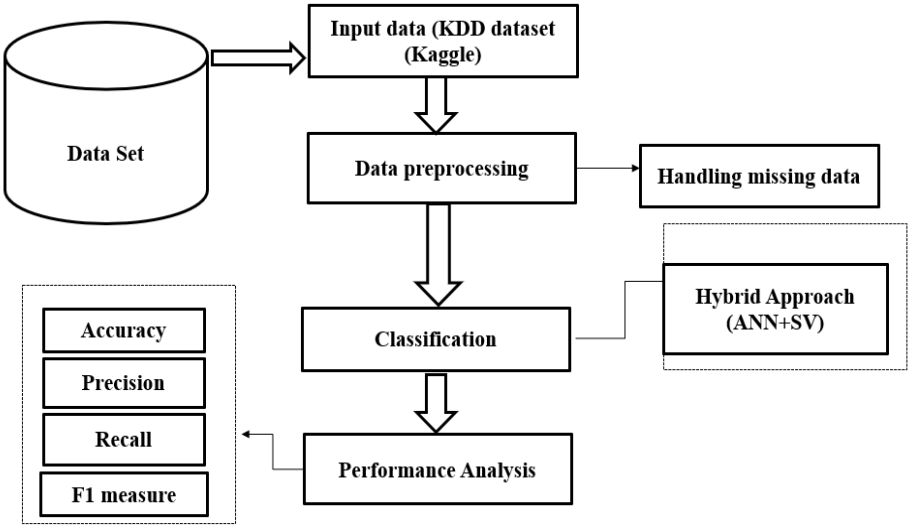


Figure 1: Flow Chart

1. Dataset

- The process begins with a dataset containing information about Intrusion Detection. This dataset typically consists of labeled instances of normal and malicious network activities.

2. Input Data

- The dataset is loaded into the system for processing and analysis. The data may contain raw network traffic features, such as packet size, timestamps, protocols, and IP addresses.

3. Preprocessing

- Data preprocessing involves preparing the dataset for effective training and testing. Key steps in this stage include:
- **Handling Missing Values**: Missing or incomplete data is filled, removed, or replaced to ensure consistency.

- **Label Encoding**: Categorical data (e.g., attack types) is converted into numerical formats to make it compatible with machine learning algorithms.

4. Data Splitting

- The pre-processed dataset is divided into two subsets:

Training Set: Used to train the ANN model to recognize patterns and learn classifications.
Testing Set: Used to evaluate the trained model's performance on unseen data.

5. Hybrid Classification

Artificial Neural Network (ANN) and Support Vector (SV)

The ANN component of the hybrid approach is responsible for learning complex and non-linear patterns within the intrusion dataset. This involves:

1. **Input Layer**: Features from the intrusion dataset (e.g., KDD dataset) are fed into the ANN.
2. **Hidden Layers**:

The ANN uses multiple hidden layers to extract higher-level patterns.
These layers capture relationships between input features, such as anomalies in network traffic indicative of intrusions.

3. **Output Layer**:

The ANN outputs intermediate feature representations instead of final classifications.
These features represent a distilled version of the input data, emphasizing patterns that differentiate between normal and malicious behavior.

Support Vector Machine (SVM)

The SVM acts as the classifier in the hybrid approach, leveraging the feature representations extracted by the ANN.

1. **Input**: The distilled features from the ANN are fed into the SVM for classification.
2. **Kernel Function**: A suitable kernel (e.g., radial basis function) is selected to transform data into a higher-dimensional space if needed, ensuring linear separability of intrusion and normal traffic classes.
3. **Hyperplane and Margin**: The SVM identifies the optimal hyperplane that separates intrusion classes from normal ones with the maximum margin.

Integration of ANN and SVM

1. **Feature Extraction by ANN**:

The ANN preprocesses the data and extracts non-linear features, reducing irrelevant noise in the dataset.

2. **Classification by SVM**: These features are passed to the SVM, which uses its robust margin-based classification to finalize predictions.
3. **Benefits of the Combination**:

ANN handles complex, high-dimensional data efficiently.
SVM ensures precise classification, particularly for datasets with clear class boundaries.
This hybrid approach leverages ANN's capability to learn from raw data and SVM's strength in robust classification, ensuring high accuracy and reliability in intrusion detection.

6. Classification

Based on the proposed model's predictions, the system classifies network traffic into two categories:
Positive Review: Indicates malicious activity (e.g., an attack).
Negative Review: Indicates normal or safe network behavior.
This classification helps identify intrusions in real time, enabling swift countermeasures to protect the cloud environment.

7. Performance Metrics

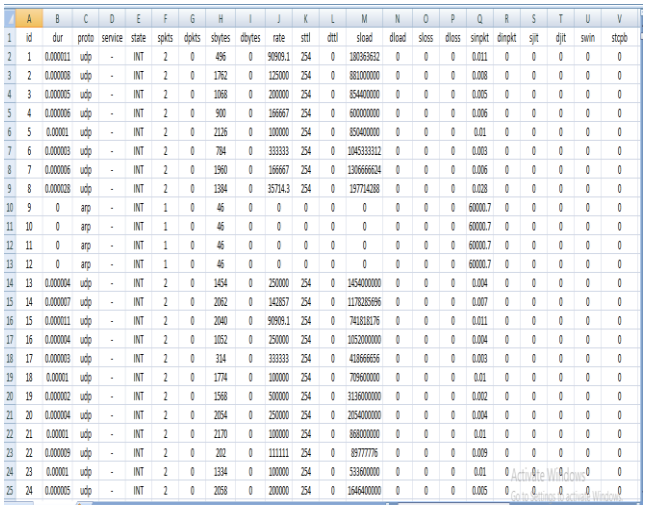
To evaluate the system's effectiveness, several performance metrics are computed, including:
Accuracy: Measures the proportion of correctly classified instances (both positive and negative) out of the total dataset.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.4, January : 2025

Precision: Evaluates how well the model identifies positive instances without including false positives.

III. Simulation and results

Python Spyder integrated development environment (IDE) version 3.7 is used to carry out the simulation.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	id	dur	proto	service	state	spkts	dpkts	sbytes	dbytes	rate	sttl	dttl	sload	dload	sloss	dloss	sinpkt	dinpkt	sjit	djit	swin	stcpb
2	1	0.000011	udp	-	INT	2	0	496	0	90909.1	254	0	180363632	0	0	0	0.011	0	0	0	0	0
3	2	0.000008	udp	-	INT	2	0	1762	0	125000	254	0	881000000	0	0	0	0.008	0	0	0	0	0
4	3	0.000005	udp	-	INT	2	0	1068	0	200000	254	0	854400000	0	0	0	0.005	0	0	0	0	0
5	4	0.000006	udp	-	INT	2	0	900	0	166667	254	0	600000000	0	0	0	0.006	0	0	0	0	0
6	5	0.00001	udp	-	INT	2	0	2126	0	100000	254	0	850400000	0	0	0	0.01	0	0	0	0	0
7	6	0.000003	udp	-	INT	2	0	784	0	333333	254	0	1045333312	0	0	0	0.003	0	0	0	0	0
8	7	0.000006	udp	-	INT	2	0	1960	0	166667	254	0	1306666624	0	0	0	0.006	0	0	0	0	0
9	8	0.000028	udp	-	INT	2	0	1384	0	35714.3	254	0	197714288	0	0	0	0.028	0	0	0	0	0
10	9	0	arp	-	INT	1	0	46	0	0	0	0	0	0	0	0	60000.7	0	0	0	0	0
11	10	0	arp	-	INT	1	0	46	0	0	0	0	0	0	0	0	60000.7	0	0	0	0	0
12	11	0	arp	-	INT	1	0	46	0	0	0	0	0	0	0	0	60000.7	0	0	0	0	0
13	12	0	arp	-	INT	1	0	46	0	0	0	0	0	0	0	0	60000.7	0	0	0	0	0
14	13	0.000004	udp	-	INT	2	0	1454	0	250000	254	0	1454000000	0	0	0	0.004	0	0	0	0	0
15	14	0.000007	udp	-	INT	2	0	2062	0	142857	254	0	1178285696	0	0	0	0.007	0	0	0	0	0
16	15	0.000011	udp	-	INT	2	0	2040	0	90909.1	254	0	741818176	0	0	0	0.011	0	0	0	0	0
17	16	0.000004	udp	-	INT	2	0	1052	0	250000	254	0	1052000000	0	0	0	0.004	0	0	0	0	0
18	17	0.000003	udp	-	INT	2	0	314	0	333333	254	0	418666656	0	0	0	0.003	0	0	0	0	0
19	18	0.00001	udp	-	INT	2	0	1774	0	100000	254	0	709600000	0	0	0	0.01	0	0	0	0	0
20	19	0.000002	udp	-	INT	2	0	1568	0	500000	254	0	3136000000	0	0	0	0.002	0	0	0	0	0
21	20	0.000004	udp	-	INT	2	0	2054	0	250000	254	0	2054000000	0	0	0	0.004	0	0	0	0	0
22	21	0.00001	udp	-	INT	2	0	2170	0	100000	254	0	868000000	0	0	0	0.01	0	0	0	0	0
23	22	0.000009	udp	-	INT	2	0	202	0	111111	254	0	89777776	0	0	0	0.009	0	0	0	0	0
24	23	0.00001	udp	-	INT	2	0	1334	0	100000	254	0	533600000	0	0	0	0.01	0	0	0	0	0
25	24	0.000005	udp	-	INT	2	0	2058	0	200000	254	0	1646400000	0	0	0	0.005	0	0	0	0	0

Figure 2: Dataset

As can be seen in figure 2, the data set is shown in the Python environment. There may be a significant amount of variation in the row and column numbers of the dataset. Every column has a name that is used to identify the features.

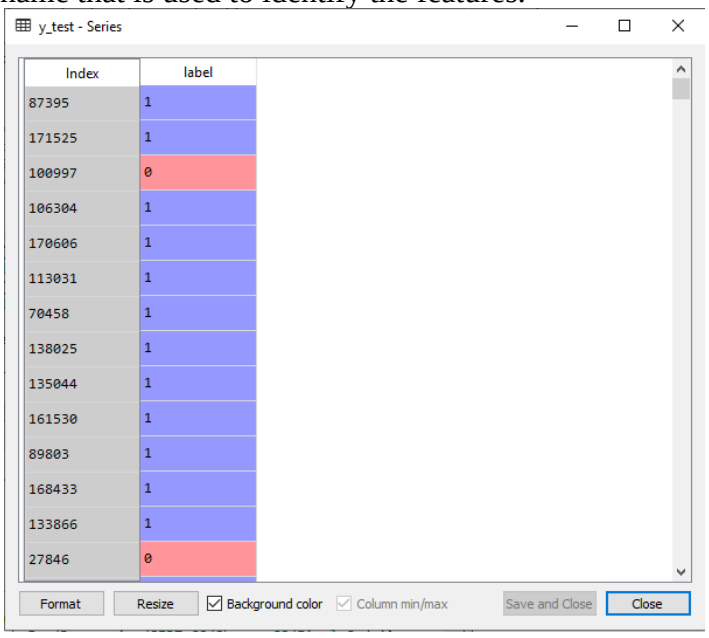


Figure 3: Y test

Y test results for this dataset may be seen in figure 3. Twenty-thirty percent of the initial dataset is used as the primary dataset for training purposes.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.4, January : 2025

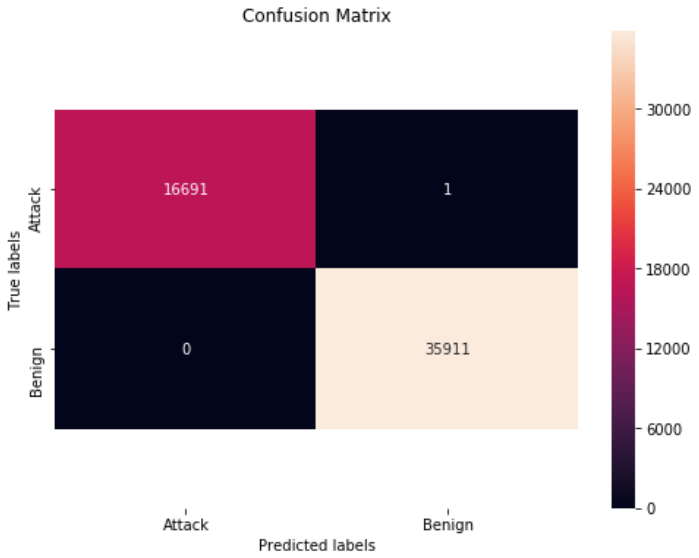


Figure 4: Confusion matrix heat map

The confusion matrices for heat maps that were produced by the ANN deep learning classification approach are shown in figure 4. In order to evaluate how successfully a categorization model performs its function, a matrix with dimensions of N by N is used.

Table 1: Simulation Results

Sr. No.	Parameters	Value (%)
1	Precision	99.99
2	Recall	99.99
3	F_Measure	99.99
4	Accuracy	99.99
5	Error Rate	0.01
6	Sensitivity	99.99
7	Specificity	99.99

Table 2: Result Comparison

Sr No	Parameter	Previous Work [1]	Proposed Work
1	Accuracy	99.94%	99.99%
2	Classification Error	0.06%	0.01%

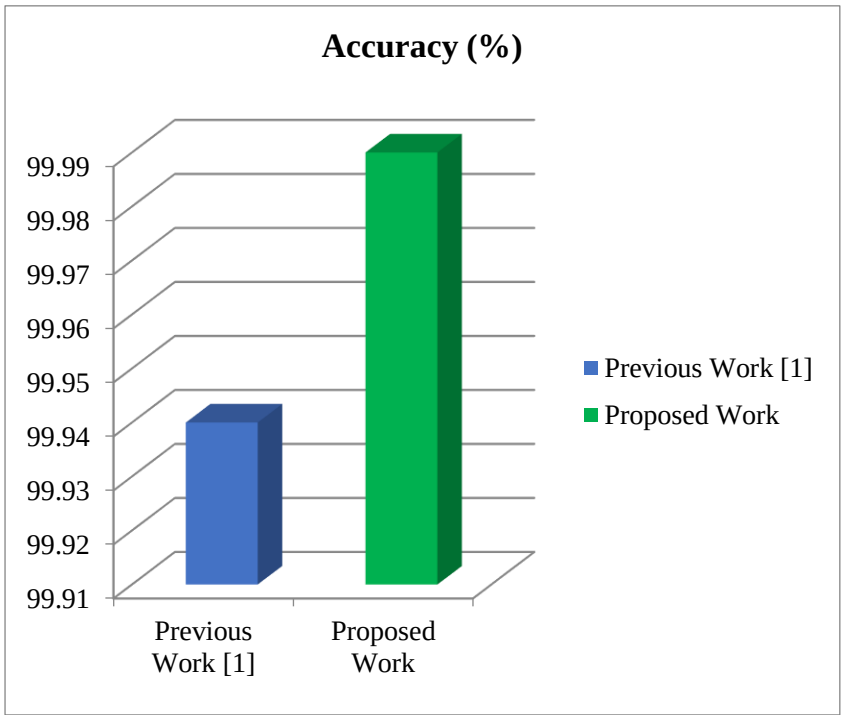


Figure 5: Accuracy Result graph

This graphical depiction of the accuracy is shown in figure 5, which may be found here. The correctness of the suggested work was superior to that of the work that was already done.

IV. Conclusion

A hybrid learning approach combining Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for cloud-based intrusion detection systems. By leveraging the pattern recognition capabilities of ANN and the classification precision of SVM, the model effectively addresses the challenges of detecting complex and evolving cyber threats in cloud environments. The experimental results demonstrated the superiority of the proposed hybrid model in terms of detection accuracy, false positive rate reduction, and scalability compared to traditional IDS solutions. This approach highlights the potential of hybrid learning techniques in enhancing cloud security while maintaining efficiency in real-time operations. Future work can focus on optimizing the model for real-world deployment, incorporating advanced feature selection methods, and extending the framework to detect emerging threats in multi-cloud and edge computing environments. The Python Spyder software is used in order to carry out simulation. The ANN method that was suggested achieves an overall accuracy of 99.99% with a classification error of 0.01% during the course of its operation.

References

[1]H. Attou, A. Guezzaz, S. Benkirane, M. Azrour and Y. Farhaoui, "Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques," in *Big Data Mining and Analytics*, vol. 6, no. 3, pp. 311-320, September 2023, doi: 10.26599/BDMA.2022.9020038.

[2]S. Ho, S. A. Jufout, K. Dajani and M. Mozumdar, "A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Convolutional Neural Network," in IEEE Open Journal of the Computer Society, vol. 2, pp. 14-25, 2022, doi: 10.1109/OJCS.2021.3050917.

[3]A. Sonkar, S. K. Sahu, A. Nayak, D. Sahu, P. Verma and R. Tiwari, "An Efficient Privacy-Preserving Machine Learning for Blockchain Network," *2024 4th International Conference on Intelligent Technologies (CONIT)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/CONIT61985.2024.10627061.

[4]S. Thirimanne, L. Jayawardana, P. Liyanaarachchi and L. Yasakethu, "Comparative Algorithm Analysis for Machine Learning Based Intrusion Detection System," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 2021, pp. 191-196, doi: 10.1109/ICIAfS52090.2021.9605814.
[5]T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3121870.
[6]W. Xu, J. Jang-Jaccard, A. Singh, Y. Wei and F. Sabrina, "Improving Performance of Autoencoder-Based Network Anomaly Detection on NSL-KDD Dataset," in IEEE Access, vol. 9, pp. 140136-140146, 2021, doi: 10.1109/ACCESS.2021.3116612.
[7]K. Cao, J. Zhu, W. Feng, C. Ma, M. Liu and T. Du, "Network Intrusion Detection based on Dense Dilated Convolutions and Attention Mechanism," 2021 International Wireless Communications and Mobile Computing (IWCMC), 2021, pp. 463-468, doi: 10.1109/IWCMC51323.2021.9498652.
[8]I. Ullah and Q. H. Mahmoud, "Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks," in IEEE Access, vol. 9, pp. 103906-103926, 2021, doi: 10.1109/ACCESS.2021.3094024.
[9]D. Park, S. Kim, H. Kwon, D. Shin and D. Shin, "Host-Based Intrusion Detection Model Using Siamese Network," in IEEE Access, vol. 9, pp. 76614-76623, 2021, doi: 10.1109/ACCESS.2021.3082160.
[10] I. Siniosoglou, P. Radoglou-Grammatikis, G. Efstathopoulos, P. Fouliras and P. Sarigiannidis, "A Unified Deep Learning Anomaly Detection and Classification Approach for Smart Grid Environments," in IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 1137-1151, June 2021, doi: 10.1109/TNSM.2021.3078381.
[11] https://www.unb.ca/cic/datasets/ids-2017.html